ISSN 0321-3900

Theory of Stochastic Processes

7(23)•2001 № 1-2

scientific
TBiMC
publishers

Proceedings of the International School on Mathematical and Statistical Applications in Economics

January 15-19, 2001, Västerås, Sweden

Edited by
Dmitrii Silvestrov, Västerås, Sweden
Anders Martin-Löf, Stockholm, Sweden
Oleksandr Borysenko, Kyiv, Ukraine

C O N T E N T S

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.xi

PREFACE

The International School on Mathematical and Statistical Applications in Economics was organised in the frame of the Tempus Tacis Joint European Project *Statistical Aspects of Economics* under auspices of Mälardalen University (Sweden), Umeå University (Sweden), Stockholm University (Sweden), University of Helsinki (Finland) and Kyiv Taras Shevchenko University (Ukraine) involved in the project.

The programme covered the following topics: financial and actuarial mathematics; survey sampling in economics; mathematical models in micro- and macro-economics; computer based modelling in economics; educational programmes and teaching in mathematical economics.

The participants of the school, 95 in total, were professors, lecturers and postgraduate students as well practitioners in business and industry from Estonia, Finland, Germany, Latvia, Poland, Sweden, Russia and Ukraine.

There were given 17 invited lectures and 23 communications at the school. Invited lecturers at the School were Tomas Björk (Stockholm), Kimmo Eriksson (Västerås), Jan Grandell (Stockholm), Mats Gyllenberg (Turku), Bengt Janson, Erik Ricknell and Jan Röman (OM Technology AB, Stockholm), Sune Karlsson (Stockholm), Anders Klevmarken (Uppsala), Gunnar Kulldorff (Umeå), Mathias Lanner and Jens Roslin (SPSS Sweden AB, Stockholm, Sweden), Anders Martin-Löf (Stockholm), Harri Nyrhinen (Helsinki), Lars-Erik Öller (Stockholm), Bengt Rosén (Stockholm), Dmitrii Silvestrov (Västerås) and Alexander Kukush (Kyiv), Bengt Swensson (Örebro), Imbi Traat (Tartu), Mikhail Yadrenko (Kyiv) and Nadiya Zinchenko (Kyiv).

The Proceedings of the School includes 32 papers based on invited lectures, communications and posters presented at the School. The Proceedings are published as a double issue of the journal *Theory of Stochastic Processes*. All papers have been reviewed, and we are grateful to the journal's Editorial Board for inviting to publish the Proceedings in the journal and arranging the refereeing of the papers.

The Proceedings of the School will be distributed to the Ukrainian and Scandinavian universities, financial and insurance institutions, and libraries that will contribute to the dissemination programme realising within the EU Tempus Tacis project *Statistical Aspects of Economics* as well as to the development of international co-operation in the area of higher education and science.

The editors

SPONSORS

The International School on Mathematical and Statistical Applications in Economics is organised in the framework of the Tempus Tacis Joint European Project "Statistical Aspects of Economics" under the auspices of:

- **Mälardalen University**
- **Umeå University**
- **Stockholm University**
- **University of Helsinki**
- **Kyiv National Taras Shevchenko University**

The following institutions have given financial support to the school:

- **City of Västerås**
- **Mälardalen University**
- **Department of Mathematics and Physics (Mälardalen University)**

The Organising Committee and the Scientific Programme Committee are very grateful to these institutions for their support of the school.

COMMITTEES

Organising Committee: **Dmitrii Silvestrov (Chairman, Västeräs), Clas Nordin (Västeräs), Evelina Silvestrova (Secretary, Umeå/Västeräs), Ingrid Westerberg-Eriksson (Umeå), Nadiya Zinchenko (Kyiv).**

Scientific Programme Committee: **Dmitrii Silvestrov (Co-chairman, Västerås), Anders Martin-Löf (Co-chairman, Stockholm), Oleksandr Borisenko (Kyiv), Kimmo Eriksson (Västerås), Gunnar Kulldorff (Umeå), Evelina Silvestrova (Secretary, Umeå/ Västerås), Esko Valkeila (Helsinki), Mikhail Yadrenko (Kyiv).**

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.1-4

International School on Mathematical and Statistical Applications in Economics

January 15-19, 2001, Västerås, Sweden

Programme

SUNDAY, January 14

18.00 - 19.00. *Opening ceremony* [**Omega** hall, the main building of the university]
19.00 - 21.00. *Welcome reception* [cafe **Origo**, the main building of the university]

MONDAY, January 15

Invited lectures [**Pi** hall, the main building of the university]

9.00 - 9.15. *Openning session*
9.15 - 10.05. Anders Martin-Löf (Stockholm): "On the application of control theory in insurance"
10.05 - 10.25.] *Refreshments*
10.25 - 11.15. Gunnar Kulldorff (Umeå): "Should we repeat the sampling design in repeated surveys?"
11.20 - 12.10. Lars-Erik Öller (Stockholm): "The accuracy of European growth and inflation forecasts"
12.10 - 13.30. *Lunch* [cafe **Origo**]

Communications [**Pi** hall]

13.30 - 13.55. Maciej Klimek (Uppsala), Gustaf Strandell (Uppsala) and Johan Tysk (Uppsala): "Testing limitations of the random walk hypothesis"
14.00 - 14.25. Erik Dotzauer (Västerås), Henrik Jönsson (Västerås) and Hans F. Ravn (Balleru): "Optimal unit commitment by branch-and-bound exploiting dual optimality conditions"
14.30 - 14.55. Andrzej Malawski (Cracow): "Dynamic Arrow-Debreu model - construction, some properties and applications"

15.00 - 15.25. Oleksandr Chernyak (Kyiv) and Andriy Stavytskyy (Kyiv): "Forecasting of the balance of payments of Ukraine"
15.25 - 15.40. *Refreshments*
15.45 - 16.10. Dmytro Gusak (Kyiv): "The distribution of extreme for risk processes on finite Markov chain"
16.15 - 16.40. Isabella Huber (Karlsruhe): "New approach to portfolio choice theory"
16.45 - 17.10. Johan Tysk (Uppsala): "Absence of arbitrage in markets with infinitely many assets"
17.15 - 17.40. Yurii V. Chernikov (Kyiv): "Goodness-of-fit test in Nevzorov's record model"

TUESDAY, January 16

Invited lectures [**Pi** hall]

9.15 - 10.05. Bengt Rosén (Stockholm): "Pareto sampling - a method for drawing samples with inclusion probabilities proportional to given sizes"
10.05 - 10.25. *Refreshments*
10.25 - 11.15. Imbi Traat (Tartu): "Distribution-based inference in survey sampling"
11.20 - 12.10. Kimmo Eriksson (Västerås): "Two-sided matching markets"
12.10 - 13.30. *Lunch* [cafe **Origo**]

Communications [**Pi** hall]

13.30 - 13.55. Torgöt Berling (Västerås) and Dmitrii Silvestrov (Västerås): "New master programme in analytical finance"
14.00 - 14.25. Richard Bonner (Västerås) and Violetta Galant (Wroclaw): "Allocation of computational resource in economic search"
14.30 - 14.55. Christer Nilsson (Västerås): "Energy use in Swedish households"
15.00 - 15.25. Alexander V. Mertens (Kyiv): "Stochastic quasi-gradient techniques in Var-based ALM models"
15.25 - 15.40. *Refreshments*
15.45 - 16.10. Andriy Kaminsky (Kyiv): "Statistical approach to portfolio optimisation on Ukrainian secondary stock market"
16.15 - 16.40. Kenneth Holmström (Västerås): "Global optimization of costly nonconvex functions, with financial applications"
16.45 - 17.10. Jörgen Hansson (Linkoping): "The use of optimization in finance"
17.15 - 18.00. *Poster session*

WEDNESDAY, January 17

Invited lectures [**Pi** hall]

9.15 - 10.05. Tomas Björk (Stockholm): "On the term structure of futures and forward prices"
10.05 - 10.25. *Refreshments*
10.25 - 11.15. Bengt Swensson (Örebro): "Reflections on unequal probability sampling strategies"
11.20 - 12.10. Mats Gyllenberg (Turku): "Approximating ruin probabilities of insurance companies. The perturbed renewal equation approach"
12.10 - 13.30. *Lunch* [cafe **Origo**]

Invited lectures [**Pi** hall]

13.30 - 14.20. Dmitrii Silvestrov (Västerås) and Alexander Kukush (Kyiv): "Optimal Monte Carlo option pricing"
14.25 - 15.15. Jan Röman, Bengt Janson and Erik Ricknell (OM Group, Stockholm): "Clearing and exchanges as a business"
15.15 - 15.40. *Refreshments*
15.40 - 16.30. Mathias Lanner and Jens Roslin (SPSS Sweden AB, Stockholm): "Demonstration of SPSS products"

Communications [**Pi** hall]

16.35 - 17.30. *Poster session*
19.00 *Conference dinner* [Student restaurant, Kårhuset, Gustavsborgsgatan 6]

THURSDAY, January 18

Invited lectures [**Pi** hall]

10.05 - 10.25. *Refreshments*
10.25 - 11.15. Anders Klevmarken (Uppsala): "Microsimulation - a tool for economic analysis"
11.20 - 12.10. Harri Nyrhinen (Helsinki): "On the ruin probabilities in an economic environment"
12.10 - 13.30. *Lunch* [cafe **Origo**]
14.00 - 16.00. *A walking tour in the centre of Västerås* [from the entrance to the main building of the university]
17.00. *An information meeting followed by a dinner in the City Hall* [**City Hall**]

FRIDAY, January 19

9.15 - 10.05. Jan Grandell (Stockholm): "Simple approximations of ruin probabilities"
10.05 - 10.25. *Refreshments*
10.25 - 11.15. Sune Karlsson (Stockholm): "Modelling autocorrelation in panel data models"
11.20 - 12.10. Michailo Yadrenko (Kyiv) and Nadiya Zinchenko (Kyiv): "EU Tempus-Tasis project *Statistical Aspects of Economics* and new educa-

tional programmes on applied statistics, financial and actuarial mathematics"
12.10 - 13.30. *Lunch* [cafe **Origo**]

Communications [**Pi** hall]

13.30 - 13.55. Olexandr Ponomarenko (Kyiv): "The master educational program in the mathematical economics and econometrics at Kyiv National Taras Shevchenko University"
14.00 - 14.25. Esko Valkeila (Helsinki / Turku): "Some properties of geometric fractional Brownian motions"
14.30 - 14.55. Yuliya S. Mishura (Kyiv) and Aleksey M. Gorelov (Kyiv): "Optimal stopping time for making irreversible"
15.00 - 15.25. Arne Frennelius (Västerås): "A stock selection guide"
15.25 - 15.40. *Refreshments*
15.45 - 16.10. Henrik Jönsson (Västerås): "Monte Carlo studies of American call options with discrete time"
16.15 - 16.40. Myroslav Drozdenko (Kyiv): "Explicit solutions for the ruin probability problem in the case of classical risk model"
16.45 - 17.10. Denys Kerekesha (Kyiv): "Some generalization of the ruin probability problem in the classical risk theory"
17.15 - 17.40. Dmitrii Silvestrov (Västerås), Viktor Galochkin (Kyiv / Västerås) and Anatoliy Malyarenko (Kyiv / Västerås): "OptAn - a pilot program system for analysis of options"
17.45 - 18.00. *Closing session*

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.5-12

RICHARD F. BONNER AND ANNA FEDYSZAK-KOSZELA

WHEN TO STOP LEARNING? BOUNDING THE STOPPING TIME IN THE PAC MODEL

The idea to include the cost of economic decision in an economic decision model is classical. It is the theme of the economics of information, economic search, decision theory, etc., and it is symbolic for Simon's concept of bounded rationality. The theme has also a computational counterpart coming from computational games. One of the broad standing problems in the border area of economics and artificial intelligence, is to adapt computational models of learning, such as Valiant's PAC model, to the context of economic games. To this category of problems belongs the question of stopping time for a learning process, a notion extending that of sample complexity. In the present note, we use the PAC convergence rates to bound the optimal stopping time in passive supervised eager learning.

2000 *Mathematics Subject Classifications.* 91B44, 91B70.

Key words and phrases. Stopping time, PAC learning.

1. Introduction

Learning theory is today a vast and diverse subject with a rapidly growing field of application. To set the scene, let us very briefly scan the landscape of its theoretical foundations. Roughly, two (nondisjoint) approaches to learning are current, statistical and computational, say, both of interest for economic theory. The statistical learning models start with Bush and Mosteller (1955) and Norman (1968) and extend to Vapnik's recent work (1995,1998). A computational perspective was added to these in the 80's by Valiant and others, mainly through a notion of Probably Approximately Correct (PAC) convergence, see Kearns and Vazirani (1994). The statistical learning models have strong pragmatic roots, see Valiant (1984), and hence place well in Economics. However, as visible from a review paper by Sobel (2000), only the early models have found a place in economic theory, leaving the application of the more recent computational versions, to

exemplify, Mitchell (1997), Nakhaeizadeh and Taylor (1997), Anthony and Barlett (1999), Poznyak and Najim (1997), or Scholkopf et al (1999), yet to be explored.

The strictly computational view on learning began with Gold's concept of identification in the limit, see the monograph of Jain et al (1999). Although theoretically basic, its practical application seems so far limited. The same can be said about the fundamental Minimum Description Length (MDL) principle of learning, derived from the notion Kolmogorov complexity, see Li and Vitanyi (1997), the applications of which are only recent, Gao et al (2000).

Pragmatic motivations of learning theories come forth in the framework of economic games as reinforcement learning, see Fudenberg and Levine (1998), and Sutton and Barto (1999). Note however, that the focus here is on the learning of the game by agents, and learning theory enters more as a tool than as an object of study. Furthermore, it is normally assumed that earning is active: the agent interactively chooses the information to acquire. An important example of active learning is economic search, Kohn and Shavell (1974), perhaps best known in Weitzman's (1979) phrasing as Pandora's Problem.

The question we take up presently is formally akin to the optimal stopping problem of economic search. We ask about the optimal stopping time in the framework of PAC learning, so as to balance of the cost of learning against an imposed pragmatic purpose. Unlike economic search, the learning we consider is passive in the usual sense that during learning the samples arrive randomly. Furthermore, roughly speaking - and this is the main point of the PAC theory as explained in Valiant (1984) - the knowledge of the probability measure generating the samples is not required to bound the learning rate, provided the size of the object to be learned is not excessive in a well-defined technical sense.

We note several previous papers linking the PAC learning model to a decision-theoretic context, for example, Haussler (1992), Haussler et al (1991, 1994), Devroye and Lugosi (1995), Freund and Schapire (1997). These, and especially, the paper of Haussler (1992), could be useful, should a more systematic study of the stopping problem for the PAC model be undertaken. It goes without saying that the problem may equally well be posed for learning models other than the PAC model.

2. Stopping time

Consider a learning process in discrete time $t = 0, 1, 2, ...$, with information arriving at unit cost. Assume that at a certain moment $t = s$ the learning stops and the resulting knowledge is from then on recalled indefinitely, each time returning an expected *reward* $r(s)$.

The behaviour of the sequence $r(s)$, $s = 0, 1, \ldots$, will obviously depend on the specifics of the learning model, but we assume throughout that $r(s)$ is eventually positive, bounded and non-decreasing with limit r_∞. We write

$$r(s) = r_\infty - q(s), r_\infty > 0, q(s) \searrow 0. \tag{1}$$

Assuming a fixed discounting factor $0 < \lambda < 1$, the expected value of the combined learning and recalling process is then

$$-\sum\nolimits_{0 \le j < s} \lambda^j + r(s) \sum\nolimits_{s \le j} \lambda^j = \frac{1}{1-\lambda} \left[(r(s)+1)\, \lambda^s - 1 \right] \tag{2}$$

$s = 0, 1, \ldots$. We refer to the integer s as the *stopping* time, and say that the stopping time is *optimal* if it maximizes (2). Clearly, by the made assumptions, a finite optimal stopping time exists. If it is not unique, we settle for the least of its values.

Note that the maximizing set of (2) is the same as that of the sequence

$$(1 - \tilde{q}(s))\, \lambda^s, s = 0, 1, \ldots \tag{3}$$

with $\tilde{q}(s) = \frac{q(s)}{r_\infty + 1}$. Note further, that for a sequence $\tilde{q}$ with a smooth extention $\tilde{Q}$ to the positive reals such that $(1 - \tilde{Q}(t))\, \lambda^t$ has a single stationary point t_0 there, the maximizing set (3) consists of no more than two integers closest to t_0. Note finally that if $\bar{q}(s)$ can be given bounds

$$\bar{q}_1(s) \le \bar{q}(s) \le \bar{q}_2(s), s = 0, 1, \ldots \tag{4}$$

with extensions $\tilde{Q}_1$ and $\tilde{Q}_2$ as above, then, up to the closest integer, the optimal stopping time is bounded by the two solutions to the equation

$$(1 - \tilde{Q}_1(t))\, \lambda^t = \max \left\{ (1 - \tilde{Q}_2(t))\, \lambda^t, t \ge 0 \right\}. \tag{5}$$

3. The PAC model

We briefly recall the PAC learning model in its most basic, the so-called restricted form. See, for example, Anthony and Barlett (1999) for details and extensions.

The learner is to determine a function $f : X \to Y$ given that f belongs to a class $F = F(X, Y)$. The information arrives sequentially during learning in the form of function values $f(x_t)$, $t = 0, 1, 2, ...$, the points $x_t \in X$ appearing randomly and independently according to a probability measure μ on X, unknown to the learner. The general (forecast) question is then: *how well can the learner guess the forthcoming value $f(x_s)$ given the past values $f(x_t)$, $t < s$, and the prior $f \in F$?*

Assume henceforth $Y = \{0, 1\}$, so functions in $F(X, Y)$ may also be viewed as subsets of X. At time s the learner forms a hypothesis $h_{x_0, \ldots, x_{s-1}} \in$

F consistent with f on the past data $h_{x_0,\ldots,x_{s-1}}(x_t) = f(x_t),\ t < s$. The probability of a wrong guess $h_{x_0,\ldots,x_{s-1}}(x_s) \neq f(x_s)$ at that moment is clearly equal to the $L^1(X,\mu)$-distance between $h_{x_0,\ldots,x_{s-1}}$ and f, also equal to the measure $\mu(h_{x_1,\ldots,x_s}\Delta f)$ of the symmetric difference of $h_{x_1,\ldots,x_s}$ and f considered as sets. This distance is random. It is less than a confidence parameter $\epsilon > 0$ with some probability

$$p_s(\epsilon) = \mu^s \left\{ x_0, \ldots, x_{s-1} : \int_X \left| h_{x_0,\ldots,x_{s-1}}(x) - f(x) \right| \, d\mu(x) < \epsilon \right\}, \quad (6)$$

intuitively expected to be large for large s, $p_s(\epsilon) > 1 - \delta$ if $1 > \delta > 0$, $s \geq s(\epsilon, \delta)$. If this estimate holds uniformly in μ, f and h, one says that the class F is *learnable* and we then assume that $s(\epsilon, \delta)$ denotes the least such integer. The starting point of the PAC learning theory is the following result; see, for example, Theorem 5.6 in Anthony and Barlett (1999).

Theorem 3.1. *The class F is learnable if and only if its Vapnik-Chervonenkis dimension d is finite. Furthermore, there exist constants $c_1, c_2 > 0$ such that*

$$\frac{c_1}{\epsilon}(d + \log\frac{1}{\delta}) \leq s(\epsilon,\delta) \leq \frac{c_2}{\epsilon}(d\log\frac{1}{\epsilon} + \log\frac{1}{\delta}), \quad 0 < \epsilon, \delta < 1. \quad (7)$$

The Vapnik-Chervonenkis dimension $d = d(F)$ is a number defined in terms of the separation properties of points in X by the functions in F. Specifically, $d(F)$ is the largest cardinality of a finite set $S \subset X$ such that the map $F \to 2^S$ of restriction to S is surjective.

For example, any non-trivial class F of binary functions on which the order induced from $\{0, 1\}$ is linear has Vapnik-Chervonenkis dimension equal to one, see Wenocur and Dudley (1981). This is in particular the case if F is the class of all non-decreasing binary functions on the unit interval $[0, 1]$.

4. Bounding the stopping time

Suppose a correct guess $h_{x_0,\ldots,x_{s-1}}(x_s) = f(x_s)$ at time s is rewarded while a wrong one $h_{x_0,\ldots,x_{s-1}}(x_s) \neq f(x_s)$ is penalised with a and b monetary units, respectively. The expectation of a wrong guess at time s is then given by

$$e_{h,f,\mu}(s) = \int_X \left| h_{x_0,\ldots,x_{s-1}}(x_s) - f(x_s) \right| \, d\mu(x_s)\ d\mu(x_0)\ldots d\mu(x_{s-1}). \quad (8)$$

Hence the expected reward $r_{h,f,\mu}(s)$ from recall at time s is

$$r_{h,f,\mu}(s) = a(1 - e_{h,f,\mu}(s)) - be_{h,f,\mu}(s) = a - (a+b)\, e_{h,f,\mu}(s). \quad (9)$$

This is of form (1), so there will clearly be an optimal stopping time $s_{opt} = s_{opt}(h, f, \mu)$. We would like to bound this number on the basis of the prior $f \in F$ only. Obviously, only the trivial lower bound $s_{opt} \geq 0$ is possible unless the possibilty of immediate lucky guess $h_{x_0} = f$ has been excluded. Leaving out the interesting but involved intermediate cases, let us place ourselves at the other end of the scale, assuming the learner is consistently making the 'least lucky' guesses $h_{x_0,\ldots,x_{s-1}}$ of f over all choices of $f \in F$ and over all probability measures μ.

For this 'least lucky' learner $\check{h}$ learning the least favorable function $\check{f}$ in the least favorable state of the world $\check{\mu}$, the expectation of a wrong guess at time s will be $e(s) = e_{\check{h},\check{f},\check{\mu}}(s) = \sup_{h,f,\mu} e_{h,f,\mu}(s)$. It follows by Theorem 3.1 that, in the case F is of finite Vapnik-Chervonenkis dimension, the sequence $e(s)$ is bounded

$$\frac{B_1}{s} \leq e(s) \leq \frac{B_2(\epsilon)}{s^\epsilon} \tag{10}$$

uniformly in μ, h and $f \in F$; here $B_1$is a positive constant depending on the set F only, and so is $B_2 = B_2(\epsilon)$ with any fix $\epsilon > 0$. The factor $\log \frac{1}{\epsilon}$ in the upper estimate in Theorem 3.1 cannot apparently be skipped, see Haussler et al (1994), hence we cannot in general have (10) with $\epsilon = 0$.

To link with our previous notation, $r_\infty = a$, $q(s) = (a + b)e(s)$, $\tilde{q}(s) = \frac{a+b}{a+1}e(s)$, and $C_i = \frac{a+b}{a+1}B_i$, $i = 1, 2$. Equation (5) with $\bar{q}_1(s) = C_1 s^{-1}$ and $\bar{q}_1(s) = C_2(\epsilon)s^{-(1+\epsilon)}$ now gives bounds for the optimal stopping time. For real numbers $x \leq y$, we write $\lfloor x, y \rceil$ for the shortest interval with integer end-points containing x and y.

Theorem 4.1 *For any $\epsilon > 0$, the optimal stopping time for the least favorable case of the* restricted *binary PAC learning model of finite Vapnik-Chervonenkis dimension, is contained in the interval $\lfloor \alpha, \beta \rceil$ where $\alpha \leq \beta$ are the two solutions of the equation*

$$(1 - \frac{C_1}{t})\,\lambda^t = \max\left\{(1 - \frac{C_2(\epsilon)}{t^{1+\epsilon}})\,\lambda^t, t \geq 0\right\}, \tag{11}$$

with $C_1 = \frac{a+b}{a+1}B_1$ and $C_2(\epsilon) = \frac{a+b}{a+1}B_2(\epsilon)$, and constants B_1 and $B_2(\epsilon)$ from (10) depending on the learned concept class F only.

We note that for the extended PAC model, which, roughly speaking, does not learn functions but instead fits functions to stochastic data, there is a similar to (10) but tighter bound

$$\frac{B_1'}{\sqrt{s}} \leq e(s) \leq \frac{B_2'}{\sqrt{s}} \tag{12}$$

for the expectation $e(s)$ of a wrong guess at time s in the least favorable case, cf Theorem 5.5 in Anthony and Barlett (1999). Hence, we also have a

corresponding result.

Theorem 4.2 *The optimal stopping time for the least favorable case of the* extended *binary PAC learning model of finite Vapnik-Chervonenkis dimension, is contained in the interval* $\lfloor \alpha, \beta \rceil$ *where* $\alpha \leq \beta$ *are the two solutions of the equation*

$$(1 - \frac{C_1}{\sqrt{t}})\,\lambda^t = \max\left\{(1 - \frac{C_2}{\sqrt{t}})\,\lambda^t, t \geq 0\right\}, \tag{13}$$

with $C_i = \frac{a+b}{a+1} B_i'$, $i = 1, 2$, *and constants* B_i' *from (12) depending on the admitted concept class only.*

To establish the Theorems, we only need to verify that functions of the form $(1 - Ct^{-p})\,\lambda^t$ with C, $p > 0$, are unimodal, i.e. increase up to some point and then fall off to zero, and that equations (11) and (13) indeed have solutions. Checking the first claim is elementary. A stationary point of the function $(1-Ct^{-p})\,\lambda^t$, $t > 0$, is a maximum and satisfies $t^{p+1} - Ct - \frac{pC}{\log \frac{1}{\lambda}} = 0$, and it is immediate that for any C, $p > 0$ and $1 > \lambda > 0$ there is exactly one such point. For $p = 1$, moreover, this point is easily identified as $\frac{C}{2}\left(1 + \sqrt{1 + \frac{4}{C \log \frac{1}{\lambda}}}\right)$. The existence of solutions to (11) and (13) is then immediate by (4) and (5). Unfortunately, we cannot solve (11) or (13) explicitly, but we expect numerical methods to be effective for fix values of parameters.

5. In conclusion

Not to complicate matters, we have looked only at the simplest of questions for the simplest of PAC learning models, and used the simplest of tools in the estimations. More general questions could allow varying learning cost, finite time horison, variable discounting rates, etc. More general PAC models, as in Anthony and Barlett (1999), would admit real-valued functions, complexity bounds, active learning, etc. Non-elementary estimating procedures may then be needed, for example, Kiefer's (1953) algorithms for optimizing unimodal functions. Note finally, questions other than the stopping time have been considered by Haussler (1992) for PAC models in a decision-theoretic context. Further work on these should be of interest to both the learning theory and its economic applications.

Bibliography

1. R. Ahlswede and I. Wegener, *Search Problems*, Wiley, (1987).

2. M. Anthony and P. L. Barlett, *Neural Network Learning: Theoretical Foundations*, Cambridge University Press, (1999).

3. R. R. Bush and F. Mosteller, *Stochastic models for learning*, Wiley, (1955).

4. L. Devroye and G. Lugosi, *Lower bounds in pattern recognition and learning*, Pattern Recognition, **28**(7), (1995), 1011-1018.

5. Y. Freund and R. E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting*, Journal of Computer and System Sciences, **55**(1), (1997), 119-139.

6. D. Fudenberg and D. K. Levine, *The Theory of Learning in Games*, MIT Press, (1998).

7. Q. Gao, M. Li and P. Vitanyi, *Applying MDL to learn best model granularity*, Artificial Intelligence, **121**, (2000), 1-29.

8. D. Haussler, *Decision theoretic generalizations of the PAC model for neural nets and other learning applications*, Information and Computation, **100**, (1992), 78-150.

9. D. Haussler, M. J. Kearns, N. Littlestone and M. K. Warmuth, *Equivalence of models for polynomial learnability*, Information and Computation, **95**(2), (1991), 129-161.

10. D. Haussler, N. Littlestone and M. K. Warmuth, *Predicting* $\{0,1\}$ *functions on randomly drawn points*, Information and Computation, **115**(2), (1994), 284-293.

11. S. Jain, D. Osherson, J. S. Royer and A. Sharma, *Systems that learn, Second edition*, MIT Press, (1999).

12. M. J. Kearns and U. V. Vazirani, *An introduction to computational learning theory*, MIT Press, (1994).

13. J. Kiefer, *Sequential minimax search for a maximum*, Proc. American Mathematical Society, **4**(2), (1953), 502-506.

14. M. Kohn and S. Shavell, *The theory of search*, Journal of Economic Theory, **4**(2), (1974), 593-123.

15. M. Li and P. M. B. Vitanyi, *An introduction to Kolmogorov complexity and its applications, 2nd edition*, Springer, (1997).

16. T. M. Mitchell, *Machine learning*, McGraw-Hill, (1997).

17. G. Nakhaeizadeh and C. C. Taylor, *Machine learning and statistics: the interface*, Wiley, (1997).

18. M. F. Norman, *Some convergence theorems for stochastic learning models with distance diminishing operators*, Journal of Mathematical Psychology, **5**, (1968), 61-101.

19. A. S. Poznyak and K. Najim, *Learning automata and stochastic optimization*, Springer, (1997).

20. B. Scholkopf, C. J. C. Burges and A. J. Smola, *Advances in kernel methods: support vector learning*, MIT Press, (1999).

21. J. Sobel, *Economists' Models of Learning*, Journal of Economic Theory, **94**, (2000), 241-261.

22. R. Sutton and A. Barto, *Reinforcement Learning*, MIT Press, (1999).

23. L. G. Valiant, *A theory of the learnable*, Comm. ACM, **27**(11), (1984), 1134-114.

24. V. N. Vapnik, *The nature of statistical learning theory*, Springer, (1995).

25. V. N. Vapnik, *Statistical learning theory*, Wiley Interscience, (1998).

26. M. Weitzman, *Optimal Search for the Best Alternative*, Econometrica, **47**(3), (1979), 641-654.

27. R. S. Wenocur and R. M. Dudley, *Some special Vapnik-Chervonenkis classes*, Discrete Mathematics, **33**, (1981), 313-318.

Department of Mathematics and Physics, Mälardalen University, SE-721 23 Västerås, Sweden. *E-mail*:rbr@mdh.se *E-mail*:aka@mdh.se

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.13-29

RICHARD F. BONNER AND VIOLETTA GALANT

ALLOCATION OF COMPUTATIONAL RESOURCE IN ECONOMIC SEARCH

We discuss the general economic search problem from the perspective of expended computational resource. We start with a brief survey of the field, including Weitzman's solution to the Pandora problem. The search problem is then put into geometric form with complexity measured by the linear dimension of the spaces of sampled variables. Falling back on some recent work on incremental approximation in Hilbert spaces, we ascertain the convergence of incremental search schemes. This corresponds, roughly, to search situations which admit iteratively computable strategies with a fixed memory bound, so that the successively computed strategies do not invalidate the previous ones but rather extend them by adding on detail. Allocation of computational resource for the search is then not essential as little computation is ever wasted. Finally, we give some simple examples when strict incrementality of search strategies fails to hold.

2000 *Mathematics Subject Classifications.* 91B34, 91B44, 91B06.

Key words and phrases. Economic search, incremental approximation.

1. Introduction

It is hard to dispute that a poorly informed decision is unlikely to be as good as a well-informed one. A process of improving the information base for a decision, in so far it consumes non-negligible economic resources, is called economic search. The cost of search may then be weighed against the increase in reward expected from a better-researched decision.

That search problems are of interest in Economics was, according to Varian (1999), apparently first observed by Stigler (1961), in effect starting a new branch of decision theory, see Kohn and Shavell (1974). This added to an already vast theory of search of information-theoretic origins, as presented in Ahlswede and Wegener. The key idea of economic search - that

investment may be delayed while waiting for information - is the point of departure of the option approach of financial theory, refining the classical net present value criteria. See Dixit and Pindyck (1994), for example. All this fits nicely in the big picture of Simon's (1982-1997) 'bounded rationality' and of 'information economics', Laffont (1989), in which search for information and its processing is part and parcel of economic activity.

Thus, in principle, the economic aspects of search are today well understood. Understanding also its computational aspects, however, is another matter. We note that only the simplest computational models of agents have been thoroughly treated in economic theory, to mention Ulam's cellular automata, see Simon (2000) for a bird's view. By contrast, the recent developments in computational games, as apparent in Fraenkel (2000), Sutton and Barto (1999), or Nebel (1996), put the agent's computational nature at the fore; see Ambroszkiewicz and Komar (1999) for agent models. This provides an unlimited reservoir of computational economic search problems. But no convincing framework to handle these is in place, notwithstanding recent technical work such as Bernardo et al (2000), Zadrozny and Elkan (2001), Tesauro and Kephart (1998), Kang (1999), or Saito (1999).

A similar view appears through the looking glass of learning theory. Indeed, although economic search is but a learning process within an economic game, see Fudenberg and Levine (1998), if to judge from a recent review by Sobel 200), learning theories seem to live in a subuniverse of Economics apparently quite disjoint from that of economic search problems. Furthermore, the early statistical models of Bush and Mosteller (1955) and Norman (1968) dominate, leaving the more recent computational theories virtually untapped. Of these, we only mention Gold's identification in the limit, see the monograph of Jain et al (1999), the statistical models of Vapnik (1998), the PAC models, see Kearns and Vazirani (1994), and the approaches based on Kolmogorov complexity, see Li and Vitanyi (1997). The latter, we note, are recently finding interesting application, see Gao et al (2000) or Smith (2000). Also, feed-back from economics to models of learning should not be unexpected. To the extent that an anchoring in economic theory is a formal expression of the pragmatic roots of a learning formalism, as apparent in Valiant (1984), economic search problems speak for a theory of more goal-oriented learning models.

Clearly, the task of merging the computational and the economic approaches to search is in many respects basic, and part of the quest to better understand the computational nature of economic decision. With this panorama as background, we presently put forward a question raised earlier in a related context in Bonner and Galant (2000). To explain, the problem of building optimal search trees was shown to be NP-complete by Hyafil and Rivest (1976), but there do exist incremental 'tree growing' algorithms of Quinlan (1986) and others that work well in practice, and sometimes

do yield optimal trees. We want to understand when is incremental search optimal, and when it is nearly so. Questions of optimality of (or approximation by) myopic search strategies in the applied context are of course not new, see Arkin (1964), for example, but apparently have not been settled in general.

We approach the question as follows. First, in Section 2, the economic search problem is recalled, exemplified, and put into geometric form. The notions of resource bounds and incrementality of strategies get form in Section 3, where we also link with the work of Barron (1993), Jones (1992) and Kurkova (1997), on incremental approximation. In this case the allocation problem trivialises in the sense that an optimal search strategy at any given resource level may be obtained - or approximated - by 'uppgrading' a strategy from a lower resource level. Section 4 provides some simple examples on strict incrementality or the lack of it in linearly ordered sets. The final Section 5 hints that the economic search problem be viewed as part of a larger Markov decision process. Finally, some of the many directions for further work are suggested in the Conclusions.

2. Economic search

2.1. The basic set-up. As a generic concept, economic search is nothing but active learning with inference and costs, and hence its orthodox place is in economic game theory. See Grenander (1981), Dixit and Pindyck (1994), Fudenberg and Tirole (1991), and Fudenberg and Levine (1998) for the fundamentals.

We begin, however, with a more open set-up. By an *abstract economic search scheme* we will mean the following structure.

- (i) a probability space $(\Omega, \mathcal{M}, \mu)$, representing the agent's prior knowledge about state of the world, where $\mathcal{M}$ is a sigma algebra on which μ is a probability measure, writing $\mu \in \mathcal{P}$,
- (ii) a family $\mathcal{G}$ of real random variables g on Ω, representing the rewards from the actions available to the agent,
- (iii) a family $\mathcal{F}$ of real random variables f on Ω, representing the observables available to the agent,
- (iv) for each $f \in \mathcal{F}$ a set M_f of random probability measures ν such that $\delta_\omega << \nu(\omega) << \mu$, representing more accurately than μ the state ω of the world,
- (v) a discrete time scale $t = 1, 2, \ldots$,

- (vi) two random non-negative functions c and τ on $\mathcal{F} \times \mathcal{P}$ prescribing the cost $c_{f,\nu}$ and the length of time $\tau_{f,\nu}$ required to evaluate an integral $\int f \, d\nu$,
- (vii) a discounting factor $0 \leq \lambda \leq 1$ assumed constant.

The *search* consists of a sequence of decisions: either

- (a) choose $f \in \mathcal{F}$ and $\nu \in M_f$, pay a price $c_{f,\nu}$ and wait $\tau_{f,\nu}$ units of time to learn the expected value $\int f \, d\nu$, or,
- (b) choose $g \in \mathcal{G}$, collect the reward $g(\omega)$, and end the search.

The *economic search problem* consists in finding a *search strategy*, in the sense of dynamic programming, see Puterman (1994), yielding the highest expected present value of the collected reward minus the search costs.

We are presently not overly concerned with extra technical assumptions required for solvability of the search problem, the absolute continuity of point evaluations, or working with random measures, see Kallenberg (1986). Some such conditions will be imposed below when passing to a geometric formulation.

2.2. Some examples. We recall some better known instances - applied or abstract - of the general search problem to remind of its universal nature; some known search problems though, for example, the apartment problem in Ciesielski and Zabczyk (1979), do not easily fall in the present framework. In each instance, the problem of bounding the complexity of search strategy opens a reservoir of specific questions. Everything starts of course with Claude Shannon (1949).

Example 2.1. *Search for a point (Shannon's noiseless coding). [See e.g. Ch 2 in Ahlswede and Wegener (1987).]* The problem is to find the shortest on the average encoding of points in a set as sequences of values of selected attributes at that point.

To see this as economic search, take Ω finite, put $\mathcal{F} = \mathcal{M} = 2^{\Omega}$, and let $\mathcal{G}$ consist of all random variables bounded by some sufficiently large number. Assume the cost and time resource functions, and the discounting factor all identically equal to one. Finally, put $M_f = \{\delta_\omega, \omega \in \Omega\}$. Shannon's 'noiseless coding theorem' then gives lower bounds on the expected value of search in terms of the entropy of the measure μ on $\mathcal{M}$.

Example 2.2. *Search for an object. [See e.g. Ch. 11 & 12 in Ahlswede and Wegener (1987).]* The problem is to locate a physical object, for example,

a lost ship in the ocean divided into n search sectors, or a misplaced pair of glasses in a house with n rooms. It may be required to search through a sector or room more than once.

Indeed, if for each of the rooms $f \in \mathcal{F}$, a single inspection $f(\omega)$ determines if the object searched for is there, this is Pandora's problem with only one of the boxes holding a reward.

Example 2.3. *Search for an alternative (Pandora's Boxes). [See Weitzman (1979).]* The problem is to choose one among a number of independent investments with known probability distributions, given the option to buy information, that is, to remove the uncertainty of an investment at a cost.

This is economic search proper in the sense of Kohn and Shavell (1974). Here $\mathcal{G} = \mathcal{F}$ is a finite set of independent random variables and $M_f = \{\delta_\omega, \omega \in \Omega\}$. Further, the agent's choice in (b) is limited to f already sampled. Weitzman (1979) showed that the optimal search strategy is determined by a function z on $\mathcal{F}$ called *reservation price*, given for each $f \in \mathcal{F}$ by the equation

$$c_f = e^{-\lambda\tau_f} \int (f - z_f)_+ \, d\mu - \left(1 - e^{-\lambda\tau_f}\right) z_f, \tag{1}$$

in the following way. If the maximum sampled reward exceeds the maximum unsampled reservartion price, collect the maximum sampled reward and stop; otherwise, sample the variable with the highest reservation price. (Notation: x_+ in (1) stands for the positive part of a real number x.)

Following Weitzman (1979), let us briefly interpret his rule in the case the rewards are binary and there is no discounting. Assume that each $f \in \mathcal{F}$ takes on a single non-zero value r_f with probability p_f and that the expected net gain $\Delta_f = r_f p_f - c_f$ is positive. Equation (1) gives in this case the reservation price $z_f = \frac{\Delta_f}{p_f}$ which, with the expected net gain kept constant, is *decreasing* as function of the probability of success (but *increasing* if the rewards r_f and prices c_f are kept constant).

Example 2.4. *Search for a set (PAC restricted binary model, see Anthony and Barlett (1999)).* Roughly, one seaks a function h in a given class $H \subset 2^X$ by sampling $h(x_t)$, $t = 0, 1, 2, ...$, $x_t \in X$.

Technical points aside, put $\Omega = 2^X$, let a prior probability μ express the 'learning bias' $h \in H$, and let $\mathcal{G} = \mathcal{F}$ consist of point evaluation functionals δ_x, $x \in X$. In an 'active' learning mode, the learner sequentially chooses the variables δ_x to sample h with. By contrast, in a 'passive' learning mode, the functionals δ_{x_t}, $t = 0, 1, 2, ...$, arrive as independent identically distributed

variables, with *unknown* probability distribution P. The PAC theory then gives lower bounds on learning time uniformly in the distribution P.

See Bonner and Koszela (2001) for a brief economic view on the PAC learning model. For details and various extentions of the PAC model, see Anthony and Barlett (1999). There are clearly possibilities to extend along present lines the PAC model to abstract inference schemes, but we do not elaborate.

Example 2.5. *Search for an approximant. [See e. g. Shapiro (1971).]* Let (X, d) be a metric space, $S \subset X$, $x \in X$. The problem is to find a minimizing sequence x_k, $k = 1, 2, \ldots$, for the distance $d(S, x) = \inf_{y \in S} d(y, x)$.

A probability space structure on X converts this problem in an obvious way into a search problem, with the reward variables of the form $g(x) = G(d(S, x))$ for some suitable decreasing function G. See Benveniste et al (1990) for interesting concretisations.

2.3. Geometric formulation. Putting the search problem into a Hilbert space follows standard procedures in stochastic theory. We do this to link with incremental approximation theory in Hilbert spaces. Two points, however, may need a comment.

First, concerning the inference. During search, the agent's evolving knowledge may be represented by an evolving probability measure μ_n, the knowledge at the outset being $\mu_0 = \mu$. The situation can be handled on the level of sigma algebras if at time n the new knowledge arrives in the form of a probability measure ν_n on a subalgebra $\mathcal{M}_n \subset \mathcal{M}$ and $\nu_n << \check{\mu}_n = \mu_n|_{\mathcal{M}_n}$ (this is the case, for example, if $\mathcal{G} = \mathcal{F}$ and $M_f = \{\delta_\omega, \omega \in \Omega\}$). One may then update μ_n putting $\mu_{n+1}(E) = \int_E \frac{d\nu_n}{d\check{\mu}_n} d\mu_n$, $E \in \mathcal{M}$, which is the unique extention of ν_n maximizing the entropy $-\int \frac{d\mu_{n+1}}{d\mu_n} \log \frac{d\mu_{n+1}}{d\mu_n} \, d\mu_n$ relative to μ_n, see e.g. Gudder and Marchand (1972). Let us write for brevity $\mu_{n+1} = \mu_n \vee \nu_n$.

Now, recall, by lifting the analysis to $L^2(\Omega, \mu)$ the inference operation lifts to a simple gluing procedure on orthogonal subspaces. Suppose α and β are positive linear functionals, defined on H and on a closed subspace V of H, respectively. Let $\alpha \vee \beta$ be the normalised linear functional simultaneously extending α from V and β from V's orthogonal complement $V^\perp$. Thus

$$\langle \alpha \vee \beta, f \rangle = \frac{\langle \alpha, P_V f \rangle + \langle \beta, P_{V^\perp} f \rangle}{\langle \alpha, P_V 1 \rangle + \langle \beta, P_{V^\perp} 1 \rangle}, f \in H, \tag{2}$$

with P_W denoting the orthogonal projection operator onto $W \subset H$. Note that this view of inference extends the previous one and, once measures have been lifted to $L^2(\Omega, \mu)$, allows to directly infer from knowledge in form of

an expected value $\int f\, d\nu_n$ with an otherwise unknown but consistent with μ_n measure ν_n.

The second point concerns the lifting of probability measures to fuctionals in $L^2(\Omega, \mu)$, which, by the Riesz thoerem, are then also elements of $L^2(\Omega, \mu)$. To avoid standard technicalities, assume that all data of the search problem can be contained within a suitably small Hilbert subspace $H \subset L^2(\Omega, \mu)$, so that the information in the search arrives as positive linear functionals ν on H of norm one, $< \nu, 1 >= \int \nu 1\, d\mu = 1$, and make no distinction between a functional and its Riesz representative. Note that if this assumption is to hold for the point masses δ_ω, $\omega \in \Omega$, the subspace H must have an Aronszajn-Bergman reproducing kernel. This opens up some interesting technical possibilities, see Ch. 6 in Shapiro (1971) and Scholkopf et al (1999), but we leave it at that.

3. Search with bounds

3.1. Resource bounds. We wish to bound the 'amount of information' which the agent may acquire during search, and study the beviour of optimal search strategies subject to bounds as the bounds increase. The 'amount of information' may be measured in various ways, for example, by the number of inspections (queries, samples, etc), the size of an underlying sigma algebra, the dimension of an associated linear space, etc. The situation is similar in most approximation schemes, from classical theory of degree of approximation, cf Ch. 8 in Shapiro (1971), to recent applied work in neural network approximation, Karny et al (1998), problems of model granularity, Gao et al (2000), Bayesian modelling, Berger (1985), etc.

All these schemes balance the accuracy of approximation against various bounds on the size of the approximating structure. We have found the following language helpful in dealing with the general situation.

Definition 3.1. *A* selection scheme with bounds *is a tuple* $(X, \leq; \phi : \mathcal{A} \to \mathcal{A}; \mathcal{E})$, *where (i)* $(X, \leq)$ *is a partially ordered set, (ii)* $\mathcal{A}$ *is a family of subsets of* X *closed under the operation* $E \cap F^\uparrow$, *with* $F^\uparrow = \cup_{x \in F} \{y : y \geq x\}$ *denoting the upper set of* F, *(iii) the map* $\phi : \mathcal{A} \to \mathcal{A}$ *is idempotent,* $\phi(E) \subset E$ *for* $E \in \mathcal{A}$, *and* $\phi(E) \neq \emptyset$ *unless* $E = \emptyset$, *(iv)* $\mathcal{E}$ *is a nested sequence of sets* $E_n \in \mathcal{A}$, $n = 1, 2, \ldots,$

$$\emptyset \neq E_1 \subset \ldots \subset E_n \subset \ldots \subset E_\infty = \cup_n E_n \subset X \tag{3}$$

consistent with the partial order $E_{n+1} \subset E_n^\uparrow$. *For* $n = 1, 2, \ldots,\ k \leq n$, *define maps* $U_n : \mathcal{A} \to \mathcal{A}$ *by* $U_n(E) = \phi(E_n \cap E^\uparrow)$ *and* $\breve{U}_{n,k} = U_n \circ U_{n-1} \circ \ldots \circ U_k$; *refer to these as maps of* global- *and* incremental uppgrade, *respectively.*

For intuition, think of points in X as mathematical structures, such as sets, planar graphs, sigma-algebras, linear spaces, etc., saying if $x \leq y$ that x is a substructure of y and y is a superstructure of x. The map ϕ represents a selection procedure, typically the computation of the optimizing subsets in an optimization scheme. The choice of the family $\mathcal{A}$ is in practice adapted to the selection procedure. The sequence (3) symbolises a graded resource bound, the elements of E_n being the structures available for selection at resource level n. Finally, think of U_n as extending structures directly to resource level n, to contrast with the incremental uppgrade $\breve{U}_n$ which extends structures successively through all the resource levels $k = 1, \ldots, n$.

Definition 3.1. *Further to previous definition, call a scheme* stable *if* $\breve{U}_{n,1}(X) \cap U_n(X) \neq \emptyset$ *and* strongly stable *if* $\breve{U}_{n,1}(X) \subset U_n(X)$ *for* $n = 1, 2, \ldots$. *If, furthermore,* $\mathcal{A}$ *has a (pseudo-) metric* d *and* $d(\breve{U}_{n,1}(X) \cap U_n(X)) \to 0$ *as* $n \to \infty$, *let us say that the scheme* converges *in this (pseudo-) metric.*

These are working definitions, yet to be put on topological grounds, cf Michael (1951) and Gierz et al (1980), but we pass as only two instances of schemes are presently of interest. Moreover, there would be no harm to assume in either case that the set X of admissible structures is finite.

In the first instance, X will be the lattice of σ- sub-algebras of a given σ-algebra $\mathcal{M}$ of subsets of a set Ω. We use this case for the examples in Section 4 and we may assume for simplicity that Ω and hence X is finite; we skip the prefix 'σ-' then. We let the family $\mathcal{A}$ consist of all finite sets of algebras in X. The resource bound (3) is defined with respect to a given set $\mathcal{F}$ of $\mathcal{M}$-measurable functions: for $n = 1, 2, \ldots$ let $E_n = E_n(\mathcal{F})$ consist of the algebras generated using sequential queries (search strategies) of length at most n from $\mathcal{F}$. The selection map ϕ will then be the map of picking minimizing subsets for the relative entropy function or some L^p- distance.

In the second case, X will be the lattice of the closed linear subspaces in a given Hilbert space $(H, \|\|)$. We identify subspaces $x \in X$ with the corresponding orthogonal projection operators P_x and endow X with the topology of weak operator convergence. Let S be a compact subset of H and let $E_n = E_n(S)$ consist of all the subspaces of X each of which is generated by at most n elements of S. Fix a point $h \in H$ let $\phi_h(E)$ be the minimizing set for the norm $\|(1 - P_x)(h)\|$ as $x \in E$. One gets a varation on the theme by taking convex hulls as $E_n(S)$ instead, with obvious modifications.

3.2. Incremental approximation. It is immediate but instructive to observe that the second scheme is stable if the set S consists of pairwise orthogonal vectors. Indeed, assume S normalised and extend it to an orthonormal basis $B \subset H$. Then $H \cong l^2(B)$ and the elements of $E_n(S)$ corre-

spond to functions on B with support in S and not containing more than n points. An element of $\phi_h(E_n(S))$ corresponds then to the space of functions vanishing outside a set of n points $s \in S$ where $|< h, s >|^2$ takes on its n largest values.

Moreover, the second scheme converges for *any* bounded S in the following sense. Consider the convex case. It was first shown by Maurey and Pisier (1980-81) that any h in the closed convex hull of S may be approximated in norm by convex combinations h_n of no more than n points in S with accuracy $O(n^{-\frac{1}{2}})$. Subsequently, motivated by applications in neural computation, Barron (1993) and Jones (1992) extended this result to incremental approximants h_n; explicitely, the combination h_{n+1} is formed from a superset of points which generate h_n. See also Kurkova (1997) and Ch 12 in Karny et al (1998).

We apply this result to the search problem as follows. For simplicity, restrict attention to the case $\mathcal{F} = \mathcal{G}$ with only point masses δ_ω, $\omega \in \Omega$, admitted in the search. Let H be the least closed linear subspace of $L^2(\Omega, \mu)$ containing the set $\mathcal{G}$. Define the following function on the lattice X of its closed subspaces

$$\rho(x) = \int \sup_{g \in \mathcal{G}} \frac{\left\langle (\mu \vee \delta_{\omega|x}), g \right\rangle}{\left\langle (\mu \vee \delta_{\omega|x}), 1 \right\rangle} \, d\mu(\omega) = \int \sup_{g \in \mathcal{G}} \frac{P_x g(\omega) + \int P_{x^\perp} g \, d\mu}{P_x 1(\omega) + \int P_{x^\perp} 1 \, d\mu} \, d\mu(\omega); \quad (4)$$

it is the expected least upper bound of the reward when variables generating x have been sampled. Note that ρ is monotone on X,

$$\sup_{g \in \mathcal{G}} \int g \, d\mu = \rho(0) \leq \rho(x) \leq \rho(y) \leq \rho(H) = \int \sup_{g \in \mathcal{G}} g \, d\mu, x \leq y \in X. \quad (5)$$

One may now formally define a selection scheme on X by an appropriate choice of a family $\mathcal{A}$ so that $\phi(E)$ is the maximizing set for ρ on $E \in \mathcal{A}$. For $n = 1, 2, \ldots$ let $E_n = E_n(\mathcal{G})$ be the collection of subspaces generated by at most n elements of $\mathcal{G}$. Denote by R_n and $\breve{R}_n$ the maximum of ρ on E_n and the incremental maximum of ρ on E_n, respectively (so $R_n = \rho(U_n(X))$ and $\breve{R}_n = \rho(\breve{U}_{n,1}(X))$ in the notation of Definition 3.1). It is clear that both sequences are monotone, and that $\breve{R}_n \leq R_n$. One can show that the function ρ is Lipschitz in the Hausdorff distance on X, $|\rho(x) - \rho(H)| \leq C d_h(x, H)$, the infimum of which over $x \in E_n$ is of the order $O(n^{-\frac{1}{2}})$ by the Barron-Jones estimates. We have thus sketched the proof of the following result.

Theorem 3.1. *Assume in the search problem that (i) the reward variables and the search variables are the same, $\mathcal{F} = \mathcal{G}$, with bounded second moments, (ii) the cost function is identically equal to one, (iii) the discounting factor is identically equal to one (no discounting). Then the expected net reward $\breve{R}_n$ from incremental search with bound n on the number of samples, and the expected net reward R_n from optimal nonsequential search with the*

same bound, are for large n of the same order, $R_n = \breve{R}_n + O(n^{-\frac{1}{2}})$.

4. Example: querying the interval

All material in this section concerns the classical Shannon search problem of Example 2.1. As discussed in Bonner and Galant (2000), the computed information does not in general increase in an additive fashion with the computational resource. We illustrate this now in the case of a finite totally ordered probability space, modelled by a finite partition of the unit interval $[0, 1]$ with the Lebesgue measure. The fact that the Borel σ-algebra of the unit interval has infinite entropy is not a problem, as one may always pass to a sufficiently large finite subalgebra.

Take $S \subset [0, 1]$ and let $\mathcal{F} = \mathcal{F}(S)$ be the family of all indicator functions f_x of the interval $[x, 1]$ with $x \in S$. Every finite query strategy in $\mathcal{F}$, sequential or not, will then output an algebra $\mathcal{B}$ given by a partition of $[0, 1]$ by an increasing sequence of points $x_k \in S$

$$0 = x_0 < x_1 < \ldots < x_n < x_{n+1} = 1. \tag{6}$$

The set $\mathcal{A}_n$ of all such algebras $\mathcal{B}$ is thus parametrised by a 'prism' $\Delta_n(S) \subset R^n$, and a subset $\phi(\mathcal{A}_n) \subset \mathcal{A}_n$ of 'optimal' algebras can be thought of as a subset $\phi(\Delta_n(S))$ of $\Delta_n(S)$. Stability and strong stability conditions are now expressed in terms of standard projections $\pi_{mn} : R^m \to R^n$, $m > n$, in the respective form

$$\pi_{mn}(\phi(\Delta_m(S)) \cap \phi(\Delta_n(S)) \neq \emptyset \tag{7}$$

and

$$\pi_{mn}(\phi(\Delta_m(S)) \subset \phi(\Delta_n(S)). \tag{8}$$

4.1. Stability in entropy. Assume, for simplicity, $\mathcal{G} = \mathcal{F}([0, 1])$. To maximize relative entropy, we minimize the function

$$h(x_1, \ldots, x_n) = \sum_{0 \leq k \leq n} (x_{k+1} - x_k) \log(x_{k+1} - x_k) \tag{9}$$

over $\Delta_n(S)$. By straightforward analysis, the minimum occurs when each of the points x_k lies closest to the midpoint of the interval $[x_{k-1}, x_{k+1}]$. For $S = [0, 1]$ the points x_k are then uniformly spaced, to remind that entropy is maximised by the uniform distribution.

Let now m_n be an increasing sequence of positive integers, $n \geq 1$, and let $\mathcal{A}_n$ in the grading (3) be the set of algebras given by partitions of the form (6) with n replaced by m_n. It then follows immediately that

Proposition 4.1. *In the case* $S = [0,1]$, *a* non-sequential *entropy maximizing query strategy is stable if and only if the number* m_{n+1} *is an integer multiple of* m_n *for* $n \geq 1$, *and, consequently, every* sequential *entropy maximizing query strategy is then (strongly) stable.*

The situation may be quite different when $S \neq [0,1]$.

Example 4.1. Let $0 \leq \delta < \frac{1}{4}$ and put $S = \left\{\frac{1}{4}+\delta, \frac{1}{2}, \frac{1}{2}+\delta, \frac{3}{4}+\delta\right\}$. While the partition given by the point $\frac{1}{2}$ is obviously entropy maximizing among all strategies of length one, it is easily verified that the optimal partition amomg all strategies of length two corresponds to the three remaining points in S.

Hence, in this case, the optimal question with only one question allowed is *not* the first of the two optimal questions with two questions allowed. The problem: for which $S \subset [0,1]$ is every sequential entropy maximazing strategy stable? seems to be open. We note in passing an elementary example in the other direction. For S consisting of five evenly spaced points, the suboptimal questions corresponding to points $\frac{1}{3}$ and $\frac{2}{3}$ can each be extended to an optimal strategy of two questions.

4.2. Stability in Euclidean metric. Let g be a square-integrable function on $[0,1]$ and let $\mathcal{B} \in \mathcal{A}_n$, parametrised by $\Delta_n(S)$ as above. The square of the L^2 distance of g to the linear space of all $\mathcal{B}$-measurable functions is then given by the function

$$\rho_g(x_1, \ldots, x_n) = \sum_{0 \leq k \leq n} \int_{x_k}^{x_{k+1}} (g(t) - E_{[x_k, x_{k+1}]}(g))^2 \, dt \tag{10}$$

where $E_{[x_k,x_{k+1}]}(g) = \frac{1}{x_{k+1}-x_k} \int_{x_k}^{x_{k+1}} g(t)\, dt$ is the mean of g over the interval $[x_k, x_{k+1}]$. The points $(x_1, \ldots, x_n) \in \Delta_n(S)$ which minimize (10) determine the optimal algebras $\mathcal{B} \in \mathcal{A}_n$, and the best approximation f to g is then $f(x) = E_{[x_k,x_{k+1}]}(g)$ for $x \in [x_k, x_{k+1}]$.

In the case $S = [0,1]$, if g is assumed continuous, the function ρ_g is continously differentiable in the open set $\Delta_n(S)$, and, by a straightforward computation, its stationary points are the solutions to the system

$$E_{[x_{j-1},x_j]}(g) = E_{[x_j,x_{j+1}]}(g), g(x_j) = \frac{1}{2}\left(E_{[x_{j-1},x_j]}(g) + E_{[x_j,x_{j+1}]}(g)\right) \tag{11}$$

for $1 \leq j \leq n$. For the identity function $g(x) = x$, for example, this system gives a uniform spacing of the points x_j; hence the observations made for relative entropy in Proposition 1 also apply to the present case (with the

same notation).

Proposition 4.2. *Let $g(x) = x$ and $S = [0,1]$. A* non-sequential *query strategy in $\mathcal{F}(S)$ minimizing the Euclidean distance (10) is stable if and only if the number m_{n+1} is an integer multiple of m_n for $n \geq 1$, and, consequently, every* sequential *query strategy minimizing the Euclidean distance is (strongly) stable.*

Observe that Example 1 holds also here for $g(x) = x$; there is also a corresponding question: for which $g \in L^2[0,1]$ and $S \subset [0,1]$ is every sequential strategy in $\mathcal{F}(S)$ minimizing the Euclidean distance (10) stable?

4.3. Stability in probability. Approximation in probability is the approximation in L^1 metric for binary functions (since for indicator functions, the distance $\|\chi_{E_1} - \chi_{E_2}\|_{L^1(\mu)}$ equals the measure $\mu(E_1 \Delta E_2)$ of the symmetric difference of their sets). For convenience, pass from $\mathcal{F}(S)$ as above to a family $\mathcal{H}(S)$ consisting of function of the form $f_y - f_x$ with $x < y$, $x, y \in S$. For technical reasons, we assume here that each f_x is the indicator function of an *open* subinterval of $[0,1]$. Informally, every question in $\mathcal{H}(S)$ corresponds to two questions in $\mathcal{F}(S)$.

Let $K \subset [0,1]$ be the standard ternary Cantor set, as in Kuratowski (1977), i.e. the set of all points in $[0,1]$ having ternary expansions involving the digits 0 and 2 only. Let $K_n \supset K$ be the closed set approximating K from outside in the standard Cantor construction, $n \geq 0$. Thus $K_0 = [0,1]$, $K_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$, etc. Finally, let $\mathcal{B}_n$ be the algebra generated by all the intervals building up K_n. Note that the algebras $\mathcal{B}_n$ are nested, $\mathcal{B}_n \subset \mathcal{B}_{n+1}$, $n \geq 0$.

The Cantor construction is a sequencial query strategy of the following kind. We ask for the value of a function $f_y - f_x$, $0 < x < y < 1$, in $\mathcal{H}(S)$: (i) if the value is one, we terminate the query and output zero, (ii) if the value is zero, we ask for the value of *two* functions $f_{y_1} - f_{x_1}$ and $f_{y_2} - f_{x_2}$ with $y_1 < x$ and $y < x_2$. At every node, the search tree thus branches out into a final leaf and two new nodes.

Proposition 4.3. *The Cantor ternary construction is stable in probability. Specifically, among all query strategies of length n in $\mathcal{H}([0,1])$ which seek the indicator function of K_N, $N >> 1$, the sequential strategy outputting $\mathcal{B}_n$ is optimal in probability, hence stable.*

There are trivial examples of non-stability in probability of non-sequential query strategies in general spaces; essentially, it suffices to limit the *a priori* family of questions to (the indicator functions of) three sets, such that measure of the first set is greater than the measure of the intersection

of the remaining two.

Example 4.2. (Non-stable non-sequential strategy.) Let $0 < a < b < c < \frac{1}{2}$. Let g be the indicator function $\chi_{[c,1-c]}$, and let $\mathcal{H}$ consist of the indicator functions of the intervals $[0, 1-b]$, $[a, 1-a]$, and $[b, 1]$. The question in $\mathcal{H}$ closest in probability to g is clearly $\chi_{[a,1-a]}$, while $\chi_{[0,1-b]}$ and $\chi_{[b,1]}$ are the pair that generates the best approximant $\chi_{[b,1-b]}$.

4. Decision processes with search

For future reference, consider the following - promising we think - compromise between the one-pass 'learn and act' model considered above, and the all-encompassing 'learn by action' reinforcement models as in Sutton and Barto (1999).

Recall the standard model of economic decision - the discrete-time Markov decision process (MDP), see Puterman (1994). Assume the process stationary. It is defined by a tuple (X, A, P, R), with X representing a set of states, A a finite set of actions, P a transition probability, and R a real-valued reward function. The set X here is normally assumed finite, but no such assumption is required at the level of general discussion. At any time $t = 0, 1, 2, ...$, an agent registers a state $x = x(t)$, picks an action $a = a(t)$ whch results in a transition probability $P_{a,x}$ to a new state $y = x(t+1)$, and receives an immediate reward $R_t = R(x, a, y)$. Thus, the agent's immediate reward depends randomly on the chosen action a; if in state $x(t)$ the expected value of the reward is $ER_t = \int R(x(t), a, y)\, dP_{a,x(t)}(y)$, and the agent's goal is to choose the actions $a = a(t)$, $t = 0, 1, 2, ...$, so as to maximize the total expected reward $\sum_{t\geq 0} \gamma^{-t} ER_t$ with respect to a discounting factor γ.

Consider the transition probability $P_{a,x}$ as the agent's prior knowledge about the response of the environment to an action a taken in a state x. Suppose a more accurate such knowledge $Q_{a,x} << P_{a,x}$ is available to the agent at a price in the following sense. There is a set F of random variables (queries) f on X which the agent may test against $Q_{a,x}$ in a sequential manner, so that a query f returns its expected value $\int f(y)\, dQ_{a,x}(y)$ at a price $c(f)$ and takes some time $\tau(f)$ to perform. The time $\tau(f)$ is measured by an external clock to be synchronised with the MDP time. In this way, the agent may at any time t insert a process of *economic search* into the MDP. The extended decision process falls within a framework of generalised Markov decision processes, cf Bertsekas (1995), Szepesvari and Littman (1996,1999).

5. In conclusion

We have put forward a problem and some arguments in justification of its further study. A number of directions come to mind.

In terms of theoretical fundamentals, it is clear that 'incremental' is a qualifier of complexity. How then does it relate to other such qualifiers, say, recursiv, iterative, ergodic, Markov, etc? Behind this question, there lurks a language problem: is it possible to talk of computation, complexity, information, semanitcs, economic systems, rational agents, etc, without multiple foirmalisms and linguistic frames? Perhaps the 'back to the (physics) basics' approach in the information sciences, in the spirit of empirical logic in Marlow (1978) may give guidance?

A better understanding of incrementality of search strategies - and algorithms in general - could come to immediate use at the level of technical theory, where notions of incremental algorithms are well-established but not necessarily well-understood or even well-defined. The Barron-Jones-Kurkova theory on incremental construction of a neural network may serve as inspiration. In pragmatic perspective, incrementality fits well in a picture of interplay of computational resource bounds of agents and their economic performance. We see the work of Szepesvari and Littman (1999) on Markov games especially promising, but as advocated, it would be interesting to include an information processing (search) model within their decision model. These largely structural aspects of search must of course be supported by hard statistical theory, as in Berger (1985) and Leadbetter et al (1983).

Finally, on the practitioner's side, it is common knowledge that the long-standing mental congestion in applied informatics is much due to difficulties with the notion of the value of information, see Kauffman and Riggins (1998). Perhaps the search perspective (= option approach) may help here. For example, following Varian's (1999) suggestions, we are now looking in this way at the problem of information retrieval Bonner et al (2001).

Bibliography

1. R. Ahlswede and I. Wegener, *Search Problems*, Wiley, (1987).
2. S. Ambroszkiewicz and J. Komar, *A Model of BDI-Agent in Game-Theoretic Framework*, In: J.-J. Ch. Meyer and P.-Y. Schobbens (eds), Formal Models of Agents, Sringer, (1999).
3. M. Anthony and P. L. Barlett, *Neural Network Learning: Theoretical Foundations*, Cambridge University Press, (1999).
4. V. I. Arkin, *Uniformly optimal strategies in search problems*, Theory of Probability and Applications, **9**, (1964), 159-160.
5. A. Barron, *Universal approximation bounds for superpositions of a sigmoidal function*, IEEE Transactions on Information Theory, **39**, (1993), 930-946.
6. A. Benveniste, M. Metivier and P. Priouret, *Adaptive algorithms and stochastic approximations*, Springer-Verlag, (1990).
7. J. O. Berger, *Statistical decision theory and Bayesian analysis*, Springer-Verlag, (1985).

8. F. P. Bernardo, P. Saraiva and E. N. Pistikopoulos, *Inclusion of information costs in process design optimization under uncertainty*, Computers and Chemical Engineering, **24**, (2000), 1695-1701.
9. D. P. Bertsekas, *Dynamic Programming and Optimal Control*, Athena Scientific, (1995).
10. R. Bonner and A. Fedyszak-Koszela, *When to stop learning: on optimal stopping time in the PAC model*, Present proceedings, (2001).
11. R. Bonner and V. Galant, *Encoding Knowledge in Tree Structures: To Grow or to Build? Incremental vs. spiral algorithms*, In: A. Baborski (ed.), Knowledge acquisition from databases, Proc. Conf. Turawa 19-21 May 2000, Wroclaw University, (2001).
12. R. Bonner and V. Galant, *Getting an answer before the question: Fast adaptive information retrieval*, In: A. Baborski (ed.), Knowledge acquisition from databases, Proc. Conf. Karpacz 18-20 May 2001, Wroclaw University, (2001).
13. R. R. Bush and F. Mosteller, *Stochastic models for learning*, Wiley, (1955).
14. Z. Ciesielski and J. Zabczyk, *A note on a selection problem*, In: Z. Ciesielski (ed.), Probability theory, PWN-Polish Scientific Publishers, (1979), 47-51.
15. A. K. Dixit and R. S. Pindyck, *Investment under uncertainty*, Princeton University Press, (1994).
16. A. S. Fraenkel, *Recent results and questions in combinatorial game complexities*, Theoretical Computer Science, **249**, (2000), 265-288.
17. D. Fudenberg and D. K. Levine, *The Theory of Learning in Games*, MIT Press, (1998).
18. D. Fudenberg and J. Tirole, *Game Theory*, MIT Press, (1991).
19. Q. Gao, M. Li and P. Vitanyi, *Applying MDL to learn best model granularity*, Artificial Intelligence, **121**, (2000), 1-29.
20. G. Gierz and others, *A Compendium of Continuous Lattices*, Springer, (1980).
21. U. Grenander, *Abstract inference*, Wiley, (1981).
22. S. Gudder and J. P. Marchand, *Non-commutative probability on von Neuman algebras*, Journal of Math. Physics, **13**, (1972), 799-806.
23. L. Hyafil and R. L. Rivest, *Constructing optimal binary decision trees is NP-complete*, Information Processing Letters, **5**(1), (1976), 15-17.
24. S. Jain, D. Osherson, J. S. Royer and A. Sharma, *Systems that learn, Second edition*, MIT Press, (1999).
25. L. K. Jones, *A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training*, Annals of Statistics, **20**(1), (1992), 608-613.
26. O. Kallenberg, *Random measures*, Akademie-Verlag, (1986).
27. B-K. Kang, *Optimal stopping problem with recall costs*, European Journal of Operational Research, **117**, (1999), 222-238.
28. M. Karny and K. Warwick and V. Kurkova (eds), *Dealing with Complexity. A Neural Network Approach*, Springer, (1998).
29. R. J. Kauffman and F. J. Riggins, *Information Systems and Economics*, Communications of the ACM, **41**(4), (1998), 32-34.
30. M. J. Kearns and U. V. Vazirani, *An introduction to computational learning theory*, MIT Press, (1994).

31. M. Kohn and S. Shavell, *The theory of search*, Journal of Economic Theory, **4**(2), (1974), 593-123.
32. K. Kuratowski, *Introduction to set theory and topology*, Pergamon Press, (1977).
33. V. Kurkova, *Dimension-independent rates of approximation by neural networks*, In: M. Karny and K. Warwick (eds.), Computer-Intensive Methods in Control and Signal Processing: Curse of Dimensionality, Birkhauser, (1997).
34. J-J. Laffont, *The economics of uncertainty and information*, MIT Press, (1989).
35. M. R. Leadbetter, G. Lindgren and H. Rootzen, *Extremes and related properties of random sequences and processes*, Springer, (1983).
36. M. Li and P. M. B. Vitanyi, *An introduction to Kolmogorov complexity and its applications, 2nd edition*, Springer, (1997).
37. A. R. Marlow (ed.), *Mathematical Foundations of Quantum Theory*, Academic Press, (1978).
38. E. Michael, *Topologies on spaces of subsets*, Transactions of the American Mathematical Society, **71**, (1951), 151-182.
39. B. Nebel, *Artificial Intelligence: A Computational Perspective*, In: G. Brewka (ed.), Principles of knowledge representation, CSLI, Stanford, California, (1996).
40. M. F. Norman, *Some convergence theorems for stochastic learning models with distance diminishing operators*, Journal of Mathematical Psychology, **5**, (1968), 61-101.
41. G. Pisier, *Remarques sur un resultat non publie de B. Maurey*, Seminaire d'analyse fonctionelle, **1-12**, (1980-81).
42. M. L. Puterman, *Markov Decision Processes - Discrete Stochastic Programming*, Wiley, (1994).
43. J. R. Quinlan, *Induction of decision trees*, Machine Learning, **1**(1), (1986), 81-106.
44. T. Saito, *Optimal stopping time with finite-period reservation*, European Journal of Operational Research, **118**, (1999), 605-619.
45. B. Scholkopf, C. J. C. Burges and A. J. Smola, *Advances in kernel methods: support vector learning*, MIT Press, (1999).
46. C. E. Shannon and W. Weaver, *The mathematical theory of communication*, University of Illinois Press, (1949).
47. H. S. Shapiro, *Topics in approximation theory*, Springer, (1971).
48. H. Simon, *Review: Barriers and bounds to Rationality*, Structural Change and Economic Dynamics, **9**, (2000), 243-253.
49. H. Simon, *Models of bounded rationality, Vols. I-III*, MIT Press, (1982-1997).
50. J. Smith, *Kolmogorov Complexity and Economic Specification*, Master's Thesis, University of Oxford, (2000).
51. J. Sobel, *Economists' Models of Learning*, Journal of Economic Theory, **94**, (2000), 241-261.
52. G. J. Stigler, *The Economics of Information*, Journal of Political Economy, **69**, (1961), 213-225.
53. R. Sutton and A. Barto, *Reinforcement Learning*, MIT Press, (1999).

54. C. Szepesvari and M. L. Littman, *A Unified Analysis of Value-Function-Based Reinforcement-Learning Algorithms*, Neural Computation, **8**, (1999), 217-259.
55. C. Szepesvari and M. L. Littman, *Generalised Markov Decision Process: Dynamic-programming and Reinforcement-learning Algorithms*, Technical Report CS-96-11, Brown University, (1996).
56. G. J. Tesauro and J. O. Kephart, *Forsight-based pricing algorithms in an economy of software agents*, In: Proc. ACM.ICE'98, Charleston, SC, (1999), 37-44.
57. L. G. Valiant, *A theory of the learnable*, Comm. ACM, **27**(11), (1984), 1134-114.
58. V. N. Vapnik, *Statistical learning theory*, Wiley Interscience, (1998).
59. H. R. Varian, *Economics and Search*, In: ACM-SIGIR Forum, ACM, (1999).
60. M. Weitzman, *Optimal Search for the Best Alternative*, Econometrica, **47**(3), (1979), 641-654.
61. B. Zadrozny and C. Elkan, *Learning and Making Decisions When Costs and Probabilities are Both Unknown*, Technical Report CS2001-0664, UCSD (2001).

Department of Mathematics and Physics, Mälardalen University, SE-721 23 Västerås, Sweden.
E-mail:rbr@mdh.se

Department of Artificial Intelligence Systems, Wroclaw University of Economics, Wroclaw, Poland.
E-mail:galant@manager.ae.wroc.pl

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.30-36

OLEKSANDR BORYSENKO AND IGOR MALYSHEV

THE LIMIT BEHAVIOUR OF INTEGRAL FUNCTIONAL OF THE SOLUTION OF STOCHASTIC DIFFERENTIAL EQUATION DEPENDING ON SMALL PARAMETER

We consider the behaviour of integral functional of the solution of stochastic differential equation with coefficients contained small parameter. The dependence on the order of small parameter in every term of equation with Wiener process and Poisson measure term is studied.

2000 *Mathematics Subject Classifications.* 60H10

Key words and phrases. Limit behaviour, integral functional, stochastic differential equation, small parameter.

1. Introduction

The investigation of limit behaviour of integral functionals of solution of stochastic differential equation is stimulated, in particular, by interest to behaviour of long-term return $(1/t)\int_0^t r(u)\,du$, as $t \to \infty$, where short interest rate $r(u)$ satisfies the stochastic differential equation of diffusion type [1], [2].

In this paper we study the behaviour, as $\varepsilon \to 0$, of functional $\eta_\varepsilon(t) = (\varepsilon^k/t)\int\limits_0^{t/\varepsilon^k} d(s,\xi(s))ds$, where $\xi(t)$ is the solution of stochastic differential equation

$$d\xi(t) = \varepsilon^{k_1} f(t,\xi(t))dt + \varepsilon^{k_2} g(t,\xi(t))dw(t) + \varepsilon^{k_3}\int_{\mathrm{R}^d} q(t,\xi(t),y)\tilde{\nu}(dt,dy), \quad (1)$$

$$\xi(0) = \xi_0;$$

$\varepsilon > 0$ is the small parameter; $k > 0,\ k_i > 0, i = 1,2,3$; $d(t,x)$ is non-random function; $f(t,x) = \{f_i(t,x), i = \overline{1,d}\}$, $q(t,x,y) = \{q_i(t,x,y), i = \overline{1,d}\}$ are non-random vector-valued functions; $g(t,x) = \{g_{ij}(t,x), i,j = \overline{1,d}\}$ is non-random matrix-valued function; $t \in [0,T], x,y \in \mathrm{R}^d$; $w(t)$ is d-dimensional

Wiener process; $\tilde{\nu}(dt,dy) = \nu(dt,dy) - \Pi(dy)dt$, $\nu(dt,dy)$ is the Poisson measure independent on $w(t)$, $E\nu(dt,dy) = \Pi(dy)dt$; $\Pi(\cdot)$ is a finite measure on Borel sets in R^d; ξ_0 is the random vector independent on $w(t)$ and $\tilde{\nu}(t,\cdot)$.

2. Auxiliary result

We need the following result.

Lemma. *Let for each $x \in \mathrm{R}^d$ there exists $\lim\limits_{T\to\infty} (1/T)\int_A^{T+A} b(t,x)\,dt = \bar{b}(x)$ uniformly with respect to A, the function $\bar{b}(x)$ is bounded and continuous, function $b(t,x)$ is bounded and continuous in x uniformly with respect to (t,x) in any region $t \in [0,\infty), |x| \le C$, and stochastic process $\xi(t)$ is stochastically continuous, then*

$$\mathrm{P}-\lim_{\varepsilon\to 0}\int_0^t b(s/\varepsilon,\xi(s))\,ds = \int_0^t \bar{b}(\xi(s))\,ds$$

for all arbitrary $t \in [0,T]$.

Proof. Since the process $\xi(t)$ is stochastically continuous then for any $\delta_1 > 0$ there exists such constant $C > 0$ that

$$\sup_{t\in[0,T]} \mathrm{P}\{|\xi(t)| > C\} \le \delta_1 \tag{2}$$

and for arbitrary $\delta_1 > 0$ and $\delta_2 > 0$ there exists such $\delta_3 > 0$ that

$$\mathrm{P}\{|\xi(t_1) - \xi(t_2)| > \delta_2\} \le \delta_1 \tag{3}$$

for all $|t_1 - t_2| < \delta_3, t_1, t_2 \in [0,T]$. We choose δ_2 such that $|\bar{b}(x) - \bar{b}(y)| < \delta_1$ and $|b(t,x) - b(t,y)| < \delta_1$ for all $t \in [0,T]$, as $|x-y| \le \delta_2$, $|x| \le C, |y| \le C$.

Let us consider partition $0 = t_0 < t_1 < \ldots < t_n = t, t \in [0,T]$ such that $\max\limits_{0\le k\le n-1} |t_{k+1} - t_k| < \delta_3$. We have for any $\delta > 0$

$$\mathrm{P}\left\{\left|\int_0^t b(s/\varepsilon,\xi(s))\,ds - \int_0^t \bar{b}(\xi(s))\,ds\right| > \delta\right\} =$$

$$= \mathrm{P}\left\{\left|\sum_{k=1}^n \int_{t_{k-1}}^{t_k} [b(s/\varepsilon,\xi(s)) - b(s/\varepsilon,\xi(t_{k-1}))]\,ds\right| > \delta/3\right\} +$$

$$+\mathrm{P}\left\{\left|\sum_{k=1}^n \int_{t_{k-1}}^{t_k} [b(s/\varepsilon,\xi(t_{k-1})) - \bar{b}(\xi(t_{k-1}))]\,ds\right| > \delta/3\right\} +$$

$$+\mathrm{P}\left\{\left|\sum_{k=1}^n \int_{t_{k-1}}^{t_k} [\bar{b}(\xi(s)) - \bar{b}(\xi(t_{k-1}))]\,ds\right| > \delta/3\right\} = \mathrm{P}_1 + \mathrm{P}_2 + \mathrm{P}_3.$$

For estimation of P_1 and P_3 we use the Chebyshev inequality, the properties of chosen partition, inequalities (2), (3) and boundedness of functions

$b(t,x), \bar{b}(x)$. Therefore we obtain $\mathrm{P_i} \le Ct\delta_1\delta^{-1}$, $i = 1,3$, where we use notation C for any constant independent on ε.

For each $k = \overline{1,n}$ from conditions of lemma we have

$$\lim_{\varepsilon\to 0}\int_{t_{k-1}}^{t_k}[b(s/\varepsilon,\xi(t_{k-1})) - \bar{b}(\xi(t_{k-1}))]\,ds = 0 \qquad \text{a.s.}$$

Therefore $\lim\limits_{\varepsilon\to 0}\mathrm{P}_2 = 0$, and for arbitrary $\delta_1 > 0, \delta > 0$

$$\lim_{\varepsilon\to 0}\mathrm{P}\left\{\left|\int_0^t b(s/\varepsilon,\xi(s))\,ds - \int_0^t \bar{b}(\xi(s))\,ds\right| > \delta\right\} \le Ct\delta_1\delta^{-1}.$$

The proof of lemma is completed.

Remark. Let $q(t,x,y)$ is bounded and uniformly continuous in x with respect to $t \in [0,\infty)$ and $y \in \mathrm{R}^d$ in every compact set $|x| \le C$. Let $\Pi(\cdot)$ be a finite measure on the σ-algebra of Borel sets in R^d and let

$$\lim_{T\to\infty}\frac{1}{T}\int_A^{T+A} q(t,x,y)\,dt = \bar{q}(x,y),$$

uniformly with respect to A for each $x \in \mathrm{R}^d, y \in \mathrm{R}^d$, where $\bar{q}(x,y)$ is bounded, uniformly continuous in x with respect to $y \in \mathrm{R}^d$ in every compact set $|x| \le C$. Then for any stochastically continuous process $\xi(t)$ we have

$$\mathrm{P}-\lim_{\varepsilon\to 0}\int_0^t\int_{\mathrm{R}^d} q(s/\varepsilon,\xi(s),y)\,\Pi(dy)ds = \int_0^t\int_{\mathrm{R}^d}\bar{q}(\xi(s),y)\,\Pi(dy)ds.$$

The proof of this statement is similar to proof of lemma.

3. Main result

Let us consider the functional $\eta_\varepsilon(t) = (\varepsilon^k/t)\int\limits_0^{t/\varepsilon^k} d(s,\xi(s))ds$, where $\xi(t)$ is the solution of equation (1). We suppose that coefficients of equation (1) satisfy the following conditions:

1) $|f(t,x)|^2 + \|q(t,x)\|^2 + |q(t,x,y)|^2 \le C$, where $|f|^2 = \sum_{i=1}^d f_i^2$, $\|g\|^2 = \sum_{i,j=1}^d g_{ij}^2$;

2) For any $N > 0$ there exists $L_N > 0$ such that

$$|f(t,x_1) - f(t,x_2)|^2 + \|g(t,x_1) - g(t,x_2)\|^2 +$$

$$+\int_{\mathrm{R}^d}|q(t,x_1,y) - q(t,x_2,y)|^2\,\Pi(dy) \le L_N|x_1 - x_2|^2,$$

for all $x_i \in \mathrm{R}^d, i = 1,2$ such that $|x_i| \le N, i = 1,2$.

3) Functions $f(t,x), g(t,x), q(t,x,y)$ are continuous in x uniformly with respect to $t \in [0,\infty), y \in \mathrm{R}^d$ and x in every set $|x| \le C$. Uniformly with respect to A for each $x \in \mathrm{R}^d, y \in \mathrm{R}^d$ there exists the following limits

$$\lim_{T\to\infty} \frac{1}{T} \int_A^{T+A} f(t,x)\,dt = \bar f(x), \lim_{T\to\infty} \frac{1}{T} \int_A^{T+A} g(t,x) g^*(t,x)\,dt = \bar G(x)$$

$$\lim_{T\to\infty} \frac{1}{T} \int_A^{T+A} q(t,x,y) q^*(t,x,y)\,dt = \bar Q(x,y).$$

Here g^* is the matrix (vector) transpose to g, therefore for vector-valued function $q(t,x,y)$ the product $q(t,x,y)q^*(t,x,y)$ is the $d \times d$-matrix-valued function.

4) The functions $\bar f(x), \bar G(x)$ are bounded, continuous in x, function $\bar Q(x,y)$ is bounded, continuous in x uniformly with respect to $y \in \mathrm{R}^d$. Matrix $\bar B(x) = \bar G(x) + \int_{\mathrm{R}^d} \bar Q(x,y)\,\Pi(dy)$ is uniformly parabolic.

Theorem. *Let conditions 1)-4) be fulfilled, $k = \min(k_1, 2k_2, 2k_3)$ and function $d(t,x)$ is bounded, continuous in x uniformly with respect to (t,x) in any region $t \in [0,\infty), |x| \le C$, there exists $\lim\limits_{T\to\infty} (1/T) \int_A^{T+A} d(t,x)\,dt = \bar d(x)$ uniformly with respect to A, the function $\bar d(x)$ is bounded and continuous.*
1. If $k_1 = 2k_2 = 2k_3$, then stochastic process $\eta_\varepsilon(t)$ converges in law, as $\varepsilon \to 0$, to stochastic process $\bar\eta(t) = (1/t) \int_0^t \bar d(\bar\xi(s))\,ds$, where process $\bar\xi(t)$ is the solution of stochastic differential equation

$$d\bar\xi(t) = \bar f(\bar\xi(t))dt + \bar\sigma(\bar\xi(t))d\bar w(t), \quad \bar\xi(0) = \xi_0, \tag{4}$$

$\bar\sigma(x) = \bar B^{1/2}(x)$; $\bar w(t)$ is some d-dimensional Wiener process.
2. If $k < k_1$, then in equation(4) the drift coefficient $\bar f(x)$ is absent; if $k < 2k_2$, then in equation (4) the diffusion matrix $\bar B(x)$ does not depend on $\bar G(x)$; and if $k < 2k_3$, then $\bar B(x)$ does not contain the term $\int_{\mathrm{R}^d} \bar Q(x,y)\,\Pi(dy)$.

Proof. We can rewrite $\eta_\varepsilon(t)$ in the form $\eta_\varepsilon(t) = (1/t) \int_0^t d(s/\varepsilon^k, \xi(s/\varepsilon^k))\,ds$. Let us denote $\xi_\varepsilon(t) = \xi(t/\varepsilon^k)$, $w_\varepsilon(t) = \varepsilon^{k/2} w(t/\varepsilon^k)$, $\tilde\nu_\varepsilon(t,\cdot) = \nu(t/\varepsilon^k,\cdot) - (t/\varepsilon^k)\Pi(\cdot)$. It worth to note that for any $\varepsilon > 0$ $w_\varepsilon(t)$ is the Wiener process and $\tilde\nu_\varepsilon(t,\cdot)$ is the centered Poisson measure. From equation (1) we obtain the following stochastic differential equation for $\xi_\varepsilon(t)$

$$\xi_\varepsilon(t) = \xi_0 + \varepsilon^{k_1-k} \int_0^t f(s/\varepsilon^k, \xi_\varepsilon(s))\,ds + \varepsilon^{k_2-k/2} \int_0^t g(s/\varepsilon^k, \xi_\varepsilon(s))\,dw_\varepsilon(s) +$$

$$+\varepsilon^{k_3} \int_0^t \int_{\mathrm{R}^d} q(s/\varepsilon^k, \xi_\varepsilon(s), y)\,\tilde\nu_\varepsilon(ds,dy). \tag{5}$$

It follows from conditions 1), 2) that the solution of equation (5) exists and unique for each $\varepsilon > 0$.

Let us check that following conditions are fulfilled:

a) $\lim\limits_{h\downarrow 0}\overline{\lim\limits_{\varepsilon\to 0}}\sup\limits_{|t-s|<h}\mathrm{P}\{|\xi_\varepsilon(t)-\xi_\varepsilon(s)|>\delta\}=0$ for any $\delta>0, t,s\in[0,T]$;

b) $\lim\limits_{N\to\infty}\overline{\lim\limits_{\varepsilon\to 0}}\sup\limits_{t\in[0,T]}\mathrm{P}\{|\xi_\varepsilon(t)|>N\}=0.$

Using the boundedness of coefficients of equation (1) and properties of stochastic integrals, we can obtain the estimates

$$E|\xi_\varepsilon(t)|^2\le C[E|\xi_0|^2+(\varepsilon^{2(k_1-k)}T+\varepsilon^{2k_2-k}+\varepsilon^{2k_3-k})t], \tag{6}$$

$$E|\xi_\varepsilon(t)-\xi_\varepsilon(s)|^2\le C[\varepsilon^{2(k_1-k)}|t-s|+\varepsilon^{2k_2-k}+\varepsilon^{2k_3-k}]|t-s|.$$

From the Chebyshev inequality and obtained estimates we have fulfillment of conditions a) and b). Similarly we can check conditions a) and b) for stochastic process

$$\zeta_\varepsilon(t)=\varepsilon^{k_2-k/2}\int_0^t g(s/\varepsilon^k,\xi_\varepsilon(s))\,dw_\varepsilon(s)+\varepsilon^{k_3}\int_0^t\int_{\mathrm{R}^d}q(s/\varepsilon^k,\xi_\varepsilon(s),y)\,\tilde\nu_\varepsilon(ds,dy).$$

Therefore [3], for any sequence $\varepsilon_n\to 0, n=1,2,\dots$ there exists a subsequence $\varepsilon_m=\varepsilon_{n_m}\to 0, m=1,2,\dots$, probability space, stochastic processes $\tilde\xi_{\varepsilon_m}(t),\tilde\zeta_{\varepsilon_m}(t)$, $\bar\xi(t),\bar\zeta(t)$ defined on this space, such that $\tilde\xi_{\varepsilon_m}(t)\to\bar\xi(t),\tilde\zeta_{\varepsilon_m}(t)\to\bar\zeta(t)$ in probability, as $\varepsilon_m\to 0$, and finite-dimensional distributions of $\tilde\xi_{\varepsilon_m}(t),\tilde\zeta_{\varepsilon_m}(t)$ are coincide with finite-dimensional distributions of $\xi_{\varepsilon_m}(t),\zeta_{\varepsilon_m}(t)$. Since we interesting in limit behaviour of distributions, we can consider processes $\xi_{\varepsilon_m}(t)$, and $\zeta_{\varepsilon_m}(t)$ instead of $\tilde\xi_{\varepsilon_m}(t),\tilde\zeta_{\varepsilon_m}(t)$. From (5) we obtain equation

$$\xi_{\varepsilon_m}(t)=\xi_0+\varepsilon_m^{k_1-k}\int_0^t f(s/\varepsilon_m^k,\xi_{\varepsilon_m}(s))\,ds+\zeta_{\varepsilon_m}(t). \tag{7}$$

From this point we will omit the sub-index m in ε_m for simplicity of notation. It worth to note that processes $\xi_\varepsilon(t)$ and $\zeta_\varepsilon(t)$ are stochastically continuous without discontinuity of second kind. For processes $\xi_\varepsilon(t)$ and $\zeta_\varepsilon(t)$ we have estimates

$$E|\xi_\varepsilon(t)-\xi_\varepsilon(s)|^4\le C[\varepsilon^{4(k_1-k)}|t-s|^4+E|\zeta_\varepsilon(t)-\zeta_\varepsilon(s)|^4], \tag{8}$$

$$E|\zeta_\varepsilon(t)-\zeta_\varepsilon(s)|^4\le C[(\varepsilon^{4k_2-2k}+\varepsilon^{4k_3-2k})|t-s|^2+$$
$$+\varepsilon^{4k_3-3k/2}|t-s|^{3/2}+\varepsilon^{4k_3-k}|t-s|], \tag{9}$$

$$E|\xi_\varepsilon(t)-\xi_\varepsilon(s)|^8\le C, E|\zeta_\varepsilon(t)-\zeta_\varepsilon(s)|^8\le C. \tag{10}$$

Since $\xi_\varepsilon(t)\to\bar\xi(t),\zeta_\varepsilon(t)\to\bar\zeta(t)$ in probability, as $\varepsilon\to 0$, then, using (10), from (8) and (9) we obtain estimates

$$E|\bar\xi(t)-\bar\xi(s)|^4\le C(|t-s|^4+|t-s|^2),\quad E|\bar\zeta(t)-\bar\zeta(s)|^4\le C|t-s|^2.$$

Therefore processes $\bar{\xi}(t)$ and $\bar{\zeta}(t)$ satisfy the Kolmogorov's continuity condition [4]. It should be noted that process $\zeta_\varepsilon(t)$ is the vector-valued square integrable martingale with matrix characteristic

$$\langle\zeta_\varepsilon,\zeta_\varepsilon\rangle(t)=\varepsilon^{2k_2-k}\int_0^t g(s/\varepsilon^k,\xi_\varepsilon(s))g^*(s/\varepsilon^k,\xi_\varepsilon(s))\,ds+$$

$$+\varepsilon^{2k_3-k}\int_0^t\int_{\mathrm{R}^d} q(s/\varepsilon^k,\xi_\varepsilon(s),y)q^*(s/\varepsilon^k,\xi_\varepsilon(s),y)\,\Pi(dy)ds. \tag{11}$$

For any $\delta>0$

$$\mathrm{P}\left\{\left|\int_0^t d(s/\varepsilon^k,\xi_\varepsilon(s))\,ds-\int_0^t\bar{d}(\bar{\xi}(s))\,ds\right|>\delta\right\}\le$$

$$\le\frac{2}{\delta}E\left|\int_0^t[d(s/\varepsilon^k,\xi_\varepsilon(s))-d(s/\varepsilon^k,\bar{\xi}(s))]\,ds\right|+$$

$$+\mathrm{P}\left\{|\int_0^t d(s/\varepsilon^k,\bar{\xi}(s))\,ds-\int_0^t\bar{d}(\bar{\xi}(s))\,ds|>\delta/2\right\}=\frac{2}{\delta}I_1+I_2.$$

Since the function $d(t,x)$ is continuous in x uniformly with respect to (t,x) in any region $t\in[0,\infty),|x|\le N$, then for any $\delta_1>0$ there exists $\delta_2>0$ such, that $\sup_{t\ge0}|d(t,x)-d(t,y)|\le\delta_1$ as $|x-y|\le\delta_2,|x|\le N,|y|\le N$. Therefore from (6) and boundedness of $d(t,x)$ we have

$$I_1\le E\int_0^t|d(s/\varepsilon^k,\xi_\varepsilon(s))-d(s/\varepsilon^k,\bar{\xi}(s))|\chi\{|\xi_\varepsilon(s)-\bar{\xi}(s)|\le\delta_2\}\times$$

$$\times\chi\{|\xi_\varepsilon(s)|\le N,|\bar{\xi}(s)|\le N\}\,ds+C\left(\int_0^t\mathrm{P}\{|\xi_\varepsilon(s)-\bar{\xi}(s)|>\delta_2\}\,ds+\right.$$

$$\left.+\int_0^t\mathrm{P}\{|\xi_\varepsilon(s)|>N\}\,ds+\int_0^t\mathrm{P}\{|\bar{\xi}(s)|>N\}\,ds\right)\le$$

$$\le\delta_1+\frac{C}{N^2}+C\int_0^t\mathrm{P}\{|\xi_\varepsilon(s)-\bar{\xi}(s)|>\delta_2\}\,ds.$$

Since $\mathrm{P}-\lim\limits_{\varepsilon\to0}\xi_\varepsilon(s)=\bar{\xi}(s)$, $\delta_1>0$ and $N>0$ are arbitrary, then $\lim\limits_{\varepsilon\to0}I_1=0$.

The process $\bar{\xi}(s)$ is continuous and function $d(t,x)$ satisfies the conditions of lemma. Therefore $\lim\limits_{\varepsilon\to0}I_2=0$ and

$$\lim_{\varepsilon\to0}\int_0^t d(s/\varepsilon^k,\xi_\varepsilon(s))\,ds=\int_0^t\bar{d}(\bar{\xi}(s))\,ds \tag{12}$$

in law (because the distributions of $\xi_{\varepsilon_m}(t),\zeta_{\varepsilon_m}(t)$ coincide with distributions of stochastic processes $\tilde{\xi}_{\varepsilon_m}(t),\tilde{\zeta}_{\varepsilon_m}(t)$ and in fact we have proved that $\mathrm{P}-\lim_{\varepsilon_m\to0}\int_0^t d(s/\varepsilon_m^k,\xi_{\varepsilon_m}(s))\,ds=\int_0^t\bar{d}(\bar{\xi}(s))\,ds$).

Let us consider the case $k_1 = 2k_2 = 2k_3$. From (7) we obtain

$$\xi_\varepsilon(t) = \xi_0 + \int_0^t f(s/\varepsilon^k, \xi_\varepsilon(s))\, ds + \zeta_\varepsilon(t),$$

where martingale $\zeta_\varepsilon(t)$ has a matrix characteristic

$$\langle \zeta_\varepsilon, \zeta_\varepsilon \rangle(t) = \int_0^t g(s/\varepsilon^k, \xi_\varepsilon(s)) g^*(s/\varepsilon^k, \xi_\varepsilon(s))\, ds +$$

$$+ \int_0^t \int_{\mathrm{R}^d} q(s/\varepsilon^k, \xi_\varepsilon(s), y) q^*(s/\varepsilon^k, \xi_\varepsilon(s), y)\, \Pi(dy) ds.$$

Using the lemma it is easy to show that $\mathrm{P}-\lim_{\varepsilon\to 0} \int_0^t f(s/\varepsilon^k, \xi_\varepsilon(s))\, ds = \int_0^t \bar{f}(\bar{\xi}(s))\, ds$, and $\mathrm{P}-\lim_{\varepsilon\to 0} \langle \zeta_\varepsilon, \zeta_\varepsilon \rangle(t) = \int_0^t \bar{B}(\bar{\xi}(s))\, ds$. Hence $\bar{\zeta}(t)$ is a vector-valued continuous square integrable martingale with matrix characteristic $\langle \bar{\zeta}, \bar{\zeta} \rangle(t) = \int_0^t \bar{B}(\bar{\xi}(s))\, ds$. It follows from [5] that there exist a d-dimensional Wiener process $\bar{w}(t)$ such that $\bar{\zeta}(t) = \int_0^t \bar{\sigma}(\bar{\xi}(s))\, d\bar{w}(s)$, where $\bar{\sigma}(x)\bar{\sigma}^*(x) = \bar{B}(x)$. Therefore the process $\bar{\xi}(t)$ is the solution of stochastic differential equation

$$\bar{\xi}(t) = \xi_0 + \int_0^t \bar{f}(\bar{\xi}(s))\, ds + \int_0^t \bar{\sigma}(\bar{\xi}(s))\, d\bar{w}(s). \tag{13}$$

From condition 4) and [6] it follows that the equation (13) has unique weak solution. Hence for any sequence $\varepsilon_m \to 0$ the stochastic process $\xi_{\varepsilon_m}(t)$ converges in probability to the solution $\bar{\xi}(t)$ of equation (13). From this and (12) we have proof of statement 1) of theorem.

When $k < k_1$ the boundedness of $f(t,x)$ implies that $E\left|\int_0^t f(s/\varepsilon^k, \xi_\varepsilon(s))\, ds\right| \le C$, therefore the second term in the right side of (5) converges to 0 in probability, as $\varepsilon \to 0$, and we obtain the first statement in 2). From boundedness of $g(t,x)$ and $q(t,x,y)$ we obtain that either first or second term in the right side of (11) converges to 0 in probability (respectively to the cases $k < 2k_2$ or $k < 2k_3$) as $\varepsilon \to 0$. Then we can complete the proof of the statement 2) of the theorem as the proof of statement 1). Theorem is proved.

Bibliography

1. Deelstra, G. and Delbaen, F., *Long-term returns in stochastic interest rate models: different convergence results,* Applied Stoch. Models and Data Analysis, 13, (1998), 401-407.

2. Deelstra, G. and Delbaen, F. *Long-term returns in stochastic interest rate models,* Insurance: Mathematics and Economics, 17, (1995), 163-169.

3. Skorokhod, A.V., *Studies in the Theory of Random Processes,* Addison-Wesley, (1965).

4. Gikhman, I.I., Skorokhod, A.V., *The Theory of Stochastic Processes,* v.I, Springer-Verlag, Berlin, (1974).
5. Ikeda, N. and Watanabe, S., *Stochastic Differential Equations and Diffusion Processes,* North Holland, Amsterdam and Kadansha, Tokyo, (1981).
6. Strook, D.W. and Varadhan, S.R.S., *Multidimensional Diffusion Processes,* Springer-Verlag, Berlin, (1979).

Department of Probability Theory and Mathematical Statistics, Kyiv National Taras Shevchenko University, Kyiv, Ukraine
E-mail: odb@mechmat.univ.kiev.ua
Department of Mathematics and Computer Science, San Jose University, San Jose, CA 95192, USA
E-mail: ifea@mathcs.sjsu.edu

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.37-44

ILYA R. BUGAYAN, MEZHLUN A. SUMBATYAN
AND ISABELLA HUBER

ABOUT INFLUENCE OF DYNAMICS OF MANUFACTURE ON STRUCTURE OF EMPLOYMENT OF THE POPULATION

Now we are the witnesses of structural modifications of the forms of employment in world scales.

The factor of labor plays very important role, as it determines a relation and dynamics of the forms of employment.

The labor function can be presented as: $L = f(E, Se, U, D)$, where E - paid employment, Se - self-employment, U - unemployment, D - employment in personal home facilities- its factors. The degree index under each factor determines a measure of the given form of employment, and function a relation of the forms of employment: $L = A_l \cdot E^{\beta_1} \cdot Se^{\beta_2} \cdot U^{\beta_3} \cdot D^{\beta_4}$, $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 1$.

The passage of production function (PF) of factor of elasticity under the factor the capital through the critical value $\alpha = 0,5$ both employers, and the hired workers, lose economic interest to the further development of manufacture; the process of development of self-employment on a basis and in an orb of creation both application of new and information technology, transition of business in virtual area is very fast.

2000 *Mathematics Subject Classifications.* 91B40.

Key words and phrases. Factor of labor, dynamics of the forms of employment, labor function, production function.

1. Introduction

The forms of employment include a paid employment, self-employment, employment in personal home facilities and unemployment. The degree of optimization of a relation of the enumerated forms of employment determines a level of a production efficiency of the public boons and rendered services as in frameworks separately of taken state, and all global community as a whole. Now we can be the witnesses of structural modifications of the forms of employment in world scales. To the extremity of 20 century

the business in the developed countries began all more brightly to exhibit itself as the dominating factor of manufacture, in others by those continues to remain the capital, in the third land : [2].

The additional motives and reasons for an amplification of the tendency to prevalence of business are connected to development of the global market calling origin of the states occupied mainly by the businessmen, and, as a corollary it, - states - paid workers, states - unemployed and countries, which do not discover to themselves places in the international division of labor or leave from global community (Iraq, Cuba). All this is a corollary of non-simultaneity of passage in the different countries of dominant properties from one factor of manufacture to another.

The factor labor plays the very important role, as it determines a relation and dynamics of the forms of employment. One forms of employment are concentrated in one countries, for example, modern self-employment on the basis of new and information technology in USA, Europe, and paid employment, personal part-time farm - in other countries.

Marked above optimization of structure of the factor labor in frameworks separately of taken state allows to raise effectiveness of use and remaining of the factors of manufacture - business, capital, land, that, in turn, reduces in economic growth.

The singularity of Russia consists in origin of variety of the forms of the market. It was by a corollary that in different territories of Russia as dominant the various factors of manufacture appear. On Northern Caucasus - labor, in capitals (Moscow, St.-Petersburg) - capital, places new and information technology or introduced, here and there land and everywhere is spread intermediary business. For this reason in different territories of Russia the relation of the every possible forms of employment has striking differences far from optimum for the given region of a level.

The labor function can be presented as: $L = f(E, Se, U, D)$, where E - paid employment, Se - self-employment, U - unemployment, D - employment in personal home facilities- its factors. The degree index under each factor determines a measure of the given form of employment, and function a relation of the forms of employment: $L = A_l \cdot E^{\beta_1} \cdot Se^{\beta_2} \cdot U^{\beta_3} \cdot D^{\beta_4}$, $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 1$.

The structural modifications of a relation of the forms of employment influence process of manufacture and consumption of the public boons. Thus the structure of employment influences effectiveness of use of the factors of manufacture, that in turn testifies to necessity of regulating of streams of the forms of employment and optimization of their relation, which is possible on the basis of the analysis of labor function on each of the factors.

At passage of dominant properties from one factor of manufacture to another the labor function varies also, i.e. the relation of the forms of employment and unemployment in the market of a transactions varies.

Here we shall be limited to detailed reviewing of correlation only of two of four factors of manufacture (capital and labor) - forms, appropriate to them, of employment (paid employment and employer).

Let at the moment t volume of the capital is equal K, and amount of the paid workers necessary for reduction them in an operation equal L. Then the production function $F(K, L)$ shows volume of output F depending on the factors K and L for any phase of time (for example for one year): $F = F(K, L)$.

Let parameter $0 < s < 1$ characterizes, what part again of created yield goes on the investment in manufacture, and parameter $0 < \mu < 1$, what part of a fixed capital (i.e. capital) constantly leaves in view of amortization. Then for Δt the modification of the capital is equal:

$$\Delta K = (sF(K, L) - \mu K)\Delta t. \tag{1}$$

If Δt aspires to zero, we come to known classical model of economic growth Ramsey-Solow [1]:

$$\frac{dK}{dt} = s_1 F(K, L) - \mu K. \tag{2}$$

Is proved, that if $t \to +\infty$ accumulation of the capital at a constant process engineering can not exceed some marginal level: $K(t) \to K(\infty)$ at $t \to +\infty$. Then on the outflow of the large time interval $dK/dt \to 0$, and we come to a limiting relation $sF = \mu K$. If for simplicity to accept production function (PF) as Cobb-Douglas (basic outcomes of work remain fair for arbitrary of neoclassical function, homogeneous degree 1), it is equivalent to equality

$$sAK^{\alpha}L^{\beta} = \mu K, \quad (\alpha + \beta = 1) \sim K = L\left(\frac{sA}{\mu}\right)^{1/\beta}. \tag{3}$$

In 1961 Phelps investigated optimum structure of manufacture, proceeding from a condition of a maxim of consumption (see [2]). As the consumption is determined by a part of the produced product which has stayed after deductions on the investment in manufacture, this magnitude is equal $F - sF = (1 - s)F$. The condition of a maximization of this magnitude in a case PF of Cobb-Douglas, at constant number of the paid workers, is

$$\max\left[(1 - s)AK^{\alpha}L^{\beta}\right] \sim \max\left[(1 - s)s^{\alpha/(1-\alpha)}\right]. \tag{4}$$

Here we have taken advantage of expression (3) for magnitude K. Obviously, the maxima of function (4) is reached in a point s^*, in which the expression in square brackets is equal to zero, that would give $s^* = \alpha$. But then it

turns out, that the income from the factor of the capital equal as is known, KdF/dK (see [1]), is

$$K\frac{dK}{dt} = \alpha F = s^* F \tag{5}$$

and, hence, coincides with volume of the investments in manufacture. Thus, " the gold rule " accumulation established by Phelps, states, that for a maximization of accumulation, the income from the factor of the capital is necessary for putting in the investment.

We shall estimate social consequences "of a gold rule" for this purpose we shall remark, that the found optimum consumption in a society $(1 - s^*)F = (1 - \alpha)F = \beta F = L(dF/dL)$ coincides with the income of the paid workers. What then is consumption of the employers, which appears equal to zero? Unfortunately, in works neoclassics we do not discover a solution of the given paradox.

2. New sight on "a gold rule" accumulation

The correct approach to structural distribution of an industrial yield F is, that is necessary in an explicit aspect to allocate 3 account components:

a) Investment in manufacture;

b) Consumption of the employers;

c) Consumption of the paid workers:

$$F = s_1 F + s_2 F + (1 - s_1 - s_2)F. \tag{6}$$

Where in the ratio (6) s_1 designates a share of issue, which is invested back in a fixed capital, and s_2 - the share of the produced yield, which is consumed by the employers. Thus the consumption of the paid workers, obviously, is determined in parameter $(1 - s_1 - s_2)$.

It is obvious, the employer aspires to maximize the consumption:

$$\max(s_2 F). \tag{7}$$

Thus dynamics of growth of the capital

$$\frac{dK}{dt} = s_1 F - \mu K \tag{8}$$

determines its limiting magnitude at the large times from the following equality

$$s_1 F = \mu K \sim K = L\left(\frac{s_1 A}{\mu}\right)^{1/(1-\alpha)}. \tag{9}$$

At last, we shall take into account, that the consumption of the paid workers is determined by equality known of the microeconomic theory of firm

$$(1 - s_1 - s_2)F = L\frac{dF}{dL} \sim 1 - s_1 - s_2 = 1 - \alpha \sim s_1 + s_2 = \alpha \qquad (10)$$

which is deduced from rushing the employer to maximize profit of firm [3].

The solution of a task of optimization (7) under conditions (9) and (10) is under construction as follows (PF - function of Cobb-Douglas)

$$\max(s_2 K) \sim \max(s_2 s_1^{\alpha/(1-\alpha)}) \sim \max\left[(\alpha - s_1)s_1^{\alpha/(1-\alpha)}\right] \sim s_1^* = \alpha^2,$$

$$s_2^* = \alpha - \alpha^2. \qquad (11)$$

Now it becomes clear, that the optimum size of the investments is determined in parameter $s_1^* = \alpha^2$, thus the consumption of the employers would be determine in factor $s_K = s_2^* = \alpha - \alpha^2$, and consumption of the paid workers - factor $s_L = 1 - s_1^* - s_2^* = 1 - \alpha$.

3. Social consequences, defined dynamics of manufacture

First of all we shall remark, that always $s_K < s_L$, as $0 < \alpha < 1$. Thus, maximizing the profit, the employer is forced to support the $\partial F/\partial L$ rate of the real salary of the paid workers at such level, that their summarized income $L(\partial F/\partial L) = s_L F$ exceeds the income of the businessman (employer) $s_K F$.

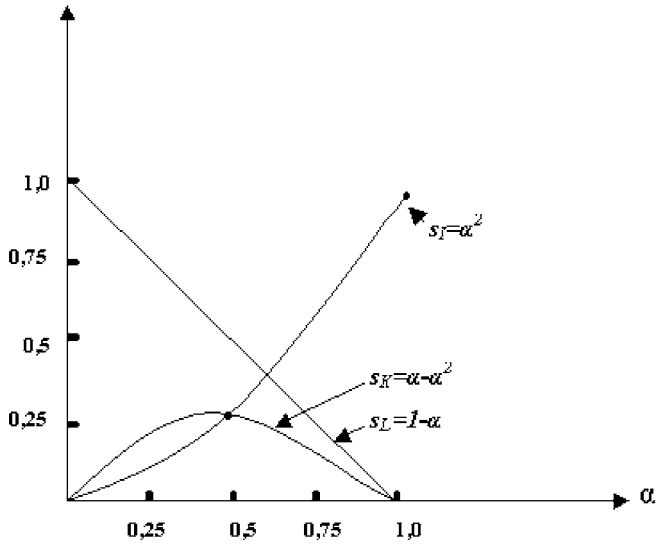


Fig.1

The further conclusions will be based on statistical given about factors of elasticity α_K - under the factor of manufacture the capital and α_L - under the factor of manufacture a labor (for PF of Cobb- Douglas $\alpha_K = \alpha$, $\alpha_L = \beta$). As is known Cobb and Douglas in 1928 for the American economy of the extremity 19 - the beginnings 20 of century have established a value $\alpha_K = \alpha = 0.25$. However later statistical data testify that with growth of the technology of manufactory factor α_K monotonically will increase and, for example, for phase 1960-1995 its average value has reached a level $\alpha_K = 0.404$ (see [4]). Apparently, the growth α_K proceeds and now. Let's trace (see Fig.1) dynamics of magnitudes $s_1 = s_1^* = \alpha^2$ (share of issue going on the investment), s_K and s_L, as functions of parameter $\alpha_K = \alpha$.

It is interesting to note, that in an initial phase of capitalism, where parameter would be still very small the summarized income of the capitalist was much less summarized income of the paid workers (probably it is explained to that employer was a little, and it is a lot of workers). With development of technology (i.e. with growth of parameter α) the share of the income of the capitalist becomes increasing in comparison with the income of the workers, however it required the increasing level of the investments. Let's remark also, that the growth rate of the income capitalist (ds_K/dL) was highest in an initial phase.

In process of growth of parameter α critical the value $\alpha = 0,5$. At passage through this value the relative share of the income of the employers begins to decrease. Developing a technology and manufacture as a whole capitalist, creates that not realizing, for itself objectively unprofitable social conditions increasing percent of the profit he is forced to put in the investment and lesser - on the consumption. It also marks an approximation of the extremity of effective capitalist manufacture. Thus, "grave-digger" of capitalism is not the working class, as supposed K. Marks, and itself capitalist. Developing capitalist manufacture, capitalist objectively leads up it to a level at which his relative income begins to fall, that predetermines crash of the basic motivation of his activity, as organizer of manufacture.

We shall add here such obvious observation that the relative income of the paid workers generally all time decreases, were monotonically decreasing function of parameter α on all interval of its modification $\alpha \in (0, 1)$.

4. Influence on structure of employment and development of new forms of production

From the previous paragraph follows, that the lass interest of the paid workers can not be driving stimulus of manufacture based on a hired transactions, as the relative long of their income constantly decreases. The basic driving force of the given form of manufacture, certainly, is the economic interest of the employers aspiring to a maximization of the profit. However

at reaching a critical value $\alpha = 0,5$ and at the employers the economic interest to development of this form of manufacture is lost, as his share of the national income begins to decrease.

Apparently, the value $\alpha = 0,5$ in the developed countries was reached in middle XX of century. It has reduced that the factor of manufacture the capital began to lose of dominant property, that became the beginning of crash of an industrial society. On change there came new information technology based on new dominant factor - business.

We shall remark, that according to the statistical data, the average value for the Soviet economy for phase with 1960 for 1994 was $\alpha = 0,539$ (see[4]). It means, that in USSR the critical value $\alpha = 0,5$ was reached in 80 years. According to a Fig.1 the further growth of the investments in a heavy industry has ceased to increase the income of manufacture. As was not found of other forms of the industrial rationes alternative to the rationes, based on development of a traditional heavy industry, it has reduced in crash of the Soviet system of manufacture. USSR even has overtaken USA on park of machine tools, however it could not prevent crash of a system.

One more interesting conclusion can be made of established dynamics, if to look at it from a position of the forms of employment of the able-bodied population. Passage at the present stage of factor of elasticity under the factor the capital through the critical value $\alpha = 0,5$ reduces that both participants of process of a hired work - both employers, and the hired workers, - lose economic interest to the further development of manufacture. Thus, from the basic forms of employment - 1) employer, 2) paid employment, 3) self-employment, 4) employment in a personal home facilities and unemployment first two forms of employment gradually remove on the second plan. Increasing percent of the able-bodied population transition from 1, 2 in 3 and 4; especially fast there is a process of development of self- employment on a basis and in an orb of creation both application of new and information technology, transition of business in virtual area. This conclusion confirm carried out in the summer of 1998 at support " of the European Organization of a raise of a standard of living and conditions of work " researches in 16 the European countries, where the level of self-employment makes on the average 13% from common number of the economically - active population. However, by results of the given researches was found out what to become self-occupied 26% from common number of the economically - active population, that is twice above existing level would like. Investigating preferences of the people, which will enter on the market of a transactions during next of five years (till 2003) only 62% would prefer to work as the hired workers, 21% has chosen self-employment, and stayed 17% have not stated any preference concerning their future work. It means, that the self-employment is the attractive form of work for a great many of the people. Also interesting that fact is, that to development and

distribution of new and information process engineerings there is a transition of workplaces in home conditions, in which 31% of the self-occupied workers work completely, and 28% fulfill a part of work of a house. [5]

Bibliography

1. Ashmanov, S.A. *Introduction in Mathematical Economics*, M: Nauka, 1984.
2. Bugayan, I.R. *Macroeconomics*, Rostov-na-Donu, "Fenics", 2000.
3. Phelps, E.S. *Golden rules of economic growth*, New York: Norton,1966.
4. Grebennikov, P.I., Leusskij, A.I., Tarasevich, L.S. *Microeconomics*, S.-P.: Publishing S.-P. GUEF, 1996.
5. Kolemaev, V.A. *Mathematical Economics*, M.: YuNITI, 1998.
6. Harald Bielenski, *Employment Options of the Future: High Demand for New Jobs in Europe - High Interest in Non-Standard Work-Forms.* Munich, February 1999.

Institute of Informatization an Control, Economical University of Rostov-on-Don, Russia
E-mail: mailto:irb@appleclub.donpac.ru,
Institute of Statistics and Mathematical Economical Theory, University of Karlsruhe, Germany
E-mail: E-mail: huber@lsoe.uni-karlsruhe.de

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.45-52

OLEKSANDR I. CHERNYAK

THE SAMPLING STRATEGY FOR BANKING SURVEY IN UKRAINE

Through different methods of survey sampling were estimate mean and total capital of commercial banks of Ukraine in 2000. There are 163 comercial banks in Ukraine. The simple random sampling (mean per bank, ratio estimate, regression estimate) and the stratified random sampling (mean per bank, separate and combined ratio estimates, separate and combined regression estimates) were used. The best result was given by stratified sampling (mean per bank).

2000 *Mathematics Subject Classifications*. 62D05.

Key words and phrases: sampling survey, stratified sampling, estimate of population mean, capital of commercial banks.

We shall estimate the mean (μ) and the total capital (τ) of commercial banks of Ukraine by state on 1.01.2000. There are $N = 163$ commercial banks in Ukraine. All information we will take in journal "Financial Risks", 2000, N1 (21). The simple random sampling with sample size n=10 gave such results (for ordinal numbers):

65 – AB "ELITA" – 20, 73 mln. UAH – y_1,
07 – AB "CITYBANK UKRAINA" – 97,96 mln. UAH – y_2,
26 – AB "ZUKB" – 58, 07 mln. UAH – y_3,
119 – AB "ROSTOK BANK" – 10,72 mln. UAH – y_4,
67 – AB "INTEGRAL" – 19, 25 mln. UAH – y_5,
60 – AB "TRANSBANK" – 23,33 mln. UAH – y_6,
10 – AB "SOSETE ZHENERAL UKRAINA" – 41,02 mln. UAH – y_7,
78 – AB "UNEX" – 17,22 mln. UAH – y_8,
12 – AB "INGBANK UKRAINA" – 32,83 mln. UAH – y_9,
35 – AKB "KYIV" – 34,52 mln. UAH – y_{10}.

$\sum_{i=1}^{10} y_i$ = 355,65 mln. UAH, $\overline{y} = {}^{355,65}\!/_{10}$ = 35,565 mln. UAH,

$\widehat{\tau} = N \cdot \overline{y} = 163 \cdot 35,565 =$ 5797,095 mln. UAH.

True value μ= 35,686 mln. UAH, τ=5816,93 mln. UAH. We are got prettily exact estimations.

Now will use stratified random sampling for estimation a mean and total capital of commercial banks of Ukraine in 2000. We divided the population into 3 strata in accordance with rating and capital. First stratum is the banks with capital more 100 mln. UAH (ratings: 1-9); second stratum is the banks with capital from 25 mln. UAH to 100 mln. UAH (ratings: 10-52). Third stratum is the banks with capital less 25 mln. UAH (ratings: 53-163). We use the following notation:

L - number of strata,
N_k - stratum size, $k = 1, ..., L$,
$W_k = \frac{N_k}{N}$ - stratum weight.
$N = 163,\ \ L = 3;\ \ N_1 = 9,\ N_2 = 43;\ \ N_3 = 111.$
Let
$n = 10,\ n_1 = 1,\ \ n_2 = 3,\ n_3 = 6,$
$W_1 = 0,055,\ \ W_2 = 0,264,\ \ W_3 = 0,681.$
From first stratum
56 – KAB "SLOVYANSKYY" – 232,16 mln. UAH.
From second stratum
11 – AB "BROKBIZNESBANK" – 94,93 mln. UAH,
45 – AKB "PIVDENKOMBANK" – 27,62 mln. UAH,
37 – AKB"PROMAYSLOVO-FINANSOVYY BANK" – 30,30 mln. UAH.
From third stratum
64 – AB "ENERGOBANK" – 22,87 mln. UAH,
119 – AB "ROSTOK-BANK" – 10,72 mln. UAH,
67 – AKB "INTERBANK" – 20,58 mln. UAH,
78 – AKB "PRYCHORNOMORYA" – 17,96 mln. UAH,
157 – AKB "ODESSA-BANK" – 5,70 mln. UAH,
153 – AKB "SLOBOZHANSCHINA" – 6,56 mln. UAH.
We use such estimators

$$y_{st} = \sum_{k=1}^{L} W_k \overline{y}_k,\ \ \widehat{\tau}_{st} = N\overline{y}_{st} = \sum_{k=1}^{L} N_k \overline{y}_k.$$

$$\overline{y_1} = 232,16,\ \ \overline{y_2} = 50,95,\ \ \overline{y_3} = 14,06,$$

$\overline{y}_{st} = 0,055 \cdot 232,16 + 0,264 \cdot 50,95 + 0,681 \cdot 14,06 = 35,792$ mln. UAH,
$\widehat{\tau} = N \cdot \overline{y}_{st} = 5834,09$ mln. UAH.

So, stratified sampling gave few better estimate, than the mean of simple random sampling.

$$|\overline{y}_{st} - \mu| = 0,106 \text{ mln. UAH}, \quad |\overline{y} - \mu| = 0,121 \text{ mln. UAH},$$

$$\left|\widehat{\tau}_{st} - \tau\right| = 17,16 \text{ mln. UAH}, \quad \left|\widehat{\tau} - \tau\right| = 19,88 \text{ mln. UAH}.$$

Now we use the ratio estimates and give information about Ukraine banks capital by state on 1.01.1999 year. So be it given for 2000 y. is Y_i and for 1999 - X_i. $\tau_x = 4278,71,\ \mu_x = 26,25$.

Ratio estimators are following

$$\widehat{\mu}_R = \widehat{R} \cdot \mu_x = \frac{\overline{y}}{\overline{x}} \cdot \mu_x, \quad \widehat{\tau}_R = N \cdot \widehat{\mu}_x = \widehat{R} \cdot \tau_x = \frac{\overline{y}}{\overline{x}} \cdot \tau_x.$$

Let $n = 10$.We use the same sample that in simple random sampling.

$$x_1 = 14,90;\ x_2 = 75,76;\ x_3 = 10,38;\ x_4 = 10,14;\ x_5 = 10,63;$$

$$x_6 = 18,31;\ x_7 = 49,40;\ x_8 = 9,24;\ x_9 = 29,40;\ x_{10} = 23,73;$$

$$\overline{x} = 25,189;\ \overline{y} = 35,565;\ \widehat{R} = 1,40;$$

$$\widehat{\tau}_R = 1,40 \cdot 4278,71 = 5990,19\ mln.\ UAH,$$
$$\widehat{\mu}_R = 1,40 \cdot 26,25 = 36,750\ mln.\ UAH,$$

$$\left|\widehat{\mu}_R - \mu_y\right| = 1,064\ mln.\ UAH, \quad \left|\widehat{\tau}_R - \tau_y\right| = 173,26\ mln.\ UAH.$$

In given example ratio estimation gave a worse result, than the mean simple random sampling and the mean of stratified random sampling. This is accounted for by not high correlation dependence of all population Y and population X.

Interestingly to spy, as changed ratings of select banks:
1.01.2000 y. - 66,10, 19, 119, 69, 61, 27, 84, 34, 31.
1.01.1999 y. - 53, 10, 70, 73, 66, 43, 12, 83, 28, 35.
Now we use separate and combined ratio estimators.
Combined estimators:

$$\overline{y}_{st} = \sum_{k=1}^{L} W_k \overline{y}_k, \quad \overline{x}_{st} = \sum_{k=1}^{L} W_k \overline{x}_k, \quad \widehat{\mu}_{R_c} = \frac{\overline{y}_{st}}{\overline{x}_{st}} \cdot \mu_x,$$

with

$$\mu_x = \sum_{k=1}^{L} W_k \mu_{x_k},\ \widehat{\tau}_{R_c} = \frac{\overline{y}_{st}}{\overline{x}_{st}} \cdot \tau_x = N \cdot \widehat{\mu}_{R_c},\ \tau_x = \sum_{k=1}^{L} \tau_{x_k}.$$

We have

$$L = 3; N_1 = 9, N_2 = 43, N_3 = 111;$$

$$n = 10, n_1 = 1, n_2 = 3, n_3 = 6,$$

$$W_1 = 0,055, W_2 = 0,264, W_3 = 0,681.$$

$$\overline{y_1} = 232,16,\ \ \overline{y_2} = 50,95,\ \ \overline{y_3} = 14,06, \overline{y}_{st} = 35,792.$$

$$\begin{array}{l} x_{11} = 152,47\,; x_{21} = 48,19;\ x_{22} = 15,78; \\ x_{23} = 15,87;\ x_{31} = 18,85;\ x_{32} = 10,14; \\ x_{33} = 10,97;\ x_{34} = 9,87;\ x_{35} = 5,61;\ x_{36} = 5,01; \\ \overline{x}_1 = 152,47;\ \overline{x}_2 = 26,61;\ \overline{x}_3 = 10,1. \end{array}$$

$$\overline{x}_{st} = 0,055 \cdot 152,47 + 0,264 \cdot 26,61 + 0,681 \cdot 10,1 = 21,74;$$

$$\mu_x = 26,25,\ \tau_x = 4278,71.$$

$$\widehat{\mu}_{R_c} = \frac{\overline{y}_{st}}{\overline{x}_{st}} \cdot \mu_x = \frac{35,792}{21,74} \cdot 26,25 = 42\ mln.\ UAH,$$

$$\widehat{\tau}_{R_c} = \frac{35,792}{21,74} \cdot 4278,71 \approx 6846\ mln.\ UAH.$$

Separate estimators:

$$\widehat{\mu}_{R_S} = \sum_{k=1}^{L} W_k \cdot \frac{\overline{y}_k}{\overline{x}_k} \cdot \mu_{x_k},$$

where $\overline{y}_k$, $\overline{x_k}$- sample mean in k-th stratum, μ_{x_k}- true mean in k-th stratum,

$$\mu_{x_k} = \frac{1}{N_k} \sum_{i=1}^{N_k} X_{ki}, \quad \widehat{\tau}_{R_s} = \sum_{k=1}^{L} \frac{\overline{y}_k}{\overline{x}_k} \cdot \tau_{x_k} = N\widehat{\mu}_{R_s},$$

with

$$\tau_{x_k} = N_k \cdot \mu_{x_k} = \sum_{i=1}^{N_k} X_{ki}.$$

$$\begin{array}{l} \tau_{x_1} = 2288,11\,,\ \mu_{x_1} = 254,23; \\ \tau_{x_2} = 1202,68\,,\ \mu_{x_2} = 27,97; \\ \tau_{x_3} = 787,92\,,\ \mu_{x_3} = 7,09. \end{array}$$

Then

$$\widehat{\mu}_{R_s} = 0,055 \cdot \frac{232,16}{152,47} \cdot 254,23 + 0,264 \cdot \frac{50,95}{26,61} \cdot 27,97 +$$

$$+0,681 \cdot \frac{14,06}{10,1} \cdot 7,09 = 41 \; mln. \; UAH,$$

$$\widehat{\tau}_{R_s} = N \cdot \widehat{\mu}_{R_s} = 41 \cdot 163 = 6683 \; mln. \; UAH.$$

These estimators gave the considerably worse results, than the mean of simple random sampling, the mean of stratified sampling and, even, simple ratio estimation. The separate estimation was better, than combined.

Now we use the regression estimators.

$$\widehat{\mu}_l = \overline{y} + \widehat{B}\left(\mu_x - \overline{x}\right), \quad \widehat{\tau}_l = N\widehat{\mu}_l, \quad \widehat{B} = \frac{\sum\limits_{i=1}^{n}\left(y_i - \overline{y}\right)\left(x_i - \overline{x}\right)}{\sum\limits_{i=1}^{n}\left(x_i - \overline{x}\right)^2}.$$

We have

$$\sum_{i=1}^{10}\left(x_i - \overline{x}\right)^2 = 4228,73, \qquad \sum_{i=1}^{10}\left(y_i - \overline{y}\right)\left(x_i - \overline{x}\right) = 4074,96,$$

$$\widehat{B} = \frac{4074,96}{4228,73} = 0,96,$$

$$\widehat{\mu}_l = 35,565 + 0,96\left(26,25 - 25,189\right) = 36,581 \; mln. \; UAH$$

$$\widehat{\tau}_l = 5962,70 \; mln. \; UAH,$$

$$\left|\widehat{\mu}_l - \mu_y\right| = 0,895 \; mln. \; UAH, \quad \left|\widehat{\tau}_l - \tau_y\right| = 145,77 \; mln. \; UAH.$$

So, the regression estimators gave a more exact result, than ratio estimators, than separate and combined ratio estimators for stratified sampling, but under this the mean of simple random sampling and the mean of stratified sampling gave a more exact result.

Finally we use separate and combined regression estimators.

Separate estimators:

$$\widehat{\mu}_{l_s} = \sum_{k=1}^{L} W_k \widehat{\mu}_{l_k},$$

with

$$\widehat{\mu}_{l_k} = \overline{y}_k + \widehat{B}_k \left(\mu_{x_k} - \overline{x}_k\right), \quad \widehat{B}_k = \frac{\sum\limits_{i=1}^{n_k} \left(y_{ki} - \overline{y}_k\right)\left(x_{ki} - \overline{x}_k\right)}{\sum\limits_{i=1}^{n_k} \left(x_{ki} - \overline{x}_k\right)^2},$$

$$\widehat{\tau}_{l_s} = N\widehat{\mu}_{l_s} = \sum_{k=1}^{L} N_k \widehat{\mu}_{l_k}.$$

$$L = 3; n_1 = 1, n_2 = 3,\ n_3 = 6,$$

$$W_1 = 0,055; W_2 = 0,264; W_3 = 0,681.$$

$$\overline{x_1} = 152,47;\ \ \overline{x_2} = 26,61;\ \ \overline{x_3} = 10,1.$$

$$\mu_{x_1} = 254,23;\ \ \mu_{x_2} = 27,97;\ \ \mu_{x_3} = 7,09.$$

$$\overline{y_1} = 232,16;\ \ \overline{y_2} = 50,95;\ \ \overline{y_3} = 14,06.$$

$$\overline{x}_{st} = 21,74;\ \ \overline{y}_{st} = 35,792;\ \ \mu_x = 26,25.$$

$$\widehat{B}_1 = 0;\ \ \widehat{B}_2 = 2,04;\ \ \widehat{B}_3 = 0,44.$$

$$\widehat{\mu}_{l_1} = 232,16;\ \ \widehat{\mu}_{l_2} = 53,72;\ \ \widehat{\mu}_{l_3} = 12,74;$$

$$\widehat{\mu}_{l_s} = 0,055 \cdot 232,16 + 0,264 \cdot 53,72 + 0,681 \cdot 12,74 = 35,527\ mln.\ UAH.$$

$$\widehat{\tau}_{l_s} = 5790,90\ mln.\ UAH.$$

Combined estimators:

$$\widehat{\mu}_{l_c} = \overline{y}_{st} + \widehat{B}_c \left(\mu_x - \overline{x}_{st}\right), \quad \widehat{B}_c = \frac{\sum\limits_{k=1}^{L} \frac{W_k^2(1-f_k)}{n_k(n_k-1)} \sum\limits_{k=1}^{n_k} \left(y_{ki} - \overline{y}\right)\left(x_{ki} - \overline{x}\right)}{\sum\limits_{k=1}^{L} \frac{W_k^2(1-f_k)}{n_k(n_k-1)} \sum\limits_{k=1}^{n_k} \left(x_{ki} - \overline{x}\right)^2}$$

$$\widehat{\tau}_{l_c} = N \cdot \widehat{\mu}_{l_c}, \quad \widehat{B}_c = 1,53;$$

$$\widehat{\mu}_{l_c} = 35,792 + 1,53\,(26,5 - 21,74) = 40,541\ mln.\ UAH,$$

$$\widehat{\tau}_{l_c} = 6608,18\ mln.\ UAH,$$

$$\left|\widehat{\mu}_{l_s} - \mu_y\right| = 0,159\ mln.\ UAH,$$

$$\left|\widehat{\tau}_{l_s} - \tau_y\right| = 26,03\ mln.\ UAH,$$

$$\left|\widehat{\mu}_{l_c} - \mu_y\right| = 4,855\ mln.\ UAH,$$

$$\left|\widehat{\tau}_{l_c} - \tau_y\right| = 791,25\ mln.\ UAH.$$

As visibly, separate estimators gave a considerably better result.
On completion, will write out all of estimators.

1. $\overline{y}_{st}$=35,792 mln.UAH, $|\overline{y}_{st} - \mu_y| = 0,106$ mln.UAH.
   $\widehat{\tau}_{st}$ = 5834,09 mln.UAH, $\left|\widehat{\tau}_{st} - \tau_y\right| = 17,16$ mln.UAH.
2. $\overline{y}$=35,565 mln.UAH, $|\overline{y} - \mu_y| = 0,121$ mln.UAH.
   $\widehat{\tau}$ = 5797,095 mln.UAH, $\left|\widehat{\tau} - \tau_y\right| = 19,88$ mln.UAH.
3. $\widehat{\mu}_{l_s}$=35,527 mln.UAH, $\left|\widehat{\mu}_{ls} - \mu_y\right| = 0,159$ mln.UAH.
   $\widehat{\tau}_{l_s}$ = 5790,90 mln.UAH, $\left|\widehat{\tau}_{st} - \tau_y\right| = 26,03$ mln.UAH.
4. $\widehat{\mu}_l$=36,581 mln.UAH, $\left|\widehat{\mu}_l - \mu_y\right| = 0,895$ mln.UAH.
   $\widehat{\tau}_l$ = 5962,70 mln.UAH, $\left|\widehat{\tau}_l - \tau_y\right| = 145,77$ mln.UAH.
5. $\widehat{\mu}_R$=36,750 mln.UAH, $\left|\widehat{\mu}_R - \mu_y\right| = 1,064$ mln.UAH.
   $\widehat{\tau}_R$ = 5990,19 mln.UAH, $\left|\widehat{\tau}_R - \tau_y\right| = 173,26$ mln.UAH.
6. $\widehat{\mu}_{l_c}$=40,541 mln.UAH, $\left|\widehat{\mu}_{l_c} - \mu_y\right| = 4,855$ mln.UAH.
   $\widehat{\tau}_{l_c}$ = 6608,18 mln.UAH, $\left|\widehat{\tau}_{l_c} - \tau_y\right| = 791,25$ mln.UAH.
7. $\widehat{\mu}_{R_s}$=41 mln.UAH, $\left|\widehat{\mu}_{R_s} - \mu_y\right| = 5,314$ mln.UAH.
   $\widehat{\tau}_{R_s}$ = 6683 mln.UAH, $\left|\widehat{\tau}_{R_s} - \tau_y\right| = 856,07$ mln.UAH.
8. $\widehat{\mu}_{R_c}$=42 mln.UAH, $\left|\widehat{\mu}_{R_c} - \mu_y\right| = 6,314$ mln.UAH.
   $\widehat{\tau}_{R_c}$ = 6846 mln.UAH, $\left|\widehat{\tau}_{R_c} - \tau_y\right| = 1029,07$ mln.UAH.

So, in our examples by most exact estimation is $\overline{y}_{st}$ with relative precision $\frac{|\overline{y}_{st}-\mu_y|}{\mu_y} \cdot 100\% = 0,29\%$ (by analogy $\widehat{\tau}_{st}$). Will direct a table of relative estimations precision.

Table 1. Comparison of estimations.

N	Sampling method	Method of estimation	Relative precision	Rating
1.	Simple random	Mean ($\overline{y}$)	0,34%	2
2.	Simple random	Ratio ($\widehat{\mu}_R$)	2,98%	5
3.	Simple random	Regression ($\widehat{\mu}_l$)	2,50%	4
4.	Stratified random	Mean ($\overline{y}_{st}$)	0,29%	1
5.	Stratified random	Separate ratio ($\widehat{\mu}_{R_s}$)	14,89%	7
6.	Stratified random	Combine ratio ($\widehat{\mu}_{R_c}$)	17,68%	8
7.	Stratified random	Separate regression ($\widehat{\mu}_{l_s}$)	0,44%	3
8.	Stratified random	Combine regression($\widehat{\mu}_{l_c}$)	13,60%	6

Department of Economic Cybernetics, Kyiv Taras Shevchenko University, 01033, Kyiv, Ukraine.
E-mail: chernyak@mail.univ.kiev.ua

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.53-73

OLEKSANDR I.CHERNYAK AND ANDRIY V.STAVYTSKYY

STATISTICAL FORECASTING OF THE BALANCE OF PAYMENTS OF UKRAINE

Statistical methods of forecasting of the Balance of Payments are investigated and compared.

2000 *Mathematics Subject Classifications.* 91B82.

Key words and phrases. Forecasting, Balance of Payments, trend, smoothing, Markov chains.

1. Time series: the main definitions

Researcher's task can be formulated very simple: to produce the best possible forecasts using any available information. But there are some problems in collecting information, creating the adequate model, which can describe economic processes correctly, and working with obtained information, which can be neither exact nor correct. While producing forecasts three main ways can be investigated: statistical, econometrical and judgmental. Econometrists try to use only that information, which has some economic sense and the usage of which can be explained by economic theory. Statisticians use any information, which may not concern economy. Experts can produce forecasts using their judgements and information that they intuitively decide to be important for forecasting. The main goal of all ways is creating final model for forecasting.

Sometimes econometrics can't explain and follow great structural changes in economy. For this reason most of econometric models are not capable to produce rather useful forecasts with minor errors.

As far as statistical way is concerned it should be noted that time series analysis is one of its main parts. Nowadays it is popular direction for forecasting. The fact is there are many economic variables, which are represented by time series. It can be mentioned that hundreds of researchers developed new methods of analysing time series.

Judgmental methods are not developed sufficiently, but referring to experts is very often needed, especially in transition economies.

Time series are sequences of numbers that indicate values of some process, which changes in every time period. As usual equal periods are

considered (for example, months, quarters, years). The order of these values is very important; that's why every value has its own index depending on a number of the period, when it is investigated. Time series are widely applied in different branches of human activity. In economics one may use time series for currency rates, stock rates, GDP and so on. In this report we'll try to demonstrate the application of methods, suggested below concerning the Balance of Payments of Ukraine.

For defining economical politics of the state one needs to forecast some articles of Balance of Payments for future periods. We've made such forecast for the most important articles such as "Export of Goods and Services", "Import of Goods and Services", "Export of Goods", "Import of Goods". These articles are the most aggregate and they give a possibility to make a conclusion about the real situation in economy. The forecasts of these articles are used by government to define the main directions of currency politics, tax politics (duties and excises) and so on.

In practice, we should make conclusions about our time series and make a proper forecast, which will be sufficient for economic activity. We are to select a statistical model, maximal error of forecast (that is sufficient for us), then make forecast, and analyse results.

Certainly, the more values we have obtained the easier to us to predict changes in time series. But sometimes there are only few observations available or the process has changed so much that no inheritance is present at all. For example, the main economic parameters of East Europe countries before changing their economies to market type and after are very different. It means that our possessing information about their GDP for 1960, for instance, cannot provide adequate knowledge of GDP for 2000. So in such a situation we can operate only 5-10 last annual variables and 10-20 last quarter variables. Now we'll discuss some methods that can give us sufficient results for the lack of data.

The next methods require more detailed analysis of time series. According to modern statistics time series consists of two main components: determinate and residual. So, the simplest model of time series is:

$$y_t = d_t + r_t,\ t = \overline{1, T}.$$

Determinate components change according to some rules that can be researched or obtained in empirical way. Usually, this component depends on time (t) and some other parameters. So, it shows the influence of some factors, which can be unknown, but their influence is steady for almost all periods of investigations.

A random component can't be predicted, no one can even state for sure in what way it changes.

In economic application the determinate component usually consist of:

1. Trend component (tr)

2. Seasonal component (s)
3. Cyclical component (c)

That's why one may put down the model of the determinate component in such a way:

$$d_t = tr_t + s_t + c_t,\ t = \overline{1, T}.$$

Now let's consider every component in detail.

The analysis of time series starts with extracting **trend component**, which is the biggest (as usual). If it is a trend then you can see it in graphic representation. Usually it goes up or down. The trend can be determined by the following factors: demographic changes, technological changes, demand structure changes and so on. The action of these factors is displayed gradually; therefore researchers prefer describing them with the help of smooth curves, which are possible to set in an analytical form.

Seasonal component shows inherent in the world and human activity recurrence of processes in time. It is frequently present at economic, meteorological and other time series. A seasonal component serves more often as a main source of short-term fluctuations of a time series; therefore its allocation considerably reduces the variation of other components. Seasonal component consists of a sequence of almost repeated cycles.

A classical example of seasonal effect is the break in even sales level of the goods in December before Christmas and New Year's Eve. The seasonal effects are inherent in many spheres of human activity: many kinds of production have a seasonal nature of manufacture; consumption of the goods has also brightly expressed seasonal prevalence.

The main idea of the analysis of the seasonal component consists of transition from comparison of all values of a time series among themselves to comparison of values through the certain period of time. For example, data of December of one year we are to compare with data of the past December, not with other months of the year.

Cyclical component occupies a somewhat intermediate place between determinate and seasonal components of time series. Trend is a smooth change, which is displayed during large time intervals. A seasonal component is a periodic function of time, which can precisely be seen, when its period is much less than a general number of investigations. A cyclical component is usually considered to be smooth but can't be included in a seasonal component. At the same time it is impossible to attribute such component to trend.

So, any time series can be presented in the form:

$$y_t = tr_t + c_t + s_t + r_t,\ t = \overline{1, T}.$$

Sometimes one uses multiplicative model:

$$y_t = tr_t \cdot c_t \cdot s_t \cdot r_t, t = \overline{1,\, T}.$$

On the basis of this model different statistical methods can be built.

While speaking of model one should define the measure of forecasting accuracy of the method. In the practice the following criteria can be used:

$MSE = \frac{1}{n}\sum\limits_t (y_t - \hat{y}_t)^2$– mean squared error for n periods,

$RMSE = \sqrt{\frac{1}{n}\sum\limits_t (y_t - \hat{y}_t)^2}$– root mean squared error for n periods,

$MAD = \frac{1}{n}\sum\limits_t |y_t - \hat{y}_t|$– mean absolute deviation for n periods,

$RMSPE = 100\sqrt{\frac{1}{n}\sum\limits_t \left(\frac{y_t - \hat{y}_t}{y_t}\right)^2}$– root mean squared percent error for n periods,

$MAPE = \frac{100}{n}\sum\limits_t \left|\frac{y_t - \hat{y}_t}{y_t}\right|$– mean absolute percentage error for n periods.

First three criteria are measured in absolute value and that's why they depend on the time series values. Two last criteria are relative and they are common. For economic research the following characteristic of the accuracy is acceptable:

$RMSE$, $MAPE$	Forecasting accuracy
Less than 10%	excellent
10% - 20%	good
20% - 40%	sufficient
40% - 50%	bad
More than 50%	awful

2. Exponential smoothing

For analysing time series there are many methods of data smoothing. First of all, we should mention exponential smoothing. The idea of this method is creating new time series on the basis of original time series with the help of such expression:

$$S_t = \alpha y_t + (1 - \alpha)\, S_{t-1}.$$

The starting value can be the first item of time series or its mean:

$S_1 = y_1$ or $S_t = \bar{y} = \frac{1}{T}\sum\limits_{t=1}^{T} y_t$.There are three ways to choose the smoothing constant. First of all it can be $\alpha = \frac{2}{T+1}$. The second way is to select it from the interval (0; 1) by empirical way. The third way is to select constant, which minimises one of the criteria of the forecasting accuracy.

The forecast is simply the last value of new time series:

$$\hat{y}_{T+1} = \hat{y}_{T+2} = \ldots = S_T.$$

There are also double and triple Brown exponential smoothing. These methods calculate new sequences twice and triple accordingly:

	Double	Triple
Formulas	$S_t^{/} = \alpha y_t + (1-\alpha)\, S_{t-1}^{/}$, $S_t^{//} = \alpha S_t^{/} + (1-\alpha)\, S_{t-1}^{//}$*.	$S_t^{/} = \alpha y_t + (1-\alpha)\, S_{t-1}^{/}$, $S_t^{//} = \alpha S_t^{/} + (1-\alpha)\, S_{t-1}^{//}$, $S_t^{///} = \alpha S_t^{//} + (1-\alpha)\, S_{t-1}^{///}$.
Forecast	$\hat{y}_{T+p} = S_T^{//}, p = 1,\ 2,\ \ldots.$	$\hat{y}_{T+p} = S_T^{///}, p = 1,\ 2,\ \ldots.$
Suggested interval for constant	(0.03, 0.16)	(0.03, 0.11)

We used a $RMSE$ criterion to choose the optimal constant. The dependence between constant α and number of observations is found. The fact is the optimal value of the constant is repeated in some way. Let's look the tables E1-E3. They represent the optimal value of the constant depending of the number of observation T and the length of forecasting horizon. It can be mentioned that there is undoubted relation. The optimal constant with some bias repeats in two periods, but for longer forecasting period.

This finding is similar for all three smoothing methods for all articles of BP. So, now we can predict the optimal constant value beforehand and that's why to obtain the most accurate forecasts (table E4-E6).

Holt's method is a developing of exponential smoothing. It also allows a trend component extracting. By this method one should build 2 new series by the following rules:

$$S_2^{/} = y_2, \quad S_2^{//} = y_2 - y_1,$$

$$S_t^{/} = \alpha y_t + (1-\alpha)\left(S_{t-1}^{/} + S_{t-1}^{//}\right),$$

$$S_t^{//} = \beta\left(S_t^{/} - S_{t-1}^{/}\right) + (1-\beta)\, S_{t-1}^{//}, t = \overline{3,\, T}.$$

The smoothing constants are suggested to select from the interval (0, 1).

The forecast by this model is:

$$\hat{y}_{T+p} = S_T^{/} + pS_T^{//}, \quad p = 1,\ 2,\ \ldots.$$

It should be noted that this model couldn't be improved like exponential smoothing.

The forecasts with the help of this method were made. Let's see the results of forecasting on the table H-1.

3. Additive-seasonal model

This model tries to define trend and seasonal components of time series. Let p is season's cycle, $s_t = s_{t+p}$ for any time period t. For example, for quarter data $p = 4$. Our task is to estimate values s_t knowing time series observations y_t and p.

The trend component is estimated by averaging:

$$\hat{tr_t} = \frac{\left(\frac{y_{t-2}+y_{t-1}+y_t+y_{t+1}}{4} + \frac{y_{t-1}+y_t+y_{t+1}+y_{t+2}}{4}\right)}{2} =$$

$$= \frac{y_{t-2} + 2y_{t-1} + 2y_t + 2y_{t+1} + y_{t+2}}{8}, t = \overline{3, T-2}.$$

Then the estimate of the seasonal component is

$$\hat{s_t} = y_t - \hat{tr_t} = y_t - \frac{y_{t-2} + 2y_{t-1} + 2y_t + 2y_{t+1} + y_{t+2}}{8} =$$

$$= \frac{6y_t - (y_{t-2} + 2y_{t-1} + 2y_{t+1} + y_{t+2})}{8}, t = \overline{3, T-2}$$

Let's define $\bar{s}_t$ is an average of all $\hat{s}_t$ for all periods $t = pk+i$, $k = 0, 1, \ldots,$ $i = \overline{1, p}$. It should be mentioned that $\overline{s_{t+p}} = \overline{s_t}$ for all $t \geq 5$. Now let's define adjusted seasonal mean component

$\overline{s_t}* = \overline{s_t} - \frac{\sum\limits_{i=1}^{p} \overline{s_i}}{p}$, so that $\sum\limits_{t=1}^{p} \overline{s_t}* = 0$.The last step is trend extracting: $tr_t = y_t - \bar{s}_t*$.

Let's suppose time series to have linear trend $tr_t = a_0 + a_1 t$, then one can obtain coefficients

$$\hat{a}_1 = \frac{T\sum\limits_{t=1}^{T} t \cdot tr_t - \sum\limits_{t=1}^{T} t \cdot \sum\limits_{t=1}^{T} tr_t}{T\sum\limits_{t=1}^{T} t^2 - \left(\sum\limits_{t=1}^{T} t\right)^2}; \hat{a}_0 = \frac{\sum\limits_{t=1}^{T} tr_t}{T} - \frac{\hat{a}_1}{T}\sum_{t=1}^{T} t.$$

So we can write our model

$$y_t = a_0 + a_1 t + \bar{s}_t*, t = \overline{1, T}.$$

The forecasts for future periods are (table D1):

$$\hat{y}_t = a_0 + a_1 t + \bar{s}_t*, t = T+1, T+2, \ldots.$$

4. Non-standard trend extraction

Many of real economic variables are changing in time in such a way. Many scientists have tried to define all components, but there is no universal method yet. Certainly, if there are a lot of observations, one can built a regression model that is relatively reliable. But very often the regression models are not good enough because they can't provide adequate insight or there are not certain factors for a regression to use. At last, if a researcher has only few observations, especially in a case of seasonal fluctuations, regression gives incorrect results and even can give a wrong direction of variable moving. That's why it is worth using another method of researching the processes with restricted number of observations.

One of these methods is known as Hodrick and Prescott Filter (1980). This method is built on the basis of real economic cycles. Let's suppose $y_t = f(t) + \varepsilon_t$ with y_t real values of the variable for research, t time; ε_t residuals; $f(t)$ any function depends on some parameters including time.

The parameters of the function f are selected to minimise:

$$S = \sum_{t=1}^{T} (y_t - f(t))^2 + \lambda \sum_{t=1}^{T} \Big((f(t+1) - f(t)) - (f(t) - f(t-1))^2 \Big) \to min$$

If only f is a linear function, then the second part of the expression is 0. We will consider only unlinear types of function f.

Using partial derivatives one can obtain necessary conditions for defining minimum of S:

$t = 1 : y_1 = f(1) + \lambda f(1) - 2\lambda f(2) + \lambda f(3)$;

$t = 2 : y_2 = f(2) - 2\lambda f(1) + 5\lambda f(2) - 4\lambda f(3) + \lambda f(4)$;

...

$t = T : y_T = f(T) + \lambda f(T-2) - 2\lambda f(T-1) + \lambda f(T)$.

That's why $\bar{y} = X\bar{f}$, with

$$\bar{y} = (y_1, y_2, \ldots, y_T)^T,$$

$$\bar{f} = (f(1), f(2), \ldots, f(T))^T,$$

and matrix X is as follows:

$$X = \begin{vmatrix} 1+\lambda & -2\lambda & \lambda & 0 & 0 & \ldots & & & \\ -2\lambda & 1+5\lambda & -4\lambda & \lambda & 0 & 0 & \ldots & & \\ \lambda & -4\lambda & 1+6\lambda & -4\lambda & \lambda & 0 & 0 & \ldots & \\ & & & & \ldots & & & & \\ & \ldots & 0 & 0 & \lambda & -4\lambda & 1+6\lambda & 4\lambda & \lambda \\ & & \ldots & 0 & 0 & \lambda & -4\lambda & 1+5\lambda & -2\lambda \\ & & & \ldots & 0 & 0 & \lambda & -2\lambda & 1+\lambda \end{vmatrix}$$

Because of X is a square and symmetric matrix

$$X^T X = XX \;\Rightarrow (XX)^{-1} = X^{-1}X^{-1} \;\Rightarrow$$

$$\Rightarrow \left(X^T X\right)^{-1} X^T = (XX)^{-1} X^T = X^{-1}; \bar{y} = X\bar{f},$$

$$\bar{f} = X^{-1}y = \left(X^T X\right)^{-1} X^T y.$$

Knowing X and $\bar{y}$one can obtain the value of f for every t. After that we can define function f or identify its parameters. The next step is to forecast value f for future period and to make a conclusion about possible value of y_{T+1}.

For annual data one may use $\lambda = 100$ and $\lambda = 1600$ for quarter data. But every macroeconomics time series needs a special value of to define. It should be noted that it isn't necessary to define function f completely, but it's possible to forecast its values with standard methods.

After extracting trend component we may extract seasonal component with standard regression model:

$$s_t = a_0 + a_1 q_1 + a_2 q_2 + a_3 q_3 + \varepsilon_t,\ t = \overline{1,\ T}.$$

Variables q_1, q_2, q_3 are called dummy. This means that, for instance, $q_1 = 1$, if we investigate first quarter of the year, and $q_1 = 0$ otherwise.

The cyclical component can be researched only if we have a lot of observations. We have only 20, so we haven't found it yet.

The last component of time series is a random component. It includes everything that was not observed. The values of this component is usually rather small, so you can neglect it or analyse it as AR(p)-, MA(q)- or ARMA(p,q)-processes.

Now let's have a look at forecasts made by this method (table P1).

5. Modeling of time series subject to changes in regime

Many variables undergo episodes in which the behaviour of the series seems to change quite dramatically. Diagram A1 provides a striking example, which demonstrates great decreasing of the trade volumes. Similar dramatic breaks will be seen if one follows almost any macroeconomic or financial time series for a sufficiently long period. Such apparent changes in the time series process can be the result of different events such as wars, financial panics or significant changes in government policies.

The main idea of the method is dividing all data into groups; each of them represents one economic regime. There are some regimes possible. For each regime its own model is built. The model is based on Markov chains.

Now let's consider the model. Let it be N regimes, y_t is $(n \times 1)$ vector of observed endogenous variables, x_t is $(k \times 1)$ vector of exogenous variables,

$\Im_t = \left(y_t^{/}, y_{t-1}^{/}, \ldots, y_{t-m}^{/}, x_t^{/}, x_{t-1}^{/}, \ldots, x_{t-m}^{/}\right)^T$ contains all observations obtained through date t. If the process is governed by regime $s_t = j$ at date t, then the conditional density of y_t is assumed to be given by

$$f(y_t|s_t = j, x_t, \Im_{t-1}; \alpha), \qquad (*)$$

with α is a vector of parameters characterising the conditional density. We will consider the $AR\,(m)$-process

$$y_t = z_t^{/}\beta_{s_t} + \varepsilon_t,$$

with $\varepsilon_t \sim N\,(0,\, \sigma^2)$, z_t is a vector of explanatory variables that could include lagged values of y.

Let all conditional densities (*) for all regimes are collected in vector η_t. And at last, we assume that conditional density depends only on the current regime s_t and not on past regimes:

$$f\,(y_t|x_t, \Im_{t-1}, s_t = j, \alpha) = \\ = f\,(y_t|x_t, \Im_{t-1}, s_t = j, s_{t-1} = i, s_{t-2} = k, \ldots, \alpha)\,,$$

though this is not really restrictive, as it was shown by many researchers.

So we can assume that transition matrix is satisfied

$$P\,\{s_t = j|s_{t-1} = i, s_{t-2} = k, \ldots, s_1 = z,\, x_t, \Im_{t-1}\} = \\ = P\,\{s_t = j|s_{t-1} = i\} = p_{ij}.$$

We collect all unknown parameters in vector $\theta = \{\alpha,\ P\}$ and our task is to estimate it on the basis of $\Im_T$.

Let $P\,\{s_t = j|\Im_t; \theta\}$ denote the analyst's inference about the value of s_t based on data obtained through date t and based on knowledge of the population parameters θ. This inference takes the form of a conditional probability that the analyst assigns to the possibility that the tth observation was generated by regime j. Collect these conditional probabilities $P\,\{s_t = j|\Im_t; \theta\}$for $j = 1,\, 2,\, \ldots, N$ in a $(N \times 1)$ vector $\hat{\xi}_{t|t}$.

One could also imagine forming forecasts of how likely the process is to be in regime j in period $t + 1$ given observations obtained through date t. Collect these forecasts in a $(N \times 1)$ vector $\hat{\xi}_{t+1|t}$, which is a vector whose jth element represents $P\,\{s_{t+1} = j|\Im_t; \theta\}$.

The optimal inference and forecast for each date t in the sample can be found by iterating on the following pair of equations:

$$\begin{aligned} \hat{\xi}_{t|t} &= \frac{\left(\hat{\xi}_{t|t-1} \otimes \eta_t\right)}{\mathrm{I}^{/}\left(\hat{\xi}_{t|t-1} \otimes \eta_t\right)}, \\ \hat{\xi}_{t+1|t} &= P \cdot \hat{\xi}_{t|t}. \end{aligned} \qquad (**)$$

Here η_t represents the $(N \times 1)$ vector whose jth element is the conditional density in (*), P represents the $(N \times N)$ transition matrix, I represents a $(N \times 1)$ vector of 1s, and the symbol denotes element-by-element multiplication. Given a starting value $\hat{\xi}_{1|0} = N^{-1}\mathrm{I}$ and an assumed value for the population parameter vector θ, one can iterate on (**) for $t = 1, 2, \ldots, T$ to calculate the values of $\hat{\xi}_{t|t}$, $\hat{\xi}_{t+1|t}$ for each date t in the sample.

The log likelihood function $L(\theta)$ for the observed data $\Im_T$ evaluated at the value of θ that was used to perform the iterations can also be calculated as a by-product of this algorithm from

$$L(\theta) = \sum_{t=1}^{T} \ln f\left(y_t | x_t, \Im_{t-1}; \theta\right),$$

with

$$f\left(y_t | x_t, \Im_{t-1}; \theta\right) = 1^{/}\left(\hat{\xi}_{t|t-1} \otimes \eta_t\right).$$

Hamilton (1994) has shown that if the transition probabilities are restricted only by the conditions $p_{ij} \geq 0, (p_{i1} + p_{i2} + \ldots + p_{iN}) = 1$ for all i and j, then the maximum likelihood estimates for the transition probabilities satisfy

$$\hat{p}_{ij} = \frac{\sum_{t=2}^{T} P\left\{s_t = j, s_{t-1} = i | \Im_T; \hat{\theta}\right\}}{\sum_{t=2}^{T} P\left\{s_{t-1} = i | \Im_T; \hat{\theta}\right\}},$$

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{N} \left(y_t - z_t^{/} \hat{\beta}_j\right)^2 P\left\{s_t = j | \Im_T; \hat{\theta}\right\},$$

$$\hat{\beta}_j = \left[\sum_{t=1}^{T} \left[zm_t(j)\right] \left[zm_t(j)\right]^{/}\right]^{-1} \left[\sum_{t=1}^{T} \left[zm_t(j)\right] \left[ym_t(j)\right]^{/}\right],$$

with

$$ym_t(j) = y_t \sqrt{P\left\{s_t = j | \Im_T; \hat{\theta}\right\}},$$

$$zm_t(j) = z_t \sqrt{P\left\{s_t = j | \Im_T; \hat{\theta}\right\}},$$

with $\hat{\theta}$ denotes the full vector of maximum likelihood estimates.

So, one should use iterative algorithm. Assuming starting $\theta^{(0)}$ one can evaluate $\theta^{(1)}$, $\theta^{(2)}$ and so on. See Hamilton (1994) for details.

We used such method for analysing balance of payments of Ukraine. For choosing a number of regimes and lags following criteria was used:

1) unchanging of the coefficients;
2) forecasting error;
3) most likely forecasting error.

For main articles of the balance of payments transition matrix and coefficients of the model were calculated (table C1). Also we obtained the probabilities of each regime for each time period (table C2). Almost for all articles the hypothesis about two regimes was correct. The best models requested one, two or three lagged variables. At last, average forecasts (table C3), current regime (table C4) and most likely scenario for four periods (table C5) were estimated.

6. Conclusions

Table F-1 summarizes the accuracy of forecasts made by described above methods. One can see that for each article of the Balance of Payments the most appropriate method can be found.

Due to so completed analysis of forecasting methods the following problems are to be solved:

1. Suppose that in the period t the process was influenced by some factors, which are not to become more in future. How it is possible to exclude the shock value from time series to produce "cleaner" forecasts using statistical models?

2. Suppose that in the period T the government changes laws in the market. In what way can one predict the changes in time series using knowledge about such changes? How can we predict the best and the worst scenarios of time series development?

3. Suppose that we'd like to use experts to produce more accurate forecasts. We can combine or correct statistical and judgmental forecasts. What role should judgement play and if it possible to automate this work?

The solution of these problems may help a to researcher produce more accurate forecasts.

Table E1. Optimal constant for exponential smoothing
Export of goods and services

n	(0)	(-1)	(-2)	(-3)	(-4)	(-5)	(-6)	(-7)
1	1.040	0.414	-0.142	-0.075	-0.098	0.644	-0.061	1.195
2	-1.125	1.270	-0.086	0.983	-0.026	0.118	1.448	0.375
3	-0.114	1.020	-0.058	0.080	0.809	0.375	0.282	1.175
4	-0.064	1.770	0.216	0.375	0.465	0.408	0.400	2.000

Import of goods and services

n	(0)	(-1)	(-2)	(-3)	(-4)	(-5)	(-6)	(-7)
1	0.534	1.750	1.327	1.584	1.777	1.348	-0.485	-0.855
2	1.330	1.750	1.177	1.691	-0.485	-0.084	1.570	1.247
3	1.180	1.800	-0.485	1.771	1.133	-0.881	-0.710	-0.615
4	-0.485	1.860	-0.300	0.139	0.718	-0.563	0.389	2.000

Export of goods

n	(0)	(-1)	(-2)	(-3)	(-4)	(-5)	(-6)	(-7)
1	1.060	0.384	-0.139	0.706	-0.088	0.374	-0.035	1.111
2	-0.118	1.260	-0.074	0.967	0.021	1.584	1.336	0.333
3	-0.107	1.040	-0.031	0.168	0.887	0.333	0.228	1.294
4	-0.040	0.038	0.799	0.333	0.372	1.220	0.378	2.000

Import of goods

n	(0)	(-1)	(-2)	(-3)	(-4)	(-5)	(-6)	(-7)
1	0.718	1.770	1.233	1.434	1.189	1.367	0.577	0.254
2	1.230	1.670	1.189	1.625	0.577	1.610	1.032	1.331
3	1.190	1.720	0.577	1.731	1.161	1.292	0.380	0.099
4	0.577	1.820	1.160	1.440	0.380	0.197	2.000	2.000

Export of services

n	(0)	(-1)	(-2)	(-3)	(-4)	(-5)	(-6)	(-7)
1	-0.150	0.674	-0.153	0.592	-0.134	1.003	-0.131	1.215
2	-0.150	1.390	1.149	1.010	-0.132	0.009	1.749	0.824
3	-0.140	0.989	-0.133	1.958	0.151	0.824	0.702	0.629
4	-0.130	-0.094	0.001	1.460	0.989	0.629	0.492	2.000

Import of services

n	(0)	(-1)	(-2)	(-3)	(-4)	(-5)	(-6)	(-7)
1	1.620	0.147	1.450	0.103	1.814	1.447	1.027	1.054
2	1.080	0.083	1.943	0.936	1.027	1.286	1.074	1.771
3	2.000	0.413	1.027	1.434	1.074	1.962	0.835	1.386
4	1.030	1.380	1.070	1.960	0.835	1.390	1.930	2.000

Table E2. Optimal constant for double exponential smoothing
Export of goods and services

n	(0)	(-1)	(-2)	(-3)	(-4)	(-5)	(-6)	(-7)
1	0.530	0.612	0.891	0.823	0.000	0.809	0.000	-0.186
2	0.891	1.140	0.000	0.991	0.000	-0.119	1.236	-0.225
3	0.000	1.010	0.000	-0.106	0.909	-0.240	0.446	1.089
4	0.000	0.000	0.332	-0.200	0.600	0.573	1.930	2.000

Import of goods and services

n	(0)	(-1)	(-2)	(-3)	(-4)	(-5)	(-6)	(-7)
1	0.701	1.430	-0.499	1.291	-0.555	1.164	-0.505	1.243
2	-0.509	1.410	-0.568	1.358	-0.530	-0.311	1.349	0.574
3	-0.582	1.460	-0.575	1.502	0.292	0.752	0.805	-1.978
4	-0.608	-0.557	-0.423	0.391	0.856	-1.960	0.624	2.000

Export of goods

n	(0)	(-1)	(-2)	(-3)	(-4)	(-5)	(-6)	(-7)
1	1.030	0.558	0.902	0.839	0.000	0.227	0.000	-0.183
2	0.902	1.130	0.000	0.985	-0.061	0.311	1.170	-0.222
3	0.000	1.020	0.000	0.302	0.934	-0.236	0.390	0.441
4	0.000	0.117	0.903	-0.203	0.518	0.507	0.615	2.000

Import of goods

n	(0)	(-1)	(-2)	(-3)	(-4)	(-5)	(-6)	(-7)
1	0.839	1.450	-0.203	1.213	-0.235	1.175	-0.251	0.449
2	-0.213	1.360	-0.240	1.316	-0.264	1.381	1.015	1.175
3	-0.250	1.390	-0.283	1.456	-0.284	1.154	-0.352	-0.298
4	-0.297	1.510	-0.319	1.230	-0.345	0.439	2.000	2.000

Export of services

n	(0)	(-1)	(-2)	(-3)	(-4)	(-5)	(-6)	(-7)
1	0.890	0.818	0.814	0.725	1.001	1.002	0.000	1.103
2	0.814	1.240	0.913	1.005	0.000	-0.045	1.473	-0.238
3	1.000	0.994	0.000	1.651	0.275	-0.256	0.843	-0.428
4	0.000	1.660	0.013	0.323	0.995	-0.453	0.701	2.000

Import of services

n	(0)	(-1)	(-2)	(-3)	(-4)	(-5)	(-6)	(-7)
1	-0.101	-0.087	-0.079	0.215	1.511	1.234	-0.207	1.027
2	1.040	0.187	0.353	0.967	-0.211	1.137	-0.290	1.331
3	0.289	0.588	-0.207	1.209	-0.296	1.947	1.930	1.257
4	-0.197	1.180	-0.303	1.480	-0.476	-0.685	1.390	2.000

Table E3. Optimal constant for triple exponential smoothing
Export of goods and services

n	(0)	(-1)	(-2)	(-3)	(-4)	(-5)	(-6)	(-7)
1	0.620	0.710	-0.141	0.877	-0.139	0.129	-0.142	1.103
2	-0.132	1.070	-0.132	0.994	-0.108	0.326	1.159	0.620
3	-0.148	1.01	-0.140	0.271	0.940	0.620	0.547	0.604
4	-0.145	-0.037	0.414	0.62	0.677	0.667	0.737	2.000

Import of goods and services

n	(0)	(-1)	(-2)	(-3)	(-4)	(-5)	(-6)	(-7)
1	0.780	1.290	1.171	1.191	1.089	1.107	-0.093	-0.410
2	1.170	1.270	1.089	1.231	-0.093	1.341	-0.432	0.683
3	1.090	1.310	-0.093	1.376	0.422	1.037	0.869	-0.591
4	-0.093	1.410	-1.960	-0.295	-0.700	-0.591	0.730	2.000

Export of goods

n	(0)	(-1)	(-2)	(-3)	(-4)	(-5)	(-6)	(-7)
1	0.588	0.691	-0.141	-0.106	-0.135	0.767	-0.118	1.057
2	-0.130	-0.136	-0.126	0.990	0.148	0.400	1.113	0.594
3	-0.145	1.010	-0.114	0.390	0.963	0.594	0.496	0.556
4	-0.124	0.190	0.936	0.594	0.608	1.080	0.723	2.000

Import of goods

n	(0)	(-1)	(-2)	(-3)	(-4)	(-5)	(-6)	(-7)
1	1.480	1.310	1.120	1.139	1.096	1.114	0.781	0.569
2	1.120	1.240	1.096	1.205	0.781	1.279	1.010	1.119
3	1.100	1.250	0.781	1.333	1.086	1.105	0.686	0.455
4	0.781	1.370	1.090	1.160	0.686	0.576	1.650	2.000

Export of services

n	(0)	(-1)	(-2)	(-3)	(-4)	(-5)	(-6)	(-7)
1	-0.119	0.875	-0.140	0.791	-0.154	1.001	-0.192	1.068
2	-0.138	1.170	0.938	1.003	-0.192	0.110	1.348	0.915
3	-0.157	0.996	-0.192	-0.161	1.112	0.915	0.895	0.815
4	-0.191	-0.183	0.043	0.429	0.996	0.815	0.789	2.000

Import of services

n	(0)	(-1)	(-2)	(-3)	(-4)	(-5)	(-6)	(-7)
1	1.400	0.289	0.223	0.303	0.619	1.157	1.013	1.018
2	0.263	0.274	0.456	0.978	1.013	1.090	1.924	1.215
3	0.389	0.682	1.013	1.137	1.927	1.258	1.675	1.125
4	1.010	1.120	1.930	1.300	1.690	1.190	1.250	2.000

Table E4. Forecast using exponential smoothing

Articles of the BOP	Optimal constant	1-2000	2-2000	3-2000	4-2000	2000
Export of goods and services	1.050	4503	4503	4503	4503	18012
Import of goods and services	1.900	5414	5414	5414	5414	21656
Export of goods	1.050	3502	3502	3502	3502	14008
Import of goods	1.800	4664	4664	4664	4664	18656
Export of services	0.970	995	995	995	995	3980
Import of services	0.400	604	604	604	604	2416

Table E5. Forecast using double exponential smoothing

Articles of the BOP	Optimal constant	1-2000	2-2000	3-2000	4-2000	2000
Export of goods and services	1.100	4565	4565	4565	4565	18260
Import of goods and services	1.560	5246	5246	5246	5246	20984
Export of goods	1.050	3518	3518	3518	3518	14072
Import of goods	1.400	4679	4679	4679	4679	18716
Export of services	0.990	996	996	996	996	3984
Import of services	0.200	539	539	539	539	2156

Table E6. Forecast using triple exponential smoothing

Articles of the BOP	Optimal constant	1-2000	2-2000	3-2000	4-2000	2000
Export of goods and services	1.050	4544	4544	4544	4544	18176
Import of goods and services	1.400	5322	5322	5322	5322	21288
Export of goods	1.050	3536	3536	3536	3536	14144
Import of goods	1.280	4796	4796	4796	4796	19184
Export of services	0.990	995	995	995	995	3980
Import of services	0.650	598	598	598	598	2392

Table H1. Forecast using Holt-Winters method

Articles of the BOP	α	β	1-2000	2-2000	3-2000	4-2000	2000
Export of goods and services	0,4	0,6	4156	4282	4408	4533	17379
Import of goods and services	0,4	0,8	3835	4107	4379	4651	16972
Export of goods	0,5	0,5	3317	3424	3531	3638	13910
Import of goods	0,5	0,6	3374	3557	3740	3923	14594
Export of services	0,4	0,6	953	979	1005	1031	3968
Import of services	0,4	0,5	583	592	600	609	2384

Table D1. Forecast using seasonal additive model

Articles of the BOP	1-2000	2-2000	3-2000	4-2000	2000
Export of goods and services	4477	5034	4946	5166	19623
Import of goods and services	4831	4748	4794	5021	19394
Export of goods	3298	3811	3727	3904	14740
Import of goods	4141	4004	4004	4353	16502
Export of services	1179	1222	1219	1262	4882
Import of services	690	744	790	768	2992

Table P1. Forecast using Hodrick-Prescott Filter

Articles of the BOP	1-2000	2-2000	3-2000	4-2000	2000
Export of goods and services	3575	4018	3838	3929	15360
Import of goods and services	3512	3466	3337	3632	13947
Export of goods	2705	3173	3068	3177	12123
Import of goods	3079	2975	2766	3072	11892
Export of services	881	858	787	772	3298
Import of services	621	653	686	616	2576

Diagram A1. Dynamics of the Balance of Payments of Ukraine

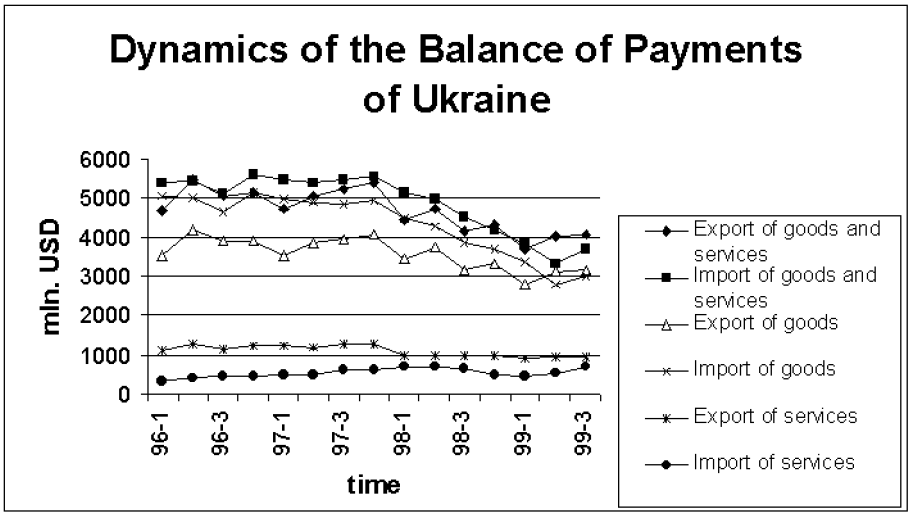


Table C1. Calculation results

Articles of the BOP	Transition matrix P	Coefficients β_{s_t}
Export of goods and services	$\begin{pmatrix} 0.514 & 0.486 \\ 0.228 & 0.772 \end{pmatrix}$	$\begin{pmatrix} 0.036 & 0.783 & 0.171 & 0.763 \\ -0.143 & -0.971 & 1.57 & 0.763 \end{pmatrix}$
Import of goods and services	$\begin{pmatrix} 0.613 & 0.130 & 0.257 \\ 0.489 & 0.159 & 0.352 \\ 0.420 & 0.366 & 0.214 \end{pmatrix}$	$\begin{pmatrix} -0.0638 & 1.084 & 0.171 \\ 0.0762 & -0.569 & 0.852 \\ -0.0001 & 1.198 & -0.626 \end{pmatrix}$
Export of goods	$\begin{pmatrix} 0.701 & 0.299 \\ 0.222 & 0.778 \end{pmatrix}$	$\begin{pmatrix} -0.123 & -0.091 & 1.12 & 0.118 \\ 0.058 & 0.630 & 0.359 & -0.090 \end{pmatrix}$
Import of goods	$\begin{pmatrix} 0.703 & 0.297 \\ 0.705 & 0.295 \end{pmatrix}$	$\begin{pmatrix} -0.019 & 0.859 & 0.565 \\ -0.067 & 1.548 & -1.150 \end{pmatrix}$
Export of services	$\begin{pmatrix} 0.795 & 0.205 \\ 0.250 & 0.750 \end{pmatrix}$	$\begin{pmatrix} -0.116 & -0.002 & 0.079 \\ 0.015 & 0.395 & 0.579 \end{pmatrix}$
Import of services	$\begin{pmatrix} 0.245 & 0.755 \\ 0.749 & 0.251 \end{pmatrix}$	$\begin{pmatrix} 0.061 & 0.401 \\ 0.213 & 0.401 \end{pmatrix}$

Table C2. Probability of the regime for each period

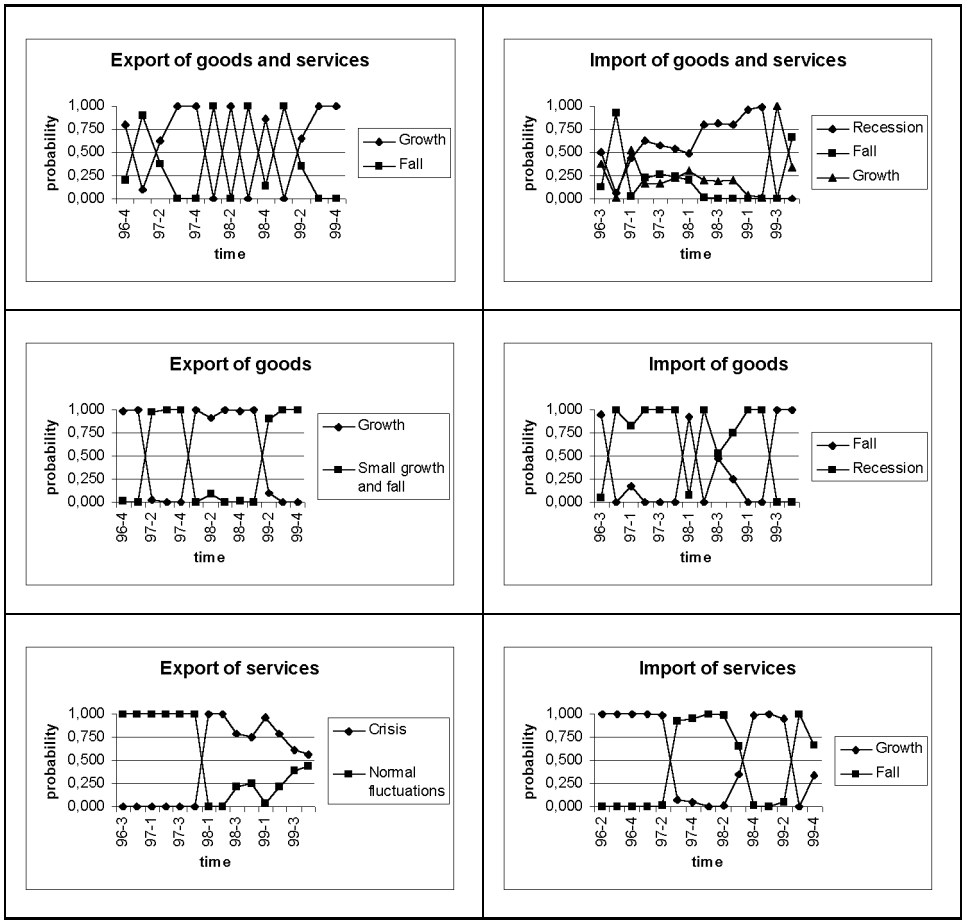


Table C3. Average forecast

Articles of the BOP	2000-1	2000-2	2000-3	2000-4	2000
Export of goods and services	4660	4660	4660	4660	18640
Import of goods and services	4839	4839	4839	4839	19356
Export of goods	3570	3570	3570	3570	14280
Import of goods	4293	4293	4293	4293	17172
Export of services	1011	1013	1021	1016	4061
Import of services	623	602	597	587	2409

Table C4. Current regime

Articles of the BOP	Regime	Probability
Export of goods and services	Up	0.999
Import of goods and services	Small up	0.660
Export of goods	Up and down	0.999
Import of goods	Small down	0.999
Export of services	Crisis	0.568
Import of services	Down	0.665

Table C5. The most likely scenario

Articles of the BOP	2000-1	2000-2	2000-3	2000-4	2000
Export of goods and services	4406	2852	4656	5215	0.157
Import of goods and services	4558	4017	4177	3592	0.074
Export of goods	3463	3624	3716	3782	0.367
Import of goods	3891	2948	3464	2624	0.247
Export of services	951	950	956	953	0.227
Import of services	662	662	662	661	0.209

Table F1. Analyzing of forecasts for 1999

Articles of BOP	1	2	3	4	5	6	7	8
Export of goods and services	16234	14976	19281	16994	16812	16679	17239	19441
Import of goods and services	15237	13297	20637	17699	19064	13938	18059	20729
Export of goods	12463	11731	14313	13239	13615	13564	13412	14888
Import of goods	12945	11702	17664	14174	14148	12705	15057	18583
Export of services	3771	3046	4968	3758	4224	3960	3827	4553
Import of services	2292	2325	2972	1581	2209	2338	3002	2146
Export of goods and services		7,75%	18,77%	4,68%	3,56%	2,74%	6,19%	19,75%
Import of goods and services		12,73%	35,44%	16,16%	25,12%	8,53%	18,52%	36,04%
Export of goods		5,87%	14,84%	6,23%	9,24%	8,83%	7,61%	19,46%
Import of goods		9,60%	36,45%	9,49%	9,29%	1,85%	16,32%	43,55%
Export of services		19,23%	31,74%	0,34%	12,01%	5,01%	1,49%	20,74%
Import of services		1,44%	29,67%	31,02%	3,62%	2,01%	30,98%	6,37%

1 - Real value; 2 - Holt-Winters model; 3 - Additive-seasonal model; 4 - Trend extraction; 5 - Changing regimes, average; 6 - Changing regimes, the most likely; 7 - Linear trend; 8 - Seasonal fluctuations.

Bibliography

1. Hodrick, R. J. and E. Prescott. *"Post-war U. S. Business-Cycles: An Empirical Investigation."* Working Paper. Carnegie-Mellon University, 1980.
2. Hamilton, James D. *Time Series Analysis*. Princeton: Princeton University Press, 1994.

Economic Cybernetics Department, Faculty of Economics, Kyiv National Taras Shevchenko University, Kyiv, Ukraine.
E-mail: chernyak@mail.univ.kiev.ua
Economic Cybernetics Department, Faculty of Economics, Kyiv National Taras Shevchenko University, Kyiv, Ukraine.
E-mail: stavytskyyandriy@ukrpost.net

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.74-87

ERIK DOTZAUER, HENRIK JÖNSSON AND HANS F. RAVN

OPTIMAL UNIT COMMITMENT BY BRANCH-AND-BOUND EXPLOITING DUAL OPTIMALITY CONDITIONS

Within the power industry, the unit commitment problem, in combination with the economic dispatch problem, is a challenging mixed integer nonlinear optimization problem. In this paper, a branch-and-bound algorithm, solving a basic version of the problem, is presented. Applying Lagrangian relaxation will generate a convex but non-smooth dual problem. Lower bounds on the optimal function value are computed from the dual objective function, and branching variables are chosen exploiting the optimality conditions of the dual problem.

AMS 1991 subject classifications. 90B35, 90C11, 90C46, 90C57.

Key words and phrases. Unit commitment, Lagrangian relaxation, branch-and-bound, optimality conditions.

1. Introduction

Since large operational costs are involved, efficient operation of the generating units in the power system is desirable. Finding the optimal production of power for the near future is known as the 'short-term planning' problem, the 'unit commitment and economic dispatch' problem or the 'unit commitment' problem. Typically a time horizon of up to one week is considered. The problem may be characterized as a nonlinear mixed integer optimization problem.

The computations may conceptually be divided into two subproblems; the 'pure' unit commitment problem is the problem to determine when a unit should be producing or not, and the economic dispatch problem is to find the optimal production levels given which units are producing in each time interval. Obviously the two problems are interdependent, i.e. they must be solved simultaneously.

For the solution of the unit commitment and economic dispatch problem, a number of optimization techniques have been suggested and implemented, including priority lists, dynamic programming, branch-and-bound, Lagrangian relaxation and expert systems. Surveys are given in Sen and Kothari (1998) and Sheble' and Fahd (1994). In recent years, methods based on Lagrangian relaxation have become the dominant ones.

In this paper, the branch-and-bound algorithm developed in Dotzauer (2001), solving a basic version of the unit commitment problem, is presented. Applying Lagrangian relaxation will generate a convex but non-smooth dual problem. Lower bounds on the optimal function value are computed from the dual objective function, and branching variables are chosen exploiting the necessary and sufficient optimality conditions of the dual problem.

In Section 2 the problem to be considered is formulated, and in Section 3 an algorithm solving the dual problem is presented. The branch-and-bound algorithm is presented in Section 4. Branching strategies are discussed in Section 5. Section 6 gives some computational results, and finally in Section 7, some conclusions are given.

2. Primal and dual problem statements

Let K be the number of production units and I the number of one-hour time intervals over which the problem is to be solved. Define $p_{i,k}$ as the power production for unit k in time interval i. Moreover, let $u_{i,k}$ be a binary variable indicating if unit k in time interval i is producing or not producing power. If the unit is producing (on), then $u_{i,k}$ is equal to one, and if the unit is not producing (off), then $u_{i,k}$ is zero.

The cost for producing power in a production unit is modeled using a second-order polynomial,

$$c_{i,k} = (\alpha_k^2 (p_{i,k})^2 + \alpha_k^1 p_{i,k} + \alpha_k^0) u_{i,k}, \tag{1}$$

and the start-up cost is modeled as constant, $c_{i,k}^{start} = (1 - u_{i-1,k}) u_{i,k} \gamma_k$. Initial states are given by $u_{0,k}$, $k = 1, ..., K$. The parameters α_k^j, $j = 0, 1, 2$, and $\gamma_k > 0$ are estimated separately for each unit. Further, the expression in equation (1) is assumed to be strictly convex, i.e. $\alpha_k^2 > 0$.

There are restrictions on the production level described as the inequality bounds

$$\underline{p}_k u_{i,k} \leq p_{i,k} \leq \overline{p}_k u_{i,k}. \tag{2}$$

It is assumed that $0 < \underline{p}_k \leq \overline{p}_k$.

The demand and reserve constraints that must be fulfilled in time interval i are

$$\sum_{k=1}^{K} p_{i,k} = p_{i,D}, \tag{3}$$

and

$$\sum_{k=1}^{K} \overline{p}_k u_{i,k} \geq p_{i,R}, \tag{4}$$

respectively. Here $p_{i,D}$ is the power demand and $p_{i,R}$ is the reserve requirement.

To summarize, define the unit commitment problem as the following nonlinear mixed integer mathematical programming problem,

$$\begin{aligned}
\min_{p,u} \quad & \left[\sum_{i=1}^{I}\sum_{k=1}^{K}(\alpha_k^2(p_{i,k})^2+\alpha_k^1 p_{i,k}+\alpha_k^0)u_{i,k}+\sum_{i=1}^{I}\sum_{k=1}^{K}(1-u_{i-1,k})u_{i,k}\gamma_k\right] \\
s.t. \quad & \sum_{k=1}^{K} p_{i,k} = p_{i,D} \\
& \sum_{k=1}^{K} \overline{p}_k u_{i,k} \geq p_{i,R} \\
& \underline{p}_k u_{i,k} \leq p_{i,k} \leq \overline{p}_k u_{i,k} \\
& u_{i,k} \in \{0,1\}.
\end{aligned} \tag{5}$$

Here $p_k = (p_{1,k}, ..., p_{I,k})$, $p = (p_1, ..., p_K)$, $u_k = (u_{1,k}, ..., u_{I,k})$ and $u = (u_1, ..., u_K)$. Problem (5) might be partitioned into two subproblems; the 'pure' unit commitment problem and the economic dispatch problem. To determine the optimal unit commitment is to compute the optimal binary variables $u_{i,k}$. Given a unit commitment, the economic dispatch problem is solved to yield the production for each individual unit. Observe that the unit commitment problem and the economic dispatch problem are interdependent, i.e. they must be solved simultaneously. The formulation (5) of the unit commitment problem is the classical version, which is also the core of any extension and refinement, see Sen and Kothari (1998) and Sheble' and Fahd (1994).

Lagrangian relaxation is performed by introducing multipliers $\lambda = (\lambda_1, ..., \lambda_I)$ and $\mu = (\mu_1, ..., \mu_I)$. Combining these with (3) and (4), respectively, and adding to the objective in (5) gives the relaxed problem

$$\begin{aligned}
\Phi(\lambda,\mu) = \min_{p,u} \quad & \left[\sum_{i=1}^{I}\sum_{k=1}^{K}(\alpha_k^2(p_{i,k})^2+\alpha_k^1 p_{i,k}+\alpha_k^0)u_{i,k}+\right. \\
& +\sum_{i=1}^{I}\sum_{k=1}^{K}(1-u_{i-1,k})u_{i,k}\gamma_k+\sum_{i=1}^{I}\lambda_i(p_{i,D}-\sum_{k=1}^{K}p_{i,k})+ \\
& \left.+\sum_{i=1}^{I}\mu_i(p_{i,R}-\sum_{k=1}^{K}\overline{p}_k u_{i,k})\right] \\
s.t. \quad & \underline{p}_k u_{i,k} \leq p_{i,k} \leq \overline{p}_k u_{i,k} \\
& u_{i,k} \in \{0,1\},
\end{aligned} \tag{6}$$

where $\Phi(\lambda,\mu)$ is the dual objective function. The dual problem is defined as

$$\begin{aligned}
\max_{\lambda,\mu} \quad & [\Phi(\lambda,\mu)] \\
s.t. \quad & \mu \geq 0.
\end{aligned} \tag{7}$$

Problem (5) is called the primal problem. By duality theory, the optimal value of the dual objective is a lower bound on the objective function value

of any feasible solution of the primal problem. When the primal problem (5) is a mixed integer problem, as in our case, a zero duality gap is not guaranteed, i.e. the optimal primal objective may be strictly larger than the optimal dual objective. Further, the dual problem is in general non-smooth, which normally makes the problem difficult to solve.

3. A solution algorithm for the dual problem

Solution strategies for the unit commitment problem (5) based on decomposition and duality were initiated in Muckstadt and Koenig (1977) and are well described in the literature, Sen and Kothari (1998), Sheble' and Fahd (1994). The intention in most of these is to solve the dual problem (7) in order to obtain a near-optimal unit commitment of the primal problem (5). In this section, a solution algorithm by Dotzauer and Ravn (2001) for the dual problem is presented.

The calculations are performed by alternately solving the relaxed problem (6) and updating the dual variables λ and μ. First consider the solution of the relaxed problem. Given a set of dual variables λ and μ, the relaxed problem decomposes into K independent subproblems; one for each production unit k,

$$\begin{aligned}\min_{p_k,u_k} \Bigg[&\sum_{i=1}^{I} (\alpha_k^2 (p_{i,k})^2 + (\alpha_k^1 - \lambda_i) p_{i,k} + \alpha_k^0 - \mu_i \overline{p}_k) u_{i,k} + \\ &+ \sum_{i=1}^{I} (1 - u_{i-1,k}) u_{i,k} \gamma_k \Bigg] \\ &s.t. \quad \underline{p}_k u_{i,k} \le p_{i,k} \le \overline{p}_k u_{i,k} \\ &\qquad u_{i,k} \in \{0,1\}.\end{aligned} \tag{8}$$

A complicating issue is the time-coupling between the binary variables $u_{i,k}$ due to the start-up cost, but this difficulty can be mastered using network or dynamic programming algorithms. Here we apply a dynamic programming solution of (8) which may be described as follows. Let $s_{i,k}$ denote the state of unit k at the beginning of stage (time interval) i, then $s_{i,k}$ may take two values; $s_{i,k} = 1$ if the unit is on in time interval $i-1$, and $s_{i,k} = 0$ if the unit is off in time interval $i-1$. Furthermore, define the *conditional reduced cost* of unit k in time interval i, $c_{i,k}^{crc}(\lambda_i, \mu_i)$, as

$$\begin{aligned}c_{i,k}^{crc}(\lambda_i, \mu_i) = \min_{p_{i,k}} &[\alpha_k^2 (p_{i,k})^2 + (\alpha_k^1 - \lambda_i) p_{i,k} + \alpha_k^0 - \mu_i \overline{p}_k] \\ &s.t. \quad \underline{p}_k \le p_{i,k} \le \overline{p}_k,\end{aligned} \tag{9}$$

where the optimal solution is found as

$$p_{i,k}^o = \begin{cases} \underline{p}_k, & \text{if } \widetilde{p}_{i,k} < \underline{p}_k \\ \widetilde{p}_{i,k}, & \text{if } \underline{p}_k \le \widetilde{p}_{i,k} \le \overline{p}_k \\ \overline{p}_k, & \text{if } \overline{p}_k < \widetilde{p}_{i,k}, \end{cases} \tag{10}$$

where $\widetilde{p}_{i,k} = (\lambda_i - \alpha_k^1)/(2\alpha_k^2)$, giving

$$c_{i,k}^{crc}(\lambda_i, \mu_i) = \alpha_k^2 (p_{i,k}^o)^2 + (\alpha_k^1 - \lambda_i) p_{i,k}^o + \alpha_k^0 - \mu_i \overline{p}_k. \tag{11}$$

Then, defining $V_{i,k}(s_{i,k})$ as the optimal cost-to-go function, the recursive dynamic programming formula is

$$\begin{aligned} V_{i,k}(s_{i,k}) = \min_{u_{i,k}} & \left[c^{crc}_{i,k}(\lambda_i, \mu_i)u_{i,k} + (1 - s_{i,k})u_{i,k}\gamma_k + V_{i+1,k}(s_{i+1,k})\right] \\ & s.t. \quad u_{i,k} \in \{0, 1\}. \end{aligned} \tag{12}$$

The recursion (12) is initialized with $V_{I+1,k}(0) = V_{I+1,k}(1) = 0$. The optimal solution at state $s_{i,k}$ is denoted $u^*_{i,k}(s_{i,k})$. It is understood that $s_{i+1,k} = 0$ if $u^*_{i,k} = 0$ and $s_{i+1,k} = 1$ if $u^*_{i,k} = 1$. It is also understood that $u^*_{i,k} = 0$ and $u^*_{i,k} = 1$ if and only if $p_{i,k} = 0$ and $p_{i,k} = p^o_{i,k}$, respectively.

Now consider the updating of the dual variables. By the fact that the dual problem (7) is non-smooth, methods for non-smooth optimization must be used. An algorithm frequently used for the solution of such problems is the subgradient method, which is a direct generalization of the steepest descent algorithm for unconstrained smooth optimization. In each iteration n of the algorithm, a new set of dual variables $[\lambda, \mu]^{n+1}$ is computed from $[\lambda, \mu]^n$, such that we get an 'improving direction' of the dual objective $\Phi(\lambda, \mu)$. The direction chosen is defined from the subgradient $[g^\lambda, g^\mu]^n$ with elements g^λ_i and g^μ_i given by equation (3) and (4), respectively, i.e.

$$g^\lambda_i = p_{i,D} - \sum_{k=1}^{K} p_{i,k} \tag{13}$$

and

$$g^\mu_i = p_{i,R} - \sum_{k=1}^{K} \overline{p}_k u_{i,k}, \tag{14}$$

where $(p_{i,k}, u_{i,k}) \in \left\{(0, 0), (p^o_{i,k}, 1)\right\}$ is given from (12). Given the step length α^n, each element i in the new set of dual variables $[\lambda, \mu]^{n+1}$ is computed by

$$[\lambda_i, \mu_i]^{n+1} = [\lambda_i, \mu_i]^n + \alpha^n [g^\lambda_i, g^\mu_i]^n. \tag{15}$$

Normally, the subgradient is normed to stabilize the algorithm, see e.g. Shor (1985).

Other commonly used methods for non-smooth optimization are cutting plane methods and bundle methods, Shor (1985). Similar to the subgradient method, also these methods compute a new update $[\lambda, \mu]^{n+1}$ in each iteration, but with the difference that more than one subgradient is taken into account.

The subgradient method, with the step in each iteration given by (15), treats the dual problem (7) as any non-smooth problem, without any consideration of the specific problem structure. The major point in the algorithm by Dotzauer and Ravn (2001) is to improve the algorithm by exploiting the necessary and sufficient optimality conditions of the dual problem. These conditions are derived from the insight that the optimal solution of the relaxed problem (6) might be non-unique at the dual optimum. In fact, given the optimal set of λ and μ, it will be the exception that the solution to (6) is unique.

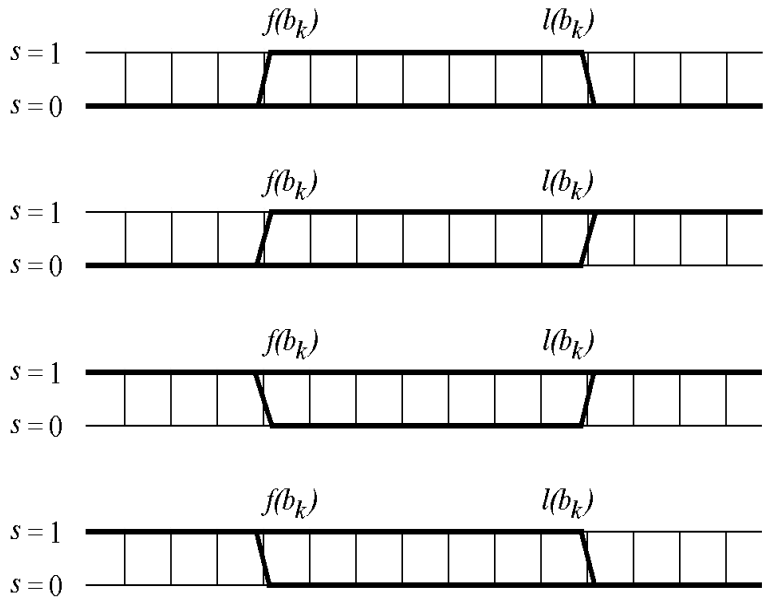


Figure 1: The four possible types of simple bivalence.

This motivates the following definitions. If, for particular (λ, μ), the unit commitment in the optimal solution of the relaxed problem in time interval i is not unique for unit k, we say that the unit is *bivalent* in this time interval. If it is not bivalent, we say that the unit is *monovalent*. A bivalent *sequence*, denoted b_k, is characterized by four parameters; the parameters $f(b_k)$ and $l(b_k)$ denote the first and last time interval, respectively, of the bivalent sequence, the parameter $s^f(b_k)$ denotes the optimal state immediately before the bivalent sequence, and the parameter $s^{l+1}(b_k)$ denotes the optimal state immediately after the bivalent sequence.

If and only if the bivalent sequence b_k fulfills

$$\sum_{i=f(b_k)}^{l(b_k)} c_{i,k}^{crc}(\lambda_i, \mu_i) = \begin{cases} \gamma_k, & \text{if } s^f(b_k) = 1 \text{ and } s^{l+1}(b_k) = 1 \\ -\gamma_k, & \text{if } s^f(b_k) = 0 \text{ and } s^{l+1}(b_k) = 0 \\ 0, & \text{otherwise,} \end{cases} \tag{16}$$

the sequence is referred to as a *simple* bivalence, or to as a simple bivalent sequence. Or equivalently, a sequence b_k is a simple bivalence if for all time intervals in $[f(b_k), \ldots, l(b_k)]$ the optimal unit commitments are identical, i.e. either all of them are zero or all of them are one. The four possible types of simple bivalence are illustrated in Figure 1.

Observe that simple bivalent sequences may follow immediately after each other, see the three top illustrations in Figure 2. Such sequence of simple bivalences is referred to as a *linked* bivalence, or to as a linked bivalent sequence. However, situations like the one at the bottom illustration in Figure 2 will not occur under the assumption that $\gamma_k > 0$.

Briefly, the algorithm from Dotzauer and Ravn (2001) solving the dual problem (7) is performed as follows. Initially starting values for λ and μ are chosen. In each iteration n, for given $[\lambda, \mu]^n$ and during the solution

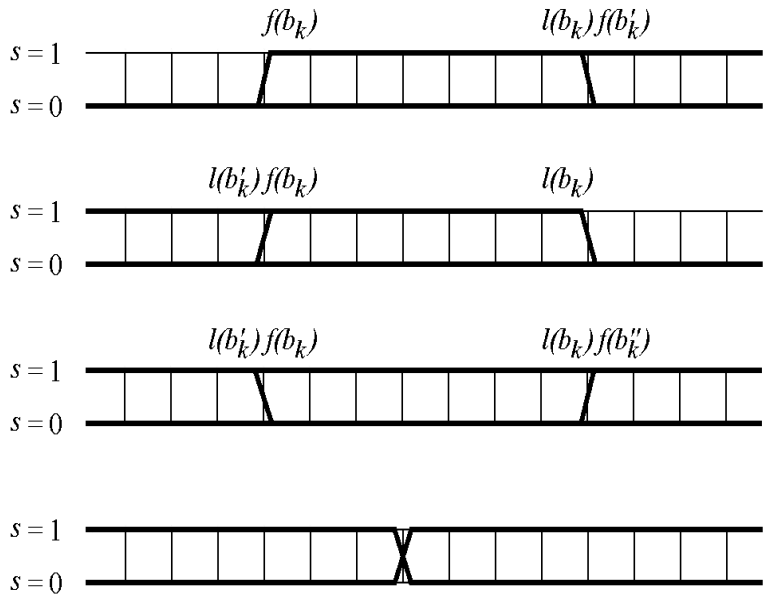


Figure 2: Examples of linked bivalence. Situations like the one at the bottom illustration will not occur under the assumption that $\gamma_k > 0$.

procedure of the relaxed problem (6), the algorithm tries to identify which units are bivalent. Given the set of assumed bivalent units, we attempt to fulfill the optimality conditions directly by solving a series of systems of nonlinear equations. If a detected bivalence is valid, this will generate the new update $[\lambda_i, \mu_i]^{n+1}$ for the relevant time interval. In time intervals without detected bivalence, λ_i and μ_i are updated using (15). The algorithm converges when the optimality conditions are fulfilled.

4. A branch-and-bound algorithm

The algorithm presented in Section 3 will solve the dual problem (7). To provide an optimal solution of the original problem, i.e. the primal problem (5), the algorithm can be embedded in a branch-and-bound framework, see e.g. Wolsey (1998). This section presents the branch-and-bound algorithm developed in Dotzauer (2001).

The methodology is to successively in each iteration of the algorithm fix some of the binary variables $u_{i,k}$ during the solution procedure of the relaxed problem (6). As the algorithm proceeds this will generate a tree structure where each node corresponds to a specific set of $u_{i,k}$ being fixed. Initially, in the root node, no variables are fixed. Moving further down in the tree a new variable $u_{i,k}$ is fixed at each node, implying that the current node will branch into two new nodes; one node with the new variable fixed to one and one node with the new variable fixed to zero. The node from which branching is performed at the current iteration is referred to as the *branching node*, and the corresponding new variable fixed is referred to as the *branching variable*. A node at the bottom of the tree is referred to as a *leaf node*. As a consequence, three decisions must be made in an iteration where branching is performed. First the choice of branching node, then the choice of branching variable, and finally, the choice if the branching variable shall be fixed to one or zero. Such decision strategies, or branching strategies, are discussed in Section 5.

The bounding part of the algorithm considers the fact that the dual objective $\Phi(\lambda, \mu)$ defines a lower bound on the objective function value of every feasible solution of the primal problem. The relation is exploited to decide when further branching from a leaf node only will generate primal solutions that are worse than the best found so far. This is possible since $\Phi(\lambda, \mu)$ is non-decreasing while moving down in the tree. When a dual value greater than the best primal solution found so far is computed, the end of the current branch in the branch-and-bound tree is reached. A leaf node where further branching is meaningful is referred to as *open*, and a leaf node where no further branching is meaningful is determined to be *closed*.

Branch-and-bound will by itself provide an optimal solution of the primal problem (5). However, since the full branch-and-bound tree will normally include an enormous number of nodes the convergence may be extremely slow. To achieve good performance, efficient branching strategies must be used. Also efficient heuristics that construct 'good' primal feasible solutions from the solution of the relaxed problem are desirable. Heuristics used for this purpose are called *Lagrangian heuristics* and are developed for the specific problem structure. Usually the unit commitment is first corrected and fixed, then the remaining continuous problem is solved to yield the economic dispatch.

The branch-and-bound algorithm for the solution of problem (5) is stated as follows:

Step 0. Let $n = 0$ and let η be the root node. Choose $[\lambda, \mu]^n_\eta$.

Step 1. Choose a node η in the branch-and-bound tree.

Step 2. Given $[\lambda, \mu]_\eta^n$, solve the relaxed problem with relevant $u_{i,k}$ being fixed.

Step 3. Construct a primal feasible solution.

Step 4. If all leaf nodes are closed, then stop.

Step 5. Compute $[\lambda, \mu]_\eta^{n+1}$, let $n = n + 1$ and go to Step 1.

First, in Step 0 of the algorithm, the dual variables λ and μ are initialized. In Step 1 the node to process in the branch-and-bound tree is chosen, i.e. the variables $u_{i,k}$ that should be fixed are chosen. The choice may either be to stay at the same node as in the previous iteration, to continue at an open leaf node or to construct two new nodes by branching and then move to one of them. In the latter alternative three decisions must be made, viz. the choice of branching node, the choice of branching variable and the choice if to move to the node corresponding to the branching variable fixed to one or zero.

At the first iteration at a new node the dual variables $[\lambda, \mu]_\eta^n$ are chosen equal to their values at the preceding (branching) node. As in the algorithm of Section 3, the assumed bivalent units are identified during the solution procedure of the relaxed problem (6) in Step 2. By using heuristics a feasible solution to the primal problem (5) is constructed in Step 3. In this paper we use a simple heuristic which combines the optimal solution of the relaxed problem from Step 2 with the corresponding bivalent units, which are all set on. Variants of the algorithm may omit Step 3 in some iterations. The algorithm terminates in Step 4 when all leaf nodes are closed.

The dual variables λ and μ are updated in Step 5. As described in Section 3, new updates of λ_i and μ_i corresponding to a time interval i with detected bivalence are computed by exploiting the dual optimality conditions. In time intervals without detected bivalence, λ_i and μ_i are in this paper updated using a cutting plane method.

5. Branching strategies

The strategies to choose branching nodes and branching variables are essential for the efficiency of a branch-and-bound algorithm. This section will discuss such strategies, both more general ones and suited strategies that exploit the special structure of the problem considered. Especially, for the algorithm presented in Section 4, we suggest how to choose branching variables by exploiting information from units identified as bivalent.

Common techniques for choosing branching nodes are depth-first strategies and breadth-first strategies, Wolsey (1998). Using a depth-first strategy the algorithm will stay in the same branch until the bottom of the branch-and-bound tree is reached. This will hopefully more rapidly generate a good primal feasible solution. However, the disadvantage is that when 'wrong'

branch is chosen in an early stage, it may take long time before the algorithm leaves the branch. This may be avoided by using a breadth-first strategy, which will process all nodes at the same level of the tree before the algorithm moves down to the next level. The disadvantage of using this method is that it may take long time before a good primal solution is found. As a consequence, the breadth-first strategy normally is used only when a primal feasible solution is found in an early iteration of the algorithm. In this case branch-and-bound is used to prove the optimality of this solution.

A third strategy is to branch on the open leaf node with the highest dual objective value. Since the dual objective will not decrease while moving down in the tree, this simple strategy will obviously generate high lower bounds, and thereby close branches more rapidly. Consequently, there is a corresponding strategy which chooses the open leaf node with the lowest dual objective value. This strategy is motivated since it will increase the lowest lower bound, which is a lower bound on the primal optimal value.

Given the branching node, the branching variable is chosen. Since normally the dual objective is non-smooth at optimum, and therefore bivalent units for the optimal dual variables exist, it is obvious to choose branching variables from the variables corresponding to such units. From a simple bivalence b_k, a branching variable is chosen due to (16), i.e. the branching variable is chosen from the set $\{u_{i,k} : i = f(b_k), ..., l(b_k)\}$. Here a natural choice is $u_{j,k}$, where j is the integer closest to $(f(b_k) + l(b_k))/2$.

By exploiting that some on-off combinations for units corresponding to linked bivalences are irrelevant (see the bottom illustration in Figure 2), other branching strategies may be formulated. The optimality conditions stated in Dotzauer and Ravn (2001) conclude that both (on, off) and (off, on) can not occur in the same two time intervals within a linked bivalence. Here $(u_{j,k}, u_{j+1,k})$ denotes on-off combinations in two consecutive time intervals j and $j + 1$, where $j = l(b_k)$ is the last time interval in a bivalent sequence b_k and $j + 1 = f(b'_k)$ is the first time interval in another bivalent sequence b'_k. A possible strategy may then be to avoid the irrelevant combinations in the branching procedure.

Another suggestion is to choose a variable $u_{i,k}$ corresponding to the unit with the highest production capacity $\overline{p}_k$. The motive is that when such variable is fixed to zero it will have the highest possible influence on the dual objective.

6. Performance analysis

This section examines the behavior of the algorithm presented in Section 4. The presented results are discussed briefly. For a more detailed examination, see Jönsson (2000). First, consider the data set defined in Virmani et. al. (1989), which models a system with 20 units over a time horizon of 24 hours. The performance of the algorithm is illustrated in Figure 3. The broken line shows the dual objective when the algorithm is not adopting the branch-and-bound methodology. This variant of the algorithm, which is the algorithm of Section 3, solves the dual problem (7) in 44 iterations.

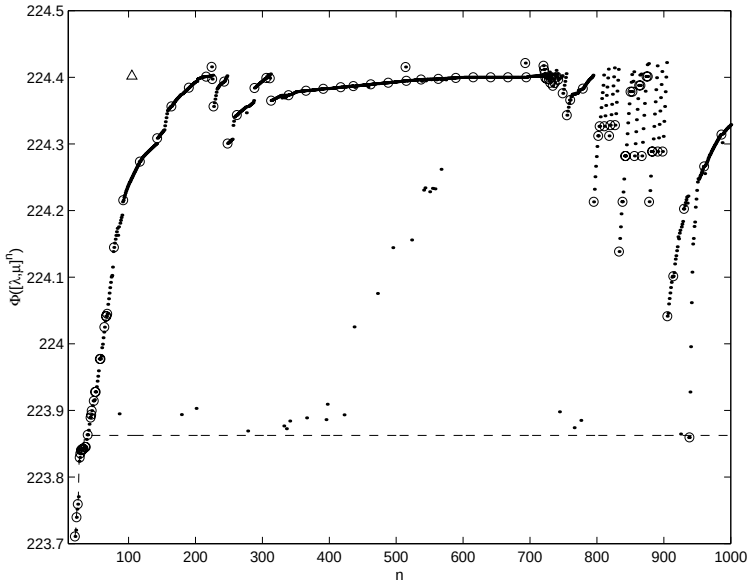


Figure 3: The dual objective in each iteration when the algorithm adopts (dots) and not adopts (broken line) branch-and-bound. A dot with a circle indicates the first iteration at a new node in the branch-and-bound tree. In iteration 105, for the first time the optimal primal solution is successfully constructed, marked with a triangle.

Applying branch-and-bound, the dual objective will increase. Here the algorithm adopts the depth-first strategy, and the simple rule to chose branching variables from detected bivalences and then move to the node with the branching variable fixed to one. If no bivalence is found during 20 consecutive iterations, the algorithm chooses the new branching variable corresponding to the unit which is 'closest to' bivalent, i.e. the unit which is closest to fulfill (16). The value of the dual objective is in each iteration marked with a dot. A dot with a circle indicates the first iteration at a new node in the branch-and-bound tree. In each iteration, the algorithm tries to generate a primal feasible solution (Step 3). The first time the optimal primal solution is successfully constructed is in iteration 105, marked with a triangle in Figure 3.

During the first two-hundred iterations the dual objective at the current node increases rapidly, and in iteration 223 for the first time the dual exceeds the best primal feasible solution found. In the following iterations, the algorithm starts to back-trace the tree to examine other branches, and this continues until the calculations are interrupted after 1000 iterations.

Remarkable is that the algorithm constructs the optimal primal solution in the first selected branch in the branch-and-bound tree. This may be a good argument for using the strategy which chooses branching variables from units identified as bivalent.

Now consider three algorithm variants, all using the depth-first strategy to chose branching nodes, but using different rules to chose branching vari-

Strategy		**Variables Fixed to 1**		**Variables Fixed to 0**	
	Iteration	Correct	Incorrect	Correct	Incorrect
Biv $\rightarrow$ 0	100	3	0	6	30
	200	4	0	12	47
	300	5	0	24	58
Biv $\rightarrow$ 1	100	43	0	0	0
	200	64	1	0	0
	300	73	1	1	1
Rand $\rightarrow$ Rand	100	9	6	4	5
	200	15	9	10	17
	300	25	12	18	22

Table 1: Number of variables fixed correctly and incorrectly compared to the optimal primal solution, after 100, 200 and 300 iterations, using three different strategies. The results are the total number of fixed variables for the three data sets considered.

ables. Three sets of data, defined in Bard (1988), Fardanesh and Villaseca (1986), and Virmani et. al. (1989), are considered. The problems consider a time horizon I of 24 hours, and the number of units K is 10, 20 and 20, respectively. Table 1 shows the number of variables fixed to the correct and to the wrong value after 100, 200 and 300 iterations. Here correct means that the variable is fixed to its value appearing in the optimal primal solution. The results are the total number of variables fixed for the three data sets considered.

In the first algorithm variant (Biv $\rightarrow$ 0), branching variables are chosen corresponding to units detected as bivalent, and the new node to process corresponds to the branching variable fixed to zero. After 100 iterations totally 39 binary variables are fixed at the current node, three are fixed to one and 36 are fixed to zero. The variables fixed to one are all correctly fixed (row one). For the 36 variables fixed to zero, only six are fixed correctly and 30 are fixed incorrectly (row one). After 200 iterations 63 variables are fixed. This time four variables are fixed to one (row two). Again, the variables fixed to one are all correct. The variables fixed to zero are 59 in total. Only 12 of these are fixed correctly and 47 are fixed incorrectly (row two). After 300 iterations 87 variables are fixed; five fixed to one and 82 fixed to zero. Again, the variables fixed to one are all correct (row three). For the variables fixed to zero, 24 are fixed correctly and 58 are fixed incorrectly (row three).

In the second variant (Biv $\rightarrow$ 1), branching variables are chosen corresponding to units detected as bivalent, and the new node to process corresponds to the branching variable fixed to one. The results are given at row four to six.

In the last variant (Rand $\rightarrow$ Rand), branching variables are chosen by random, and the new node to process is chosen randomly between the

nodes corresponding to the branching variable fixed to zero and one. A new branching variable is here chosen every tenth iteration. The approach represents an algorithm which do not use any sophisticated branching strategy. The results are given at row seven to nine in Table 1.

Comparing the three strategies, the second (Biv $\rightarrow$ 1) is superior. Here, 100%, 98.5% and 97.4% of the fixed variables are fixed correctly after 100, 200 and 300 iterations, respectively. In the first case (Biv $\rightarrow$ 0), the corresponding results are 23%, 25% and 33%. The third case (Rand $\rightarrow$ Rand) gives about fifty percent chance to fix a variable correct.

In conclusion, the strategy to choose the branching variable from the set of variables corresponding to units identified as bivalent, and then branch to the node corresponding to the branching variable fixed to one, seems to be a promising strategy.

7. Conclusions

A basic version of the unit commitment problem, in combination with the economic dispatch problem, was considered. The algorithm from Dotzauer and Ravn (2001), solving the dual problem, was embedded in a branch-and-bound framework. The algorithm adopts a strategy choosing branching variables corresponding to units identified as bivalent, that is units identified to have a non-unique unit commitment. The performance of the algorithm was analyzed, and from the considered test cases, the methodology to choose branching variables corresponding to bivalences seems to be a promising strategy.

Bibliography

1. Bard, J.F., *Short-Term Scheduling of Thermal-Electric Generators Using Lagrangian Relaxation,* Operations Research **36**, (1988), 756-766.

2. Dotzauer, E., and Ravn, H.F., *Dual Optimality Conditions for the Unit Commitment Problem,* In thesis: Energy System Operation by Lagrangian Relaxation (by Dotzauer), Department of Mathematics, Linköping University, Sweden, (2001), 33-53.

3. Dotzauer, E., *A Branch-and-Bound Algorithm Exploiting Dual Optimality Conditions in Solution of the Unit Commitment Problem,* In thesis: Energy System Operation by Lagrangian Relaxation (by Dotzauer), Department of Mathematics, Linköping University, Sweden, (2001), 75-86.

4. Fardanesh, B., and Villaseca, F.E., *Two-Step Optimal Thermal Generation Scheduling,* Automatica, **22**, (1986), 361-366.

5. Jönsson, H., *A Branch-and-Bound Algorithm for Solution of the Unit Commitment Problem*, Master Thesis, Department of Mathematics and Physics, Mälardalen University, Västerås, Sweden, (2000).

6. Muckstadt, J.A., and Koenig, S.A., *An Application of Lagrangian Relaxation to Scheduling in Power-Generation Systems,* Operations Research **25**, (1977), 387-403.

7. Sen, S., and Kothari, D.P., *Optimal Thermal Generating Unit Commitment: a Review,* Electrical Power & Energy Systems **20**, (1998), 443-451.

8. Sheble', G.B., and Fahd, G.N., *Unit Commitment Literature Synopsis,* IEEE Trans. on Power Systems **9**, (1994), 128-135.

9. Shor, N.Z., *Minimization Methods for Non-Differentiable Functions*, Springer-Verlag, (1985).

10. Virmani, S., Imhof, K., and Mukherjee, S., *Implementation of a Lagrangian Relaxation Based Unit Commitment Problem,* IEEE Trans. on Power Systems **4**, (1989), 1373-1379.

11. Wolsey, L.A., *Integer Programming*, John Wiley and Sons, (1998).

Department of Mathematics and Physics, Mälardalen University, SE-721 23 Västerås, Sweden.
E-mail: erik.dotzauer@mdh.se, henrik.jonsson@mdh.se
Elkraft System, Lautruphøj 7, DK-2750 Ballerup, Denmark.
E-mail: hrv@elkraft.dk

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.88-94

ARNE FRENNELIUS

THE STOCK SELECTION GUIDE AND SOME IMPROVEMENTS OF IT

The Stock Selection Guide developed in the USA is described. Two improvements to the model are presented. The first is a determination of what future yearly growth of earnings is required to make a stock worth buying. The other is an adaptation of this formula so that the market interest rate is taken into account.

2000*AMS Mathematics Subject Classifications*. 91B28.

Key words and phrases. The Stock Selection Guide, earnings-loss ratio, price-earnings ratio, multiple regression.

1. Introduction

The Stock Selection Guide was developed in the USA during the 1950s since when it has been used by members of the National Associations of Investors Corporation in the USA. The model is based upon fundamental analysis. It was introduced to Sweden at the end of the 1980s by the Swedish Shareholders Association which popularized the model through books, courses etc.

2. Problems and purpose

The Stock Selection Guide has been used by many investors for a long time in the USA and many other countries, among them Sweden. One problem with the model in its original form is that it takes long time to perform the analysis without a computer. However computer software has been developed to apply the model, by myself Sweden and by the Swedish company Delphi Economics.

The user of the model should state his anticipated yearly growth in earnings from the company. Based upon this figure the model will give three zones, a buy zone, where the stock is worth buying, a neutral zone and a sell zone. A central concept in the model is the earnings-loss ratio defined by

$$\frac{H-P}{P-L}$$

where P = stock price now, H = optimistic highest stock price after 5 years and L = lowest anticipated price after 5 years. How H and L are determined is described later in the paper.

The stock is considered to be worth buying if this ratio exceeds 2 (still better is if it exceeds 3). The anticipated yearly growth in earnings per share is used to determine H. I propose that this should be inverted. Instead I pose the question: What is the required yearly growth in the earnings in order to make the stock fall into the buy zone, i e to make the earnings-loss ratio exceed 2?

Another weakness with the model is that it does not take the market interest rate into account. There is a strong connection between the market interest rate and the stock market. I have made a multiple regression study of the average price-earnings-ratio in the Swedish stock market and some market interest rates.[1] This yields a method of determining the justified average price-earnings-ratio.

I intend to further adjust the required yearly growth in the earnings in order to place the stock in the buy zone so that the prevailing market interest rate is taken into account.

3. The stock selection guide

The general goal for the investor should be to double his stock value after 5 years. In order to achieve this he should

- think long term
- invest regularly
- limit his risks
- buy stocks in growth companies at a reasonable price
- understand what he buys

The model for analyzing a single stock covers two pages. The first page contains a lin-log diagram. This should be used to plot bars over historic values on earnings/share, sales, stock prices and dividends. On the assumption of the future yearly growth in earnings per share, an extrapolation can then be made in order to achieve predicted earnings per share after 5 years.

The second page consists of four sections, numbered 2-5. Here data about profitability, past p/e-ratios, judgement of risks and return the next 5 years and expected total return should be entered. This results in the stock price being placed in a buy zone, a neutral zone and a sell zone. The stock price is placed in the buy zone if the earnings/loss ratio (see above) is at least 2, and the rating is even higher if it exceeds 3.

The H and the L in the earnings/loss ratio are determined in the following way.

[1] The price/earnings-ratio is defined as stock price/earnings per share.

From historical data the mean of the highest price/earnings ratio for the last 5 years is calculated. In this paper $\bar{H}$ is used to denote this mean. Using the lin-log graph it is possible to predict the optimistic earnings per share after 5 years. Multiplying these 2 numbers gives an optimistic highest price after 5 years, here denoted by H.

For judgement of the lowest price after 5 years, denoted here by L, 4 alternatives exist

a) Historical data are used to calculate the mean of the lowest price/earnings ratio for the last 5 years. Here this mean is denoted by $\bar{L}$. An estimation of lowest earnings per share after 5 years is then made. As a default value one might use this year's earnings per share. These 2 numbers are multiplied, which gives the L in alternative a).

b) An average of the lowest stock price for the last five years is calculated. This is L according to alternative b).

c) L = lowest price the last 3 years.

d) L = lowest price based upon dividend = $\dfrac{\text{dividend now}}{\text{highest yield the last five years}}$.

It is up to the investor to select one of these 4 alternatives. One common procedure is to select the highest figure as an estimate of L.

My purpose is to derive a formula for required yearly growth in earnings per share over the next 5 years needed to place the stock price in the buy zone. This is obtained by using the following three equations and solving them with respect to G.

Notations:

G = optimistic expected yearly growth in earnings per share

G_2 = required yearly growth in earnings per share in order to place the stock in the buy zone, i.e. with an earnings/loss-ratio exceeding 2 with the price it has now

G_3 analogously

E_0 = current earnings per share

E_5 = optimistic expected earnings per share after 5 years

$\bar{H}$ = highest p/e-ratio after 5 years

H = highest price after 5 years based upon an optimistic judgement

L = lowest price after 5 years

P = stock price now

Equations:

$$E_5 = E_0 \cdot \left(1 + \frac{G}{100}\right)^5$$

$$H = \bar{H} \cdot E_5 = \bar{H} \cdot E_0 \left(1 + \frac{G}{100}\right)^5$$

$$\frac{H - P}{P - L} = 2$$

Solution:

$$H - P = 2P - 2L$$

$$H = 3P - 2L$$

$$\bar{H} \cdot E_0 \left(1 + \frac{G}{100}\right)^5 = 3P - 2L$$

$$\left(1 + \frac{G_2}{100}\right)^5 = \frac{3P - 2L}{\bar{H} \cdot E_0}$$

$$1 + \frac{G_2}{100} = \left(\frac{3P - 2L}{\bar{H} \cdot E_0}\right)^{0.2}$$

$$G_2 = 100 \cdot \left(\left(\frac{3P - 2L}{\bar{H} \cdot E_0}\right)^{0.2} - 1\right)$$

Analogously we obtain

$$G_3 = 100 \cdot \left(\left(\frac{4P - 3L}{\bar{H} \cdot E_0}\right)^{0.2} - 1\right)$$

4. Connection between rate of interest and stock prices

Let us begin with some quotations:

> "A rule of thumb that is used in the stock market is that a change of the rate of interest by 1% affects stock prices by 10 %." Mats Jonnerhag, Brsinsikt, Sweden.
>
> "The model that Alan Greenspan uses is simple. He compares the interest rate on 10-year Treasury Notes with the return in the stock market. The interest rate on these 10-year American Treasury notes is now at 5.5%. Today's stock prices only yield a return of 4.5 to 4.9% based upon the p/e-ratio on 20–22 that the earnings of this year correspond to. The return on the more risky stock market is therefore now 10–20% below the long term interest rate." Hans Westerberg in Svenska Dagbladet, Sweden, 25 Jan, 1998.
>
> "Today the average p/e-ratio is between 19 and 20. It is the highest valuation that we have noted. At the same time the average market interest rate is 5.3%. This indicates that the Stockholm Stock Exchange should have an average p/e ratio in the range 18–19." Peter Malmqvist in Market report from Aragon Fondkommission, 30 Sept, 1997.

These quotations show that the actors in the stock market are familiar with the fact that there is a clear connection between rates of interest and stock prices and that varying models exist to describe this connection.

Alan Greenspan's formula is thus

$$\text{jp/e} = \frac{1}{10 - \text{year interest rate}}$$

where jp/e means justified p/e-ratio.

Peter Malmqvist's formula is

$$\text{jp/e} = \frac{1}{\frac{10 - \text{year interest rate} + 3 - \text{month interest rate}}{2}}.$$[2]

Many other stock analysts also use the same formula as Peter Malmqvist.

Both Greenspan and Malmqvist thus invert the market interest rate to obtain the jp/e. If we use i to denote the market interest rate we obtain

$$jp/e = \frac{1}{i}$$

or

$$je/p = i$$

where je/p means justified earnings/price ratio.

I have performed a multiple regression analysis using data from Sweden in the 1990s in order to analyze the connection between je/p and different market interest rates. The model specification was

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$$

where
y = average e/p for the whole Swedish Stock Market
x_1 = 10-year interest rate in Sweden
x_2 = 5-year interest rate in Sweden
x_3 = 180 days interest rate in Sweden
x_4 = 90 days interest rate in Sweden

This gave the following result.
e/p = 3.51 90 days -3.32 180 days + 0.326 5-years + 0.449 10-years

Predictor	Coef	Stdev	t-ratio	p
Noconstant				
90 days	3.5071	0.7357	4.77	0.000
180 days	-3.3158	0.8272	-4.01	0.001
5-years	0.3261	0.4845	0.67	0.508
10-years	0.4485	0.3132	1.43	0.166

[2]Oral information from Peter Malmqvist

$s = 0.4477$

There is an apparent multicollinearity in the material. The $\hat{\beta}_2$ is negative which is in contrary to logic. And the t-ratio connected with β_3 is not significant. Thus the variables x_2 and x_3 were removed from the model. The subsequent analysis thus included only x_1 and x_4 as predictors. This gave the following result.
e/p = 0.524 90 days + 0.491 10-years

Predictor	Coef	Stdev	t-ratio	p
Noconstant				
90 days	0.52443	0.08708	6.02	0.000
10-years	0.49123	0.06874	7.15	0.000

$s = 0.5677$

An inversion of the result gives the following equation.

$$\frac{p}{e} = \frac{1}{0.524 \cdot 90 \text{ days interest rate} + 0.491 \cdot 10 - \text{years interest rate}}.$$

A calculation of the Mean Absolute Deviation for the obtained formula ("Regr"), for Greenspan's formula ("Alan") and for Malmqvist's formula ("Peter") was made with the following result.

Model	Mean Absolute Deviation
Regr	1.02
Alan	2.58
Peter	1.03

We see that the formula obtained by regressions analysis is very similar to that obtained by Malmqvist. Our choice of model will therefore be Malmqvist's, i.e.

$$jp/e = \frac{1}{\frac{10 - \text{year interest rate} + 3 - \text{month interest rate}}{2}}.$$

5. The new model

The purpose in now to adjust the G_2 formula so that the market interest rate (defined as in sec 4) is taken into account. I have not made any trials to adjust L. The only adjustment needed is therefore a new way of determining $\bar{H}$.

In the formula derived in section 3, $\bar{H}$ was an average of the highest p/e-ratio for the stock over the last 5 years. Let us denote these values by $H_1, H_2, \ldots, H_5$. The first step is to adjust these values so that the market interest rate during these 5 years is taken into account. An average of the

market interest rate for each was calculated and denoted by i_1, i_2, ..., i_5. The justified p/e-ratio for these years is denoted by j_1, j_2, ..., j_5 and is found using the formula

$$j_k = \frac{1}{i_k}, \qquad k = 1, \ldots, 5.$$

A correction factor for H_k is denoted by F_k and is obtained using

$$F_k = \frac{H_k}{j_k}, \qquad k = 1, \ldots, 5.$$

The average of F, ..., F_5 is denoted by $\bar{F}$.

The p/e-ratio today, justified by the present interest rate is denoted j_0 and is in the same manner as above obtained using

$$j_0 = \frac{1}{i_0}.$$

We now substitute $\bar{H}$ in the original formula for G_2 with $\bar{F} \cdot j_0$ which gives the adjusted G_2-formula

$$G_2 = 100 \cdot \left(\left(\frac{3P - 2L}{\bar{F} \cdot j_0 \cdot E_0} \right)^{0.2} - 1 \right).$$

Analogously we obtain the adjusted G_3-formula

$$G_3 = 100 \cdot \left(\left(\frac{4P - 3L}{\bar{F} \cdot j_0 \cdot E_0} \right)^{0.2} - 1 \right).$$

Bibliography

1. *The Stock Selection Guide.* Includes instruction manual and NAIC's Official Guide. National Association of Investors Corporation.

2. *Hitta kursvinnare med Aktiespararnas Aktieanalys. En praktisk handledning i konsten att bygga upp en lönsam aktieportfölj.* Aktiespararnas Aktieskola. ISBN 91-85524-53-0. 1997.

Department of Economics and Informatics, Mälardalen University, Box 883, S 721 23 Västerås, Sweden.
E-mail: afs@mdh.se

Theory of Stochastic Processes
Vol.7 (23), no.1-2 ,2001, pp.95-101

OLEKSIY M. GORELOV

DISCRETE APPROXIMATION OF OPTIMAL STOPPING TIME IN THE PROBLEM OF IRREVERSIBLE INVESTMENT

Let $X_1(t)$,$X_2(t)$ geometrical Brownian motions, possible correlated. We study the problem of optimal stopping : finding of stopping time $\tau^* \in [0,T]$ such that

$$\sup_{\tau\in[0,T]} E^x\{X_1(\tau) - X_2(\tau)\} = E^x\{X_1(\tau^*) - X_2(\tau^*)\},$$

where sup being taken all over all finite stopping times τ, and E^x denotes the expectation when $(X_1(0), X_2(0)) = x = (x_1, x_2)$.

2000 *Mathematics Subject Classifications* 60G40.

Key words and phrases: Optimal stopping, geometric Brownian motion, discrete approximation .

1. Introduction

Let $(\Omega, \mathcal{F}, P)$be a complete probability space,$\{\mathcal{F}_t, t \in [0,\infty)\}$ be increasing right continuous flow of σ -algebras, $\mathcal{F}_t \in \mathcal{F}$, $B(t) = (B_1(t), B_2(t))$ be $\mathcal{F}_t$ adapted two dimensional Brownian motion, $t \in [0,\infty)$.Also let $X_1(t)$ and $X_2(t)$ be a random processes , which are the solution of a system of stochastic differential equations :

$$\begin{cases} dX_1(t) = b_1X_1(t)dt + X_1(t)[q_{11}dB_1(t) + q_{12}dB_2(t)], X_1(0) = x_1 \\ dX_2(t) = b_2X_2(t)dt + X_2(t)[q_{21}dB_1(t) + q_{22}dB_2(t)], X_2(0) = x_2 \end{cases}$$

where b_i,q_{ij} are constants, $1 \le i,j \le 2$, $q_i = (q_{i1}, q_{i2}) \in \mathbf{R^2}$, $i = 1,2$.

We can write solutions of these equations as

$$X_i(t) = x_i exp\{(b_i - \frac{1}{2}a_{ii})t + q_i \cdot B(t)\},$$

where $a_{ij} = q_i \cdot q_j, 1 \le i,j \le 2$.

It is necessary to find stopping time $\tau^* \in [0,T]$ - such that

$$E\{X_1(\tau^*) - X_2(\tau^*)\} = \sup_{\tau} E\{X_1(\tau) - X_2(\tau)\}.$$

This problem was considered by Yaozhong Hu and Bernt Oksendal for $T = \infty$, they found τ^* only for $0 > b_2 > b_1$, applying one variation inequality. They proved that, under certain conditions, the stopping set S contains a halfspace :

$$S = \{(x_1, x_2) :\in R^2_+; x_1 > \mu x_2\}$$

We find an optimal stopping time on finite interval. It is more difficult problem. As it was impossible to indicate exact stopping time we construct a sequence of stopping times's that is easy to calculate and approximates an optimal moment.

2. Main results

Let $X(t)$ be a.s. continuous on $[0,T]$ process such that $E\{\sup\limits_{t\in[0,T]} |X(t)|\} < \infty$. Let us consider following uniform partition of interval $[0,T]$ $\lambda_n := \{i\triangle t, i = 0..n\}$ and set of stopping times $D := \{\tau : P\{\tau \in \lambda_n\} = 1\}$.
Consider also stopping times

$$\begin{aligned}
&\tau_n(n) := T \\
&\tau_k(n) := \begin{cases} k\triangle t, X(k\triangle t) \geq E\{X(\tau_{k+1}(n))|\mathcal{F}_{k\triangle t}\} \\ \tau_{k+1}(n), X(k\triangle t) < E\{X(\tau_{k+1}(n))|\mathcal{F}_{k\triangle t}\} \end{cases} \\
&\tau_0(n) := \begin{cases} \triangle t, X(0) \geq E\{X(\tau_1(n))|\mathcal{F}_{\triangle t}\} \\ \tau_1(n), X(0) < E\{X(\tau_1(n))|\mathcal{F}_{\triangle t}\} \end{cases}
\end{aligned} \tag{1}$$

for all $n \geq 1$. Following theorems demonstrate that we can use this stopping times to approximat τ^*.

Theorem 1. *Stopping time $\tau_0(n)$ is optimal in λ_n.*

Proof. Let $\tau \in \lambda_n$ a.s. Now, we assume for $0 \leq i \leq n$

$$\tilde{\tau_i} := \tau \cdot I\{\tau < i\triangle t\} + \tau_i(n) \cdot I\{\tau \geq i\triangle t\}.$$

and show that $E\{X(\tau_0(n))\} = E\{X(\tilde{\tau_0})\} \geq E\{X(\tilde{\tau_1})\} \ldots \geq E\{X(\tilde{\tau_n})\} = E\{X(\tau)\}$. Since $\tilde{\tau_0} = \tau_0(n)$ and $\tilde{\tau_n} = \tau_n(n)$ then first and last equalities are true. For $0 < i < n$

$$\begin{aligned}
&E\{X(\tilde{\tau_i}) - X(\tilde{\tau}_{i+1})\} = \\
&= E\{[X(\tilde{\tau_i}) - X(\tilde{\tau}_{i+1})] \cdot [I\{\tau < i\triangle t\} + I\{\tau \geq i\triangle t, \tau_i = \tau_{i+1}\} + \\
&\qquad + I\{\tau = i\triangle t, \tau_i = i\triangle t\} + I\{\tau > i\triangle t, \tau_i = i\triangle t\}]\}.
\end{aligned}$$

On $\{\tau < i\triangle t\}$, $\{\tau \geq i\triangle t, \tau_i = \tau_{i+1}\}$ and $\{\tau = i\triangle t, \tau_i = i\triangle t\}$ $\tilde{\tau}_i$ and $\tilde{\tau}_{i+1}$ coincide and therefore $E\{X(\tilde{\tau}_i) - X(\tilde{\tau}_{i+1})\} = E\{[X(\tilde{\tau}_i) - X(\tilde{\tau}_{i+1})]\times$ $\times[I\{\tau > i\triangle t, \tau_i = i\triangle t\}]\} \geq 0$. We obtain that $E\{X(\tau_0(n))\} \geq E\{X(\tau)\}$ for any $\tau \in \lambda_n$.
In addition we can conclude that for $k = 1..n$ $E\{X(\tau_k(n))\} \geq E\{X(\tau)\}$ for any $\tau : P\{\tau \in \{i\triangle t, k \leq i \leq n\}\} = 1$ and $\tau_k(n)$ is optimal on $\{i\triangle t, k \leq i \leq n\}$.

Theorem 2. *$E\{X(\tau^*) - X(\tau_0(n))\} \to 0$ a.s. when $n \to \infty$*

Proof. Let consider

$$\tau_0^{'}(n) := \sum_{i=k+1}^{n-1} i\triangle t \cdot I\{\tau^* \in [(i-1)\triangle t, i\triangle t)\} + T \cdot I\{\tau^* \in [T - \triangle t, T]\}.$$

Then

$$\begin{aligned}
&E\{X(\tau^*) - X(\tau_0(n))\} = \\
&= E\Bigg\{ \sum_{i=k+1}^{n-1} (\tau^* - i\triangle t) \cdot I\{\tau^* \in [(i-1)\triangle t, i\triangle t)\} + \\
&\qquad\qquad + (\tau^* - T) \cdot I\{\tau^* \in [T - \triangle t, T]\} \Bigg\} \leq \\
&\leq E\Bigg\{ \sum_{i=k+1}^{n-1} \left(\sup_{t\in[(i-1)\triangle t, i\triangle t]} X(t) - \inf_{t\in[(i-1)\triangle t, i\triangle t]} X(t) \right) \times \\
&\qquad\qquad \times I\{\tau^* \in [(i-1)\triangle t, i\triangle t)\} + \\
&+ \left(\sup_{t\in[T-\triangle t, T]} X(t) - \inf_{t\in[T-\triangle t, T]} X(t) \right) \cdot I\{\tau^* \in [T - \triangle t, T]\} \Bigg\} \leq \\
&\leq E\left\{ \max_i \left(\sup_{t\in[(i-1)\triangle t, i\triangle t]} X(t) - \inf_{t\in[(i-1)\triangle t, i\triangle t]} X(t) \right) \right\} \to 0 \quad a.s.
\end{aligned}$$

when $n \to \infty$. Since on $\{i\triangle t, k \leq i \leq n\}$ $\tau_0(n)$ is optimal from Theorem 1 then $E\{X(\tau_0^{'}(n))\} \leq E\{X(\tau_0(n))\} \leq E\{X(\tau^*)\}$. Therefore $E\{X(\tau^*) - X(\tau_0(n))\} \to 0$ $a.s.$ when $n \to \infty$.

3. Calculation of optimal time.

We shall consider only the cases when $b_1 > 0, b_2 > 0$ and $b_1 < 0, b_2 < 0$. Really we have that $E\{X_1(T) - X_2(T)|\mathcal{F}_{i\triangle t}\} = X_1(i\triangle t)e^{b_1(n-i)\triangle t} -$ $- X_2(i\triangle t)e^{b_2(n-1)\triangle t}$. Therefore, for $b_1 > 0, b_2 < 0$

$$X_1(i\triangle t) - X_2(i\triangle t) < X_1(t)e^{b_1(n-i)\triangle t} - X_2(t)e^{b_2(n-1)\triangle t}$$

$$X_1(i\triangle t)(1-e^{b_1(n-i)\triangle t}) < X_2(i\triangle t)(1-e^{b_2(n-1)\triangle t})$$

for any $X_1(i\triangle t), X_2(i\triangle t)$ thus $\tau_0(n) = T$ a.s, and for
if $b_1 < 0, b_2 > 0$

$$X_1(i\triangle t) - X_2(i\triangle t) > X_1(t)e^{b_1(n-i)\triangle t} - X_2(t)e^{b_2(n-1)\triangle t}$$

$$X_1(i\triangle t)(1-e^{b_1(n-i)\triangle t}) > X_2(i\triangle t)(1-e^{b_2(n-1)\triangle t})$$

for any $X_1(i\triangle t), X_2(i\triangle t)$ thus $\tau_0(n) = 0$ a.s.

In order to know discrete stopping time $\tau_0(n)$ for process $X(t) = X_1(t) - X_2(t)$ we need to calculate $E\{X_1(\tau_{i+1}) - X_2(\tau_{i+1}))|\mathcal{F}_{i\triangle t}\}$ as a function of a process value for any $0 \le i \le n-1$. It is easy to calculate that $E\{X_1(\tau_n) - X_2(\tau_n))|\mathcal{F}_{(n-1)\triangle t}\} = X_1((n-1)\triangle t)\cdot e^{b_1\triangle t} - X_2((n-1)\triangle t)\times$ $\times e^{b_2\triangle t}$.

Further,

$$\begin{aligned} &E\Big\{X_1(\tau_{n-1}) - X_2(\tau_{n-1})|\mathcal{F}_{(n-2)\triangle t}\Big\} = \\ &= E\Big\{[X_1(T-\triangle t) - X_2(T-\triangle t)]\cdot I_A + \\ &\qquad +[X_1(T) - X_1(T)]\cdot I_{\bar{A}}|\mathcal{F}_{T-2\triangle t}\Big\}. \end{aligned} \tag{2}$$

Where

$$\begin{aligned} A := \Big\{\omega : \quad & X_1(T-\triangle t) - X_2(T-\triangle t) > \\ & > X_1(T-\triangle t)e^{b_1\triangle t} - X_2(T-\triangle t)e^{b_2\triangle t}\Big\}. \end{aligned} \tag{3}$$

Let us rewrite (2) as

$$\begin{aligned} &X_1(T-2\triangle t)C_1E\{e^{B_1q_{11}+B_2q_{12}}I_A|\mathcal{F}_{T-2\triangle t}\} - \\ &-X_2(T-2\triangle t)C_2E\{e^{B_1q_{21}+B_2q_{22}}I_A|\mathcal{F}_{T-2\triangle t}\} + \\ &+X_1(T-2\triangle t)C_1\cdot e^{b_1\triangle t}E\{e^{B_1q_{11}+B_2q_{12}}I_{\bar{A}}|\mathcal{F}_{T-2\triangle t}\} - \\ &-X_2(T-2\triangle t)C_2\cdot e^{b_2\triangle t}E\{e^{B_1q_{21}+B_2q_{22}}I_{\bar{A}}|\mathcal{F}_{T-2\triangle t}\}, \end{aligned} \tag{4}$$

where

$$\begin{aligned} C_1 &:= e^{[b_1-\frac{1}{2}(q_{11}^2+q_{12}^2)]\triangle t}, \quad C_2 := e^{[b_2-\frac{1}{2}(q_{21}^2+q_{22}^2)]\triangle t}, \\ B_1 &:= B_1(T-\triangle t) - B_1(T-2\triangle t), \quad B_2 := B_2(T-\triangle t) - B_2(T-2\triangle t). \end{aligned}$$

And (3) can be written as

$$\left\{\omega : \frac{X_1(T-\triangle t)}{X_2(T-\triangle t)} > \frac{1-e^{b_2\triangle t}}{1-e^{b_1\triangle t}}\right\} = \left\{\omega : B_1q_{11} + B_2q_{12} > \right.$$
$$> \triangle t\left[(b_2-b_1) - \frac{1}{2}(q_{21}^2+q_{22}^2-q_{11}^2-q_{12}^2)\right] +$$
$$\left. + log\left(\frac{(1-e^{b_2\triangle t})X_2(T-2\triangle t)}{(1-e^{b_1\triangle t})X_1(T-2\triangle t)}\right)\right\}.$$

The set A is half-plane in (B_1, B_2) space. Since $X_1(t) - X_2(t)$ is Markov's process

$$E\{e^{B_1q_{11}+B_2q_{12}}I_A|\mathcal{F}_{T-2\triangle t}\} = \int\int_A \frac{1}{2\pi\triangle t}e^{B_1q_{11}+B_2q_{12}}e^{-\frac{B_1^2+B_2^2}{2\triangle t}}dB_1dB_2.$$

After integration we obtain that $E\{e^{B_1q_{11}+B_2q_{12}}I_A|\mathcal{F}_{T-2\triangle t}\} = e^{\frac{\triangle t}{2}(q_{11}^2+q_{12}^2)}\cdot R$, where

$$Q := \frac{\triangle t}{2}[(q_{11}-q_{21})^2 - (q_{12}-q_{22})^2]$$
$$R := \frac{1}{2} + \frac{1}{2}\cdot erf\left(\frac{\triangle t[b_1-b_2] + Q + log\left(\frac{(1-e^{b_1\triangle t})X_1(T-2\triangle t)}{(1-e^{b_2\triangle t})X_2(T-2\triangle t)}\right)}{2\sqrt{Q}}\right).$$

Other three expectations in (4) may be obtained in the same way. Then the right-hand side of (2) equals

$$E\{X_1(\tau_{n-1}) - X_2(\tau_{n-1}))|\mathcal{F}_{(n-2)\triangle t}\} =$$
$$= \left[X_1(T-2\triangle t)\cdot e^{b_1\triangle t} - X_2(T-2\triangle t)\cdot e^{b_2\triangle t}\right]\cdot R +$$
$$+ \left[X_1(T-2\triangle t)\cdot e^{2b_1\triangle t} - X_2(T-2\triangle t)\cdot e^{2b_2\triangle t}\right]\cdot(1-R)$$

and continuation region at the moment $T-2\triangle t$ is determined by the inequality

$$\frac{X_1}{X_2} > \frac{1-e^{b_2\triangle t}R - e^{2b_2\triangle t}(1-R)}{1-e^{b_1\triangle t}R - e^{2b_1\triangle t}(1-R)}, b_1 < 0, b_2 < 0 \qquad (5)$$
$$\frac{X_1}{X_2} < \frac{1-e^{b_2\triangle t}R - e^{2b_2\triangle t}(1-R)}{1-e^{b_1\triangle t}R - e^{2b_1\triangle t}(1-R)}, b_1 > 0, b_2 > 0.$$

where X_1, X_2, R mean $X_1(T-2\triangle t), X_2(T-2\triangle t)$ and $R\left(\frac{X_1(T-2\triangle t)}{X_2(T-2\triangle t)}\right)$ respectively. Note that $R\left(\frac{X_1}{X_2}\right)$ is monotone increasing function of $\frac{X_1}{X_2}$, and

the right-hand side of (5) can be transformed as

$$\begin{aligned}
&\frac{1-e^{b_2\triangle t}R-e^{2b_2\triangle t}(1-R)}{1-e^{b_1\triangle t}R-e^{2b_1\triangle t}(1-R)}=\\
&=\frac{e^{b2\triangle t}\left(e^{b2\triangle t}-1\right)}{e^{b1\triangle t}\left(e^{b1\triangle t}-1\right)}+\frac{-e^{b_1\triangle t}+e^{b_1\triangle t}e^{b_2\triangle t}-e^{2b_2\triangle t}+e^{b_2\triangle t}}{e^{2b_2\triangle t}(e^{b_1\triangle t}-1)\left[R-\frac{e^{b_1\triangle t}+1}{e^{b_1\triangle t}}\right]}= \qquad (6)\\
&=\frac{(e^{b_1\triangle t}-e^{b_2\triangle t})(e^{b_2\triangle t}-1)}{e^{2b_2\triangle t}(e^{b_1\triangle t}-1)\left[R-\frac{e^{b_1\triangle t}+1}{e^{b_1\triangle t}}\right]}.
\end{aligned}$$

If $b_1<b_2$ then (6) is monotone increasing and if $b_1>b_2$ then (6) is monotone decreasing function of $\frac{X_1}{X_2}$, so if $b_1>b_2$ continuation region has a form $\frac{X_1}{X_2}<\alpha$ or $\frac{X_1}{X_2}>\alpha$ depending on sign of b_1,b_2, where α is a root of equation

$$\alpha=\frac{1-e^{b_2\triangle t}R(\alpha)-e^{2b_2\triangle t}(1-R(\alpha))}{1-e^{b_1\triangle t}R(\alpha)-e^{2b_1\triangle t}(1-R(\alpha))}.$$

At the next step we need to calculate

$$\begin{aligned}
&E\{X_1(\tau_{n-2})-X_2(\tau_{n-2})|\mathcal{F}_{T-3\triangle t)}\}= \qquad (7)\\
&=E\Big\{[X_1(T-2\triangle t)-X_2(T-2\triangle t)]\cdot I_A+\\
&\qquad +[X_1(\tau_{n-1})-X_1(\tau_{n-1})]\cdot I_{\bar{A}}|\mathcal{F}_{T-3\triangle t}\Big\}.
\end{aligned}$$

Where

$$\begin{aligned}
A:=\Big\{&\omega: X_1(T-2\triangle t)-X_2(T-2\triangle t)>\\
&>E\{X_1(\tau_{n-1})-X_2(\tau_{n-1})|\mathcal{F}_{T-2\triangle t}\}\Big\}.
\end{aligned}$$

As in (2) represent (7) in such way

$$\begin{aligned}
&X_1(T-3\triangle t)C_1E\{e^{B_1q_{11}+B_2q_{12}}I_A|\mathcal{F}_{T-3\triangle t}\}-\\
&-X_2(T-3\triangle t)C_2E\{e^{B_1q_{21}+B_2q_{22}}I_A|\mathcal{F}_{T-3\triangle t}\}+\\
&+X_1(T-3\triangle t)C_1E\{e^{B_1q_{11}+B_2q_{12}}(e^{b_1\triangle t}R+e^{2b_1\triangle t}(1-R))I_{\bar{A}}|\mathcal{F}_{T-3\triangle t}\}-\\
&-X_2(T-3\triangle t)C_2E\{e^{B_1q_{21}+B_2q_{22}}(e^{b_2\triangle t}R-e^{2b_2\triangle t}(1-R))I_{\bar{A}}|\mathcal{F}_{T-3\triangle t}\},
\end{aligned}$$

And again obtain that stopping set has a form

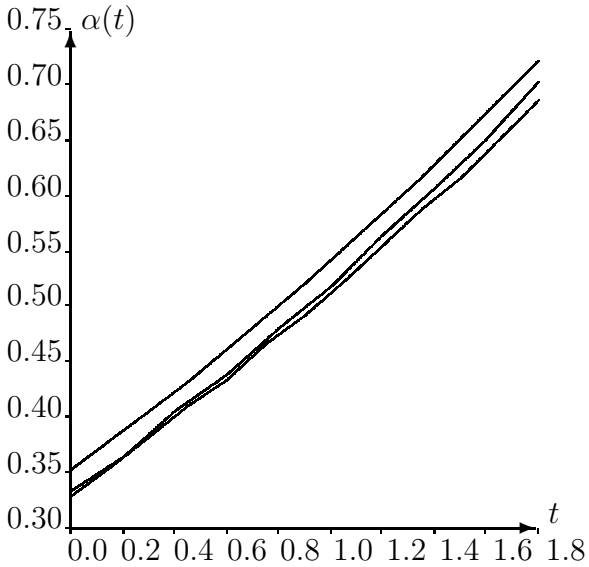
$$F\left(\frac{X_1(T-3\triangle t)}{X_2(T-3\triangle t)}\right)>0. \qquad (8)$$

The method of induction permits to calculate that that for all $0 \leq k \leq n-1$ continuation region has a form

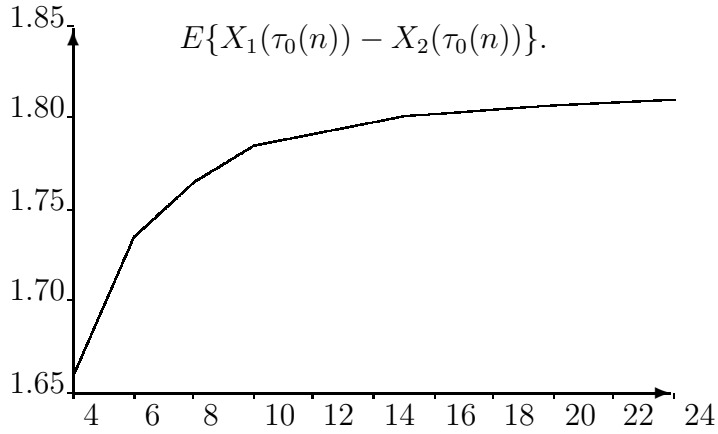
$$F\left(\frac{X_1(k\triangle t)}{X_2(k\triangle t)}\right) > 0. \tag{9}$$

4. Simulation results

We have constructed simple model for simulation geometric Brownian motion and finding discrete stopping times in double precision arithmetic. In this model expectation (1) are computed by interpolation as surfaces $E(X_1(i\triangle t), X_2(i\triangle t))$ in rectangular area to which process belong with probability $1-10^{-12}$ on uniform grid 150×150 items. In all the cases continuation region was a half-plane $\frac{X_1(i\triangle t)}{X_2(i\triangle t)} > \alpha(i\triangle t)$. More over function $\alpha(i\triangle t)$ with computation error accuracy is linear function from $i\triangle t$.
For example for $x_1 = 0.9$, $x_2 = 1.9$, $b_1 = 1.5$, $b_2 = 1.1$, $q_{11} = 0.5$, $q_{12} = 0.5$, $q_{21} = 0.2$, $q_{22} = 0.6$, $T = 2$. On the figure $n = 5$, 10 and 15 respectively from top to bottom.



When n is increasing then $E\{X_1(\tau_0(n)) - X_2(\tau_0(n))\}$ is also increasing and converges to $E\{X_1(\tau^*)) - X_2(\tau^*)\}$. On the following figure $n = 4, 6, 8, 10, 16, 20, 24$.



Bibliography

1. Yaozhong, Hu and Bernt Oksendal, *Optimal time to invest when the price processes are geometric Brownian motions*, Finance and Stochastics, Springer, (1998), V.**2**, 295-310.

Department of Mechanics and Mathematics, Kyiv National Taras Shevchenko University, 01033, Kyiv, Ukraine.
E-mail: grolm@mail.ru

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.102-108

JAN GRANDELL

SIMPLE APPROXIMATIONS OF RUIN PROBABILITIES[1][2]

A "simple approximation" of a ruin probability is an approximation using only some moments of the claim distribution and not the detailed tail behaviour of that distribution. Such approximations may be based on limit theorems or on more or less *ad hoc* arguments. The most successful simple approximation is certainly the De Vylder approximation, which is based on the idea to replace the risk process with a risk process with exponentially distributed claims such that the three first moments coincide. That approximation is known to work extremely well for "kind" claim distributions. The main purpose of this paper is to analyse the De Vylder approximation and other simple approximations from a more mathematical point of view and to give a possible explanation why the De Vylder approximation is so good.

2000 *Mathematics Subject Classifications.* 62P05, 90A46, 62E17.

Key words and phrases. Ruin probability, Approximations.

1. Introduction

We will consider the classical model of an insurance risk business, i.e. where the claim occur according to a Poisson process $N = \{N(t);\, t \geq 0\}$ with intensity α and the costs of the claims are described by a sequence $\{Z_k\}_1^\infty$ of independent and identically distributed random variables, having the common distribution function F.

The total amount of claims paid by the company in the interval $(0, t]$ is then described by the *claim process*

$$Y(t) = \sum_{k=1}^{N(t)} Z_k, \quad \left(\sum_{k=1}^{0} Z_k \stackrel{\text{def}}{=} 0 \right).$$

[1]Invited lecture.

[2]This paper was first published in the Proceedings of the conference "Probabilistic Analysis of Rare Events: Theory and Problems of Safety, Insurance and Ruin" ed. by V. V. Kalashnikov and A. M. Andronov. Riga Aviation University, Riga, Latvia, (1999). It's reprinting was approved by the editors of this Proceedings.

The *risk process*, X, is defined by

$$X(t) = ct - Y(t),$$

where c is a positive real constant corresponding to the premium income. We will here only treat the case with only positive risksums, i.e. we assume that $F(0) = 0$.

The *ruin probability* $\Psi(u)$ of a company facing the risk process X and having initial capital u is defined by

$$\Psi(u) = P\{u + X(t) < 0 \text{ for some } t > 0\}.$$

We have (this is, in fact, the *Pollaczek–Khinchine formula*)

$$\Psi(u) = \Big(1 - \frac{\alpha\mu}{c}\Big) \sum_{n=0}^{\infty} \Big(\frac{\alpha\mu}{c}\Big)^n \overline{F}_I^{n*}(u) = \frac{\rho}{1+\rho} \sum_{n=0}^{\infty} \Big(\frac{1}{1+\rho}\Big)^n \overline{F}_I^{n*}(u),$$

where

$$F_I(z) \stackrel{\text{def}}{=} \frac{1}{\mu} \int_0^z (1 - F(x))\, dx \quad \text{and} \quad \rho \stackrel{\text{def}}{=} \frac{c - \alpha\mu}{\alpha\mu}.$$

This presentation is based on Grandell (2000), from which the abstract above is taken. The rest of this paper may be looked upon as an extended abstract of Grandell (2000).

2. Approximations

The most famous approximation is certainly the *Cramér–Lundberg approximation*

$$\Psi(u) \sim \Psi_{\text{CL}}(u) \stackrel{\text{def}}{=} \frac{\rho\mu}{h'(R) - c/\alpha}\, e^{-Ru}, \quad u \to \infty, \tag{1}$$

i.e.

$$\lim_{u\to\infty} e^{Ru}\, \Psi(u) = \frac{\rho\mu}{h'(R) - c/\alpha},$$

where the *Lundberg exponent* R is the positive solution of

$$h(r) \stackrel{\text{def}}{=} \int_0^\infty (e^{rz} - 1)\, dF(z) = cr/\alpha.$$

This approximation, which goes back to Cramér (1930), is very accurate for large values of u. The approximation requires that the tail of F decreases at least exponentially fast, and thus for instance the lognormal and the Pareto distributions are excluded.

In order to include that last mentioned distributions it is usual to consider distributions F such that $F_I \in \mathcal{S}$, i.e.

$$\overline{F}_I^{2*}(z) \sim 2\overline{F}_I(z), \quad z \to \infty.$$

In that case

$$\Psi(u) \sim \frac{1}{\rho}\overline{F}_I(u), \quad u \to \infty. \tag{2}$$

does hold *exactly* when $F_I \in \mathcal{S}$. The approximation (2) has a much slower speed of convergence than (1).

Both (1) and (2) are practically somewhat difficult to apply, since they require full knowledge of the claim distribution.

We will here concentrate on "simple" approximations, by which we mean that the approximations only depend on some moments of F. Let

$$\zeta_k = E[Z_j^k], \quad k = 1, 2, 3$$

and note that

$$\mu \stackrel{\text{def}}{=} E[Z_j] = \zeta_1 \quad \text{and} \quad \operatorname{Var}[Z_j] = \zeta_2 - \zeta_1^2.$$

The simplest such approximation seems to be the *diffusion approximation*

$$\Psi(u) \approx \Psi_D(u) \stackrel{\text{def}}{=} e^{-2\rho\zeta_1 u/\zeta_2} \tag{3}$$

which goes back to Hadwiger (1940). It is nowadays derived by application of weak convergence of the compound Poisson process to a Wiener process, from which it follows that (3) may be used if ρ is small and u is large in such a way that u and ρ^{-1} are of the same order.

The *De Vylder approximation*, proposed by De Vylder (1978), is based on the idea to replace the risk process X with a risk process $\widetilde{X}$ with exponentially distributed claims such that

$$E[X^k(t)] = E[\widetilde{X}^k(t)] \quad \text{for } k = 1, 2, 3.$$

We are led to the approximation

$$\Psi_{\mathrm{DV}}(u) = \frac{3\zeta_2^2}{3\zeta_2^2 + 2\zeta_1\zeta_3\rho} \exp\left\{-\frac{6\zeta_1\zeta_2\rho u}{3\zeta_2^2 + 2\zeta_1\zeta_3\rho}\right\}.$$

Other simple approximations are:

The Rényi approximation:

$$\Psi_R(u) \stackrel{\text{def}}{=} \frac{1}{1+\rho} \exp\left\{-\frac{2\rho\zeta_1 u}{\zeta_2(1+\rho)}\right\}.$$

It is shown by Kalashnikov (1997) that

$$\sup_u |\Psi(u) - \Psi_R(u)| \le \frac{4\rho\zeta_1\zeta_3}{3\zeta_2^2(1+\rho)} \quad \text{for all } \rho > 0.$$

The exponential approximation:

$$\Psi_E(u) \stackrel{\text{def}}{=} \exp\left\{-1 - \frac{2\rho\zeta_1 u - \zeta_2}{\sqrt{\zeta_2^2 + \frac{4}{3}\rho\zeta_1\zeta_3}}\right\}.$$

The Lundberg approximation:

$$\Psi_L(u) \stackrel{\text{def}}{=} e^{-2\zeta_1\rho u/\zeta_2}\left[1 + \left(\rho u - \frac{\zeta_2}{2\zeta_1}\right)\frac{4\rho\zeta_1^2\zeta_3}{3\zeta_2^3}\right],$$

where the index L stands for Ove Lundberg – the son of Filip Lundberg;

The Beekman–Bowers approximation:

$$\Psi_{\text{BB}}(u) \stackrel{\text{def}}{=} \frac{1}{1+\rho}\int_{\beta u}^{\infty} \frac{x^{\gamma-1}}{\Gamma(\gamma)}\, e^{-x}\, dx,$$

where

$$\beta = \frac{2\zeta_1\rho}{\zeta_2 + (\frac{4\zeta_1\zeta_3}{3\zeta_2} - \zeta_2)\rho} \quad \text{and} \quad \gamma = \frac{1+\rho}{1 + (\frac{4\zeta_1\zeta_3}{3\zeta_2^2} - 1)\rho}\,.$$

3. Comparisons of the approximations

For any approximation $\Psi_A(u)$ of $\Psi(u)$ we consider the *relative error* $\mathcal{E}_A(u)$ given by

$$\mathcal{E}_A(u) \stackrel{\text{def}}{=} \frac{\Psi_A(u) - \Psi(u)}{\Psi(u)} \tag{4}$$

or, when $\Psi(u)$ is unknown,

$$\mathcal{E}_{A,\text{CL}}(u) \stackrel{\text{def}}{=} \frac{\Psi_A(u) - \Psi_{\text{CL}}(u)}{\Psi_{\text{CL}}(u)}. \tag{5}$$

We will always regard $\mathcal{E}_{A,\text{CL}}(u)$ for a fixed value of ρu and small values of ρ and then consider the leading term in the Taylor expansion around $\rho = 0$.

The result of those expansions will "often" be of the form

$$\mathcal{E}_{A,\text{CL}}(u) = c_A(\rho u - d_A)\rho^k + O(\rho^{k+1}), \quad k = 1, 2.$$

Routine calculations yield:

$$\mathcal{E}_{D,\mathrm{CL}}(u) = -\frac{4\zeta_1^2\zeta_3}{3\zeta_2^3}\left(\rho u - \frac{\zeta_2}{2\zeta_1}\right)\rho + O(\rho^2);$$

$$\mathcal{E}_{R,\mathrm{CL}}(u) = -\frac{4\zeta_1^2\zeta_3 - 6\zeta_1\zeta_2^2}{3\zeta_2^3}\left(\rho u - \frac{\zeta_2}{2\zeta_1}\right)\rho + O(\rho^2);$$

$$\mathcal{E}_{E,\mathrm{CL}}(u) = -\frac{2\zeta_1^3}{9\zeta_2^5}(3\zeta_2\zeta_4 - 2\zeta_3^2)\left(\rho u - \frac{\zeta_2}{\zeta_1}\right)\rho^2 + O(\rho^3);$$

$$\mathcal{E}_{\mathrm{DV},\mathrm{CL}}(u) = -\frac{2\zeta_1^3}{9\zeta_2^5}(3\zeta_2\zeta_4 - 4\zeta_3^2)\left(\rho u - \frac{\zeta_2}{\zeta_1}\right)\rho^2 + O(\rho^3);$$

$$\mathcal{E}_{L,\mathrm{CL}}(u) = -\frac{8\zeta_1^4\zeta_3^2}{9\zeta_2^6}(\rho u - d_{L-})(\rho u - d_{L+})\rho^2 + O(\rho^3),$$

where

$$d_{L\pm} = -\frac{3\zeta_2(\zeta_2\zeta_4 - 4\zeta_3^2)}{8\zeta_1\zeta_3^2} \pm \frac{\zeta_2}{8\zeta_1\zeta_3^2}\sqrt{9\zeta_2^2\zeta_4^2 - 24\zeta_2\zeta_3^2\zeta_4 + 48\zeta_3^4}$$

Thus it seems natural to regard

$$c_{L-} = -\frac{8\zeta_1^4\zeta_3^2}{9\zeta_2^6}(d_{L-} - d_{L+}) \quad \text{and} \quad c_{L+} = -c_{L-}$$

as the correspondences to c_E or c_{DV}.

For the Beekman–Bowers approximation the Taylor expansion of $\mathcal{E}_{\mathrm{BB},\mathrm{CL}}(u)$ is more complicated. Nevertheless routine calculations yield

$$\mathcal{E}_{\mathrm{BB},\mathrm{CL}}(u) = \frac{2\zeta_1\zeta_3 - 3\zeta_2^2}{3\zeta_2^2} \cdot b(2\zeta_1\rho u/\zeta_2) \cdot \rho + O(\rho^2),$$

where

$$b(x) = x - 2\log(x) - 2E_1(x)e^x - 2\tilde{\gamma} + 1,$$

$$E_1(x) \stackrel{\mathrm{def}}{=} \int_x^\infty e^{-t}/t\,dt \text{ is the Exponential Integral,}$$

$$\tilde{\gamma} \approx 0.57721 \text{ is Euler's constant.}$$

Let $x_{\mathrm{BB}-}$ and $x_{\mathrm{BB}+}$ be the two solutions of $b(x) = 0$. By numerical solution of the equation we get

$$x_{\mathrm{BB}-} \approx 0.38435, \quad x_{\mathrm{BB}+} \approx 2.7273$$

and

$$b'(x_{\mathrm{BB}-}) \approx -1.1423, \quad b'(x_{\mathrm{BB}+}) \approx 0.43376.$$

With similar arguments as for the Lundberg approximation it is natural to let

$$d_{\mathrm{BB}\pm} = \frac{\zeta_2 x_{\mathrm{BB}\pm}}{2\zeta_1} \quad \text{and} \quad c_{\mathrm{BB}\pm} = \frac{(4\zeta_1^2\zeta_3 - 6\zeta_1\zeta_2^2)b'(x_{\mathrm{BB}\pm})}{3\zeta_2^3}.$$

Notice that this differ from the "general rule" for the other approximations regarding the relation between the order of ρ and the number of moments used. This implies that $c_{\mathrm{BB}\pm}$ is naturally compared with c_D and c_R while the involved moments and numerical comparison indicate that Ψ_{BB} ought to be compared with Ψ_L, Ψ_E and Ψ_{DV}.

Example 1 (Γ-distributed claims.) We consider the case with $\alpha = 1$, $\rho = 10\%$ and where the claims are Γ-distributed with mean 1 variance 100. Then we have $\zeta_1 = 1$, $\zeta_2 = 101$, $\zeta_3 = 20301$, and $\zeta_4 = 6110602$.

In this case we have

$$d_D = d_R = 50.5, \quad c_D = -0.0263, \quad c_R = -0.0065,$$

$$d_E = d_{\mathrm{DV}} = 101, \quad c_E = -0.0217, \quad c_{\mathrm{DV}} = -0.0043,$$

$$d_{L-} = 23.09, \quad d_{L+} = 166.47, \quad |c_{L\pm}| = 0.0495,$$

$$d_{\mathrm{BB}-} = 19.41, \quad d_{\mathrm{BB}+} = 137.73, \quad c_{\mathrm{BB}-} = -0.00739, \quad c_{\mathrm{BB}+} = 0.00281.$$

Thus the approximations Ψ_D and Ψ_R ought to work reasonably well for $u \approx 505$. Further Ψ_R ought to be better than Ψ_D. Similarly Ψ_E and Ψ_{DV} ought to work best for $u \approx 1010$ and Ψ_L for $u \approx 231$ and 1665. Finally, by comparing $c_{L\pm}$, c_E, and c_{DV}, the De Vylder approximation ought to be best. All these considerations about the simple approximations are quite in agreement with the figures given in the Table 2. The Beekman–Bowers approximation ought to work best for $u \approx 194$ and 1377. For $u \approx 194$ the picture might be "disturbed" by the fact that $\Psi_{\mathrm{BB}}(0) = \Psi(0)$. According to the figures $\mathcal{E}_{\mathrm{BB}}(u) \approx 0$ and increasing for $u \approx 1600$ which, by some good will, may be regarded as being in agreement with the figures.

u	$\Psi(u)$	$\mathcal{E}_D(u)$	$\mathcal{E}_R(u)$	$\mathcal{E}_L(u)$	$\mathcal{E}_E(u)$	$\mathcal{E}_{\mathrm{BB}}(u)$	$\mathcal{E}_{\mathrm{DV}}(u)$
0	0.90909	10.0%	0.0%	– 4.6%	– 1.6%	0.0%	– 2.8%
300	0.52114	5.9%	1.7%	0.2%	1.3%	– 0.1%	0.3%
600	0.30867	– 1.3%	0.0%	1.2%	0.8%	– 0.8%	0.2%
900	0.18287	– 8.0%	– 1.6%	1.6%	0.4%	– 0.9%	0.1%
1200	0.10834	– 14.3%	– 3.3%	1.4%	– 0.1%	– 0.7%	– 0.0%
1500	0.06418	– 20.1%	– 4.8%	0.8%	– 0.5%	– 0.2%	– 0.1%
1800	0.03803	– 25.5%	– 6.4%	– 0.2%	– 1.0%	0.3%	– 0.2%
2100	0.02253	– 30.6%	– 7.9%	– 1.5%	– 1.5%	1.0%	– 0.3%
2400	0.01335	– 35.4%	– 9.5%	– 3.2%	– 1.9%	1.8%	– 0.4%
2700	0.00791	– 39.8%	– 11.0%	– 5.0%	– 2.4%	2.7%	– 0.5%
3000	0.00468	– 43.8%	– 12.3%	– 7.0%	– 2.7%	3.6%	– 0.5%

Table 2: Γ-distributed claims.

Bibliography

1. Cramér, H., *On the Mathematical Theory of Risk.* Skandia Jubilee Volume, Stockholm, (1930). Reprinted in *Harald Cramér Collected works Vol. I.,* Ed. by A. Martin-Löf, Springer-Verlag, Berlin, (1994), 601–678.
2. De Vylder, F. E., A practical solution to the problem of ultimate ruin probability. *Scand. Actuarial J.*, (1978), 114–119.
3. Grandell, J., Simple approximations of ruin probabilities. *Insurance: Mathematics and Economics,* **26**, (2000), 157–173.
4. Hadwiger, H., Über die Wahrscheinlichkeit des Ruins bei einer grossen Zahl von Geschäften. *Arkiv für mathematische Wirtschaft- und Sozialforschung,* **6**, (1940), 131–135.
5. Kalashnikov, V., *Geometric Sums: Bounds for Rare Events with Applications.* Kluwer Academic Publishers, Dordrecht, (1997).

Department of Mathematics, Royal Institute of Technology, SE-10044 Stockholm, Sweden.
E-mail: gran@math.kth.se

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.109-120

DMYTRO GUSAK

THE DISTRIBUTION OF EXTREMA FOR RISK PROCESSES ON THE FINITE MARKOV CHAIN

Last time risk processes in Markov environments are intensively studied. Similar processes could be considered as the Poisson process on the homogeneous irreducible Markov chain $\{x(t), t \geq 0\}$ with m states. These processes $\{\xi(t), x(t)\}$ are two-dimensional homogeneous Markov processes, where the first component $\{\xi(t), t \geq 0, \xi(0) = 0\}$ is the processes with conditionally independent increments.

On the base of some factorization results for such processes the relations for the distributions of extrema are precised. Under corresponding conditions the relation for the distribution of the absolute minimum of $\xi(t)$ is established. This distribution defines the ruin probability for corresponding risk processes on Markov chain.

2000*Mathematics Subject Classifications.* 60J50, 60J70, 60K10, 60K15.

Key words and phrases. Risk process in Markov environments, ruin probability, semi-continuous Poisson process, extrema of process, distribution of absolute minimum, the infinitely divisible factorization identity, Pollachek-Khinchin formula.

Many papers (see Grandell (1981), Asmussen (1989, 1994), Shmidly (1998)) are devoted to risk processes in Markov environments and to the distributions of their functionals. Similar processes could be considered as the Poisson processes on the finite irreducible Markov chain $\{x(t),\ t \geq 0\}$ with infinitesimal matrix $\mathbb{Q} = \Lambda[\mathbb{P} - \mathbb{I}]$ and with the transition probabilities $p_{kr}(t)\ \ (k, r = \overline{1, m})$

$$\mathbb{P}(t) = \|p_{kr}(t)\| = e^{t\mathbb{Q}};\ \mathbb{P}_0 = \lim_{s\to 0} s(s\mathbb{I} - \mathbb{Q})^{-1} = \begin{pmatrix} \pi_1 & \dots & \pi_m \\ \dots & \dots & \dots \\ \pi_1 & \dots & \pi_m \end{pmatrix}.$$

These processes $Z(t) = \{\xi(t), x(t)\}$ (see. Gusak (1995)) are 2-dimentional homogeneous Markov processes, where $\{\xi(t), t \geq 0, \xi(0) = 0\}$ is the process

with conditionally independent increments; $\xi(t)$ is composed of Poisson processes $\xi_k(t), k = \overline{1,m}$ with the cumulant function

$$\psi_k(\alpha) = i\alpha a_k - \frac{b_k^2}{2}\alpha^2 + \lambda_k' \int_{-\infty}^{0} (e^{i\alpha x} - 1) dF_k(x). \tag{1}$$

The evolution of $Z(t)$ is defined by the ch.f.

$$\Phi_t(\alpha) = \|E[e^{i\alpha\xi(t)}, x(t) = r | x(0) = k]\| = e^{t\Psi(\alpha)},$$

$$\Psi(\alpha) = \|\delta_{kr}\psi_k(\alpha)\| + \mathbb{Q}. \tag{2}$$

The ruin probabilities for the considered processes are discribed by the distributions of their extrema $\xi^{\pm}(t) = \sup\limits_{0\le u\le t} (\inf)\xi(u)$. Denote θ_s the random variable (r.v.): $\mathcal{P}\{\theta_s > t\} = e^{-st}, s > 0$;

$$\Phi(s,\alpha) = \|E_{kr}[e^{i\alpha\xi(\theta_s)}]\| = s\int_0^{\infty} e^{-st}\Phi_t(\alpha)dt = s(s\mathbb{I} - \Psi(\alpha))^{-1},$$

$$\Phi_{\pm}(s,\alpha) = \mathbb{E}e^{i\alpha\xi^{\pm}(\theta_s)}. \tag{3}$$

The ch.f. $\Phi_{\pm}(s,\alpha)$, are defined by the components of the factorization identity for $\Phi(s,\alpha)$. The relation for distributions of absolute extrema

$$\xi^{\pm} = \lim_{t\to\infty} \xi^{\pm}(t) = \lim_{s\to 0} \xi^{\pm}(\theta_s)$$

are establshed under corresponding conditions.

To established the results about distributions of extrema for process $\{\xi(t), x(t)\}$ on Markov chain we remind some of them for usual processes with independent increments $\{\xi(t), \xi(0) = 0; t \ge 0\}$. We denote the characteristic function (ch.f.) and the corresponding cumulant function (c.f.)

$$\varphi_t(\alpha) = Ee^{i\alpha\xi(t)} = e^{t\psi(\alpha)}, \tag{4}$$

$$\psi(\alpha) = i\alpha a - \frac{\sigma^2}{2}\alpha^2 + \int_{-\infty}^{\infty} [e^{i\alpha x} - 1 - i\alpha x\chi(|x| \le 1)]\Pi(dx). \tag{5}$$

According to the identity of the infinitely divisible factorazation for ch.f.

$$\varphi(s,\alpha) = Ee^{i\alpha\xi(\theta_s)} = \frac{s}{s - \psi(\alpha)} \tag{6}$$

the following decomposition is true

$$\varphi(s,\alpha) = \varphi_+(s,\alpha)\varphi_-(s,\alpha), \tag{7}$$

where $\varphi_{\pm}(s,\alpha) = Ee^{i\alpha\xi^{\pm}(\theta_s)} = Ee^{i\alpha(\xi(\theta_s)-\xi^{\mp}(\theta_s))}$. If $\Pi(dx) \equiv 0$ for $x > 0$ in (5), then $\xi(t)$ has only negative jumps and $\xi(t)$ is upper semi-continuous. In this case we denote

$$\rho(s) = P'(s,+0)\overline{P}^{-1}(s,0),$$

$$P(s,x) = P\{\xi(\theta_s) < x\},\ \overline{P}(s,x) = 1 - P(s,x),$$

$$P'(s,x) = \frac{\partial}{\partial x}P(s,x), x \neq 0,$$

$$P_{\pm}(s,x) = P\{\xi^{\pm}(\theta_s) < x\}\ (\pm x \geq 0).$$

If $\sigma^2 = 0, a > 0,\ P'(s,+0) - P'(s,-0) = sa^{-1}$.

The assertion about distribution of $\xi^{\pm}(\theta_s)$ and $\xi^{\pm}$ follows from (Korolyuk (1975), Bratijchuk, Gusak (1990), Gusak (1985)).

Proposition. *If $\xi(t)$ is upper semi-continuous process, then the distribution of $\xi^{+}(\theta_s)$ is exponential*

$$\varphi_{+}(s,\alpha) = \frac{\rho(s)}{\rho(s) - i\alpha},\ P\{\xi^{+}(\theta_s) > x\} = e^{-\rho(s)x}, x > 0. \tag{8}$$

The distrtibution of $\xi^{-}(\theta_s)$ is defined by the relation

$$P\{\xi^{-}(\theta_s) < x\} = P\{\xi(\theta_s) < x\} + \rho^{-1}(s)P'(s,x), x < 0. \tag{9}$$

If $\sigma^2 = 0, a > 0$, then

$$p_{-}(s) = P\{\xi^{-}(\theta_s) = 0\} = s(a\rho(s))^{-1}, s > 0. \tag{10}$$

From this proposition the relations for the distributions of $\xi^{\pm}$ follow,

$$\xi^{\pm} = \sup_{t<\infty}(\inf)\xi(t) = \lim_{t\to\infty}\xi^{\pm}(t) = \lim_{s\to 0}\xi^{\pm}(\theta_s).$$

There are the following three cases:

1. $m_1 = E\xi(1) > 0$. In this case

$$m_1\rho(s) \approx s, s \to 0;\ s^{-1}\rho(s) \underset{s\to 0}{\longrightarrow} \rho'(0) = m_1^{-1},$$

$$\varphi_{+}(s,\alpha) \underset{s\to 0}{\longrightarrow} 0,\ P\{\xi^{+} = +\infty\} = 1. \tag{11}$$

The absolute minimum ξ^{-} has the distribution

$$P\{\xi^{-} < x\} = \frac{1}{\rho'(0)}\frac{d}{dx}\int_{-\infty}^{0} P\{\xi(t) < x\}dt,\ x < 0. \tag{12}$$

For $\sigma^2 = 0, a > 0$

$$p_- = P\{\xi^- = 0\} = \frac{m_1}{a} > 0 \text{ (if } \sigma^2 > 0, p_- = 0). \tag{13}$$

If $0 < \lambda = \int_{-\infty}^{0} \Pi(dx) < \infty$, $\sigma^2 = 0, a > 0$, $m_1 > 0$, then

$$\varphi_-(\alpha) = Ee^{i\alpha\xi^-} = \frac{a - \lambda\tilde{F}(0)}{a - \lambda\tilde{F}(\alpha)}, \quad F(x) = P\{\xi_k < x\},$$

$$\tilde{F}(\alpha) = \int_{-\infty}^{0} e^{i\alpha x} F(x)dx. \tag{14}$$

Let $\tilde{\varphi}(\alpha) = \frac{\tilde{F}(\alpha)}{\tilde{F}(0)} = Ee^{i\alpha\tilde{\xi}}$ $(\tilde{\xi} \le 0)$, then Pollachek-Khinchin formula is true

$$Ee^{i\alpha\xi^-} = \frac{p_-}{1 - q_-\tilde{\varphi}(\alpha)} = p_- \sum_{k=0}^{\infty} (q_-\tilde{\varphi}(\alpha))^k, q_- = P\{\xi^- < 0\} = 1 - p_-. \tag{15}$$

2. $m_1 < 0$. In this case $\rho(s) \underset{s\to 0}{\longrightarrow} \rho > 0$ (ρ is the root of the equation: $\lambda \int\limits_{-\infty}^{0} e^{\rho x} F(x)dx = a + \rho\frac{\sigma^2}{2}$). Absolute maximum ξ^+ has the distribution

$$\varphi_+(\alpha) = Ee^{i\alpha\xi^+} = \lim_{s\to 0} \varphi_+(s, \alpha) = \frac{\rho}{\rho - i\alpha}, \quad P\{\xi^+ > x\} = e^{-\rho x}, x > 0; \tag{16}$$

$$\varphi_-(s, \alpha) \to 0 \quad s \to 0, \ P\{\xi^- = -\infty\} = 1.$$

Hence ξ^- has the degenerate distribution.

3. $m_1 = 0$. In this case

$$\rho^2(s) \approx \frac{2s}{\mathcal{D}\xi(1)}, \ \rho(s) \approx \sqrt{\frac{2s}{\mathcal{D}\xi(1)}}, \ \varphi_\pm(s, \alpha) \underset{s\to 0}{\longrightarrow} 0. \tag{17}$$

Hence $P\{\xi^\pm = \pm\infty\} = 1$.

If $\pi = \text{diag}\,(\pi_1, \ldots, \pi_m)$, then $\mathbb{P}_0 = \|1\|\pi$. For risk processes in Markov environments ($b_k^2 = 0, a_k > 0$) in [2; 3] the averaged process $\xi_0(t)$ was introduced with the help of

$$a_0 = \sum_{k=1}^{m} \pi_k a_k, \ \lambda_0 = \sum_{k=1}^{m} \lambda_k\pi_k, F_0(x) = \frac{1}{\lambda_0} \sum_{k=1}^{m} \pi_k\lambda_k F_k(x).$$

The ruin probabilities for processes in Markov environments are defined by distributions of ladder heights of $\xi_0(t)$.

For general homogeneous processes with independent increments on Markov chain $\{\xi(t), x(t)\}$ instead of the identity (7) the dual identies (see [5]) are true on $\operatorname{Im}\alpha = 0$

$$\Phi(s,\alpha) := \mathbb{E}e^{i\alpha\xi(\theta_s)} = \begin{cases} \Phi_+(s,\alpha)\mathbb{P}_s^{-1}\check{\Phi}_-(s,\alpha), \\ \Phi_-(s,\alpha)\mathbb{P}_s^{-1}\check{\Phi}_+(s,\alpha), \end{cases}$$

where

$$\begin{aligned} \Phi_\pm(s,\alpha) &= \mathbb{E}e^{i\alpha\xi^\pm(\theta_s)},\ \mathbb{P}_s = s(s\mathbb{I}-\mathbb{Q})^{-1}, \\ \check{\Phi}_\pm(s,\alpha) &= \mathbb{E}e^{i\alpha(\xi(\theta_s)-\xi^\pm(\theta_s))}, \end{aligned}$$

are the ch.f.of extrema and their complements.

We introduce the notations (by means of an exponential r.v. θ_s, which does not depend on $\xi(t)$ and $x(t)$)

$$\left.\begin{aligned} \mathbb{P}(s,x) &= \|P\{\xi(\theta_s) < x, x(\theta_s) = r/x(0) = k\}\|, \\ \Phi(s,\alpha) &= \textstyle\int_{-\infty}^{\infty} e^{i\alpha x} d\mathbb{P}(s,x) = s(s\mathbb{I}-\Psi(\alpha))^{-1}, \\ \Psi(\alpha) &= i\alpha\mathbb{A} - \tfrac{1}{2}\mathbb{B}^2\alpha + \Lambda' \textstyle\int_{-\infty}^{0}(e^{i\alpha x}-1)d\mathbb{F}(x) + \mathbb{Q}, \\ \mathbb{A} &= \|\delta_{kr}a_k\|,\ \mathbb{B}^2 = \|\delta_{kr}b_k^2\|,\ \mathbb{F}(x) = \|\delta_{kr}F_k(x)\|, \end{aligned}\right\} \tag{19}$$

$$a_k > 0,\ \text{if } b_k^2 = 0,$$

$$\mathbb{P}'(s,x) = \frac{\partial}{\partial x}\mathbb{P}(s,x), x \neq 0, \mathbb{R}_+(s) = \mathbb{P}'(s,+0)\overline{\mathbb{P}}^{-1}(s,0),$$

$$\overline{\mathbb{P}}(s,x) = \mathbb{P}_s - \mathbb{P}(s,x),\ \check{\mathbb{R}}(s) = \overline{\mathbb{P}}^{-1}(s,0)\mathbb{P}'(s,+0). \tag{20}$$

We suppose that $\xi(t)$ is upper semi-continuous process.

Let $\tau^+(z)$ be the time of the first crossing of $z > 0$ by $\xi(t)$

$$\tau^+(z) = \inf\{t > 0 : \xi(t) > z\}, z > 0.$$

The distribution of $\tau^+(z)$ is tightly connected with the distribution of $\xi^+(t)$. In the semi-continuous case the positive (negative) components of (18) could be precised. Firstly we precise the relations for the positive components.

Theorem 1. *Let $\{\xi(t), x(t)\}$ be upper semi-continuous process with ch.f. (19). Then $\{\tau^+(z), x(\tau^+(z))\}$ is a non-decreasing homogeneous "process" with respect to $z \geq 0$ on Markov chain $y(z) = x(\tau^+(z))$ with m states. The generating function (g.f.)*

$$\mathbb{T}_+(s,z) = \mathbb{E}e^{-s\tau^+(z)} = \|E[e^{-s\tau^+(z)}, y(z) = r|y(0) = k]\| \tag{21}$$

is represented by an exponential matrix

$$\mathbb{T}_+(s,z) = \exp\{-z\mathbb{R}_+(s)\}, z > 0, \tag{22}$$

where $\mathbb{R}_+(s)$ *is defined in (20).*

The distribution of $\xi^+(\theta_s)$ *and* $\mathbb{T}_+(s,z)$ *are connected by the relations*

$$\mathbb{P}\{\xi^+(\theta_s) > z, x(\theta_s) = r | x(0) = k\} = \mathbb{T}_+(s,z)\mathbb{P}_s,\ z > 0, \tag{23}$$

$$\Phi_+(s,\alpha) = (\mathbb{I} - i\alpha\mathbb{R}_+^{-1}(s))^{-1}\mathbb{P}_s. \tag{24}$$

The cumulant of $\{\tau^+(z), y(z)\}$ – $\mathbb{R}_+(s)$ *satisfies the "leftside" equation*

$$\Psi(-i\mathbb{R}_+(s)) \stackrel{def}{=} \mathbb{A}\mathbb{R}_+ + \frac{1}{2}\mathbb{B}^2\mathbb{R}_+^2 + \Lambda\int_{-\infty}^{0} d\mathbb{F}(x)(e^{x\mathbb{R}_+(s)} - \mathbb{I}) + \mathbb{Q} = s\mathbb{I}. \tag{25}$$

The ch.f.of the complement of the minimum $\check{\xi}(\theta_s) = \xi(\theta_s) - \xi^-(\theta_s)$ *is defined by the relation*

$$\check{\Phi}_+(s,\alpha) = \mathbb{E}e^{i\alpha\check{\xi}(\theta_s)} = \mathbb{P}_s(\mathbb{I} - i\alpha\check{\mathbb{R}}^{-1}(s))^{-1}, \tag{26}$$

where $\check{\mathbb{R}}(s)$ *is defined in (20).*

Proof is based on the additive property of $\tau^+(z)$.

$$\tau^+(x+z) \doteq \tau^+(x) + \tau^+(z),\ x > 0,\ z > 0.$$

That means the g.f. (21) satisfies the equation

$$\mathbb{T}(s, x+z) = \mathbb{T}(s,x)\mathbb{T}(s,z),\ x > 0,\ z > 0,\ \mathbb{T}(s,0) = \mathbb{I},$$

which has the matrix resolution (22). The equation for $\mathbb{P}'(s,x)$ $(x \neq 0)$ defins $\mathbb{R}_+(s)$:

$$\mathbb{P}'(s,z) = \mathbb{E}e^{-s\tau^+(z)}\mathbb{P}'(s,+0) = e^{-z\mathbb{R}_+(s)}\mathbb{P}'(s,+0) \Rightarrow$$

$$\Rightarrow \overline{\mathbb{P}}(s,0) = \mathbb{R}_+^{-1}(s)\mathbb{P}'(s,+0).$$

It is evident that

$$P_{kr}\{\xi^+(t) > z\} = \sum_{j=1}^{m}\int_0^t P_{kj}\{\tau^+(z) \in du\}P_{jr}(t-u). \tag{27}$$

After Laplace transformation the relation (27) implies (23). From (22) and (23) after Stiltjes transformation the relation (24) is proved. The similar relation is established for the distribution of $\check{\xi}(\theta_s) = \xi(\theta_s) - \xi^-(\theta_s)$

$$\mathbb{P}\{\check{\xi}(\theta_s) > x\} = \mathbb{P}_s \exp\{-\check{\mathbb{R}}(s)x\},\ x > 0,$$

which has the ch.f.(26).

To find the distributions of

$$\xi^+ = \lim_{s\to 0}\xi^+(\theta_s),\ \check{\xi} = \lim_{s\to 0}\check{\xi}(\theta_s)$$

we introduce for $m_k = E\xi_k(1)$ and $E\xi_k^2(1)$ the averaged moments

$$m_1^0 = \sum_{k=1}^m \pi_k m_k,\ m_2^0 = \sum_{k=1}^m \pi_k E\xi_k^2(1) < \infty. \tag{28}$$

If $m_1^0 < 0$, then

$$\lim_{s\to 0}\mathbb{R}_+(s) = \mathbb{R}_+(0) =$$

$$= \frac{d}{dx}\int_0^\infty \mathbb{P}\{\xi(t) < x\}dt|_{x=+0}(\int_0^\infty \mathbb{P}\{\xi(t) > 0\}dt)^{-1}, \tag{29}$$

$$\lim_{s\to 0}\check{\mathbb{R}}(s) = \check{\mathbb{R}}(0) = (\int_0^\infty \mathbb{P}\{\xi(t) > 0\}dt)^{-1}\frac{d}{dx}\int_0^\infty \mathbb{P}\{\xi(t) < x\}dt|_{x=+0}.$$

If $m_1^0 > 0$, then

$$\mathbb{R}_+(0) = -\frac{d}{dx}\int_0^\infty \mathbb{P}\{\xi(t) < x\}dt|_{x=+0}\mathbb{Q},$$

$$\check{\mathbb{R}}(0) = -\mathbb{Q}\frac{d}{dx}\int_0^\infty \mathbb{P}\{\xi(t) < x\}dt|_{x=+0}. \tag{30}$$

If $m_1^0 = 0$, then $|\mathbb{R}_+(s)| \underset{s\to 0}{\longrightarrow} 0$, $|\check{\mathbb{R}}(s)| \underset{s\to 0}{\longrightarrow} 0$.

Corollary 1. *If $m_1^0 < 0$, then*

$$\left.\begin{array}{l}\mathbb{E}[e^{i\alpha\xi^+}, \tau^+(z) < \infty] = (\mathbb{I} - i\alpha\mathbb{R}_+^{-1}(0))^{-1}\mathbb{P}_0, \\ \mathbb{P}\{\xi^+ > z\} = \exp\{-z\mathbb{R}_+(0)\}\mathbb{P}_0,\ z > 0,\end{array}\right\} \tag{31}$$

where $\mathbb{R}_+(0)$ is defind in (29). Analogously

$$\mathbb{P}\{\check{\xi} > z\} = \mathbb{P}_0\exp\{-z\check{\mathbb{R}}(0)\}, z > 0,$$

where $\check{\mathbb{R}}(0)$ is defined in (29). In the case $m_1 < 0$

$$\mathbb{P}\{\tau^+(z) < \infty\} = \|P\{\tau^+(z) < \infty, y(z) = r|y(0) = k\}\| =$$

$$= \exp\{-z\mathbb{R}_+(0)\} \le \mathbb{P}_0. \tag{32}$$

If $m_1^0 \ge 0$, then

$$\mathbb{P}\{\xi^+ > z\} = \mathbb{P}_0, \forall z > 0.$$

If $m_1 > 0$, then $\mathbb{R}_+(0)$ (see (30)) defines the transition matrix for $\{y(z), z \ge 0\}$

$$\mathbb{P}\{y(z) = r|y(0) = k\} = \exp\{-z\mathbb{R}_+(0)\}, z > 0. \tag{33}$$

For negative components of factorization the following assertion is true.

Theorem 2. *If $\xi(t)$ is $\{\xi(t), x(t)\}$ is upper semi-continuous process with ch.f. (19), then*

$$\left.\begin{array}{l} \Phi_-(s,\alpha) = \mathbb{E}e^{i\alpha\xi^-(\theta_s)} = \overline{\mathbb{P}}(s,0) + \mathbb{E}_-[e^{i\alpha\xi(\theta_s)}](\mathbb{I} - i\alpha\check{\mathbb{R}}^{-1}(s)), \\ \mathbb{P}\{\xi^-(\theta_s) < x\} = \mathbb{P}\{\xi(\theta_s) < x\} - \mathbb{P}'(s,x)\check{\mathbb{R}}^{-1}(s), x < 0, \\ \mathbb{P}\{\xi^-(\theta_s) = 0\} = s\mathbb{A}^{-1}\check{\mathbb{R}}^{-1}(s), \ \textit{if } \mathbb{B}^2 = 0, \mathbb{A} > 0. \end{array}\right\} \tag{34}$$

If $\mathbb{B}^2 \geq 0, \mathbb{A} > 0$, then

$$\mathbb{P}\{\xi^-(\theta_s) = 0\} = s\mathbb{A}_*^{(-1)}\check{\mathbb{R}}^{-1}(s), \ \mathbb{P}'(s+0) - \mathbb{P}'(s-0) = s\mathbb{A}_*^{(-1)}, \tag{35}$$

$$\mathbb{A}_*^{(-1)} = \|\delta_{kr}a_k^{-1}\delta(a_k > 0, b_k = 0)\|; \ (\mathbb{A}_*^{(-1)} = \mathbb{O}, \ \textit{if } \mathbb{B}^2 > \mathbb{O}).$$

The distribution of $\overline{\xi}(\theta_s) = \xi(\theta_s) - \xi^+(\theta_s)$ is defined by the relation

$$\check{\Phi}_-(s,\alpha) = \mathbb{E}e^{i\alpha\overline{\xi}(\theta_s)} = \mathbb{P}(s,0) + (\mathbb{I} - i\alpha\mathbb{R}_+^{-1}(s))\mathbb{E}_-[e^{i\alpha\xi(\theta_s)}], \tag{36}$$

where

$$\mathbb{E}_-[e^{i\alpha\xi(\theta_s)}] = \mathbb{E}[e^{i\alpha\xi(\theta_s)}, \xi(\theta_s) < 0],$$
$$\mathbb{P}\{\overline{\xi}(\theta_s) = 0\} = s\mathbb{A}_*^{(-1)}\mathbb{R}_+^{-1}(s). \tag{37}$$

From the Theorem 2 the next assertion follows after the limit passage $(s \to 0)$.

Corollary 2. *If $m_1^0 > 0$, $\mathbb{B}^2 = 0$, then under conditions of Theorem 2*

$$\mathbb{P}_0^- = \mathbb{P}\{\xi^- = 0\} = \lim_{i\alpha\to\infty} m_1^0 i\alpha\Psi^{-1}(\alpha)\mathbb{P}_0 = \|p_{kr}^-\|, \tag{37}$$

$$p_{kr}^- = \frac{m_1^0\pi_r}{a_k}, \ p_k^- = \sum_{r=1}^m p_{kr}^- = \frac{m_1^0}{a_k}. \ (\textit{If } \mathbb{B}^2 > 0, \ p_{kr}^- = 0 \ \ \forall k, r).$$

The distribution of negative values for ξ^- is defined by the truncate ch.f.

$$\mathbb{E}[e^{i\alpha\xi^-}, \xi^- < 0] = m_1^0[i\alpha\Psi^{-1}(\alpha)]^-\mathbb{C}\mathbb{P}_0, \tag{38}$$

where $\mathbb{C}\mathbb{P}_0 = \|c_k^0\pi_r\|$) is defined from the condition

$$m_1^0 \lim_{\alpha\to 0}[i\alpha\Psi^{-1}(\alpha)]^-\mathbb{C}\mathbb{P}_0 = \mathbb{P}_0 - \mathbb{P}^- = \overline{\mathbb{P}}^- = \|\overline{p}_{kr}^-\|. \tag{39}$$

If $m_1^0 \leq 0$, then $\mathbb{P}\{\xi^- < z\} = \mathbb{P}_0, \forall z < 0$.

We remind that

$$[C + \int_{-\infty}^{\infty} e^{i\alpha x}G(x)dx]^{\pm} = \pm\int_0^{\pm\infty} e^{i\alpha x}G(x)dx.$$

To convert $\Psi(\alpha)$ and to realise the limit passage in (37) we put $i\alpha = r$

$$\tilde{\mathbb{F}}(r) = \int_{-\infty}^{0} e^{rx}\mathbb{F}(x)dx,$$

then we receive instead of $\Psi(\alpha)$ (with $\mathbb{B}^2 = 0$)

$$\mathbb{K}(r) = r(\mathbb{A} - \Lambda'\tilde{\mathbb{F}}(r)) + \mathbb{Q}.$$

It follows from Korolyuk, Turbin (1978)

$$\lim_{r\to 0} r\mathbb{K}^{-1}(r) = \mathbb{P}_0(m_1^0)^{-1},$$

and for $r \to \infty$ $\tilde{\mathbb{F}}(r) \to 0$, hence

$$\lim_{r\to\infty} m_1^0 r\mathbb{K}_r^{-1}(r)\mathbb{P}_0 = \mathbb{P}_0^- = m_1^0\|\frac{\pi_i}{a_k}\|,\ m_1^0 a_k^{-1}\pi_i = p_{ki}^-.$$

From the second factorization identity (18)

$$s(s\mathbb{I} - \Psi(\alpha))^{-1} = \mathbb{E}e^{i\alpha\xi^-(\theta_s)}(\mathbb{I} - i\alpha\check{\mathbb{R}}^{-1}(s))^{-1}, \tag{40}$$

taking into a count that $\det\Psi(\alpha) = i\alpha D(i\alpha)$, we obtain the relation, equivalent to (38).

$$\mathbb{E}[e^{i\alpha\xi^-}, \xi^- < 0] = [\frac{1}{\mathcal{D}(i\alpha)}\mathrm{Adj}\{\Psi(\alpha)\}]^-\mathbb{C}\mathbb{P}_0 = \|\varphi_{kr}^-(\alpha)\|. \tag{41}$$

$$\varphi_{kr}^-(\alpha) = \varphi_k^-(\alpha)\pi_r;\ k = \overline{1;m},\ r = \overline{1;m},$$

$$\varphi_k^-(\alpha) = [\frac{1}{\mathcal{D}(i\alpha)}\sum_{r=1}^{m}\psi_{kr}^{adj}(\alpha)]^- c_k^0,\ \varphi_k^-(0) = \sum_{r=1}^{m}\overline{p}_{kr}^- = \overline{p}_k^-. \tag{42}$$

Constants c_k^0 are defined from (39).

Remark. Let $\xi_{(u)}(t) = u + \xi(t)$ $(u > 0)$ be a risk process in Markov environments, generated by Markov chain $x(t)$ with m states. We consider it as the Poisson process on the chain: $\{\xi_{(u)}(t), x(t)\}$. The matrix of ruin probabilities on finite interval $[0;T]$

$$\mathbb{Q}(u,T) = \mathbb{P}\{\xi_{(u)}(t) < 0 \text{ for some } t \in [0,T]\} =$$

$$= \|P\{\xi_{(u)}(t) < 0 \text{ for some } t \in [0,T], x(T) = r|x(0) = k\}\|$$

is defined by the distribution of minimum for $\xi(t)$. $\mathbb{Q}(u,T) = \mathbb{P}\{\xi^-(T) < -u\} = \|P\{\xi^-(T) < -u, x(T) = r|x(0) = k\}\|$ $(u > 0)$. The matrix of classic ruin probabilities

$$\mathbb{Q}(u) = \lim_{T\to\infty}\mathbb{Q}(u,T) = \|q_{ik}(u)\|,\ q_{ik}(u) = q_i(u)\pi_k$$

is defined by the distribution of the absolute minimum

$$\mathbb{Q}(u) = \mathbb{P}\{\xi^- < -u\} \ (u > 0),$$

$$q_k(u) = \mathbb{P}\{\xi^- < -u, x(0) = k\}, \ k = \overline{1, m}.$$

Example. Let $\{\xi(t), x(t)\}$ be the process defined on two-state Markov chain $x(t)$:

$$\Lambda = \mathbb{I}, \ \mathbb{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ \mathbb{Q} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}.$$

The process $\xi(t)$ is composed from $\xi_j(t) = (1 + \delta_j)t - S(t)$ with corresponding cumulant

$$k_j(r) = rg_j(r), \ g_j(r) = \delta_j + 1 - \tilde{F}(r), \ \tilde{F}(0) = 1, j = 1, 2; ES(1) = 1.$$

$E\xi_j(1) = \delta_j$ $(0 < \delta_j \leq 1)$ are interpretated as the relative safety loading. For this process

$$\pi_1 = \pi_2 = \frac{1}{2}, \ m_1^0 = \frac{1}{2}(\delta_1 + \delta_2) > 0.$$

Consequently the non-degenerate distribution of ξ^- existes.

To find it we convert $\mathbb{K}(r)$

$$\mathbb{K}^{-1}(r) = \frac{1}{\det \mathbb{K}(r)} \begin{pmatrix} rg_2(r) - 1 & -1 \\ -1 & rg_1(r) - 1 \end{pmatrix},$$

$$\det \mathbb{K}(r) = r\mathcal{D}(r), \ \mathcal{D}(r) = rg_1(r)g_2(r) - 2(1 - \tilde{F}(r)) - \delta_1 - \delta_2.$$

We form the matrix, which is used in (37)

$$rm_1^0 \mathbb{K}^{-1}(r)\mathbb{P}_0 = \frac{\delta_1 + \delta_2}{4\mathcal{D}(r)} \begin{pmatrix} rg_2(r) - 2 & rg_2(r) - 2 \\ rg_1(r) - 2 & rg_1(r) - 2 \end{pmatrix}. \tag{43}$$

By limit passage $(r \to \infty)$ we find out that

$$p_1^- = P_1\{\xi^- = 0\} = \frac{\delta_1 + \delta_2}{2(1 + \delta_1)} = \frac{m_1^0}{1 + \delta_1} > 0,$$

$$p_2^- = P_2\{\xi^- = 0\} = \frac{m_1^0}{1 + \delta_2} > 0.$$

By adding elements of the matrix (43) we obtain the functions which are subjected to the proective operator:

$$f_1(r) = m_1^0 \frac{rg_2(r) - 2}{\mathcal{D}(r)}, \ f_2(r) = m_1^0 \frac{rg_1(r) - 2}{\mathcal{D}(r)},$$

$$f_j(r) \underset{r\to\infty}{\longrightarrow} p_j^- \ (j=1,2).$$

According to (42) we find truncated g.f.

$$E_i[e^{r\xi^-}, \xi^- < 0] = c_i^0[f_i(r)]^- \ (i=1,2),$$

$$c_i^0 = \bar{p}_i^-([f_i(r)]^-)^{-1}|_{r=0}.$$

In particular case, when $F(x) = e^x \ (x<0), \tilde{F}(r) = \frac{1}{1+r}$,

$$\mathcal{D}(r) = \frac{1}{(r+1)^2}[r^3(1+\delta_1)(1+\delta_2) + p_2(r)] = \frac{A(r)}{(1+r)^2},$$

$$p_2(r) = 2r^2(\delta_1\delta_2 - 1) + r(\delta_1\delta_2 - 2(1+\delta_1+\delta_2)) - \delta_1 - \delta_2,$$

$A(r)$ has two negative roots $r_{1,2} = -\rho_{1,2}$ and one positive $r_0 > 0$. That's wy

$$f_i(r) = p_i^- + \frac{A_i}{r+\rho_1} + \frac{B_i}{r+\rho_2} + \frac{C_i}{r-r_0}.$$

By the proectiv operation (see (39)) we find

$$\mathbb{E}_i[e^{r\xi^-}, \xi^- < 0] = c_i^0(\frac{A_i}{r+\rho_1} + \frac{B_i}{r+\rho_2}) \ (i=1,2),$$

$$P_i\{\xi^- < x\} = c_i^0(\rho_1^{-1}A_ie^{\rho_1 x} + \rho_2^{-1}B_ie^{\rho_2 x}), \ x<0, \tag{44}$$

$$c_i^0 = \frac{\rho_1\rho_2(1-p_i^-)}{\rho_2 A_i + \rho_1 B_i}, \ p_i^- = \frac{m_1^0}{1+\delta_i}, i=1,2.$$

If the risk process is $\xi_{(u)}(t) = u + \xi(t)$, then for the process $\{\xi_{(u)}(t), x(t)\}$ according to (44) the ruin probabilities are defined by the relations $(u>0)$

$$q_i(u) = P_i\{\xi^- < -u\} = c_i^0\left(\frac{A_i}{\rho_1}e^{-\rho_1 u} + \frac{B_i}{\rho_2}e^{-\rho_2 u}\right), i=1,2.$$

If $\delta_1 = \frac{1}{21}, \delta_2 = 1$, then $r_0 = 2$,

$$\rho_{1,2} = \frac{12 \mp \sqrt{23}}{22}; \ \rho_1 \approx 0,578; \ \rho_2 \approx 0,76,$$

$$p_1^- = \frac{1}{2}, \ p_2^- = \frac{11}{42}, \ c_1^0 \approx 0,83682, \ c_2^0 \approx 1,42669.$$

$$\begin{cases} A_1 \approx 0,00325497 \\ B_1 \approx 0,256236 \\ C_1 \approx 0,445055 \end{cases} \qquad \begin{cases} A_2 \approx -0,04555787 \\ B_2 \approx 0,248941 \\ C_2 \approx -0,0724097 \end{cases}$$

This example shows us that in the case, when jumps of $\xi_j(t)$ $(j = \overline{1.m})$ have the exponential distributions or $F_j(x) = P\{\xi_j < x\}$ are the distributions of Erlang type then elements of $\mathbb{K}^{-1}(r)$ are the fractionally-rational functions. In this case the factorization and projective procedure are not complicate for finding the matrix of ruin probabilities. If $F_j(x)$ are arbitrary continuous distributions $(F_j(0) = 1)$ then $\mathcal{D}(r)$ has some root $r_* > 0$ and elements of $\mathbb{K}^{-1}(r)$ have a form

$$f_{kj}(r) = \frac{g_k(r)}{r - r_*}\pi_j, \qquad g_k(r) = \int_{-\infty}^{\infty} e^{rx} G_k(x)\, dx.$$

In this case $g_k(r)$ is not necessarily fractionally-rational and

$$\left[\frac{g_k(r)}{r - r_*}\right]^- = \frac{1}{r - r_*}[g_k^-(r) - g_k^-(r_*)]$$

is easy inverted

$$G_k^*(x) = \int_{-\infty}^{x} G_k(y) e^{-r_*(x-y)} dy, \quad x < 0.$$

Bibliography

1. Grandell, G., *Aspects of Risk Theory*, Springer-Verlag, New York, (1991).
2. Asmussen, S., *Risk Theory in Markovian environments*, Scand. Act. J., No. 1, (1989), 66-100.
3. Asmussen, S., Nenriksen, T., Kliiperberg, C., *Large claims approximations for risk processes in a Markovian envirounments*, Stoch. Proc. and their Appl., **54**, (1994), 29-43.
4. Schmidli, H., *Distribution of the first ladder height of a stationary risk process perturbed by α-stable Levy motion*, Reserch Reports, Dep. of Theor. Stat., Univer.of Aarhus, Denmark, No. 394, (1998), 1-13.
5. Gusak, D. V., *Boundary problems for processes with independent increments on finite Markov chains*, Inst. of Mathematics, Kyiv, (1998) (in Ukrainian).
6. Korolyuk, V. S., *Boundary problems for compound Poisson processes*, Naukova Dumka, Kyiv, (1975) (in Russian).
7. Bratiychuk, N. S., Gusak, D. V., *Boundary problems for processes with independent increments*, Naukova Dumka, Kyiv, (1990) (in Russian).
8. Gusak, D. V., *Factorization method in boundary problems for homogeneous processes with independent increments*. Preprint 85.43., Inst. of Math., Kyiv, (1985) (in Russian).
9. Korolyuk, V. S., Turbin, A. F., *Mathematical foundations of the phase lumping of complex systems*, Naukova Dumka, Kyiv, (1978) (in Russian).

Institute of Mathematics of National Academy of Science of Ukraine, 01601, Kyiv, Ukraine.
E-mail: dmytro@imath.kiev.ua

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.121-141

THOMAS HELLSTRÖM and KENNETH HOLMSTRÖM

GLOBAL OPTIMIZATION OF COSTLY NONCONVEX FUNCTIONS, WITH FINANCIAL APPLICATIONS[1]

The paper considers global optimization of costly objective functions, i.e. the problem of finding the global minimum when there are several local minima and each function value takes considerable CPU time to compute. Such problems often arise in industrial and financial applications, where a function value could be a result of a time-consuming computer simulation or optimization. Derivatives are most often hard to obtain, and the algorithms discussed make no use of such information. Response surface methods are promising for global optimization of costly non-convex objective functions. We discuss our implementation of an algorithm by Powell and Gutmann based on the use of radial basis functions (RBF). Another interesting response surface method is the Efficient Global Optimization (EGO) method by Jones et al. We have implemented these two methods, together with the DIRECT and constrained DIRECT method by Jones in the TOMLAB optimization environment (Holmström (1999)). We discuss the application of these global optimization methods for parameter estimation in trading algorithms and in models for time series prediction.

AMS 2000 subject classifications. 90C26, 90C56, 91B28.

Key words and phrases. Mathematical Programming, Nonlinear Programming, Matlab, Mathematical Software, Algorithms, Nonconvex Optimization, Global Optimization, Time Series Predictions, Financial time series.

[1]This work is supported in part by the project "Stochastic modelling of insurance and finance processes and systems" funded by the Knowledge Foundation.

1. Introduction

The task of global optimization is to find the set of parameters x in the feasible region $\Omega \subset \mathbf{R^d}$ for which the objective function $f(x)$ obtains its smallest value. In other words, a point x^* is a *global optimizer* to $f(x)$ on Ω, if $f(x^*) \leq f(x)$ for all $x \in \Omega$. On the other hand, a point $\hat{x}$ is a *local optimizer* to $f(x)$, if $f(\hat{x}) \leq f(x)$ for all x in some neighborhood around $\hat{x}$. Obviously, when the objective function has several local minima, there could be solutions that are locally optimal but not globally optimal and standard local optimization techniques are likely to get stuck before the global minimum is reached. Therefore, some kind of global search is needed to find the global minimum with some reliability.

The global optimization page (http://solon.cma.univie.ac.at/~neum/glopt.html), maintained by Arnold Neumaier, contains many commented links to online information relevant to global optimization. There is also an online survey of different global optimization methods at http://www.cs.sandia.gov/opt/survey/. The emphasis in this paper is on problems with costly objective functions.

In our Applied Optimization and Modeling group at Mälardalen University we work in several applied areas with a need for global optimization techniques. One example is the problem of designing a passenger train, where the objective is to minimize the total mass with constraints on ride quality measures. The design parameters are the carbody mass and frequencies, the bogie frame mass and yaw damper attachment positions. In this problem, a single function value is computed by running an expensive (time-consuming) computer simulation. With a simplified model, one simulation takes more than three minutes and with an accurate model, about half an hour. In computational finance, we study the prediction of various kinds of quantities related to stock markets, like stock prices, stock volatility and ranking measures. These are noisy problems with several local minima. With the increasing use of high-frequency data, simulated trading and time series analysis of multiple data series results in costly global optimization problems as presented in Hellström and Holmström (1999), and further discussed in Section 4.

Previously we have made Matlab implementations of the DIRECT (Jones, Perttunen and Stuckman 1993), the new constrained DIRECT (Jones 2001), and the Efficient Global Optimization (EGO) (Jones, Schonlau, Welch 1998) algorithms. The implementations are part of the TOMLAB optimization environment, described in Holmström (1999a, 1999b, 1999c). The implementation of the DIRECT algorithm is further discussed and analyzed in Björkman and Holmström (1999). Recently Powell (1999) and Gutmann (1999) presented an algorithm of response surface type based on radial basis function approximation. The idea of the RBF algorithm is to use radial

basis function interpolation to define a utility function (Powell 1999). The next point, where the original objective function should be evaluated, is determined by optimizing on this utility function.

In Section 2 we describe the basic RBF algorithm and discuss some special features of the implementation. The other global optimization algorithms considered are briefly described in Section 3. In Section 4 the application of global optimization methods on financial problems are discussed.

2. The RBF algorithm

Our RBF algorithm is based on the ideas presented by Gutmann (1999), with some extensions and further development. The algorithm is implemented in the Matlab routine *rbfSolve* and described in more detail in Björkman and Holmström (2001). The RBF algorithm deals with box-bounded global optimization problems of the form

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ s/t & -\infty < x_L \leq x \leq x_U < \infty, \end{array} \tag{1}$$

where $f(x) \in \mathbf{R}$ and $x, x_L, x_U \in \mathbf{R^d}$. We assume that no derivative information is available and that each function evaluation is very expensive. For example, the function value could be the result of a time-consuming experiment or computer simulation.

2.1. Description of the Algorithm. We now consider the question of choosing the next point where the objective function should be evaluated. The idea of the RBF algorithm is to use radial basis function interpolation and a measure of 'bumpiness' of a radial function, σ say. A target value f_n^* is chosen that is an estimate of the global minimum of f. For each $y \notin \{x_1, \ldots, x_n\}$ there exists a radial basis function s_y that satisfies the interpolation conditions

$$\begin{array}{rcl} s_y(x_i) & = & f(x_i), \quad i = 1, \ldots, n, \\ s_y(y) & = & f_n^*. \end{array} \tag{2}$$

The next point x_{n+1} is calculated as the value of y in the feasible region that minimizes $\sigma(s_y)$. It turns out that the function $y \mapsto \sigma(s_y)$ is much cheaper to compute than the original function.

Here, the radial basis function interpolant s_n has the form

$$s_n(x) = \sum_{i=1}^{n} \lambda_i \phi\left(\|x - x_i\|_2\right) + b^T x + a, \tag{3}$$

with $\lambda_1, \ldots, \lambda_n \in \mathbf{R}$, $b \in \mathbf{R^d}$, $a \in \mathbf{R}$ and ϕ is either cubic with $\phi(r) = r^3$ or the thin plate spline $\phi(r) = r^2 \log r$. Gutmann (1999) considers other choices of ϕ and of the additional polynomial, see the table below. Later in Gutmann (2000) he concludes that the situation in the multiquadric and Gaussian cases is disappointing.

RBF	$\phi(r) > 0$	$p(x)$
cubic	r^3	$a^T \cdot x + b$
thin plate spline	$r^2 \log r$	$a^T \cdot x$ + b
linear	r	b
multiquadric	$\sqrt{(r^2 + \gamma^2)}$	
Gaussian	$\exp(-\gamma r^2)$	

The unknown parameters λ_i, b and a are obtained as the solution of the system of linear equations

$$\begin{pmatrix} \Phi & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ c \end{pmatrix} = \begin{pmatrix} F \\ 0 \end{pmatrix}, \tag{4}$$

where Φ is the $n \times n$ matrix with $\Phi_{ij} = \phi\left(\|x_i - x_j\|_2\right)$ and

$$P = \begin{pmatrix} x_1^T & 1 \\ x_2^T & 1 \\ . & . \\ . & . \\ x_n^T & 1 \end{pmatrix}, \lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ . \\ . \\ \lambda_n \end{pmatrix}, c = \begin{pmatrix} b_1 \\ b_2 \\ . \\ . \\ b_d \\ a \end{pmatrix}, F = \begin{pmatrix} f(x_1) \\ f(x_2) \\ . \\ . \\ f(x_n) \end{pmatrix}. \tag{5}$$

s_y could be obtained accordingly, but there is no need to do that as one is only interested in $\sigma(s_y)$. Powell (1992) shows that if the rank of P is $d+1$, then the matrix

$$\begin{pmatrix} \Phi & P \\ P^T & 0 \end{pmatrix} \tag{6}$$

is nonsingular and the linear system (4) has a unique solution.

σ is defined in Gutmann (2000). For s_n in (3) it is

$$\sigma(s_n) = \sum_{i=1}^{n} \lambda_i s_n(x_i). \tag{7}$$

Further, it is shown that $\sigma(s_y)$ is

$$\sigma(s_y) = \sigma(s_n) + \mu_n(y) \left[s_n(y) - f_n^*\right]^2, \quad y \notin \{x_1, \ldots, x_n\}. \tag{8}$$

Thus minimizing $\sigma(s_y)$ subject to constraints is equivalent to minimizing g_n defined as

$$g_n(y) = \mu_n(y) \left[s_n(y) - f_n^*\right]^2, \quad y \in \Omega \setminus \{x_1, \ldots, x_n\}, \tag{9}$$

where $\mu_n(y)$ is the coefficient corresponding to y of the Lagrangian function L that satisfies $L(x_i) = 0$, $i = 1, \ldots, n$ and $L(y) = 1$. It can be computed as follows. Φ is extended to

$$\Phi_y = \begin{pmatrix} \Phi & \phi_y \\ \phi_y^T & 0 \end{pmatrix}, \tag{10}$$

where $(\phi_y)_i = \phi(\|y - x_i\|_2)$, $i = 1, \ldots, n$, and P is extended to

$$P_y = \begin{pmatrix} P \\ y^T & 1 \end{pmatrix}. \tag{11}$$

Then $\mu_n(y)$ is the $(n+1)$-th component of $v \in \mathbf{R^{n+d+2}}$ that solves the system

$$\begin{pmatrix} \Phi_y & P_y \\ P_y^T & 0 \end{pmatrix} v = \begin{pmatrix} 0_n \\ 1 \\ 0_{d+1} \end{pmatrix}. \tag{12}$$

We use the notation 0_n and 0_{d+1} for column vectors with all entries equal to zero and with dimension n and $(d+1)$, respectively. The computation of $\mu_n(y)$ is done for many different y when minimizing $g_n(y)$. This requires $O(n^3)$ operations if not exploiting the structure of Φ_y and P_y. Hence it does not make sense to solve the full system each time. A better alternative is to factorize the interpolation matrix and update the factorization for each y. An algorithm that requires $O(n^2)$ operations is described in Björkman and Holmström (2001).

When there are large differences between function values, the interpolant has a tendency to oscillate strongly. It might also happen $\min s_n(y)$ is much lower than the best known function value, which leads to a choice of f_n^* that overemphasizes global search. To handle these problems, large function values are in each iteration replaced by the median of all computed function values.

Note that μ_n and g_n are not defined at $x_1, \ldots, x_n$ and

$$\lim_{y \to x_i} \mu_n(y) = \infty, \quad i = 1, \ldots, n. \tag{13}$$

This will cause problems when μ_n is evaluated at a point close to one of the known points. The function $h_n(x)$ defined by

$$h_n(x) \;=\; \begin{cases} \frac{1}{g_n(x)}, & x \notin \{x_1, \ldots, x_n\} \\ 0, & x \in \{x_1, \ldots, x_n\} \end{cases} \tag{14}$$

is differentiable everywhere on Ω, and is thus a better choice as objective function. Instead of minimizing $g_n(y)$ in (9) one may minimize $-h_n(y)$. In our implementation we instead minimize $-\log(h_n(y))$. By this we avoid a

flat minimum and numerial trouble when $h_n(y)$ is very small.

2.2. The Choice of f_n^*. For the value of f_n^* it should hold that

$$f_n^* \in \left[-\infty, \min_{y\in\Omega} s_n(y)\right]. \tag{15}$$

The case $f_n^* = \min_{y\in\Omega} s_n(y)$ is only admissible if $\min_{y\in\Omega} s_n(y) < s_n(x_i)$, $i = 1,\ldots,n$. There are two special cases for the choice of f_n^*. In the case when $f_n^* = \min_{y\in\Omega} s_n(y)$, then minimizing (9) is equivalent to

$$\min_{y\in\Omega} s_n(y). \tag{16}$$

In the case when $f_n^* = -\infty$, then minimizing (9) is equivalent to

$$\min_{y\in\Omega\setminus\{x_1,\ldots,x_n\}} \mu_n(y). \tag{17}$$

So how should f_n^* be chosen? If $f_n^* = -\infty$, then the algorithm will choose the new point in an unexplored region, which is good from a global search point of view, but the objective function will not be exploited at all. If $f_n^* = \min_{y\in\Omega} s_n(y)$, the algorithm will show good local behaviour, but the global minimum might be missed. Therefore, there is a need for a mixture of values for f_n^* close to and far away from $\min_{y\in\Omega} s_n(y)$. Gutmann (1999) describes two different strategies for the choice of f_n^*. In this paper we study one of the strategies.

The strategy, denoted **idea 1**, is to perform a cycle of length $N+1$ and choose f_n^* as

$$f_n^* = \min_{y\in\Omega} s_n(y) - W \cdot \left(\max_i f(x_i) - \min_{y\in\Omega} s_n(y)\right), \tag{18}$$

with

$$W = \left[\frac{(N - (n - n_{init}))\mathrm{mod}(N+1)}{N}\right]^2, \tag{19}$$

where n_{init} is the number of initial points. Here, $N = 5$ is fixed and $\max_i f(x_i)$ is not taken over all points, except for the first step of the cycle. In each of the subsequent steps the $n - n_{max}$ points with largest function value are removed (not considered) when taking the maximum. Hence the quantity $\max_i f(x_i)$ is decreasing until the cycle is over. Then all points are considered again and the cycle starts from the beginning. More formally, if $(n - n_{init})\mathrm{mod}(N+1) = 0$, $n_{max} = n$, otherwise

$$n_{max} = \max\left\{2, n_{max} - \mathrm{floor}((n - n_{init})/N)\right\}. \tag{20}$$

A check is performed when $(n - n_{init})\text{mod}(N+1) = N$. This is the stage when a purely local search is performed, so it is important to make sure that the minimizer of s_n is not one of the interpolation points or too close to one. The test used is

$$f_{min} - \min_{y\in\Omega} s_n(y) \leq 10^{-4} \max\{1, |f_{min}|\}, \tag{21}$$

where f_{min} is the best function value found so far, i.e. $\min_i f(x_i)$, $i = 1, \ldots, n$. If (21) is true, then

$$f_n^* = \min_{y\in\Omega} s_n(y) - 10^{-2} \max\{1, |f_{min}|\}, \tag{22}$$

otherwise f_n^* is set to 0.

2.3. A Compact RBF Algorithm Description. In the previous sections the basic RBF algorithm implemented in our Matlab routine *rbfSolve* were described in detail. We now summarize the RBF algorithm in the compact description below.

- Choose n initial points $X = \{x_i, i = 1, ..., n\}$.
  Use 2^d corner points or at least $d + 1$ points.
- Compute $f_i = f(x_i), i = 1, ..., n$, set $n_{init} = n$.
- Compute Radial basis interpolation minimizing semi-norm and interpolating points

$$\begin{aligned} & s_n = arg\min_s \quad < s, s > \\ & s/t \qquad\qquad\quad s(x_i) = f(x_i), \quad i = 1, ..., n \end{aligned} \tag{23}$$

  The optimal solution is the solution to (4).
- *While* $n < MaxFuncEval$
  Repeat Cycle $k = 0, ..., N$ (Local and global search, $N = 5$)
  1. If $k = 0$ solve the minimization problem $\min_{y\in\Omega} s_n(y)$.
  2. Compute f_n^* in (18) dependent on position k in the cycle.
  3. $x_{new} = arg\min_y - \log h_n(y)$, $h_n(y)$ defined in (14).
  4. *If* new point x_{new} *acceptable* (Not too close to $x_1, \ldots, x_n$), $n = n + 1$; $x_{new} = x_n$; $f_n = f(x_{new})$; $X = [X, x_{new}]$; end
  5. $f_{best} = \min f(x_i), x_i \in X$;

6. $x_{best} = arg\min f(x_i), x_i \in X$;
7. Update the matrix factorizations of Φ and P and find new interpolant s_n by solving (12).

- *End of while*

One problem is how to choose the points $x_1, \ldots, x_{n_{init}}$ to include in the initial set. We only consider box constrained problems, and choose the corners of the box as initial points, i.e. $n_{init} = 2^d$. Starting with other points is likely to lead to the corners during the iterations anyway. Having a "good" point beforehand, one can include it in the initial set.

The subproblem

$$\min_{y \in \Omega} \quad s_n(y) \ , \tag{24}$$

is itself a problem which could have more than one local minima. To solve (24) (at least approximately), we start from the interpolation point with the least function value, i.e. $\text{argmin} f(x_i)$, $i = 1, \ldots, n$, and perform a local search. In many cases this leads to the minimum of s_n. Of course, there is no guarantee that it does. We use analytical expressions for the derivatives of s_n and perform the local optimization using *ucSolve* TOMLAB running the inverse BFGS algorithm as described in Holmström and Björkman (1999). As an alternative we use the *NPSOL* solver by Gill, Murray, Saunders and Wright (1998) using the MEX-file interface that is part of TOMLAB.

To minimize $-\log h_n(y)$ we use our Matlab routine *glbSolve* implementing the DIRECT algorithm (see Section 3.1). We run *glbSolve* for 500 function evaluations and choose x_{n+1} as the best point found by *glbSolve*. When $(n - n_{init}) \text{mod}(N + 1) = N$ (when a purely local search is performed) and the minimizer of s_n is not too close to any of the interpolation points, i.e. (21) is not true, *glbSolve* is not used to minimize $g_n(y)$ or $f^*(y)$. Instead, we choose the minimizer of (24) as the new point x_{n+1}. The TOMLAB routine *AppRowQR* is used to update the QR decomposition.

Our experience so far with the RBF algorithm shows that the minimum is sometimes very sensitive for the scaling of the box constraints. To overcome this problem we transform the search space to the unit hypercube.

In our implementation it is possible to **restart** the optimization with the final status of all parameters from the previous run.

3. Other global optimization algorithms

In the following sections, Section 3.1 - 3.3, short descriptions of the DIRECT, constrained DIRECT and EGO algorithms are given.

3.1. DIRECT. DIRECT is an algorithm developed by Jones, Perttunen and Stuckman (1993) for finding the global minimum of a multi-variate

function subject to simple bounds, using no derivative information. The algorithm is a modification of the standard Lipschitzian approach that eliminates the need to specify a Lipschitz constant. The idea is to carry out simultaneous searches using all possible constants from zero to infinity. In Jones et al. (1993) they introduce a different way of looking at the Lipschitz constant. The Lipschitz constant is viewed as a weighting parameter that indicate how much emphasis to place on global versus local search. In standard Lipschitzian methods, this constant is usually large because it must be equal to or exceed the maximum rate of change of the objective function. As a result, these methods place a high emphasis on global search, which leads to slow convergence. In contrast, the DIRECT algorithm carries out simultaneous searches using all possible constants, and therefore operates on both the global and local level. DIRECT deals with problems of the form

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ \text{s.t.} & x_L \ \leq \ x \ \leq \ x_U \ , \end{array} \tag{25}$$

where $f \in \mathbf{R}$ and $x, x_L, x_U \in \mathbf{R^d}$. The finite box defined by the bound constraints is normalized to $[0,1]^d$. and partitioned into smaller boxes. Then it is true that the side lengths of the boxes are 3^{-k} for some $k \in \mathbf{N}$. It is guaranteed to converge to the global optimal function value, if the objective function f is continuous or at least continuous in the neighborhood of a global optimum. This could be guaranteed since, as the number of iterations goes to infinity, the set of points sampled by DIRECT form a dense subset of the unit hypercube. In other words, given any point x in the unit hypercube and any $\delta > 0$, DIRECT will eventually sample a point (compute the objective function) within a distance δ of x. However, the use of the midpoint in each box leads to the disadvantage that the boundary can only be reached in the limit, and the convergence will be slow when the minimizer lies at the boundary.

We have implemented the DIRECT algorithm in Matlab, and in Björkman and Holmström (1999), we discuss the implementation details of our Matlab implementation. The efficiency of the implementation is analyzed by a comparison to the results of Jones's implementation on nine standard test problems for box-bounded global optimization. In fifteen out of eighteen runs the results were in favor of our implementation.

One version of the DIRECT code is available as the Matlab routine *gblSolve* for download at http://www.ima.mdh.se/tom, the home page of the Applied Optimization and Modeling group. It is free for academic use. A faster version, *glbSolve*, is part of the TOMLAB v3.0 optimization environment described in Holmström (2001).

3.2. Constrained DIRECT. Jones (2001) presents an extension of the

DIRECT algorithm which handles nonlinear and integer constraints, a global mixed-integer nonlinear programming problem of the form

$$\begin{array}{lrcccl} \min\limits_{x} & f(x) & & & & \\ & x_L & \leq & x & \leq & x_U \\ \text{s.t.} & c_L & \leq & c(x) & \leq & c_U \ , \\ & x_i \in I & & \text{integer} & & \end{array} \tag{26}$$

where $f \in \mathbf{R}$, $x, x_L, x_U \in \mathbf{R^d}$, $c, c_L, c_U \in \mathbf{R^m}$ and I is the index set for the integer variables. The constrained version of DIRECT does not explicitly handle equality constraints and it works best when the integer variables describe an ordered quantity. It is less effective when the integer variables are categorical. If no constraints are present, this constrained version of DIRECT reduces to the box-bounded version, with some minor differences.

We have implemented the constrained version of the DIRECT algorithm in Matlab with a slightly more general problem formulation that explicitly handles linear constraints as

$$\begin{array}{lrcccll} \min\limits_{x} & f(x) & & & & & \\ & -\infty < \ x_L & \leq & x & \leq & x_U & < \infty \\ s/t & b_L & \leq & Ax & \leq & b_U & \\ & c_L & \leq & c(x) & \leq & c_U, & x_j \in \mathbf{N} \quad \forall \mathbf{j} \in \mathrm{I}, \end{array} \tag{27}$$

where $x, x_L, x_U \in \mathbf{R^n}$, $f(x) \in \mathbf{R}$, $A \in \mathbf{R^{m_1 \times n}}$, $b_L, b_U \in \mathbf{R^{m_1}}$ and c_L, $c(x)$, $c_U \in \mathbf{R^{m_2}}$.
The variables $x \in I$, the index subset of $1, ..., n$, are restricted to be integers. Our constrained DIRECT code is available as the Matlab routine *glcSolve* in TOMLAB. Feedback from TOMLAB users and tests we have ran show that the solver works well.

3.3. EGO. The EGO (Efficient Global Optimization) algorithm by Jones, Schonlau and Welch (1998) is also an interesting algorithm, which like the RBF algorithm belongs the class of Response Surface Methods. These models first fit a model function to data collected by evaluating the objective function at a number of initial points. Then a utility function is used to determine the new point where the objective function should be evaluated. In EGO, a nonlinear stochastic process model, the DACE (Design and Analysis of Computer Experiments) predictor, is fit by use of nonlinear regression unlike the RBF algorithm, where linear regression is used. Then, EGO balances between global and local search by choosing the new point where the objective function should be evaluated as the one which maximizes an expected improvement utility function.

We have implemented the EGO algorithm in Matlab and it is available as the Matlab routine *ego* in TOMLAB.

4. Financial applications

Trading strategies present an interesting and challenging application for global optimization without derivatives. The object function can be defined as the achieved profit when applying a trading system on historical data. It is clear that neither the object function nor the derivatives are available in analytical form. Furthermore, the function value is often the result of a simulated trading with many years of data, and takes in the order of minutes to compute with an ordinary desk top computer. The need for optimization routines suitable for costly object functions is therefore clear. In this section we will use the previously described DIRECT algorithm to optimize simple trading rules parameterized with three and four parameters. A big problem with such optimization is the estimation of out-of-sample performance for the obtained trading rules. In particular, it is very easy to jump into conclusions regarding trading rules that exhibit extremely profitable behavior, when tested on historical data. These misjudgments are often caused by the rules covering too few examples in the examined data. We will approach this problem with nonconvex global optimization of trading rules with a constraint added in the problem formulation. The effect is a regularization, where solutions covering too few examples are rejected. The modeling is performed with a sliding-window technique and generates different parameters for the optimized trading rules in each time window. For more details and results, refer to Hellström (2000b) where another data set is analyzed with the same approach.

4.1. Trading Rules. A general way to formulate strategies for stock trading is to define a trading rule as a time series $T(t)$ such as

$$T(t) = \begin{cases} \text{Buy} & : \text{ if } g(t) = 1 \\ \text{Sell} & : \text{ if } g(t) = -1 \\ \text{Do nothing} & : \text{ if } g(t) = 0 \end{cases} \tag{28}$$

where g is a function of the previous stock prices $Close$:

$$g : \{Close(t), Close(t-1), ..., Close(t-k)\} \rightarrow \{-1, 0, 1\}. \tag{29}$$

Trading rule (28) is designed to serve as decision support in actual stock trading, as indicated by the labels *Buy, Sell,* and *Do Nothing.* Function g determines the type of the trading rule. By extending expression (29) with the input variables *High* (highest-paid price), *Low* (lowest-paid price), *Open* (first price) and *Volume* (number of traded stocks), most standard technical indicators, such as the *Stochastic Oscillator*, the *Relative Strength Index (RSI), Moving Average Convergence/Divergence (MACD)* etc. (Kaufman (1998)), can be described in this fashion. Quite often the buy and sell decisions are controlled by separate expressions and the trading rules are

then denoted *Buy rule* and *Sell rule* respectively. Hereinafter we use the notation g_s to denote a trading rule applied to one specific stock s.

Function g is normally parameterized with a few parameters that can to be determined by numerical optimization. In this paper, three trading rules for generating *Buy* signals, are used to demonstrate the techniques with constrained optimization. All three are based on standard technical indicators, well-known by the trading community. For a thorough introduction to the subject, refer to Kaufman (1998). However, the standard indicators have been augmented with a term that includes the traded volume. This too is in accordance with common practice among traders. We include the traded volume as a term in all our technical trading rules. To facilitate a uniform modeling for *all* stocks in the market, a normalized measure has to be defined.

4.1.1. Gaussian Volume. The Gaussian volume $V_n(t)$ is a transformation of the traded volume (number of stocks) $V(t)$ defined as

$$V_n(t) = (V(t) - m_V(t))/\sigma_V(t), \tag{30}$$

where the mean $m_V(t)$ and the standard deviation $\sigma_V(t)$ for the volume are computed in an n days long window up to time t. V_n expresses the number of standard deviations, by which the volume differs from its running mean. The normalization makes it possible to compare values of V_n for different stocks and also for different times. In this paper the Gaussian volume V_{10} is used and is denoted by $gvol10$, since this is the name of the ASTA (Hellström (2000a)) implementation of the function.

4.1.2. Crossing Moving-Average. This is an implementation of a common trading rule based on two moving averages of different length. The trading rule signals *Buy*, if a short moving-average mav_{x_1} crosses a long moving-average mav_{x_2} from below. A *Sell* signal is issued when mav_{x_1} crosses the mav_{x_2} from above. In this paper we define the *Buy* rule mav as

$$mav(x_1, x_2, x_3) = Mavx(x_1, x_2) \wedge gvol10 > x_3, \tag{31}$$

where

$$Mavx(x_1, x_2) = mav_{x_1}(t) > mav_{x_2}(t) \wedge mav_{x_1}(t-1) \leq mav_{x_2}(t-1) \tag{32}$$

and mav_{x_1} is a x_1-day moving average of the stock prices up to time t.

4.1.3. Trading Channel Breakout. The main part of this trading rule is what is popularly known as *Bollinger Bands* (see e.g. page 91 in Kaufman (1998)). The complete trading rule is defined as

$$break(x_1, x_2, x_3) = breakout(x_1, x_2) \wedge gvol10 > x_3, \tag{33}$$

where the *breakout* function is defined as

$$breakout(x_1, x_2) = \begin{array}{l} Close(t) > (mav_{x_1}(t) + x_2 \cdot \sigma_{x_1}(t)) \quad \wedge \\ Close(t-1) \leq (mav_{x_1}(t) + x_2 \cdot \sigma_{x_1}(t)) \end{array} \tag{34}$$

and $mav_{x_1}(t)$ is is a x_1-day long moving average of the stock prices up to time t. Function $\sigma_{x_1}(t)$ computes the standard deviation of the $Close$ up to time t. The idea is to define an upper boundary for a trading channel and generate a *Buy* signal when the $Close$ penetrates this boundary from below. This upper boundary is defined as the sum of a moving average mav_{x_1} and x_2 times an estimate of the standard deviation σ_{x_1}.

4.1.4. Level of Resistance. The trading rule *Level of Resistance*, in this paper denoted *resist*, is based on a technique commonly executed by manual inspection of the stock charts. The general idea is to identify *peaks* in a window backwards, where the *Close* price is roughly the same. When such peaks are found, a *Buy* signal is generated if the *Close* price crosses from below the level for the found peaks. We define the trading rule *resist* as

$$resist(x_1, x_2, x_3, x_4) = xresist(x_1, x_2, x_3) \wedge gvol10 > x_4 \tag{35}$$

where

$$xresist(x_1, x_2, x_3) = Close(t) > plevel \wedge Close(t-1) \leq plevel \tag{36}$$

and

$$plevel = \begin{cases} l: & \text{if at least } x_2 \text{ peaks in } Close \text{ that differs by less than } x_3\% \text{ can} \\ & \text{be identified at level } l \text{ in an } x_1\text{-day long window backwards.} \\ 0: & \text{otherwise} \end{cases}$$

4.2. Performance Evaluation. Performance evaluation for a trading rule is needed in two stages of the process. First, in the optimization phase, when parameters for the trading rule have to be determined. The second stage is when the final trading rule is evaluated on the test data set previously unseen. For more information about performance evaluation of trading algorithm refer to Hellström (1999b) or Refenes (1995). Trading-rule-based methods are normally evaluated by trading simulation, where the trading rule controls the buying and selling of one or several stocks over a period of time. Examples of this approach in conjunction with optimization can be found in Hellström and Holmström (1999). However, it is also possible to evaluate a trading rule with a fixed prediction horizon, of which the advantage is that all situations where the trading rules fire (i.e.: $T(t) \neq$ *Do Nothing* in (28)) are evaluated. When performing a trading simulation, this is normally not the case, since the simulated trader is bounded

by the real-world constraint of a limited amount of money. This prevents the trader from executing some of the *Buy* signals that the trading rules produce. Since the fraction of left-out trades can be as high as 80-90%, a scheme with randomization and repeated simulations is normally required to produce reliable performance measures for the trading rules. Therefore in this study we evaluate trading rules at fixed prediction horizons. The measure of interest is the correctness of the sign of the price change from the time of the prediction to 5 days ahead. This way of evaluating predictions has gained increased interest in recent years as an alternative to the more conventional way of minimizing the error of the level prediction. A comparative study of sign and level methods can be found in Leung, Daouk and Chen (2000) where the presented experiments suggest that methods predicting the sign provide higher profits than methods predicting the level for a number of investigated stock indexes.

For a time period $[1, ..., T]$ and a set of stocks S, the h-day positive hit rate for a *Buy* rule g is defined as

$$H_g^+ = \frac{card\{(t,s)|R_h^s(t+h) > 0, g_s(t) = 1, 1 \leq t \leq T-h, s \in S\}}{card\,\{(t,s)|R_h^s(t+h) \neq 0, g_s(t) = 1, 1 \leq t \leq T-h, s \in S\}} \quad (37)$$

where g_s is the function specifying the trading rule as described in (28). The return R_h^s is the relative change in price and is defined as

$$R_h^s(t) = 100 \cdot \frac{Close_s(t) - Close_s(t-h)}{Close_s(t-h)} \quad (38)$$

where $Close_s(t)$ is the price for a stock s at the end of day t. The hit rate H_g^+ for a *Buy* rule g indicates how often a *Buy* signal is followed by a true increase in the stock price. The hit rate H_g^- for a *Sell* rule is defined correspondingly but with returns $R_h < 0$.

4.3. Optimization. The function g that defines the trading rule is normally parameterized with a few parameters x that have to be determined in order to maximize the chosen performance measure on the historical data. To express this parameterization, the notation $g[x]$ will be used.

One big problem about trading rules in general and optimizing them in particular is the statistical significance of the estimated performance. The trading rule (28) normally issues *Buy* or *Sell* signals only for a minor part of the points in the time series. This results in low levels of significance for the produced performance measures. It is often easy to find a trading rule that historically outperforms any benchmark, as long as it does not have to produce more than a few signals per year. However, the performance on previously unseen data is most often very bad in these situations. We therefore formulate a constrained optimization problem for a *Buy rule* g

(*Sell* rules can be treated in a similar way) as

$$\begin{array}{l} arg\ \max\limits_{x}\ \ H^{+}_{g[x]} \\ \text{s.t.} \\ card\{(s,t)|g_s[x](t)=1, t \leq T-h, s \in S\} \geq N_0, \\ x_L \leq x \leq x_H \end{array} \tag{39}$$

where x_L and x_H are lower and upper bounds for the unknown parameters and the other constraint is the total number of *Buy* signals. The hit rate $H^{+}_{g[x]}$ is given by definition (37). With the introduced notation, $g_s[x](t)$ denotes the trading rule g parameterized with parameters x and applied to stock s for time t. The optimization routine performs simulations up to time T to compute the hit rate and number of trading signals for a given $g[x]$. The purpose is to maximize the hit rate $H^{+}_{g[x]}$ by altering the variables x that parameterize the function g. The final performance measure is the out-of-sample hit rate $H^{+}_{g[x]}$, computed for time $t > T$ with the optimal estimated parameters x.

Using a 'hard' constraint in the optimization problem in (39) leads to a nonsmooth problem. Because of the uncertainty in the choice of the 'most' suitable value of N_0, it is reasonable to reformulate the problem using a 'soft' constraint approach that generates a smooth problem. The approach uses a sigmoid function to smoothly model the behavior of the added constraint and is inspired by the membership-function concept used in fuzzy logic (see e.g. Klir and Yuan (1995)). The new problem formulation, in which the objective function in (39) is weighted with the output of a sigmoid, is

$$\begin{array}{l} arg\ \max\limits_{x}\ \ H^{+}_{g[x]} \cdot\ support_{N_0}\left(card\{(s,t)|g_s[x](t)=1, t \leq T-h, s \in S\}\right) \\ \text{s.t.} \\ x_L \leq x \leq x_H \end{array} \tag{40}$$

where $support_{N_0}$ is given by the sigmoid function

$$support_{N_0}(n) = \frac{1}{1+e^{-\alpha(n-\beta)}}. \tag{41}$$

The parameters α and β are computed to fulfill the equations $support_{N_0}(N_0) = 0.99$ and $support_{N_0}(N_0 \cdot 0.5) = 0.01$. This ensures a smooth penalty for trading rules that generate less than N_0 trading signals. If more than N_0 trading signals are generated, the $support_{N_0}$ function returns essentially 1 and hence does not affect the search for an optimal function g. The constraint acts like a regularizer, since the search space for the function g is

reduced by requiring a minimum number of trading signals. This improves the statistical significance of the estimated performance and the generalizability of the found solution (i.e. the achieved hit rate on previously unseen data). The choice of the cut-off value N_0 is a trade-off between the achieved hit rate on the training data and the generalizability.

The optimization problem (40) is a box-bounded nonconvex global optimization problem. It is suitable to use derivative free methods, since no analytical expressions for $g[x]$ and $H^+_{g[x]}$ are available. In our tests we are using the *DIRECT* algorithm described in Section 3.1.

4.4. Experimental Design. Technical analysis of stocks is normally based on the premise that the market's behavior does not change much over time. While future movements in stock prices are never copies of the past, the market's way of responding to new situations is assumed to be similar to the way it has handled them in the past (Gencay, Stengos (1998)). Since this is not necessarily a valid assumption the optimization will be performed with a sliding window technique.

The hit rate H^+_g in the object function (39) is computed using the non-interactive version of the ASTA system, which performs market simulations of trading rules given in symbolic form. The ASTA system is written in Matlab and has a large number of technical indicators implemented. The system is thoroughly described in Hellström (2000a). Examples of usage is found in Hellström (1999a).

The test is utilizing a sliding-window technique with a 2-year training data period followed by a 1-year test period. The starting point of the training period is moved between 1990 and 1995 in 1-year steps. This results in six separate modeling/test periods. The presented performance is the total for the six test periods (1992,...,1997). The purpose of using sliding windows in the optimization is twofold. First, the stability in the performance can be studied since we get six performance measures instead of one. Second, the trading rules are allowed to adapt to time-varying market conditions such as volatility, long-term trends etc. Eighty of the largest Swedish stocks are included in the test, which provides a total number of data points of around 111000 (not all stocks have data for the entire period). The trading rules select a small fraction of these points (date and stock) as suggested opportunities to buy stocks.

The results for 5-day prediction horizon are presented in Table 1 , with positive hit rate H^+ and number of points N where a trading signal is generated. Separate measures for training data and test data are presented in the columns labeled H_{tr}, N_{tr}, H_{te} and N_{te}. The rightmost column shows the lower 90% confidence limit[2] for the hit rate H_{te}. The cut-off value N_0, used for the regularization, is set to 100. Each of the eight rows represents a

[2]The lower boundary for a 90% confidence interval.

prediction method. The first three rows show the results for the trading rules $resist_{100}$, $break_{100}$ and mav_{100} described in Section 4.1. The parameters $x_1, x_2, ...$ are optimized for best performance on the training data, using the regularization described above ($N_0 = 100$). The following three rows show the same trading rules as above, but with no regularization to control the number of generated trading signals ($N_0 = 1$): $resist_1$, $break_1$ and mav_1. Performance for the benchmark methods Naive-5_+ and Naive-ε are also reported. The Naive-5_+ predictor of the returns for a stock s asserts today's return $R_5^s(t)$ (price increase since $t-5$) as the prediction of $R_5^s(t+5)$. The Naive-ε prediction of prices for a stock s asserts today's price $Close_s(t)$ as the best estimate of $Close_s(t+5)$. To enable comparison of hit rate predictions, the naive predictor is modified so the best estimate of today's price is assumed to be $Close_s(t+5) + \varepsilon$. This means that the predicted returns R_5^s are always positive. This naive predictor is denoted below *Naive-ϵ*.

The computed optimal parameters for a specific *Buy* rule vary for the six test periods. The ones computed for test period 1992 are presented in Table 2.

Table 1: Hit rate and number of selected points for optimized trading rules. Totals from 6 1-year test periods (1992-1997) with the preceding 2 years for training. 5 days prediction horizon.

Method	H_{tr}	N_{tr}	H_{te}	N_{te}	**90%−low** H_{te}
$resist_{100}$	65.82	746	63.44	454	59.55
$break_{100}$	63.07	1075	55.64	692	52.44
mav_{100}	61.96	715	50.40	371	46.01
$resist_1$	76.84	177	59.83	117	51.82
$break_1$	64.89	786	52.04	417	47.89
mav_1	71.55	239	53.75	160	46.94
$Naive - e$	48.33	196470	50.06	102651	49.80
$Naive - 5_+$	48.83	84054	49.53	46202	49.14

4.5. Results. As expected, the optimized trading rules perform much better for the training data than for the test data. This effect is much more emphasized for the non-regularized trading rules than for the regularized ones. The difference can be understood as over-fitting of data that can be controlled by the regularization. The out-of-sample hit rates H_{te} show no systematic difference between the two kinds of predictors. The small observed differences should be seen rather as stochastic fluctuations caused by the low accuracy in the estimation of the hit rates for the non-regularized trading rules. The lower 90% confidence limit reveals how uncertain the hit

Table 2: Optimized trading rules for 1992. 5 days prediction horizon.

Method	**Optimized expression**
$resist_{100}$	$xresist(84, 4, 4.06) \wedge gvol10 > 0.67$
$break_{100}$	$breakout(38, 1.5) > 0 \wedge gvol10 > 2.94$
mav_{100}	$Mavx(5, 83) \wedge gvol10 > 0.33$
$resist_1$	$xresist(42, 6, 1.83) \wedge gvol10 > 3.78$
$break_1$	$breakout(117, 2.5) > 0 \wedge gvol10 > 2.5$
mav_1	$Mavx(11, 112) \wedge gvol10 > 1.3$

rates H_{te} are for these rules. This uncertainty comes from the low number of predictions generated. None of the non-regularized trading rules can be said to significantly outperform the benchmark predictors, while the regularized *resist* predictor has 63% hit rate, which is significantly higher than the benchmarks.

4.6. Stability of the Found Optima. The experimental setup with sliding windows gives a stable evaluation of the trading rules. In this section an additional test of the stability and relevance of the optimized trading rules is performed. In Table 3, the three regularized trading rules optimized with data from 1990-1991 are applied not only for 1992 but also for the following years up to 1997. This means that the optimized rules are regarded as globally valid instead of valid only for the year following the optimization period. Performance for the benchmark predictors are also presented for comparison. The results show that the average hit rate for the trading rules for the six years, is clearly lower than the one achieved by the sliding-window approach, as shown in Table 1 (the relevant value for comparison is shown in column H_{te}). Furthermore, the individual results for each year show no clear tendency and can be regarded as random variations. These observations give further credibility to the sliding-window results and show that the optimizations really are catching patterns and regularities in the data and not only spurious local optima in random and noisy object functions.

4.7. Summary of the Results. The constrained optimization that avoids too few selected points is essential, both for practical reasons (since we want to get assistance in our buy and sell decisions more than a few times per year), and for a reasonably safe estimate of the expected hit rate out-of-sample. Without safeguarding against too few points, the found optima gives excellent performance on the training data, but no significant improvement relative to pure chance on the test data. Furthermore, the results show that the high hit rate achieved with the *resist* trading rule, to a large extent is a result of the adaptive modeling with sliding windows.

Table 3: Hit rate for trading rules optimized with data from 1990-1991. 5-day prediction horizon.

Method	92	93	94	95	96	97	Average	H_{te}
$resist_{100}$	55.1	66.9	59.1	52.8	57.4	56.6	57.6	63.44
$break_{100}$	54.8	65.1	48.3	46.5	55.8	52.1	54.5	55.64
mav_{100}	45.8	59.8	39.3	44.3	58.7	57.1	50.3	50.40
$Naive - e$	44.1	55.6	46.7	47.4	53.8	52.1	50.1	50.06
$Naive - 5_+$	46.7	56.2	45.1	46.0	52.2	49.3	49.5	49.53

5. Conclusions and further work

Global optimization techniques can be used to improve the performance of trading algorithms and time series predictions. When the problems are costly to compute, the use of surrogate modeling techniques like the RBF algorithm is promising and should be further exploited.

In the RBF algorithm, work is needed to avoid too large condition number on the interpolation matrix for increasing number of sampled points. Also better choices of initial set must be investigated, when n is not small. Our goal is to implement a robust and fast RBF algorithm in both Matlab and Fortran.

We will further test the use of surrogate model techniques for the optimization of trading algorithms and time series model predictions. It is interesting to use the regularization techniques described for more advanced prediction methods, e.g. EXPAR (Exponential Autoregressive) models and for high frequency data.

Bibliography

1. Björkman, M. and Holmström, K., *Global Optimization with the DIRECT Algorithm in Matlab*, Advanced Modeling and Optimization, **1**(2), (1999), 17-37.
2. Björkman, M. and Holmström, K., *Global Optimization of Costly Nonconvex Functions using Radial Basis Functions,* Optimization and Engineering, **2**, (2001), To be published.
3. Blume, L., Easley, D. and O'Hara, M., *Market statistics and technical analysis: The role of volume,* Journal of Finance, **49**, (1994), 153–181.
4. Brock, W., Lakonishok, J. and LeBaron, B., *Simple technical rules and the stochastic propertites of stock returns,* Journal of Finance, **47**, (1992), 1731–1764.
5. Campbell, J. Y., Grossman, S. J. and Wang, J., *Trading volume and serial correlation in stock returns,* Quarterly Journal of Economics, **108**, (1993), 905–940.
6. Gencay, R. and Stengos, T., *Moving average rules, volume and the predictability of security returns with feedforward networks,* Journal of Forecasting, (1998), 401–414.

7. Gill, P. E., Murray, W., Saunders, M. A. and Wright, M. H., *User's guide for NPSOL 5.0: A Fortran package for nonlinear programming,* Technical Report SOL 86-2, Revised July 30, 1998, Systems Optimization Laboratory, Department of Operations Research, Stanford University, Stanford, California 94305-4022, (1998).

8. Gutmann, H-M., *A radial basis function method for global optimization,* Technical Report DAMTP 1999/NA22, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, England, (1999).

9. Gutmann, H-M., *On the semi-norm of radial basis function interpolants*, Technical Report DAMTP 2000/NA04, Department of Applied Mathematics, and Theoretical Physics, University of Cambridge, England, (2000).

10. Hellström, T., *A Random Walk through the Stock Market*, Licentiate thesis, Umeå University, Umeå Sweden, (1998).

11. Hellström, T., *ASTA - a Tool for Development of Stock Prediction Algortihms,* Theory of Stochastic Processes, **5**(21), (1999a), 22–32.

12. Hellström, T., *Data Snooping in the Stock Market,* Theory of Stochastic Processes, **5**(21), (1999b), 33–50.

13. Hellström, T., *ASTA - User's Reference Guide,* Technical Report UMINF-00.16 ISSN-0348-0542, Department of Computing Science Umeå University, Umeå Sweden, (2000a).

14. Hellström, T., *Optimization of Trading Rules with a Penalty Term for Increased Risk-Adjusted Performance*, Advanced Modeling and Optimization, (2000b), **2**(3).

15. Hellström, T. and Holmström, K., *Parameter Tuning in Trading Algorithms using ASTA,* In Abu-Mostafa, Y. S. , LeBaron, B., Lo, A. W. and Weigend, A. S., (ed), Computational Finance 1999, Cambridge, MA., MIT Press, (1999).

16. Holmström, K., *The TOMLAB Optimization Environment in Matlab*, Advanced Modeling and Optimization, **1**(1), (1999a), 47–69.

17. Holmström, K., *New Optimization Algorithms and Software.,* Theory of Stochastic Processes, **5(21)**(1-2), (1999b), 55–63.

18. Holmström, K., *The TOMLAB v2.0 Optimization Environment*, In K. Holmström and E. Dotzauer, editors, Proceedings from the 6th Meeting of the Nordic Section of the Mathematical Programming Society, Västerås, 1999. Department of Mathematics and Physics, Mälardalen University, Sweden, (1999c).

19. Holmström, K., *TOMLAB v3.0 User's Guide,* Technical Report IMa-TOM-2001-01, Department of Mathematics and Physics, Mälardalen University, Sweden, (2001).

20. Holmström, K. and Björkman, M., *The TOMLAB NLPLIB Toolbox for Nonlinear Programming*, Advanced Modeling and Optimization, **1**(1), (1999), 70–86.

21. Iglehart, D. L. and Voessner, S., *Optimization of a trading system using global search techniques and local optimization,* Journal of Computational Intelligence in Finance, **6**, (1998), 36–46.

22. Jones, D. R., *DIRECT*, Encyclopedia of Optimization, (2001). To be published.

23. Jones, D. R., Perttunen, C. D. and Stuckman, B. E., *Lipschitzian optimization without the Lipschitz constant*, Journal of Optimization Theory and Applications, **79**(1), (1993), 157–181.
24. Jones, D. R., Schonlau, M. and Welch, W. J., *Efficient global optimization of expensive Black-Box functions*, Journal of Global Optimization, **13** (1998), 455–492.
25. Karpov, J. M., *The relation between price changes and traded volume,* Journal of Financial and Quantitative Analysis, **22**, (1987), 109–126.
26. Kaufman, P. J., *Trading Systems and Methods*, John Wiley and Sons, New York, (1998).
27. Klir, G.J. and Yuan, B., *Fuzzy Sets and Fuzzy Logic. Theory and Applications*, Prentice-Hall, Inc, New Jersey, USA, (1995).
28. Kuo, G. W., *Some exact results for moving-average trading rules with applications to UK indices,* In E. Acar and S. Satchell, editors, Advanced Trading Rules, Butterworth Heinemann, Oxford,(1998), 81–102.
29. Leung, M. T., Daouk, H. and Chen, A.-S., *Forecasting stock indices: a comparison of classification and level estimation methods*, International Journal of Forecasting, **16**:1,(2000), 73–190.
30. Levich, R. M. and Thomas, L. R., *The significance of technical trading-rule profits in the foreign exchange market: a bootstrap approach,* Journal of International Money and Finance, **12**, (1993), 451–474.
31. Powell, M. J. D., *The theory of radial basis function approximation in 1990,* In W.A. Light, editor, Advances in Numerical Analysis, Volume 2, Wavelets, Subdivision Algorithms and Radial Basis Functions, Oxford University Press, (1992), 105–210.
32. Powell, M. J. D., *Recent research at Cambridge on radial basis functions,* In M. D. Buhmann, M. Felten, D. Mache, and M. W. Müller, editors, New Developments in Approximation Theory, Birkhäuser, Basel, (1999), 215–232.
33. Refenes, A.-P., *Testing strategies and metrics*, In Refenes, A.-P., editor, Neural Networks in the Capital Markets, John Wiley & Sons, Chichester, England, (1995), 67–76.
34. Zar, J. H., *Biostatistical Analysis*, Prentice-Hall, Inc, New Jersey, USA, (1999).

Department of Computing Science,
Umeå University, SE-901 87 Umeå, Sweden
E-mail: thomash@cs.umu.se; http://www.cs.umu.se/ thomash

Center for Mathematical Modeling,
Department of Mathematics and Physics,
Mälardalen University, P.O. Box 883, SE-721 23 Västerås, Sweden
E-mail: hkh@mdh.se; http://www.ima.mdh.se/tom

Theory of Stochastic Processes
Vol.7 (23), no.1-2 ,2001, pp.142-155

ANDERS HOLMBERG AND BENGT SWENSSON

ON PARETO πps SAMPLING: REFLECTIONS ON UNEQUAL PROBABILITY SAMPLING STRATEGIES[1]

In the design based approach for inference in survey sampling, unequal probability sampling is an essential part of efficient strategies, i.e. efficient combinations of sampling design and estimator. During the last five years of the 20th century, new probability (approximately) proportional-to-size sampling designs with attractive properties have been proposed. A short review of these will be given, accompanied by a few reflections on their potential usefulness in applied settings.

2000 *Mathematics Subject Classifications.* 62D05.

Key words and phrases. Anticipated variance, GREG estimator, model-based stratified sampling, optimal sampling designs, Pareto sampling, PoMix sampling, probability proportional-to-size.

1. Introduction

We will consider the following survey set-up. Let $U = \{1, ..., k, ..., N\}$ be a finite population of size N (the number of population elements). At first, we assume that there is only one study variable, denoted y. The y value for the kth population element is denoted y_k. We want to estimate the population total $t_y = \sum_{k \in U} y_k = \sum_U y_k$ from a sample survey. Furthermore, we assume that there are Q auxiliary variables, denoted $u_1, \ldots, u_q, \ldots, u_Q$. The known value of the qth auxiliary variable for the kth population element is denoted u_{qk}. Hence, we have, for each population element k, access to a known vector $\mathbf{u}_k = (u_{1k}, \ldots, u_{qk}, \ldots, u_{Qk})'$. Moreover, we assume that there is a suitable known function $h(\cdot)$ such that every $z_k = h(\mathbf{u}_k)$ is known and strictly positive ($k = 1, ..., N$), which means that the variable z can be used as a size variable in the sampling design.

[1]Invited lecture

A without-replacement probability proportional-to-size sampling design $p(\cdot)$ based on the size variable z, denoted $\pi ps(z)$, will be used to select a (set) sample $s \subseteq U$ of size n_s, i.e. the first-order inclusion probabilities π_k are given by $\pi_k = nz_k/t_z$,where $t_z = \sum_U z_k$ and $n = E_p(n_s)$ is the expected sample size. (In the rest of the paper it is assumed that $nz_k/t_z \leq 1$, $k = 1, ..., N$.)

Much of the discussion in the literature focusses on strategies, design-estimator pairs, in which a $\pi ps(z)$ design is combined with the unbiased π estimator (Horvitz-Thompson estimator, HT estimator)

$$\hat{t}_{y\pi} = \sum\nolimits_{k\in s} \frac{y_k}{\pi_k} = \sum\nolimits_s \frac{y_k}{\pi_k} \tag{1}$$

From this expression, we see that if the study variable y is exactly proportional to the size variable z, i.e. $y_k = cz_k$ $(k = 1, ..., N)$, we have

$$\hat{t}_{y\pi} = \frac{n_s}{n} t_y$$

which means that (i) for a random size design the only variation of the estimator is due to variation in the sample size, while (ii) for a design of the fixed given size n there is no variation at all, since all samples of the given size result in $\hat{t}_{y\pi} = t_y$. So, if it is possible to find a size variable which is approximately proportional to the study variable, the π estimator should perform well. These facts are the original reasons for the long-lasting search for πps designs with good properties.

However, there is no reason to restrict the discussion to strategies based on the π estimator. On the contrary, in situations suitable for πps sampling, we are apt to have access to auxiliary variables which can be used for *GREG* (generalized regression) estimation. Thus, we will now suppose that it is possible to use the Q auxiliary variables to form another set of auxiliary variables $x_1, \ldots, x_j, \ldots, x_J$. Let, for $k = 1, ..., N$, $\mathbf{x}_k = (x_{1k}, \ldots, x_{jk}, \ldots, x_{Jk})'$ be a known vector, let $\mathbf{t}_x = \sum_U \mathbf{x}_k = (t_{x_1}, \ldots, t_{x_j}, \ldots, t_{x_J})'$ and let $\hat{\mathbf{t}}_{x\pi} = (\hat{t}_{x_1\pi}, \ldots, \hat{t}_{x_j\pi}, \ldots, \hat{t}_{x_J\pi})'$ be the corresponding π estimator. The *GREG* estimator can now be defined as

$$\hat{t}_{yGREG} = \hat{t}_{y\pi} + (\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi})'\hat{\mathbf{B}} \tag{2}$$

where

$$\hat{\mathbf{B}} = \left(\sum\nolimits_s \frac{\mathbf{x}_k \mathbf{x}_k'}{c_k \pi_k} \right)^{-1} \sum\nolimits_s \frac{\mathbf{x}_k y_k}{c_k \pi_k} \tag{3}$$

where c_k is a suitably chosen constant. If there is a strong linear relationship between y and $\mathbf{x}$, the *GREG* estimator will outperform the π estimator. For a full account of the reasoning behind the *GREG* estimator and further results, see e.g. chapter 6 in Särndal, Swensson and Wretman (1992), *SSW* for short.

For the comments to be made later in this paper we need a result from chapter 12 in *SSW*, which should be consulted for more detail. Suppose that the regression model ξ underlying the *GREG* estimator can be assumed to be

$$y_k = \mathbf{x}_k'\beta + \varepsilon_k \tag{4}$$

with

$$\begin{cases} E_\xi(\varepsilon_k) = 0 \\ V_\xi(\varepsilon_k) = \sigma_k^2 \\ E_\xi(\varepsilon_k\varepsilon_l) = 0; \quad k \neq l \end{cases} \tag{5}$$

where $\sigma_1^2, ..., \sigma_N^2$ are known up to a constant multiplier. In this case, an approximation to the anticipated variance

$$E_\xi E_p[(\hat{t}_{yGREG} - t_y)^2] - [E_\xi E_p(\hat{t}_{yGREG} - t_y)]^2$$

denoted $ANV(\hat{t}_{yGREG})$ is given by

$$ANV(\hat{t}_{yGREG}) = \sum\nolimits_U (\pi_k^{-1} - 1)\sigma_k^2 \tag{6}$$

Result 12.2.1 in *SSW* now states that for a sampling design $p(\cdot)$ such that $E_p(n_s) = n$, an optimal design is such that the first-order inclusion probabilities are given by

$$\pi_k = \pi_{0k} = n\sigma_k / \sum\nolimits_U \sigma_k \tag{7}$$

Many sample selection schemes which implement πps sampling designs have been proposed over the years, the most popular in applications being the systematic sampling procedure suggested by Madow (1949). For example, Brewer and Hanif (1983) list 50 schemes. However, if we exclude random size designs, it has turned out to be hard to devise a scheme for arbitrary sample size n that has a number of desirable properties, e.g. (a) the actual selection of the sample is relatively simple, (b) all first-order inclusion probabilities are strictly proportional to the size variable, (c) the design admits (at least approximately) unbiased estimation of the design variances $V_p(\hat{t}_{y\pi})$ and $V_p(\hat{t}_{yGREG})$. If we also want to be able to base the sample selection on the technique of permanent random numbers (*PRN*), which is desirable in large survey organizations taking many surveys, some of which are repeated over time, it will be even harder. (For a fairly recent overview of the *PRN* technique, see Ohlsson (1995).)

In this paper we will give a very brief account of, and a few comments on, *Pareto πps sampling* introduced by Rosén (1997a,b) as a special case of a more general class of designs proposed by him under the name of *Order (sampling with fixed distribution shape) πps sampling,* and *Poisson mixture (PoMix) sampling* proposed by Kröger, Särndal and Teikari (1999, 2000). These designs will be briefly compared to *model-based simple random*

stratified sampling, proposed by Wright (1983), as outlined in chapter 12 of *SSW*. All these designs may be alternatives for the practitioner when the use of the *PRN* technique is desirable.

2. A brief account of selected designs

In this brief overview, no account will be given of how to implement the selected designs by using the *PRN* technique, since our comments will focus on other aspects. The reader is referred the relevant literature, e.g. Ohlsson (1995) and Kröger, Särndal and Teikari (1999) and further references in these papers. We will also assume that the reader is familiar with Poisson sampling and its special case Bernoulli sampling, e.g. as described in *SSW*.

2.1. Pareto πps sampling. Generalizing a πps sampling approach suggested by Ohlsson (1990, 1998) under the name sequential Poisson sampling, Rosén (1997a,b) introduces a family of fixed size order πps schemes, of which sequential Poisson sampling is a special case. A particularly good scheme in this family is Pareto πps sampling. A Pareto $\pi ps(z)$ sample is realized as follows: (i) Compute $\lambda_k = nz_k/t_z$, $k = 1, ..., N$, (called target inclusion probabilities). (ii) Realize independent standard uniform random variables $U_1, U_2, ..., U_N$ and form the ranking variables

$$Q_k = \frac{U_k(1-\lambda_k)}{\lambda_k(1-U_k)} \qquad (k = 1, ..., N) \tag{8}$$

(iii) The elements with the n smallest Q_k-values then constitute the sample s (of size n) to be surveyed.

As an estimator for the population total t_y Rosén suggests the so-called quasi-HT-estimator

$$\hat{t}_{yqHT} = \sum\nolimits_s \frac{y_k}{\lambda_k} \tag{9}$$

The motivation for the prefix "quasi" lies in the fact that the target inclusion probabilities λ_k do not exactly equal the corresponding true first-order inclusion probabilities, π_k, of the design. However, extensive studies show that in most applied settings the differences are negligible. (The true π_k can be calculated to any degree of precision by an algorithm given by Aires, see Aires (2000) and earlier references in this PhD thesis.) See also Traat, Bondesson and Meister (2000). Rosén also provides the asymptotic variance and an easily calculated consistent variance estimator. Furthermore, Rosén (2000) outlines results for generalized regression estimation and Pareto πps sampling.

2.2. Poisson Mixture (*PoMix*) sampling

2.2.1. Random size PoMix *sampling.* In Kröger, Särndal and Teikari (1999) Poisson Mixture (*PoMix*) sampling is introduced. It is based on the

random size Poisson sampling design. If we disregard its implementation by the use of permanent random numbers, it can simply be described as follows. Let $\pi_k^{PO} = nz_k/t_z$ be the inclusion probabilities that would be used for a Poisson $\pi ps(z)$ sampling design, let $f = n/N$ be the expected sampling fraction, and fix a value $w \in [0, f]$, where w is called the Bernoulli width. Compute the *PoMix* first-order inclusion probabilities by the linear transformation

$$\pi_k = w + (1 - w/f)\pi_k^{PO} \tag{10}$$

and draw a Poisson sample using these latter (non-πps) inclusion probabilities. The term *PoMix* is used because the inclusion probabilities alternatively can be written as

$$\pi_k = \frac{w}{f}\pi_k^{BE} + (1 - \frac{w}{f})\pi_k^{PO} \tag{11}$$

where $\pi_k^{BE} = f$ is the constant inclusion probability that would be used in Bernoulli sampling with expected size n (which, of course, is a special case of Poisson sampling.) Choosing $w = 0$ leads to Poisson $\pi ps(z)$ sampling, $w = f$ leads to Bernoulli sampling, while intermediate w values can be regarded as a Poisson-Bernoulli sampling mix.

The authors conduct a Monte Carlo study (based on slightly modified real data from 1,000 Finnish enterprises), where four different estimators for the population total t_y, the π (HT) estimator and three variants of the $GREG$ estimator, are considered, while using various Bernoulli widths. The three $GREG$ estimators behave similarly, and they outperform by far, as expected, the π estimator. The authors also note that for their particular study the choice $w = 0.3f$ seems to be optimal, and it is offered as a tentative general recommendation. However, no strong underpinning of this recommendation is given.

2.2.2. Fixed size PoMix *sampling.* Choosing a random size design like Poisson πps sampling instead of a corresponding fixed size design inflates the variance of the π estimator. However, there is essentially no such variance penalty when using $GREG$ estimation. On the other hand, we still have to put up with an unpredictable sample size, an annoyance to many practitioners. This is the reason for Kröger, Särndal and Teikari (2000) to consider a fixed size *PoMix* approach based on order πps sampling.

The approach is essentially as follows. Let $\lambda_k^{ORD} = nz_k/t_z$ be target inclusion probabilities that would be used for an order $\pi ps(z)$ sampling design (sequential Poisson sampling and Pareto sampling are used in the paper), let $f = n/N$ be the sampling fraction, and fix a value $w \in [0, f]$, where w is the width parameter. Compute the *PoMix* first-order target inclusion probabilities by the following linear transformation of the λ_k^{ORD}

$$\lambda_k = w + (1 - w/f)\lambda_k^{ORD} \tag{12}$$

and draw an order sample using these latter (non-πps) target inclusion probabilities. The target inclusion probabilities can also be written as

$$\lambda_k = \frac{w}{f}\lambda_k^{SI} + (1 - \frac{w}{f})\lambda_k^{ORD} \tag{13}$$

where $\lambda_k^{SI} = f$ is the constant target inclusion probability that would be used in simple random without-replacement sampling (SI for short) of size n (which, of course, is a special case of order sampling.) Choosing $w = 0$ leads to order $\pi ps(z)$ sampling, $w = f$ leads to SI, while intermediate w values can be regarded as an Order-SI sampling mix.

The authors undertake a Monte Carlo study, based on artificial data generated from models where the regression of y on x (one auxiliary variable) is linear through the origin with three different degrees of heteroscedasticity using two sets of x_k-values with different skewness. Two different $GREG$ estimators are included (together with the quasi-HT estimator.) Among the conclusions are a recommendation to use a value of w in the range $0.2f$ to $0.6f$ for populations of the type considered in the study. The paper leaves unanswered the question of an optimal choice of w.

2.3. Model-based stratified simple random sampling. Suppose that the regression model ξ underlying the $GREG$ estimator is as assumed by equations (4)-(5). Model-based stratified simple random sampling (mb-$STSI$), proposed by Wright (1983), is an approach which comes close to an optimal design. It is achieved as follows (borrowing from the presentation in SSW).

1. Order the values σ_k in increasing magnitude

$$\sigma_{(1)} \leq \sigma_{(2)} \leq \ldots \leq \sigma_{(N)} \tag{14}$$

2. Let H be the number of strata wanted, and calculate $\sum_U \sigma_{(k)} = N\bar{\sigma}$. In the first stratum, U_1, include the first N_1 elements ordered as in (14) so that $\sum_{U_1} \sigma_{(k)}$ is as close as possible to $N\bar{\sigma}/H$. In the second stratum, U_2, include the next N_2 elements ordered as in (14) so that $\sum_{U_2} \sigma_{(k)}$ is as close as possible to $N\bar{\sigma}/H$, and so on.

3. Allocate (as close as possible) equally the n sample elements, i.e. take $n_h = n/H$, $h = 1, ..., H$.

4. Select by simple random without-replacement sampling n_h elements from U_h, $h = 1, ..., H$.

2.4. Comments on the choice of sampling design. Suppose that the regression model (4)-(5) is such that the heteroscedasticity is given by $\sigma_k^2 = \sigma^2 z_k^\gamma$, i.e.

$$y_k = \mathbf{x}_k'\beta + \varepsilon_k$$

with

$$\begin{cases} E_\xi(\varepsilon_k) = 0 \\ V_\xi(\varepsilon_k) = \sigma_k^2 = \sigma^2 z_k^\gamma \\ E_\xi(\varepsilon_k\varepsilon_l) = 0; \quad k \neq l \end{cases} \tag{15}$$

and that we want to use the $GREG$ estimator (2). In this case, using equation (7), an (approximately) optimal design is such that

$$\pi_{0k} = n\sigma_k / \sum\nolimits_U \sigma_k = n z_k^{\gamma/2} / \sum\nolimits_U z_k^{\gamma/2} \tag{16}$$

Furthermore, suppose that we in the planning phase of the survey use γ_p, which means that we misjudge the heteroscedasticity if $\gamma_p \neq \gamma$. This may for example be the case if we use the standard approach $\gamma_p = 2$, leading to inclusion probabilities directly proportional to z. Whatever the reason, we end up with the inclusion probabilities

$$\pi_{pk} = n z_k^{\gamma_p/2} / \sum\nolimits_U z_k^{\gamma_p/2} \tag{17}$$

Let $\pi_{pk}(w)$ be the transformed inclusion probabilities used in $PoMix$ sampling, and let $\hat{t}_{yGREG0}$ and $\hat{t}_{yGREGp(w)}$ be the $GREG$ estimators based on the two sets of inclusion probabilities, given by (16) and (17), respectively. (In the case of order sampling the π_{0k} and π_{pk} are replaced by the corresponding λ_{0k} and λ_{pk}, respectively.) The two ANVs are now given by

$$ANV(\hat{t}_{yGREG0}) = \sigma^2 \sum\nolimits_U (\pi_{0k}^{-1} - 1) z_k^\gamma \tag{18}$$

and

$$ANV(\hat{t}_{yGREGp(w)}) = \sigma^2 \sum\nolimits_U [\pi_{pk}(w)^{-1} - 1] z_k^\gamma \tag{19}$$

respectively. Finally, let

$$AVP(\gamma_p, w) = \frac{ANV(\hat{t}_{yGREGp(w)})}{ANV(\hat{t}_{yGREG0})} = \frac{\sum_U [\pi_{pk}(w)^{-1} - 1] z_k^\gamma}{\sum_U (\pi_{0k}^{-1} - 1) z_k^\gamma} \tag{20}$$

be a measure of the approximate anticipated variance penalty of using the non-optimal inclusion probabilities

$$\pi_{pk}(w) = w + (1 - w/f) \frac{n z_k^{\gamma_p/2}}{\sum_U z_k^{\gamma_p/2}}$$

Note that $AVP(\gamma_p, 0)$ corresponds to no $PoMix$ transformation at all, and hence measures the approximate anticipated variance penalty of using direct $\pi ps(z^{\gamma_p/2})$ sampling (e.g. Poisson sampling or Pareto sampling).

2.5. Comments based on anticipated variances. Our comments in this subsection will be based on a numerical example, using $N = 1000$ values on a right-skewed z-variable. This set of values is essentially a blown up version of the 1975 population sizes of the 281 smallest Swedish municipalities (e.g. Stockholm, Gothenburg and Malmö are excluded), and it could be seen as a rather typical size variable when πps sampling is contemplated. The (expected) sample size is $n = 50$.

2.5.1. Effects of using nonoptimal direct $\pi ps(z^{\gamma_p/2})$ sampling. In order to find a near optimal design, it is essential to have decent a priori knowledge of the value of the heteroscedasticity parameter γ. Table 1 portrays the approximate anticipated variance penalty for a few combinations of (γ, γ_p).

Table 1. $AVP(\gamma_p, 0)$ for selected values of (γ, γ_p)

		γ_p					
		0.00	**0.50**	**1.00**	**1.50**	**2.00**	**2.50**
	0.00	1	*1.07*	*1.32*	*1.91*	*3.38*	*7.52*
	0.50	1.06	1	*1.06*	*1.27*	*1.72*	*2.73*
γ	**1.00**	1.24	1.05	1	*1.05*	*1.24*	*1.62*
	1.50	1.56	1.22	1.05	1	*1.05*	*1.22*
	2.00	2.10	1.52	1.21	1.05	1	*1.05*
	2.50	2.99	2.03	1.50	1.20	1.05	1

Table 1 gives an indication of the penalty that results from choosing an "incorrect" planning value for γ. For $|\gamma_p - \gamma| \leq 0.5$ there is only a slight variance increase, while $|\gamma_p - \gamma| > 1$ may lead to severely increased variances. For example, using $\gamma_p = 2$, corresponding to inclusion probabilities proportional to z, when γ in fact is 0.5, leads to a 72% variance increase.

Since most surveys have many survey variables which may exhibit quite different degrees of heteroscedasticity and since we can only use one set of inclusion probabilities in a specific survey, it is obvious that planning for good variance properties of one estimator, may be bought at a high price for estimators of other survey variable totals.

Table 1 also indicates that if you have good knowledge of γ, i.e. can choose a γ_p-value close to γ, there is, from a pure optimality perspective, little to win by using $PoMix$.

2.5.2. Effects of using PoMix sampling. In Kröger et al (1999) $PoMix$ sampling from an empirical population is studied using $\gamma_p = 2$ when $\gamma \approx 1.45$. In their application, the best choice of w/f seems to lie in the neighborhood of 0.3. As a tentative recommendation, they propose the use of $w = 0.3f$.

In table 2 below we give optimal choices (determined by a numerical search algorithm) of w/f for different combinations of (γ, γ_p) together with the corresponding anticipated variance penalties for $PoMix$ sampling from our population.

Table 2. $AVP(\gamma_p, w_{opt})$ indexed by w_{opt}/f for selected values of (γ, γ_p)

		γ_p					
		0.00	**0.50**	**1.00**	**1.50**	**2.00**	**2.50**
	0.00	1	*1*$_1$	*1*$_1$	*1*$_1$	*1*$_1$	*1*$_1$
	0.50	1.06	1$_0$	*1.00*$_{0.47}$	*1.01*$_{0.66}$	*1.01*$_{0.75}$	*1.01*$_{0.82}$
γ	**1.00**	1.24	1.05$_0$	1$_0$	*1.01*$_{0.30}$	*1.02*$_{0.45}$	*1.03*$_{0.58}$
	1.50	1.56	1.22$_0$	1.05$_0$	1$_0$	*1.01*$_{0.20}$	*1.03*$_{0.35}$
	2.00	2.10	1.52$_0$	1.21$_0$	1.05$_0$	1$_0$	*1.01*$_{0.14}$
	2.50	2.99	2.03$_0$	1.50$_0$	1.20$_0$	1.05$_0$	1$_0$

From table 2 it is clear that $w_{opt} = 0$ when $\gamma_p < \gamma$, i.e. no design mix. When $\gamma_p > \gamma$, it is always possible to find a $PoMix$ transformation which leads to an almost optimal design. For example, using the standard value $\gamma_p = 2$, we can for known γ always find w_{opt}. (Note that, for a given planning value γ_p, larger $(\gamma_p - \gamma)$ differences demands larger w_{opt}.) However, since this demands very good knowledge of the heteroscedasticity pattern, it will be difficult to achieve.

Remark. It seems to be difficult to find a general closed analytical exact expression for w_{opt}. In our specific application, a good approximation when $\gamma \leq \gamma_p$ is given by $w_{opt}/f \approx (1-f/2)(1-\gamma/\gamma_p)$, extending a hint by Axelson (2001).

Kröger et al (1999) recommend the use of $w = 0.3f$ as a compromise value. In table 3 the anticipated variance penalty is given when using this compromise value. To make it easier to compare $PoMix$ sampling, using the compromise value for w, with direct $\pi ps(z^{\gamma_p/2})$ sampling we also give the ratio between $ANV(\hat{t}_{yGREGp(0.3f)})$ and $ANV(\hat{t}_{yGREGp(0)})$ in table 4, which thus directly indicates when the use of $PoMix$ might be profitable.

Table 3. $AVP(\gamma_p, 0.3f)$ for selected values of (γ, γ_p).

		γ_p					
		0.00	**0.50**	**1.00**	**1.50**	**2.00**	**2.50**
	0.00	1	*1.03*	*1.11*	*1.23*	*1.38*	*1.54*
	0.50	1.06	1.01	*1.01*	*1.06*	*1.14*	*1.25*
γ	**1.00**	1.24	1.09	1.02	*1.01*	*1.03*	*1.10*
	1.50	1.56	1.29	1.13	1.04	*1.02*	*1.03*
	2.00	2.10	1.65	1.35	1.17	1.07	*1.04*
	2.50	2.99	2.24	1.74	1.41	1.22	1.11

From tables 3 and 4 it is obvious that in situations where $\gamma_p > \gamma$, we will be better off using $PoMix$ than using direct $\pi ps(z^{\gamma_p/2})$ sampling, since $PoMix$ is less sensitive to deviations of γ_p from γ.

Table 4. $ANV(\hat{t}_{yGREGp(0.3f)})/ANV(\hat{t}_{yGREGp(0)})$ for selected values of (γ, γ_p).

		γ_p					
		0.00	**0.50**	**1.00**	**1.50**	**2.00**	**2.50**
	0.00	1	*0.96*	*0.84*	*0.64*	*0.41*	*0.20*
	0.50	1	1.01	*0.95*	*0.83*	*0.66*	*0.46*
γ	**1.00**	1	1.04	1.02	*0.96*	*0.83*	*0.68*
	1.50	1	1.06	1.08	1.04	*0.97*	*0.84*
	2.00	1	1.09	1.12	1.11	1.07	*0.99*
	2.50	1	1.10	1.16	1.18	1.16	1.11

2.5.3. Effects of using mb-STSI. Good stratified simple random sampling is often a robust approach, which is one of the reasons for its popularity among practitioners. Since model based stratified simple random sampling is expected to come close to being an optimal design, one might expect this design to be a good compromise candidate in survey settings where the use of πps sampling is contemplated.

Let $ANV_{mb\text{-}STSI_H}$ denote the approximate anticipated variance of the $GREG$ estimator under $mb\text{-}STSI$ using H strata. In table 5 the ratio between $ANV_{mb\text{-}STSI_H}$ and $ANV(\hat{t}_{yGREGp(0)})$ is given for a few values of γ and γ_p. From this table we see that $mb\text{-}STSI$ seems to be less sensitive to nonoptimal choices of γ than πps. It should therefore be consider to be a serious contender to various πps designs, since it also, like $PoMix$, avoids the problem of extremely small inclusion probabilities.

Table 5. $ANV_{mb\text{-}STSI_{10}}/ANV(\hat{t}_{yGREGp(0)})$ for selected values of (γ, γ_p),

		γ_p					
		0.00	**0.50**	**1.00**	**1.50**	**2.00**	**2.50**
	0.00	1	*0.99*	*0.94*	*0.81*	*0.59*	*0.34*
	0.50	1	1	*1.00*	*0.97*	*0.88*	*0.71*
γ	**1.00**	1	1.01	1.01	*1.02*	*1.00*	*0.94*
	1.50	1	1.01	1.02	1.02	*1.03*	*1.03*
	2.00	1	1.03	1.02	1.02	1.04	*1.05*
	2.50	1	1.04	1.05	1.03	1.04	1.05

2.6. Comments based on a minor simulation study. Anticipated variances give rough guidelines for practical work. However, for any real finite population, factual conditions will deviate more or less from model assumptions. Hence, it will be valuable to get insights from actual sampling from finite populations that mirror such conditions. To this end, we will now give a few results from a minor simulation study.

2.6.1. Study design. For the simulation study we have used a finite population U of size $N = 1,000$, three study variables and one auxiliary variable. This population is a blown up version of an authentic population, *viz. MU281*, which consists of the 281 smallest Swedish municipalities according to the 1975 population size as given in *SSW* (Appendix B). Although we use an artificial population, it should closely mirror actual real-world conditions, since it has been created using a method suggested by Vale and Maurelli (1983), which essentially preserves the correlation structure, marginal means, variances, and coefficients of skewness and kurtosis of the original population with respect to study variables and auxiliary variables. In the present study *P75* (1975 population) is used as the only auxiliary variable, while *REV84* (1984 real estate values), *P85* (1985 population) and *RMT85* (1985 revenues from municipal taxation) are used as study variables. *P75* serves as auxiliary variable x in the regression estimator for each of the study variable totals, and it is also used as size variable z in the designs. The regression of each of the three study variables on the auxiliary variable is linear, deviating more or less from the origin. Summary results for ordinary (unweighted) least squares linear regressions on the population data is given in table 6, together with ML estimates of γ according to Harvey (1976).

Table 6. Summary results of population OLS linear regression and ML estimates of γ

Regression	R^2 (%)	$\hat{\gamma}$
$REV84 = 506.8 + 94.1P75$	83.5	0.6
$P85 = 0.30 + 1.03P75$	99.1	1.1
$RMT85 = -17.1 + 8.55P75$	97.5	1.5

For each $\gamma_p = 0$, 0.5, 1, 1.5, 2 and 2.5 three sampling designs were used, *viz.* (D1) Pareto $\pi ps(z^{\gamma_p/2})$, (D2) PoMix based on a Pareto-SI mix with $\lambda_k = w + (1 - w/f)\lambda_k^{Par}$, where $w = 0.3f$ and $\lambda_k^{Par} = nz_k^{\gamma_p/2}/\sum_U z_k^{\gamma_p/2}$, and (D3) *mb-STSI* with $H = 10$ strata formed by using $\sigma_k^2 = \sigma^2 z_k^{\gamma_p}$. For each of the $6 \times 3 = 18$ combinations $A = 10{,}000$ independent samples of size $n = 50$ were drawn. For sample $a = 1$, ..., A the three study variable totals were estimated using the simple regression estimator

$$\hat{t}_{yGREG} = N[\tilde{y}_s + \hat{B}(\bar{x}_U - \tilde{x}_s)]$$

where $\tilde{y}_s = \left(\sum_s y_k/\pi_k\right)/\hat{N}$, $\tilde{x}_s = \left(\sum_s x_k/\pi_k\right)/\hat{N}$, $\hat{N} = \sum_s 1/\pi_k$, and where $\hat{B} = \left(\sum_s (x_k - \tilde{x}_s)(y_k - \tilde{y}_s)/\pi_k\right)/\sum_s (x_k - \tilde{x}_s)^2/\pi_k$. (For the two designs D1 and D2 the π_k´s were replaced by λ_k´s.) For each of the 18 combinations

the variance of $\hat{t}_{yGREG}$, for each of the three estimators for the study variable population totals, was estimated by

$$S^2(\hat{t}_{yGREG}) = \sum_{a=1}^{A} \left(\hat{t}_{yGREGa} - \sum_{a=1}^{A} \hat{t}_{yGREGa}/A \right)^2 /(A-1)$$

Finally, using Pareto sampling as benchmark, the variance ratios

$$\frac{S^2(\hat{t}_{yGREG})_{PoMix}}{S^2(\hat{t}_{yGREG})_{Pareto}} \text{ and } \frac{S^2(\hat{t}_{yGREG})_{mb\text{-}STSI}}{S^2(\hat{t}_{yGREG})_{Pareto}} \tag{21}$$

were computed for each study variable.

2.6.3. Results and comments. The 18 variance ratios are given in table 7 below.

Table 7. Variance ratios according to (21) using Pareto sampling as benchmark.

		γ_p					
		0	**0.5**	**1**	**1.5**	**2**	**2.5**
REV84	*PoMix*	1	1.04	1.01	0.91	0.77	0.64
($\hat{\gamma} = 0.6$)	*mb-STSI*	0.95	0.93	0.94	0.91	0.88	0.83
P85	*PoMix*	1	1.05	1.05	0.96	0.85	0.73
($\hat{\gamma} = 1.1$)	*mb-STSI*	0.95	0.94	0.92	0.90	0.88	0.87
RMT85	*PoMix*	1	1.06	1.06	0.99	0.88	0.78
($\hat{\gamma} = 1.5$)	*mb-STSI*	0.93	0.96	0.93	0.91	0.88	0.88

Comparing tables 4, 5 and 7, we see that the simulation results by and large give the same picture as the results from the study of anticipated variances. In fact, *mb-STSI* now seems to perform even better.

3. Conclusions

During the last few years a renewed interest in unequal probability sampling has emerged among survey statisticians. In addition to the papers mentioned above, reference might be given to e.g. Sugden et al (1996), Berger (1998), Deville and Tillé (1998), and Agarwal and Kumar (1998). In the present paper we have focussed on sampling designs and schemes which admit the use of permanent random numbers. As is obvious from our brief overview, the introduction of order πps sampling and $PoMix$ sampling has equipped the survey sampling practitioner with new valuable tools. However, it is also obvious from our results, based on anticipated variances and a simulation study from a population that mirrors factual conditions, that the somewhat older model-based stratified sampling approach still is a strong alternative in applications where probability proportional-to-size sampling is considered to be appropriate.

Bibliography

1. Aires, N. (2000). *Techniques to Calculate Exact Inclusion Probabilities for Conditional Poisson Sampling and Pareto πps Sampling Designs.* PhD thesis, Chalmers University of Technology and Göteborg University, Göteborg, Sweden.
2. Agarwal, S.K., and Kumar, P. (1998). On the relative efficiency of estimators of population total in unequal probability sampling when the study variable has weak relationship with size variable. *Computational Statistics & Data Analysis,* **28**, 271-281.
3. Axelson, M. (2001). Personal communication.
4. Berger, Y.G. (1998). Variance Estimation Using List Sequential Scheme for Unequal Probability Sampling. *Journal of Official Statistics,* **14**, No. 3, 315-323.
5. Brewer, K.R.W., and Hanif, F. (1983). *Sampling with Unequal Probabilities.* Springer-Verlag, New York.
6. Deville, J-C, and Tillé, Y. (1998). Unequal probability sampling without replacement through a spitting method. *Biometrika,* **85**, 1, 89-101
7. Harvey, A.C. (1976). Estimating Regression Models with Multiplicative Heteroscedasticity. *Econometrika*, **44**, No. 3, 461-465
8. Kröger, H., Särndal, C.E., and Teikari, I. (1999). Poisson mixture sampling: a family of designs for coordinated selection using permanent random numbers. *Survey Methodology*, **25**, 3-1.
9. Kröger, H., Särndal, C.E., and Teikari, I. (2000). Poisson Mixture Sampling Combined with Order Sampling: a Novel Use of the Permanent Random Number Technique. Manuscript submitted for publication (00/08/30).
10. Madow, W.G. (1949). On the theory of systematic sampling, II. *Ann. Math. Statist.* **20**, 333-354.
11. Ohlsson, E. (1990). Sequential Poisson Sampling from a Business Register and its Application to the Swedish Consumer Price Index. Statistics Sweden, R&D Report 1990:6.
12. Ohlsson, E. (1995). Coordination of Samples Using Permanent Random Numbers. In: *Business Survey Methods.* John Wiley & Sons, New York.
13. Ohlsson, E. (1998). Sequential Poisson Sampling. *Journal of Official Statistics, **14**,* 135-158.
14. Rosén, B. (1997a). Asymptotic Theory for Order Sampling. *J. Stat. Plan. Inf.*, **62**, 159-191.
15. Rosén, B. (1997b). On Sampling with Probability Proportional to Size. *J. Stat. Plan. Inf.*, **62**, 159-151.
16. Rosén, B. (2000). Generalized Regression Estimation and Pareto *πps*. Statistics Sweden, R&D Report 2000:5.
17. Särndal, C.E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling.* Springer-Verlag, New York.

18. Sugden, R.A., Smith, T.M.F., and Brown, R.P. (1996). Chao's list sequential scheme for unequal probability sampling. *Journal of Applied Statistics,* **23**, 4, 413-421.
19. Traat, I., Bondesson, L., and Meister, K. (2000). Distribution Theory for Sampling Designs. Research Report, Vol. 6, No. 2, Dept. of Math. Stat., Umeå University.
20. Vale, C.D., and Maurelli, V.A. (1983). Simulating nonnormal distributions. *Psychometrika,* **48**, 465-471.
21. Wright, R.L. (1983). Finite population sampling with multivariate auxiliary information. *Journal of the American Statistical Association*, **78**, 879-884.

Department of Research and Development, Statistics Sweden, SE-701 89 Örebro, Sweden.
E-mail: Anders.Holmberg@scb.se
Department of Statistics (ESA), Örebro University, SE-701 82 Örebro, Sweden.
E-mail: Bengt.Swensson@esa.oru.se

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.156-162

ISABELLA HUBER

OPTIMAL PORTFOLIO THEORY FOR STABLE DISTRIBUTED RETURNS

In this paper we examine the foundations of the classical portfolio theory for assets with heavy-tailed distributed returns. We show that the optimal allocation can be significantly changed if the heavy-tailedness (non-normality) of the asset returns is taken into account. The paper is based on results obtained jointly with Sergio Ortobelli, University of Bergamo, Svetlozar (Zari) Rachev, University of Karlsruhe and UCSB, and Eduardo Schwartz, Anderson School of Management.

2000 *Mathematics Subject Classifications.* 91B28.

Key words and phrases. Classical portfolio theory, heavy-tailed distributed returns, optimal allocation.

1. Introduction

In this work we introduce a stable (non-Gaussian) model for optimal portfolio choice. While a vast number of empirical studies confirm that financial asset returns are not normally distributed, many concepts in theoretical and empirical finance that have been developed over the past decades still rest upon the assumption that asset returns follow a normal distribution.

The fundamental works of Mandelbrot (1962, 63a,b, 67) and Fama (1965) have sparked considerable interest in studying the heavy-tailedness of the empirical distributions of financial returns. The excess kurtosis found in Mandelbrot's and Fama's investigations led them to reject the normal assumption and propose the stable Paretian distribution as a statistical model for asset returns. The Fama and Mandelbrot conjecture was supported by numerous empirical investigations in the subsequent years (see the recent work of Mittnik, Rachev and Paolella (1997) and the references in Rachev and Mittnik (2000)). In fact, several empirical studies have examined the

distributional properties of stock returns and found that empirical distributions of stock returns are typically fat tailed and more peaked around the origin than the normal distribution.

The practical and theoretical appeal of the stable non-Gaussian approach is given by their attractive properties that are almost the same as the normal one. A relevant desirable property of the stable distributional assumption is that stable distributions have domain of attraction. Hence, any distribution in the domain of attraction of a specified stable distribution will have properties which are close to those of the stable distribution. The second attractive aspect of the stable- Paretian assumption (that is, the stable non-Gaussian assumption) is the stability property, i.e. stable distributions are stable with respect to summation independent identically distributed random variables. Hence, the stability governs the main properties of the underlying distribution. Detailed accounts of theoretical aspects of stable distributed random variables can be found in Janicki and Weron (1994).

The fundamental portfolio optimization problem consists of minimization of the individual risk when the mean portfolio return meets a given benchmark. The goal of this paper is to implicitly propose the mean risk analysis that facilitates the interpretation of the obtained results.

We consider as risk measure the expected value of a power of the mean absolute deviation (hence when the power is equal to two we obtain the classical quadratic optimal problem). We examine the optimal allocation between a riskless return and risky stable distributed returns, and then we compare the allocation obtained with the Gaussian and the stable non-Gaussian disributional assumption for the risky return.

In the second section we introduce the optimal portfolio allocation problem. In the following section we analyze the obtained results for different risk powers. The last section briefly summarizes the results.

2. An optimal allocation problem with stable distributed returns

Consider the problem of finding the optimal allocation in an investment portfolio $x'z$, where x is the weight vector, $z = [z_0, z_1, \ldots, z_d]'$ is the vector of asset returns in which z_0 is the risk free asset return, and $z_i,\ i = 1, 2, \ldots, d$ is the i-th risky asset return. The problem is to minimize the investor's individual risk $E\left(|x'z - E(x'z)|^r\right)$, where r is the "power-of-risk". In this framework, the investor's goal is to find an optimal portfolio that maximizes the expected mean of the increment wealth $E(x'z)$, or, alternatively, an optimal portfolio, that maximizes the utility functional

$$U(x'z) = E(x'z) - cE\left(|x'z - E(x'z)|^r\right), \tag{1}$$

where c and r are positive real numbers.

We assume that the vector of risky asset returns, z, is α-stable distributed, with index of stability $\alpha > 1$ – which implies the existence of the first moment – that is: $z \stackrel{d}{=} S_\alpha(\sigma_z, \beta_z, m_z)$, where α is the index of stability, σ_z is the scale (dispersion) parameter, β_z is the skewness parameter, and m_z is the mean vector of z.

Suppose that short selling is allowed, but it is uniformly bounded, say $-0.03 \leq x_i \leq 1, i = 0, 1, \ldots, d$. Thus we have the following minimization problem:

$$\begin{aligned} &\min_x E\left(|x'z - E(x'z)|^r\right), \\ &x'\mu = m_z, \\ &x'e = 1, \\ &-0.03 \leq x_i \leq 1, i = 0, 1, \ldots, d, \end{aligned} \tag{2}$$

where $\mu = [\mu_0, \mu_1, \ldots, \mu_d]'$ is the vector of means of the asset returns, and $e = [1, \ldots, 1]'$.

In Problem (2) the objective functional is non-differentiable. Applying in this case the subdifferential approach (see B. Morduchovich (1988) and Appendix A) we obtain a numerical solution of (2).

Notice that $r \in [1, \alpha]$ and $1 < \alpha < 2$. In fact when, which is possible only in the Gaussian case, and thus, the optimization problem (2) is equivalent to the Markowitz - Tobin mean-variance model, that is, to the following quadratic programming problem :

$$\begin{aligned} &\min_x x'\mathcal{Q}x, \\ &x'\mu = m_z, \\ &x'e = 1, \\ &-0.03 \leq x_i \leq 1, i = 0, 1, \ldots, d, \end{aligned} \tag{3}$$

where $\mathcal{Q}$ is the covariance matrix of the multivariate normal distributed vector of returns r.

In this case the utility functional is the quadratic one. As it is well-known (see the references in Ortobelli, Rachev and Schwartz (1999)) the quadratic utility is sufficient for asset choice to be completely described in terms of a preference relation defined over the mean and variance of expected returns, but quadratic utility displays the undesirable properties of satiation and increasing absolute risk aversion. Thus economic conclusions based on the assumption of quadratic utility functions are often counter-intuitive and are not applicable to individuals who always prefer more wealth to less and who treat risky investments as normal goods.

3. Analysis of the numerical results

We numerically solved problem (2) for series of index-daily returns (the same data were used by Ortobelli, Rachev, Huber and Schwartz (2000) and Ortobelli, Rachev, Schwartz (1999)).

(i) 13 risky asset returns and a risk free asset return;

(ii) 3 risky asset returns (DAX30, CAC40 and S & P500) and a risk free asset return;

(iii) one risky asset return (S & P500) and a risk free one.

The risk free return is given by the three-month-LIBOR, 6 % p.a.

In order to make a choice for the power-of-risk r in problem (2), we first estimate the stable parameters of the asset returns in the above portfolio. Below is the table of estimated parameters of the stable fit to the sample distribution of z. The estimation procedure is done via ML- method, see Rachev and Mittnik (2000) and the references there in.

Table 1: Estimated stable daily index parameters

andreas.huber@ mannheim.deAsset	α	β	μ	σ
DAX 30	1.6541	-0.3085	0.0010	0.0076
DAX 100 Perfomance	1.6311	-0.2870	0.0010	0.0070
CAC 40	1.8107	-0.4292	0.0006	0.0087
FTSE ALL SHARE	1.7453	-0.1140	0.0005	0.0053
FTSE 100	1.8066	-0.0429	0.0006	0.0062
FTSE ACTUARIES 350	1.7599	-0.1052	0.0006	0.0056
Nikkei 300 weighted stock average	1.7244	0.0293	0.0001	0.0080
Nikkei 300 simple average	1.7167	-0.0036	0.0003	0.0074
Nikkei 500	1.7190	-0.0944	0.0000	0.0075
Corn No2 Yellow cents	1.6833	-0.1907	0.0000	0.0083
Coffee Brazilian	1.5763	-0.0587	0.0000	0.0143
Dow Jones Industrials	1.7355	-0.2471	0.0009	0.0049
S & P 500 Composite	1.6976	-0.0677	0.0010	0.0046

Because $r \in [1, \alpha)$ we performed the tests for the following values $r = 1, r = 1.5$, and $r = 1.65$. We also considered the normal case when $r = 2$. The results of the numerical solution are shown on the figures 1-6.

The numerical tests showed that the best value of the risk power for the other portfolios is $r = 1$ (Fig. 1, 2, 4). This value is often used in practice because it satisfies the condition $r < \alpha$, $(\alpha \in (1, 2])$. In this case it is not necessary to estimate the index of stability for each asset return.

The efficient frontiers for $r = 1.5, 1.65$ and 2 are represented in Figures 3, 5 and 6. Zenios (1993) obtained similar results in the Gaussian case. These graphs show that model (2) is sensitive to numerical errors and the obtained real efficient frontiers are not optimal.

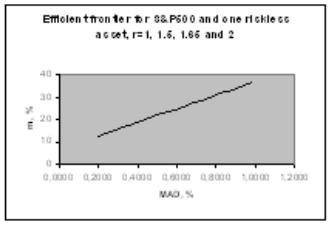


Figure 1

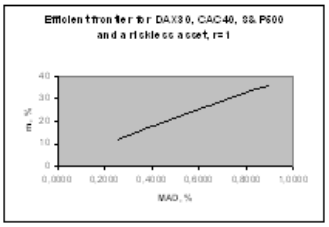


Figure 2

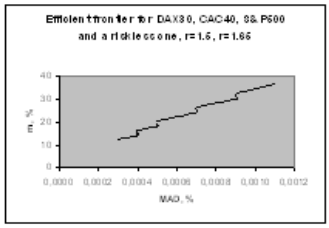


Figure 3

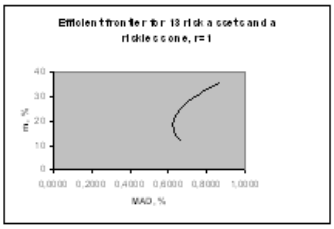


Figure 4

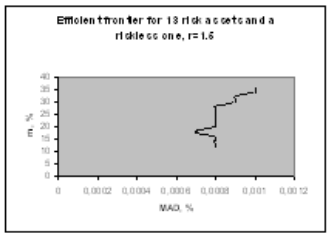


Figure 5

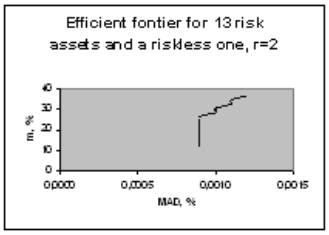


Figure 6

4. Conclusions

We introduce a numerical method for solution of an optimal allocation problem in portfolio of assets that have different individual stable parameters, thus allowing for various heavy-tailedness of the asset returns' distributions. The algorithm does not require imposing a general stable index α for all the assets in the portfolio in contrast to the analytical approach.

We believe that our approach can be a basis for software development the goal of which is (i) to perform statistical forecast of distributions of asset returns according to the historically estimated stable parameters and then (ii) to obtain an optimal portfolio using the above proposed numerical algorithm.

In our next paper we will present the new results on the model enhancement: adjusting the input data using prior information, add constraints (turnover, diversification, bounds) and testing for minimum- variance efficiency within confidence limits.

Appendix A

We approximate the theoretical minimizing functional by it sample counterpart:

$$E\left(|x'z - E(x'z)|^r\right) \sim \frac{1}{N}\sum_{j=1}^{N}\left|\sum_{i=0}^{d} x_i(z_i^{(j)} - \mu_i)\right|^r, \qquad (A1)$$

where μ_i is the mean of the i-th asset return. Denote $c_{i,j} = (z_i^{(j)} - \mu_i)/N$. Then the problem can be written in the following form

$$\begin{array}{l} \min_{x_i} \sum_{j=1}^{N}\left|\sum_{i=0}^{d} x_i c_{i,j}\right|^r, \\ \sum_{i=0}^{d} x_i\mu_i = m_z, \\ \sum_{i=0}^{d} x_i = 1, \\ -0.03 \le x_i \le 1, i = 0, 1, \ldots, d. \end{array} \qquad (A2)$$

The numerical solution of problem (A2) requires the computation of the functional and subjective gradients (Optimization Toolbox User's Guide, Copyright 1990-1997 by MathWorks, Inc.). First we find the functional gradient:

$$\frac{\partial f}{\partial x_i} = \frac{\partial}{\partial x_i}\sum_{j=1}^{N}\left|\sum_{i=0}^{d} x_i c_{i,j}\right|^r = \sum_{j=1}^{N}\left(\frac{\partial}{\partial x_i}\left|\sum_{i=0}^{d} x_i c_{i,j}\right|^r\right) = r\sum_{j=1}^{N}\left|\sum_{i=0}^{d} x_i c_{i,j}\right|^{r-1}\tilde{c}_{i,j},$$

where

$$\tilde{c}_{i,j} = \begin{cases} c_{i,j}, & x_i > 0, \\ [-c_{i,j}, c_{i,j}], & x_i = 0, \\ -c_{i,j}, & x_i < 0, \end{cases} \qquad i = 0, 1, \ldots, d,\ j = 1, 2, \ldots, N.$$

We set $g_1 = \sum_{i=1}^{d} x_i\mu_i - m_z$, $g_2 = \sum_{i=1}^{d} x_i - 1$. Then the subjective gradients are $\partial g_1/\partial x_i = \mu_i$, $\partial g_2/\partial x_i = 1$, $i = 0, 1, \ldots, d$.

Bibliography

1. Fama, E., *The behavior of stock market prices.* Journal of Business 38, (1965a), 34-105.
2. Haugen, R. A., *Modern investment theory,* Prentice Hall, Upper Saddle River, NJ07458, (1997).
3. Janicki, A. and Weron, A. *Simulation and chaotic behavior of stable stochastic processes,* New York: Marcel Dekker, (1994).
4. Mandelbrot, B. B., *Paretian distribution and income maximization,* Quartery Journal of Economics, 76, (1962a), 57-85.
5. Mandelbrot, B., *New methods in statistical economics,* Journal of Political Economy 71, (1963a), 421-440.
6. Mandelbrot, B., *The variation of certain speculative prices,* Journal of Business, 26, (1963b), 394-419.
7. Mandelbrot, B., *The variation of some other speculative prices,* Journal of Business, 40, (1967a), 393-413.
8. Mittnik, S., Rachev, S., Paolella, M., *Stable paretian modelling in finance: some empirical and theoretical aspects,* A practical guide to heavy tails: statistical techniques for analyzing heavy tailed distributions, R. Adler et al, eds, Boston: Birkhauser, (1997).
9. Morduchovich, B., *Approximations methods in problems of optimization and control.* NAUKA, Moscow, (1988), (in Russian).
10. *Optimization Toolbox: User's Guide,* Copyright 1990- 1997 by the MathWorks, Inc.
11. Ortobelli, S. L., Rachev, S. T., Schwartz, E., *On the problem of optimal portfolio with stable distributed returns,* Preprint, Chair of Econometrics and Statistics, Department of Economics, University of Karlsruhe, (1999).
12. Ortobelli, S. L., Rachev, S. T., Huber, I., Schwartz, E., *Optimal portfolio choice for assets with heavy-tailed distributed returns: evidence from the German stock market,* Preprint, Chair of Econometrics and Statistics, Department of Economics, University of Karlsruhe.
13. Rachev, S. and Mittnik, S., Stable model in finance, New York: Wiley, (2000).
14. Rachev, S. T., Racheva-Jotova, B., Hristov, B., *New approach to risk management,* The Mercury Software. Technical Documentation, Bravo Risk management Group Ltd, (2000).
15. Zenios, S. A., *Financial optimization,* Cambridge University Press, (1993).

Chair of Econometrics and Statistics, Department of Economics, University of Karlsruhe, Germany
E-mail: huber@lsoe.uni-karlsruhe.de

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.163-188

HENRIK JÖNSSON

MONTE CARLO STUDIES OF AMERICAN TYPE CALL OPTIONS WITH DISCRETE TIME[1]

The paper presents an algorithm for studies of optimal stopping domains of American type call options with discrete time. To study the stopping domain for each moment before the expiration day we use a grid structure with discrete points. The idea is to compare the profit from exercising the option with the expected profit of a future exercise for every point on the grid. The expected profit is estimated using a Monte Carlo method. Stopping domains for several different types of payoff functions are presented. The paper also presents studies of the probability of classification errors.

AMS 2000 subject classifications. 60J05, 60G40, 65C05, 91B28, 91B70.

Key words and phrases. Markov process, optimal stopping, Monte Carlo methods, Mathematical Economics, Finance, Stochastic models, American options.

1. Introduction

An option is a contract between a seller and a buyer, concerning some underlying asset, giving the buyer the right, but not the obligation, to buy (or sell) the underlying asset to a fixed price called the strike price. The underlying asset can for example be a stock, a foreign currency or a stock index. The option is only valid during a specific time period and it expires at the expiration day.

There are several different types of options; e.g. American and European. The difference between an American type option and an European type option is in the ways to exercise the option. For the American type option, it is possible to exercise the option at any given time until the expiration date. For the European type option, on the other hand, it is only possible

[1]This work is supported in part by the project "Stochastic modelling of insurance and finance processes and systems" funded by The Knowledge Foundation.

to exercise the option at the expiration date. For the holder of an European type option it is always possible to resell the option in order to generate a profit or minimize losses, and this reselling is similar to the exercising of an American type option. In this paper only American type options are considered.

For both the buyer and the seller of the option, it is important to maximize the profit of the investment. In this paper we study only optimal strategies for buyer of American type options. The buyer of the American type option is interested to determine when to exercise the option in order to maximize the profit. The problem is to find the optimal time for exercising the option.

The optimal time to exercise the option depends on which type of payoff function that is used. There is a specific payoff function for each option type.

Optimal time to exercise the option is determined by the optimal stopping domain. The stopping domain is defined for each moment as the set of all stock prices for which it is better to exercise the option than to keep it. The knowledge of the structure of the stopping domain of a payoff function can be used to decide when to exercise an American type option. The first time the price of the underlying stock enter the stopping domain, the rule has to be to exercise the option. The structure of the stopping domain varies between different types of payoff functions. The structure of the domain also depends on the volatility and the drift of the underlying stock, and the risk-free interest rate in the market.

The present paper presents studies of the structure of the stopping domain for different types of payoff functions, in particular standard linear, piecewise linear, quadratic, stepwise and logarithmic, for American type options in discrete time. The studies are based on the Monte Carlo method and the underlying stock is modelled with a geometrical random walk with multiplicative increments with log-normal distribution.

Our discussion is based on the results of Kukush and Silvestrov (2000), where theoretical studies of the stopping domains for standard, piecewise linear convex and general convex payoff functions are presented. Boyle (1977) introduced option pricing using the Monte Carlo method and a recent survey of Monte Carlo methods in finance is given in Boyle, Broadie and Glasserman (1997). A large variety of numerical methods in finance is presented in the book edited by Rogers and Talay (1997). A survey of recent numerical methods for pricing derivative securities are given in the paper by Broadie and Detemple (1997).

In Section 2 different types of payoff functions are presented, a mathematical definition of the stopping domains is also given and the model of the underlying pricing process is introduced.

The Monte Carlo algorithm used to study the structure of the optimal stopping domains are described in Section 3. To study the stopping domain,

an upper and a lower threshold value of the initial stock price are set. Between the boundaries we use a grid structure to choose the initial stock price, i.e. the initial stock price can be chosen at levels which are increased with a fixed delta from the lower threshold up to the upper threshold. The idea is to work backward from the last moment N, the expiration date of the option, until the first moment is reached and for each moment investigate the stopping domain, i.e. which prices are contained in the set. By simulating a large number of trajectories for each stock price on the grid, the expected profit of the option is evaluated. If the expected profit is less than the profit for the given stock price at the given moment, then the stock price is in the stopping domain.

In Section 4 we analyze how good the results of the algorithm are. We do this by investigating the probabilities of classification errors. There are two types of classification errors. First, if the algorithm indicates that the stock price belongs to the stopping domain but the right decision is that it is not. Second, if the algorithm indicates that the stock price does not belongs to the stopping domain, but the right decision is that it does. If we know the probability of making a classification error for one point on the grid, then we can estimate the probability of having a classification error for the grid as a whole. And this gives us a way to decide how good our results are.

Results of the experiments and several examples of the structures found in the experiments are given in Section 5.

We would like to refer to the work by Peter Westermark (1999) where some preliminary studies of the structures of optimal stopping domains for some particular payoff functions were made with analogous methods. The Bernoullian geometrical random walk was used as the model of the pricing process of the underlying asset in this work.

Our studies cover a wider class of payoff functions and are based on a more advanced model of the underlying pricing process, namely, the geometrical random walk with log-normal multiplicative increments. This model is a more realistic discrete analogue of the classical Brownian motion than the Bernoullian geometrical random walk.

2. American Type Options

The American type option is defined by three parameters; the option price $C \geq 0$, the strike price $K > 0$ and the expiration date N. Since the buyer always pays the price C for the option, the optimal stopping time does not depend of C and hence we can simply set $C = 0$ in the studies of the structure of the stopping domains.

Also, a payoff function is connected with the option. The payoff function

is a function $f : R^+ \to R^+$. We consider only payoff functions which are homogenous in time.

A general formulation of the linear payoff function with a single slope is given by

$$g(x) = a\,[x - K_n]_+ = \begin{cases} a(x-K), & if \quad x > K, \\ 0, & if \quad 0 \le x \le K, \end{cases} \tag{1}$$

where $K > 0$ and $a > 0$ is the strike price and the scale pricing coefficient, respectively. For the payoff function of the standard American call option with discrete time $a = 1$.

It is also interesting to study other types of payoff functions. The following payoff functions are also considered in this paper. For all these payoff functions the parameter K is called the strike price, by analogy with the standard American option.

The piecewise linear payoff function with two intervals with different slopes is given by

$$g(x) = \begin{cases} 0, & if \quad x < K, \\ a_1(x-K), & if \quad x \in [K, K'), \\ a_1(K'-K) + a_2(x-K'), & if \quad x \ge K', \end{cases} \tag{2}$$

where $K' > K > 0$ is the second strike price and $a_1, a_2 \ge 0$ are the scale pricing coefficients for price intervals $[K, K')$ and $[K', \infty)$, respectively.

Also, payoff functions with more than two intervals with different slopes can be considered. However, we restrict our studies in this paper to the simplest case with a payoff function with two intervals with different slopes only.

Another payoff function considered is the quadratic function. A general quadratic payoff function is given by

$$g(x) = \begin{cases} 0, & if \quad x < K, \\ (x-K)^2 & if \quad x \ge K, \end{cases} \tag{3}$$

where $K > 0$ is again the strike price.

We can also construct a payoff function with stepwise increment, where the payoff is constant on a certain level for a given price interval. For each price interval $[0, K_1)$, $[K_1, K_2)$, $[K_2, K_3)$, ..., $[K_{P-1}, K_P)$, $[K_P, \infty)$ payoff values $0 < L_1 < L_2 < ... < L_P < \infty$ are defined. The general stepwise payoff function can be defined by

$$g(x) = \begin{cases} 0, & if \quad x \in [0, K_1), \\ L_1, & if \quad x \in [K_1, K_2), \\ L_2, & if \quad x \in [K_2, K_3), \\ \vdots & \vdots \quad \vdots \\ L_P, & if \quad x \in [K_P, \infty). \end{cases} \tag{4}$$

In this paper we consider only stepwise payoff functions with two or three steps.

The logarithmic function can also be interesting as a payoff function, since it is non-convex, but monotonically increasing, and is commonly used as a utility function. The logarithmic payoff function can be given by

$$g(x) = \begin{cases} 0, & if \quad x \leq K, \\ a\ln(x-K), & if \quad x > K, \end{cases} \tag{5}$$

where again $K > 0$ and $a > 0$.

To model the underlying asset we consider a discrete time pricing process. The discrete time pricing process considered can be formulated as

$$S_n = A(S_{n-1}, Y_n), n = 0, 1, ..., N \tag{6}$$

where A is a continuous function acting on $R^+ \times R$ to R^+, S_n is the stock price at moment n and Y_n, $n = 0, 1, ..., N$, is a sequence of i.i.d. non-negative real-valued random variables. The initial value S_0 of the pricing process is a non-random value. In this paper the pricing process is considered to be a geometrical random walk with multiplicative increment, i.e.

$$S_n = S_{n-1}Y_n, n = 0, 1, ..., N \tag{7}$$

where Y_n have a log-normal distribution, i.e.

$$Y_n = e^{\mu+\sigma X_n}, n = 0, 1, ..., N, \tag{8}$$

and where μ and σ is the stocks drift and volatility, respectively, and $X_n, n = 0, 1, ..., N$, are independent standard normal random variables.

Since this paper presents studies of the structure of optimal stopping domains, we have to define optimal stopping time and optimal stopping domain.

The optimal stopping time τ^*_{opt} has to be found in the set of all Markov moments $\tau \leq N$ and is given by the moment $\tau^*_{opt} \leq N$ which maximizes the functional

$$\Phi_g(\tau) = Ee^{-r\tau}g(S_\tau), \tag{9}$$

where r is the risk less interest rate. Thus, the optimal stopping time satisfies

$$\Phi_g(\tau^*_{opt}) = \sup_{\tau \leq N} \Phi_g(\tau). \tag{10}$$

The optimal stopping time can be defined, as was proved in Shiryaev (1978), as follows

$$\tau^*_{opt} = \min\{0 \leq n \leq N : S_n \in \Gamma^*_n\}, \tag{11}$$

where Γ_n^* is the optimal stopping domain for moment $n = 0, 1, ..., N$ and $\Gamma_N^* = [0, \infty)$.

The optimal stopping domain can be found using the operator T, acting on a non-negative measurable function $f(x)$, defined as the expectation

$$Tf(x) = Ee^{-r} f(A(x, Y_1)) = \int_R e^{-r} f(A(x, y)) P\{Y_1 \in dy\}, \tag{12}$$

where e^{-r} is the discounting factor of one day. Thus, the operator T is defined as the expectation of a payoff starting from stock price x and taking one step forward.

Further, let $w_0(x) = g(x)$ and for $k = 1, 2, ..., N$ define the recursion

$$w_k(x) = \max\{g(x), Tw_{k-1}(x)\}, \tag{13}$$

where $g(x)$ is the payoff for stock price x.

Then the set

$$\Gamma_n = \{x \in R^+ : g(x) = w_{N-n}(x)\} \tag{14}$$

is defined as the stopping domain for moment n.

For the optimal stopping time τ_{opt}^*, the functional (9) is

$$\Phi_g(\tau_{opt}^*) = w_N(S_0). \tag{15}$$

To find the optimal moment to exercise the option, it is necessary to know the structure of the optimal stopping domain. The above statement of the existence of a stopping domain does not give the explicit structure of such domains. The only thing we know is that the optimal stopping domain exists and is determined by the recurrent equations (13) - (15), thus

$$\Gamma^* \in \{\Gamma_0, \Gamma_1, ..., \Gamma_{N-1}, \Gamma_N = [0, \infty)\}. \tag{16}$$

where $\Gamma_n, n = 0, 1, ..., N$ is a sequence of Borel measurable subsets on R^+. Note that $\Gamma_N = [0, \infty)$, since all stock prices at the expiration day have to belong to the stopping domain.

The structure of the stopping domain varies between different types of payoff functions. Kukush and Silvestrov (2000) investigated the structure for optimal stopping domains for different types of convex payoff functions. In the present paper we have investigated the structure of the optimal stopping domain for different types of payoff functions, in particular standard linear, piecewise linear, quadratic, stepwise and logarithmic, using an algorithm based on the results of Shiryaev (1978) and Kukush and Silvestrov (2000).

3. Description of the Algorithm

To study the structure of the stopping domain we first define an upper and a lower boundary value for the stock price. It is possible do this, since it is unlikely that the stock price will take values in the whole interval $[0, +\infty)$ during the time period of interest. Let s_u and s_l denote the upper and the lower boundary value, respectively.

For each moment $n = 0, 1, ..., N$ we define the stock price as discrete values

$$s_{n,j} = s_l + j\Delta, j = 0, 1, 2, ..., J, \tag{17}$$

such that $s_{n,0} = s_l$ and $s_{n,J} = s_u$ and Δ is chosen to give a reasonable accuracy in the approximative description of optimal stopping domains. The result is a grid structure with discrete points $(n, s_{n,j})$, since both the time and the stock price is discrete.

The idea used to study the stopping domain is to work backwards from the expiration day N and for each point $(n, s_{n,j})$ on the grid compare the profit if we exercise the option at moment n and stock price $s_{n,j}$ with the expected profit if we wait to the moment $n+1$ before exercising the option.

From the formulation of payoff function (1) - (5), only stock prices greater than the lower strike price can belong to the stopping domain, since the profit is zero for all stock prices less than the strike price. Hence, s_l can be equal to the strike price K or K_1, independent of the type of payoff function considered.

For each moment $n = 0, 1, ..., N$ we have to investigate not only the discrete values $s_{n,j}$, but the whole interval $[s_l, s_u] \subset R^+$, since the stopping domain is a Borel measurable set on R^+. This implies that we for each moment n have to decide if the stock prices between two adjacent points $(n, s_{n,j})$ and $(n, s_{n,j+1})$, $j = 0, 1, ..., J$, belongs to the stopping domain or not. Given the interval

$$I_{n,j} = [s_{n,j} - \frac{\Delta}{2}, s_{n,j} + \frac{\Delta}{2}), \tag{18}$$

we say that $I_{n,j} \in \Gamma_n$ if $s_{n,j} \in \Gamma_n$.

Since the option expires at $n = N$, all stock prices at the expiration date have to belong to the stopping domain, i.e.

$$\Gamma_N = [s_l, s_u]. \tag{19}$$

Next, moment $N-1$ is considered. For each stock price $s_{N-1,j}$ the profit $g(s_{N-1,j})$ is compared with the expected profit $Tw_0(s_{N-1,j})$. If $g(s_{N-1,j}) > Tw_0(s_{N-1,j})$ then $I_{N-1,j} \in \Gamma_{N-1}$. The approximative stopping domain for moment $N-1$ is defined by

$$\Gamma_{N-1} = \bigcup_{j:g(s_{N-1,j})>Tw_0(s_{N-1,j})} I_{N-1,j}. \tag{20}$$

We use a Monte Carlo method to estimate the value of the optimal expected continuation profit $Tw_0(s_{N-1,j})$. For each stock price $s_{N-1,j}$ at moment $N-1$ we estimate $Tw_0(s_{N-1,j})$, i.e. the expected profit of the option if we wait to the expiration day N before exercising the option, using

$$\widehat{T}_{N-1}^{(M)} g(s_{N-1,j}) = \frac{1}{M} \sum_{i=1}^{M} e^{-r} g(s_{N-1,j} Y_N^{(i)}), \tag{21}$$

where $Y_N^{(i)}$, $i = 1, ..., M$, are i.i.d. random variables with log-normal distribution. This will give an approximation of the stopping domain (20)

$$\widehat{\Gamma}_{N-1} = \bigcup_{j: g(s_{N-1,j}) > \widehat{T}_{N-1} g(s_{N-1,j})} I_{N-1,j}. \tag{22}$$

For moment $N-2$ we have to take into account that we already know the structure of the stopping domain of moment $N-1$. For each stock price $s_{N-2,j}$ we take one step ahead to moment $N-1$ in each iteration, i.e. we generate a sequence of stock prices $s_{N-1}^{(i)} = s_{N-2,j} Y_{N-1}^{(i)}$, $i = 0, 1, ..., M$. Note that the generated stock price $s_{N-1}^{(i)}$ may not be one of the discrete points defined on the grid, i.e. it is possible that $s_{N-1}^{(i)} \neq s_{N-1,j}$ for all $j = 0, 1, ..., J$. If $s_{N-1}^{(i)}$ does not belong to the stopping domain of moment $N-1$, $s_{N-1}^{(i)} \notin \widehat{\Gamma}_{N-1}$, we have to generate a new stock price $s_N^{(i)} = s_{N-1}^{(i)} Y_N^{(i)}$. Otherwise, we calculate $g(s_{N-1}^i)$.

The estimation of $Tw_1(s_{N-2,j})$, the optimal expected continuation profit for each stock price $s_{N-2,j}$, $j = 0, 1, ..., J$, in moment $N-2$, is calculated using

$$\begin{aligned} \widehat{T}_{N-2}^{(M)} g(s_{N-2,j}) \quad = \tfrac{1}{M} \textstyle\sum_{i=1}^{M} \quad & (e^{-r} g(s_{N-1}^{(i)}) I(s_{N-1}^{(i)} \in \widehat{\Gamma}_{N-1}) + \\ & e^{-2r} g(s_N^{(i)}) I(s_{N-1}^{(i)} \notin \widehat{\Gamma}_{N-1})), \end{aligned} \tag{23}$$

where $s_{N-1}^{(i)} = s_{N-2,j} Y_{N-1}^{(i)}$ and $s_N^{(i)} = s_{N-2,j} Y_{N-1}^{(i)} Y_N^{(i)}$, i.e. the generated stock price in iteration i starting from the stock price $s_{N-2,j}$ and taking one step and two steps ahead, respectively. Further, $I(x \in \Gamma)$ is equal to one if $x \in \Gamma$, otherwise equal to zero.

After the expected profit has been estimated we have to decided if the stock price $s_{N-2,j}$, and thus the interval $I_{N-2,j}$, belongs to the stopping domain for moment $N-2$ or not.

The structure of the stopping domain Γ_{N-2} is approximated by

$$\widehat{\Gamma}_{N-2} = \bigcup_{j: g(s_{N-2,j}) > \widehat{T}_{N-2}^{(M)} g(s_{N-2,j})} I_{N-2,j}. \tag{24}$$

For every moment $n < N-2$ we have to estimate the expected continuation profit $Tw_{N-n-1}(s_{n,j})$ of every stock price $s_{n,j}$, $j = 0, 1, 2, ..., J$ and

we have to use the fact that we know the structure of the optimal stopping domains of moment $n+1, n+2, ..., N-1, N$ when we estimate.

As the algorithm is stated we generate M independent series $Y_0^{(1)}, ..., Y_N^{(1)}$, ... $, Y_0^{(M)}, ..., Y_N^{(M)}$ of i.i.d. random variables with log-normal distribution. In principle this is possible, but in the program we have used independent series of i.i.d. random variables for each point $s_{n,j}$.

To investigate the structure of the stopping domains for different types of payoff function we use an algorithm based on the knowledge of the existence of optimal stopping time, described in Shiryaev (1978), and presented above. The algorithm creates a grid structure with discrete points $(n, s_{n,j})$ that is used to study the stopping domain for each moment. To decide which stock prices that belongs to the stopping domain of moment n, we use a Monte Carlo method to estimate the expected profit of the option if we exercise it in the future. The actual stopping profit if we exercise the option at moment n is compared with the estimate of the optimal expected continuation profit. If the actual stopping profit is greater than the optimal expected continuation profit, then the stock price belongs to the stopping domain of moment n.

4. Analysis of the Algorithm

To analyze the algorithm we study the probability of classification error. A classification error occur when the stock price belongs to the stopping domain, but the algorithm indicates that the stock price does not or vice versa.

There are two types of classification errors. First, if the algorithm indicates that the stock price $s_{n,j}$ belongs to the stopping domain, i.e. $g(s_{n,j}) > \widehat{T}_n^{(M)} g(s_{n,j})$, but the stock price does not, i.e. $g(s_{n,j}) < Tw_{N-n-1}(s_{n,j})$. The probability of making this type of classification error is given by

$$\begin{aligned} p_{n,j} &= P\left\{\widehat{T}_n^{(M)} g(s_{n,j}) < g(s_{n,j})\right\} \\ &= P\left\{\frac{\widehat{T}_n^{(M)} g(s_{n,j}) - Tw_{N-n-1}(s_{n,j})}{\sigma_{n,j}}\sqrt{M} < \frac{g(s_{n,j}) - Tw_{N-n-1}(s_{n,j})}{\sigma_{n,j}}\sqrt{M}\right\} \\ &\simeq 1 - \Phi\left(\frac{Tw_{N-n-1}(s_{n,j}) - g(s_{n,j})}{\sigma_{n,j}}\sqrt{M}\right), \end{aligned} \tag{25}$$

by the central limit theorem and where $\sigma_{n,j}/\sqrt{M}$ is the standard deviation of the estimate $\widehat{T}_n^{(M)} g(s_{n,j})$ and $\sigma_{n,j}$ is the standard deviation of one component in the sum defining $\widehat{T}_n^{(M)} g(s_{n,j})$.

The second type of classification error is when the algorithm indicates that the stock price does not belong to the stopping domain, i.e. $g(s_{n,j}) < \widehat{T}_n^{(M)} g(s_{n,j})$, but the stock price does, i.e. $g(s_{n,j}) > Tw_{N-n-1}(s_{n,j})$. The

probability of this error is given by

$$\begin{aligned} q_{n,j} &= P\left\{\widehat{T}_n^{(M)} g(s_{n,j}) > g(s_{n,j})\right\} \\ &= P\left\{\frac{\widehat{T}_n^{(M)} g(s_{n,j}) - Tw_{N-n-1}(s_{n,j})}{\sigma_{n,j}}\sqrt{M} > \frac{g(s_{n,j}) - Tw_{N-n-1}(s_{n,j})}{\sigma_{n,j}}\sqrt{M}\right\} \\ &\simeq 1 - \Phi\big(\frac{g(s_{n,j}) - Tw_{N-n-1}(s_{n,j})}{\sigma_{n,j}}\sqrt{M}\big). \end{aligned} \tag{26}$$

Thus, the two types of classification error have the same probability.

Note that the probability is proportional to the square root of the number of simulations used M and it also depends on the standard deviation $\sigma_{n,j}$. Thus, there are two ways of decreasing the probability, either we can increase the number of simulations or we can reduce the variance. In this paper we do not analyze any variance reduction techniques.

To calculate these probabilities we have to know the true values of $Tw_{N-n-1}(s_{n,j})$ and $\sigma_{n,j}$. Since this is not possible we have to estimate both $Tw_{N-n-1}(s_{n,j})$ and $\sigma_{n,j}$. We can use the same method as before to estimate $Tg(s_{n,j})$, the only difference is that we have to use a very large number of calculations, something that we want to escape from when we make our investigation about the stopping domains.

To get a good estimate of the standard deviation $\sigma_{n,j}$ we first estimate the second moment of the optimal continuation profit. For moment $N-1$ and $N-2$ this is done by

$$\widetilde{T}_{N-1}^{(M)} g(s_{N-1,j}) = \frac{1}{M}\sum_{i=1}^{M} e^{-2r} g^2(s_{N-1,j} Y_N^{(i)})), \tag{27}$$

and

$$\begin{aligned} \widetilde{T}_{N-2}^{(M)} g(s_{N-2,j}) &= \tfrac{1}{M}\textstyle\sum_{i=1}^{M} \quad (e^{-2r} g^2(s_{N-1}^{(i)}) I(s_{N-1}^{(i)} \in \widehat{\Gamma}_{N-1}) + \\ &\qquad e^{-4r} g^2(s_N^{(i)}) I(s_{N-1}^{(i)} \notin \widehat{\Gamma}_{N-1})), \end{aligned} \tag{28}$$

respectively, and for moment n=0,1,...,N-3, the formulas are similar.

The estimate of the standard deviation is as follows

$$\widetilde{\sigma}_{n,j} = \sqrt{\widetilde{T}_n^{(M)} g(s_{n,j}) - (\widehat{T}_n^{(M)} g(s_{n,j}))^2}. \tag{29}$$

It is easiest to study the probabilities $p_{N-1,j}$ and $q_{N-1,j}$, since all stock prices in moment N is in the stopping domain Γ_N. For moment $n = N-2, ..., 0$ we have to use the knowledge of the stopping domains of the moments $n+1, ..., N$ when we calculate the probabilities.

We can also study the loss in optimality related to the classification error of stock price $s_{n,j}$. Let $g(s_{n,j})$ be the profit if we exercise the option at moment n and let $E_{n,j}^{cont} = Tw_{N-n-1}(s_{n,j})$ be the optimal expected continuation profit, i.e. the expected value of the profit if we wait at least one day

before exercising the option. The loss in optimal profit is given by

$$L^*_{n,j} = \frac{E^{cont}_{n,j} - g(s_{n,j})}{E^{opt}_{n,j}}, \tag{30}$$

where $E^{opt}_{n,j} = \max(E^{cont}_{n,j}, g(s_{n,j}))$. Thus, $L^*_{n,j}$ is the loss or gain of using the continuation strategy relative the optimal strategy.

Instead of using the optimal strategy we can study the loss in optimality relative the stopping strategy,

$$L_{n,j} = \frac{E^{cont}_{n,j} - g(s_{n,j})}{g(s_{n,j})}. \tag{31}$$

If $L_{n,j} > 0$ we know that $E^{cont}_{n,j}$ is optimal and the gain of using the continuation strategy relative the stopping strategy is given by $L_{n,j}$. If $L_{n,j} < 0$ we know that $g(s_{n,j})$ is optimal and the relative loss using the continuation strategy is given by $L_{n,j}$.

To get a dimensionless measure of the variance we use the quotient

$$d^2_{n,j} = \frac{\sigma^2_{n,j}}{(g(s_{n,j}))^2}. \tag{32}$$

If

$$b_{n,j} = \frac{|L_{n,j}|}{d_{n,j}}, \tag{33}$$

then the probability

$$p_{n,j} \simeq 1 - \Phi(\sqrt{M} b_{n,j}). \tag{34}$$

We study only the classification errors for the standard payoff function in this paper. Let us consider in details the case of $n = N - 1$. In the limiting case, when $s_{N-1,j} \to \infty$,

$$\begin{aligned} L_{N-1,j} &= \frac{E^{cont}_{N-1,j} - g(s_{N-1,j})}{g(s_{N-1,j})} = \frac{Ee^{-r}[s_{N-1,j}Y_N - K]^+ - [s_{N-1,j} - K]^+}{[s_{N-1,j} - K]^+} \\ &= \frac{e^{-r}E[Y_N - \frac{K}{s_{N-1,j}}]^+}{[1 - \frac{K}{s_{N-1,j}}]^+} - 1 \to e^{\mu + \frac{\sigma^2}{2} - r} - 1, \end{aligned} \tag{35}$$

$$d^2_{N-1,j} \to e^{2\mu + \sigma^2 - 2r}(e^{\sigma^2} - 1), \tag{36}$$

and

$$b_{N-1,j} \to \frac{\left| e^{\mu + \frac{\sigma^2}{2} - r} - 1 \right|}{\sqrt{e^{2\mu + \sigma^2 - 2r}(e^{\sigma^2} - 1)}}. \tag{37}$$

We have studied the probability of classification error for two different sets of parameters of the underlying pricing process. First we used a small volatility and a zero drift. Second we used a larger volatility and a negative

$s_{N-1,j}$	$g(s_{N-1,j})$	$\widetilde{E}^{cont}_{N-1,j}$	$\widetilde{\sigma}^2_{N-1,j}$	$\widetilde{L}_{N-1,j}$	$\widetilde{d}^2_{N-1,j}$	$\widetilde{b}_{N-1,j}$
101.00	1.00	1.005319	0.245170	0.005319	0.245170	0.010742
102.00	2.00	2.000951	0.259712	0.000476	0.064928	0.001867
103.00	3.00	3.000908	0.263820	0.000303	0.029313	0.001767
104.00	4.00	4.000536	0.270794	0.000134	0.016925	0.001031
105.00	5.00	5.000490	0.273660	0.000098	0.010946	0.000937
106.00	6.00	6.000411	0.281912	0.000069	0.007831	0.000775
107.00	7.00	7.000502	0.285575	0.000072	0.005828	0.000939
108.00	8.00	8.000199	0.284366	0.000025	0.004443	0.000373
109.00	9.00	8.999920	0.299919	-0.000009	0.003703	0.000145
110.00	10.00	9.999982	0.302093	-0.000002	0.003021	0.000034
111.00	11.00	10.999519	0.309146	-0.000044	0.002555	0.000865
112.00	12.00	11.999368	0.317137	-0.000053	0.002202	0.001123
113.00	13.00	12.999380	0.328893	-0.000048	0.001946	0.001082
114.00	14.00	13.999405	0.328798	-0.000042	0.001678	0.001037
115.00	15.00	14.998969	0.341408	-0.000069	0.001517	0.001765

Table 1: Stock price, stopping profit and estimated values of expected continuation profit $E^{cont}_{N-1,j} = Tw_0(s_{N-1,j})$, variance $\widetilde{\sigma}^2_{N-1,j}$, $\widetilde{L}_{N-1,j}$, $\widetilde{d}^2_{N-1,j}$ and $\widetilde{b}_{N-1,j}$. The values of the parameters of the underlying pricing process are $\mu = 0.0$, $\sigma = 0.005$.

drift. In both cases we used a risk less interest rate of four percent per year, a strike price $K = 100$ and $\Delta = 1.0$. Numerical results are presented in Table 1 - Table 4.

Table 1 shows the estimated values $g(s_{N-1,j})$, $\widetilde{E}^{cont}_{N-1,j}$, $\widetilde{\sigma}^2_{N-1,j}$, $\widetilde{L}_{N-1,j}$, $\widetilde{d}^2_{N-1,j}$ and $\widetilde{b}_{N-1,j}$ for stock prices between 101 and 115. The parameters of the underlying pricing process are $\mu = 0.0$ and $\sigma = 0.005$. We used M=10^7 number of simulations to estimate the values.

The first thing that we can see is that every stock price greater than or equal to $s^*_{N-1} = 109$ is in the stopping domain, since $\widetilde{L}_{N-1,j} < 0$. Further, $\widetilde{\sigma}^2_{N-1,j}$ is increasing but $\widetilde{d}^2_{N-1,j}$ is decreasing when the stock price increases. $\widetilde{L}_{N-1,j}$ and $\widetilde{b}_{N-1,j}$ decreases, on the other hand, until $s_{N-1,j} = 110$ and then starts to increase. Thus the loss in optimality is smallest near the threshold value $s^*_{N-1} = 109$. This indicates that the difference between the stopping profit and the optimal expected continuation profit is very small near the threshold value. Note that the loss in optimality is small for all stock prices in the interval. This is due to the fact that the mean of the $Y^{(i)}_N$'s is small, since the values of μ and σ are small.

Table 2 shows the estimated probabilities of making classification error for four different the number of simulations M. The values corresponds to the estimated values shown in Table 1.

		$\widetilde{p}_{N-1,j}$			
$s_{N-1,j}$	$\widetilde{b}_{N-1,j}$	$M = 5 * 10^4$	$M = 10^5$	$M = 10^6$	$M = 10^7$
101.00	0.010742	0.008153	0.000341	0.000000	0.000000
102.00	0.001867	0.338168	0.277465	0.030953	0.000000
103.00	0.001767	0.346342	0.288108	0.038576	0.000000
104.00	0.001031	0.408879	0.372259	0.151386	0.000560
105.00	0.000937	0.417041	0.383525	0.174437	0.001527
106.00	0.000775	0.431246	0.403249	0.219292	0.007154
107.00	0.000939	0.416831	0.383235	0.173819	0.001489
108.00	0.000373	0.466785	0.453081	0.354658	0.119236
109.00	0.000145	0.487043	0.481679	0.442247	0.322974
110.00	0.000034	0.497000	0.495757	0.486586	0.457654
111.00	0.000865	0.423337	0.392250	0.193587	0.003123
112.00	0.001123	0.400891	0.361284	0.130783	0.000192
113.00	0.001082	0.404439	0.366152	0.139694	0.000312
114.00	0.001037	0.408333	0.371507	0.149915	0.000522
115.00	0.001765	0.346540	0.288368	0.038777	0.000000
116.00	0.001732	0.349278	0.291955	0.041644	0.000000
117.00	0.001823	0.341764	0.282135	0.034145	0.000000
118.00	0.002582	0.281849	0.207105	0.004911	0.000000
119.00	0.002979	0.252659	0.173078	0.001446	0.000000
120.00	0.002573	0.282523	0.207912	0.005040	0.000000
121.00	0.003375	0.225233	0.142936	0.000369	0.000000

Table 2: The stock price, the estimated values of $\widetilde{b}_{N-1,j}$ and values of probabilities for different number of simulations M. The values of the parameters of the underlying pricing process are $\mu = 0.0$, $\sigma = 0.005$.

$s_{N-1,j}$	$g(s_{N-1,j})$	$\widetilde{E}^{cont}_{N-1,j}$	$\widetilde{\sigma}^2_{N-1,j}$	$\widetilde{L}_{N-1,j}$	$\widetilde{d}^2_{N-1,j}$	$\widetilde{b}_{N-1,j}$
101.00	1.00	1.807711	8.196880	0.807711	8.196880	0.282119
102.00	2.00	2.317591	10.416028	0.158795	2.604007	0.098405
103.00	3.00	2.906708	12.789239	-0.031097	1.421027	0.026087
104.00	4.00	3.565243	15.230821	-0.108689	0.951926	0.111400
105.00	5.00	4.293338	17.660846	-0.141332	0.706434	0.168153
106.00	6.00	5.073696	19.940714	-0.154384	0.553909	0.207435
107.00	7.00	5.908342	22.061303	-0.155951	0.450231	0.232419
108.00	8.00	6.776691	24.014385	-0.152914	0.375225	0.249632
109.00	9.00	7.679068	25.689649	-0.146770	0.317156	0.260616
110.00	10.00	8.610540	27.062420	-0.138946	0.270624	0.267093
111.00	11.00	9.554690	28.417028	-0.131392	0.234851	0.271126
112.00	12.00	10.513495	29.453017	-0.123875	0.204535	0.273906
113.00	13.00	11.482011	30.408137	-0.116768	0.179930	0.275279
114.00	14.00	12.459679	31.167297	-0.110023	0.159017	0.275906
115.00	15.00	13.436818	32.012825	-0.104212	0.142279	0.276279

Table 3: Stock price, stopping profit and estimated values of expected continuation profit $E^{cont}_{N-1,j} = Tw_0(s_{N-1,j})$, variance $\widetilde{\sigma}^2_{N-1,j}$, $\widetilde{L}_{N-1,j}$, $\widetilde{d}^2_{N-1,j}$ and $\widetilde{b}_{N-1,j}$. The values of the parameters of the underlying pricing process are $\mu = -0.015$, $\sigma = 0.05$.

As we can see in Table 2 the probability is very large for every stock price when we use $M = 5 * 10^4$ simulations. The large probabilities for classification errors depends on the small values of $\widetilde{L}_{N-1,j}$. When we increase the number of simulations the probability of making classification error decreases as expected. When we use 10 million simulations at each point $s_{N-1,j}$ it is only the stock prices near the threshold value $s^*_{N-1} = 109$ that has a probability that is not negligible. The probability of making a classification error at the stock price 110 is almost 0.5 for every number of simulations used. We need more than 10^7 number of simulations to reduce the probability of classification error and in this case it is better to use variance reduction techniques instead of increasing the number of simulations.

Table 3 shows the values of $g(s_{N-1,j})$, $\widetilde{E}^{cont}_{N-1,j}$, $\widetilde{\sigma}^2_{N-1,j}$, $\widetilde{L}_{N-1,j}$, $\widetilde{d}^2_{N-1,j}$ and $\widetilde{b}_{N-1,j}$ for moment $N-1$ when the parameters of the underlying pricing process is $\mu = -0.015$ and $\sigma = 0.05$.

In Table 3 the value of the stock price for which $\widetilde{L}_{N-1,j} < 0$ for the first time is $s^*_{N-1} = 103$. We can also see that the values of $\widetilde{\sigma}^2_{N-1,j}$, $\widetilde{L}_{N-1,j}$, $\widetilde{d}^2_{N-1,j}$ and $\widetilde{b}_{N-1,j}$ are much larger than the corresponding values in Table 1. This is a result of the higher drift μ and volatility σ of the underlying pricing process, which gives a larger mean of the $Y^{(i)}_N$'s.

As before the values of $\widetilde{d}^2_{N-1,j}$ decreases when the stock price increases

		$\widetilde{p}_{N-1,j}$				
$s_{N-1,j}$	$\widetilde{b}_{N-1,j}$	$M = 10^3$	$M = 10^4$	$M = 10^5$	$M = 10^6$	$M = 10^7$
101.00	0.282119	0.000000	0.000000	0.000000	0.000000	0.000000
102.00	0.098405	0.000930	0.000000	0.000000	0.000000	0.000000
103.00	0.026087	0.204704	0.004545	0.000000	0.000000	0.000000
104.00	0.111400	0.000214	0.000000	0.000000	0.000000	0.000000
105.00	0.168153	0.000000	0.000000	0.000000	0.000000	0.000000
106.00	0.207435	0.000000	0.000000	0.000000	0.000000	0.000000
107.00	0.232419	0.000000	0.000000	0.000000	0.000000	0.000000
108.00	0.249632	0.000000	0.000000	0.000000	0.000000	0.000000
109.00	0.260616	0.000000	0.000000	0.000000	0.000000	0.000000
110.00	0.267093	0.000000	0.000000	0.000000	0.000000	0.000000

Table 4: The stock price, the estimated values of $\widetilde{b}_{N-1,j}$ and values of probabilities for different number of simulations M. The values of the parameters of the underlying pricing process are $\mu = -0.015$, $\sigma = 0.05$.

and the behavior of $\widetilde{L}_{N-1,j}$ and $\widetilde{b}_{N-1,j}$ are the same as in Table 1 and they take their smallest values at the threshold value $s^*_{N-1} = 103$.

Table 4 shows the probabilities of making classification error for four different number of simulations M. The values corresponds to the estimated values shown in Table 3.

The probability is negligible everywhere even for $M = 10^3$ except for the threshold value $s^*_{N-1} = 103$ where the probability is $p = 0.199200$ and the two stock prices closest to $s^*_{N-1} = 103$, but the probabilities for these two points are very small. This shows that it is possible to generate a good result for moment $N - 1$ even for $M = 10^3$ and we need only to use 10^4 or 10^5 number of simulations for the threshold value.

Figure 1 shows $\widetilde{d}^2_{N-1,j}$ as a function of the stock price between 100 and 1000 for the standard American call option for the same simulation as for Table 3.

We can see that $\widetilde{d}^2_{N-1,j}$ is converging to the asymptote (36) as the stock price increase in Figure 1. The convergence indicates that the algorithm behaves properly. The difference between the values of $\widetilde{d}^2_{N-1,j}$ and the asymptote is large for stock prices between 100 and 200.

Figure 2 shows the value of $\widetilde{b}_{N-1,j}$ as a function of the stock price for the same simulation as Figure 1.

The value of $\widetilde{b}_{N-1,j}$ in Figure 2 fluctuates around its asymptote (37) when the difference between the values of the stock price and the strike price is large.

We have also studied the probability of classification error for moment $n = N-5$ for standard American call option with $\mu = -0.015$ and $\sigma = 0.05$ of the underlying pricing process. As for moment $N - 1$ the probabil-

ity of classification error was negligible for every stock price except at the threshold value $s^*_{N-5} = 105$ when we used $M = 10^4$. The probability at $s^*_{N-5} = 105$ was $p = 0.003295$. When we used $M = 10^3$ the probability was nonzero for $104 \leq s_{N-5} \leq 106$ and negligible otherwise. The probabilities for stock price 104, 105 and 106 was 0.041746, 0.195126 and 0.001649, respectively. The results shows that it should be possible to generate stopping domains for all moments from $N-5$ to N using only $M = 10^3$ simulations at each stock price $s_{N-5,j}$ and use $M = 10^4$ or $M = 10^5$ for $104 \leq s_{N-5} \leq 106$ only.

The result presented in Table 4 and the results of moment N-5 indicates that it is possible to generate stopping domains for the standard payoff function with 10^3 number of simulations for all stock prices except for the stock prices at the threshold value of moment $N-5$ to $N-1$, when $\mu = -0.015$ and $\sigma = 0.05$.

In this section we have shown that the probabilities of classification errors for the standard American option depend on the values of the drift and volatility of the underlying pricing process. Large values on μ and σ give small probabilities of classification errors.

We have also shown that the probability of classification error is highest near the threshold value of the stopping domain.

It is possible to study the probability of classification error for other types of payoff functions using the same method.

5. Computational results

In this section we present some results of the investigation of the structures of stopping domains we have made. We have investigated the structure of the stopping domain for several different types of payoff functions, in particular standard linear, piecewise linear, quadratic, piecewise stepwise and logarithmic. All simulations were made using $\mu = -0.015$ and $\sigma = 0.05$ of the underlying pricing process.

Figure 3 shows the stopping domain for a standard American call option with $K = 500$.

The picture clearly shows that the stopping domain of a standard American call option consists of a number of intervals $[s^*_n, \infty]$ for $n = 1, 2, ..., 30$. For each day there exists a threshold value s^*_n such that $s^*_1 \geq s^*_2 \geq ... \geq s^*_{30} = 0$. At day one the threshold value is equal to the stock price $s^*_1 = 531$ and at the day before the expiration day the threshold value is equal to $s^*_{29} = 513$.

Figure 4, Figure 5 and Figure 6 shows the stopping domain for a option with a payoff function with two intervals with different slopes, see (2).

In Figure 4 the two slopes are $a_1 = 1.0$ and $a_2 = 2.0$, and the strike prices are $K_1 = 500$ and $K_2 = 570$. We can see a division of the stopping

domain into two domains around the second strike price K_2. The lower domain is defined by a lower and an upper threshold value. The difference between these values is less in the beginning of the period and increases close to the expiration day. The lower threshold has a similar shape as the threshold of the standard option in Figure 3, it decreases from 534 at n=1 to 515 at n=29. As for the standard option the decrease is sharper in the end of the period.

The upper domain in Figure 4 has the same structure as the stopping domain of the standard payoff function, i.e. it consists of intervals $[s_n^*, \infty]$ with threshold values $s_n^*, n = 1, 2, ..., 30$, but the change in the threshold values are smaller than for the standard option. The values change from 567 at $n = 1$ to 569 at $n = 29$. Close to the expiration day, the upper threshold of the lower domain and the threshold value of the upper domain are approaching each other.

In Figure 5 the lower stopping domain has decreased and it exist only for $n \geq 22$. Here the slope a_1 and the strike prices K_2 and K_2 are the same as for Figure 4, but the slope $a_2 = 3.0$. Also the upper domain has decreased and the threshold values are 580 in the beginning of the period and 577 in the end.

Note that the point with stock price 579 in moment $n = 23$ is classified as a continuation point and that the point with the same stock price in moment $n = 22$ is classified as a stopping point. It is probable that we have a classification error in one of these points. Either the point in moment $n = 23$ should be classified as a stopping point or the point in moment $n = 22$ should be classified as a continuation point.

We can suspect that we have classification errors at the points with stock prices around 540 in moment $n = 21$, since we have a single stopping point here.

In Figure 6 we have used a strike prices $K_1 = 500$ and $K_2 = 560$ and slope $a_1 = 1.0$ and $a_2 = 2.0$. Comparing with Figure 4 we see that the lower domain has changed. This time the domain begins at n=5 and both the upper and the lower threshold changes more than in the first case.

Note that the point with stock price 555 at moment $n = 21$ is classified as a stopping point. It is very probable that we have a classification error at this point, since the point at stock price 554 is classified as a continuation point.

As we can see the shape of the lower domain depends on the difference of K_1 and K_2 and the difference between the slopes a_1 and a_2. It is also clear that the probability of making a classification error is larger when we are close to the threshold values.

To get a sharp picture we used a larger number of simulations in the regions of the domains where the probability of classification error is high. We have not investigated the values of these probabilities but it is clear from

the pictures produced that we had classification errors when the number of simulations was 10^5, sometimes even for 10^6.

Figure 7 shows the stopping domain for a quadratic payoff function with the strike price $K = 500$.

If we compare this stopping domain with the stopping domain of the standard payoff function with the same strike price in Figure 3 we see that they have the same structure, but the threshold values are much higher for the quadratic function. We can also see that the change in the threshold values are greater for the quadratic payoff function.

In Figure 8 and Figure 9 the stopping domains of a stepwise payoff function with two and three steps, respectively, are shown.

In Figure 8, which shows the stopping domain of the stepwise payoff function with $K_1 = 500$, $K_2 = 550$, $L_1 = 10$ and $L_2 = 30$, two domains exists. For the lower domain the lower threshold value is constant and equal to the stock price 501 for $n = 1, ..., 29$. The upper threshold value is 513 in the beginning of the period and increases close to the expiration day. The threshold value of the upper domain is constant and equal to the stock price 551.

In Figure 9, which shows the stopping domain of a stepwise payoff function with $K_1 = 500$, $K_2 = 540$, $K_3 = 580$, $L_1 = 10$, $L_2 = 20$ and $L_3 = 40$, we have three well defined domains. The lower threshold value for each of the three domains are constant at the stock prices 501, 541 and 581, respectively. For the first and the second domain we have upper threshold values that increases as we get closer to the expiration day. Note that the increase is larger for the lower domain.

For the payoff function with a single step with strike price $K = 500$ the stopping domain consists of the intervals $[501, \infty]$ for $n = 1, ..., N - 1$.

Figure 10 shows the stopping domain for the logarithmic payoff function with a strike price $K = 500$.

The stopping domain has the same structure as for the standard payoff function with a decreasing threshold value close to the expiration day. The main difference is that the threshold value are larger for the standard payoff function. For day one the threshold value is $s_1^* = 507$ and for day $N-1 = 29$ $s_{29}^* = 503$.

It is possible to model a logarithmic payoff function with the piecewise linear payoff function. For example we can use the payoff function with two slopes $a_1 = 1.0$ and $a_2 = 0.25$ and strike prices $K_1 = 500$ and $K_2 = 530$. The stopping domain is shown in Figure 11.

Comparing with the logarithmic payoff function in Figure 10 the threshold values of the stopping domain are larger, e.g. $s_1^* = 519$ and $s_{29}^* = 511$, for the piecewise linear payoff function.

Comparing the different stopping domains shown above it is clear that the structure of the stopping domain depends not only on the payoff function

but also on the parameters of the function. The later was very clear in the examples with the piecewise linear payoff function with two slopes were we saw that the structure of the stopping domain was dependent on both the difference between the slopes and the difference between the strike prices.

6. Conclusions

In this paper we have studied the structure of the stopping domain for different types of payoff functions, in particular standard linear, piecewise linear, quadratic, stepwise and logarithmic, for American type call option with discrete time.

The algorithm used to study the structure is based on the idea of comparing the profit if we exercise the option at moment n and stock price $s_{n,j}$ with the expected profit if we exercise the option in the future. To estimate the expected profit we used a Monte Carlo method.

The first conclusion is that for a wide class of monotone payoff functions the optimal stopping domain have a threshold structure. This structure can be more complicated than the single threshold that exist for the standard payoff function.

Second, the structure of the optimal stopping domain depends not only on the type of payoff function studied, but also on the parameters of the payoff function. The structure of the piecewise linear payoff function with two different slopes, for example, showed a dependence on both the difference between the strike prices and the difference between the two slopes.

Third, the probabilities of making a classification error at a point are dependent on both the distance to the threshold and the distance to the expiration day. The error probabilities are large close to the threshold and far away from the expiration day.

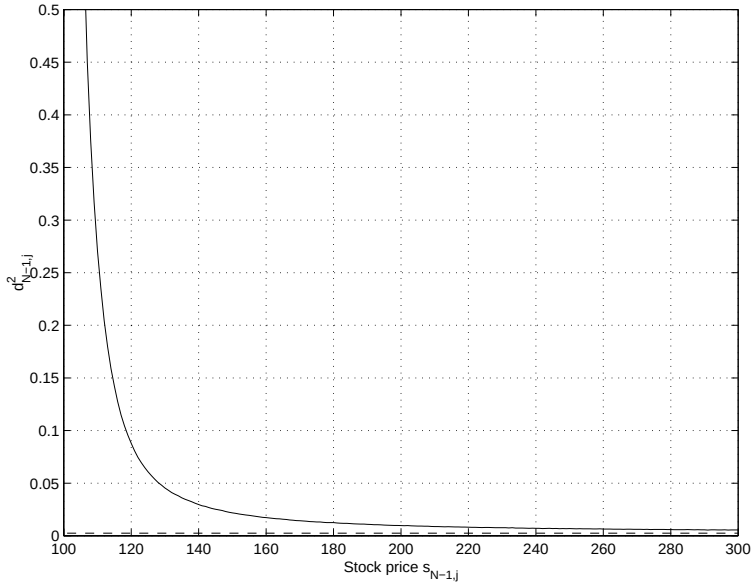


Figure 1: Values of $d^2_{N-1,j}$ as a function of the stock price $s_{N-1,j}$. The parameters of the underlying pricing process are $\mu = -0.015$ and $\sigma = 0.05$.

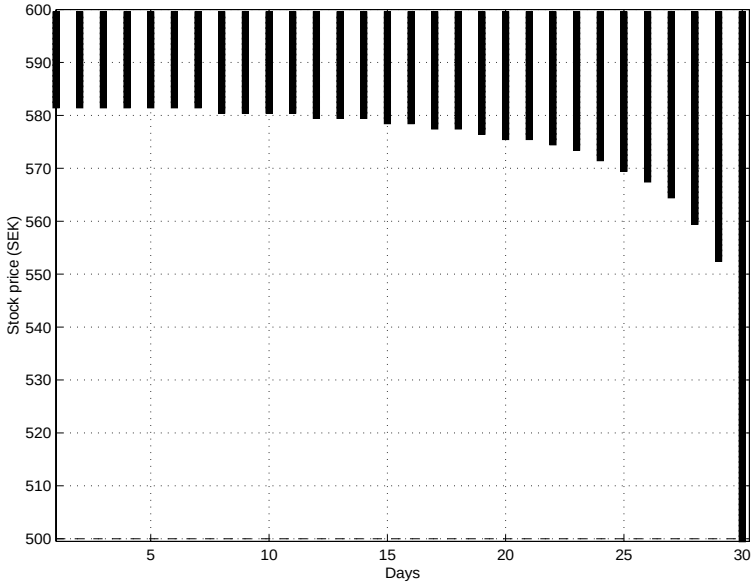


Figure 2: Values of $b_{N-1,j}$ as a function of the stock price $s_{N-1,j}$. The parameters of the underlying pricing process are $\mu = -0.015$ and $\sigma = 0.05$.

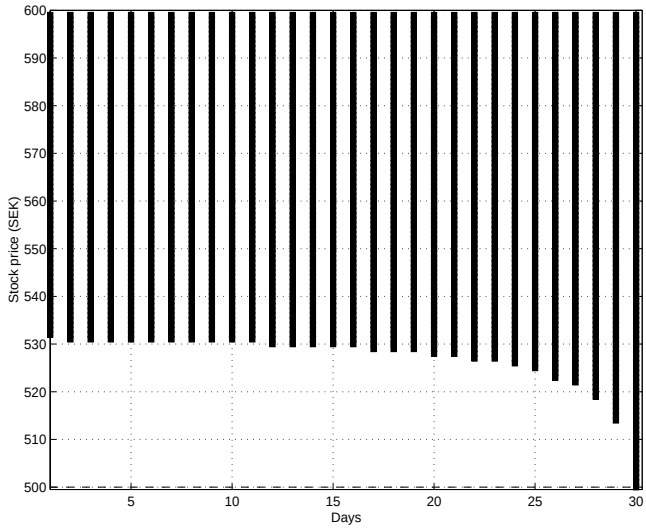


Figure 3: Stopping domain for standard American call option with strike price $K = 500$. The values of the parameters of the underlying pricing process is $\mu = -0.015$ and $\sigma = 0.05$.

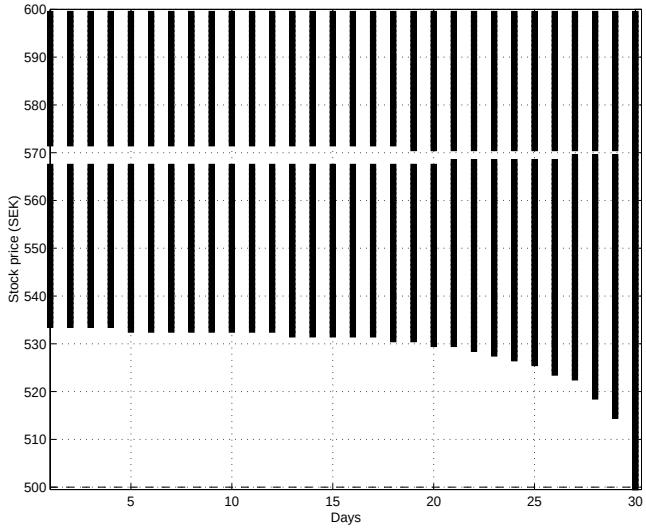


Figure 4: Stopping domain for American call option with piecewise linear payoff function with two slopes $a_1 = 1.0$ and $a_2 = 2.0$, and strike prices $K_1 = 500$ and $K_2 = 570$. Number of simulations $M = 10^6$ for $520 \leq s_{n,j} \leq 540$ and $560 \leq s_{n,j} \leq 575$, and $M = 10^5$ otherwise.The values of the parameters of the underlying pricing process is $\mu = -0.015$ and $\sigma = 0.05$.

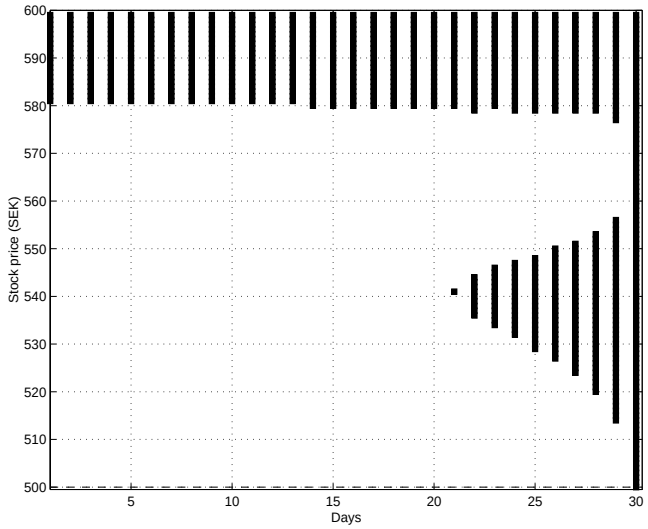


Figure 5: Stopping domain for American call option with payoff function with two slopes, $a_1 = 1.0$ and $a_2 = 3.0$. Strike price $K_1 = 500$ and $K_2 = 570$. $M = 10^6$ for $515 < S < 560$ for $n > 15$ and for $575 < S < 585$, $M = 10^5$ otherwise. The values of the parameters of the underlying pricing process is $\mu = -0.015$ and $\sigma = 0.05$.

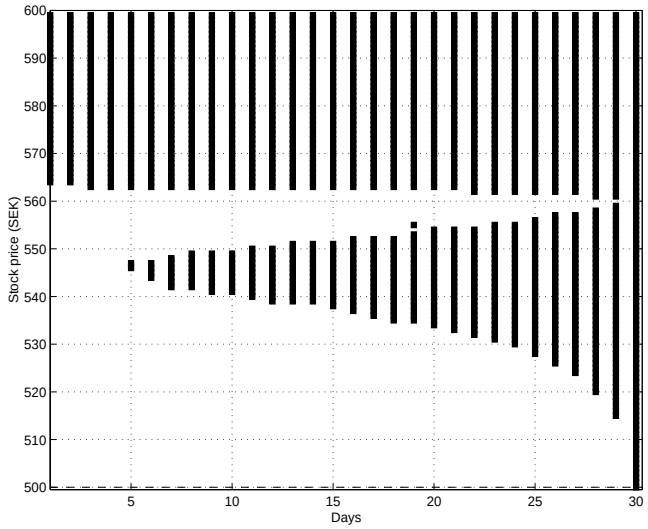


Figure 6: Stopping domain for American call option with payoff function with two slopes, $a_1 = 1.0$ and $a_2 = 2.0$. Strike price $K_1 = 500$ and $K_2 = 560$. $M = 10^6$ for $535 < S < 555$ for $n < 20$ and for $530 < S < 570$, $M = 10^5$ otherwise. The values of the parameters of the underlying pricing process is $\mu = -0.015$ and $\sigma = 0.05$.

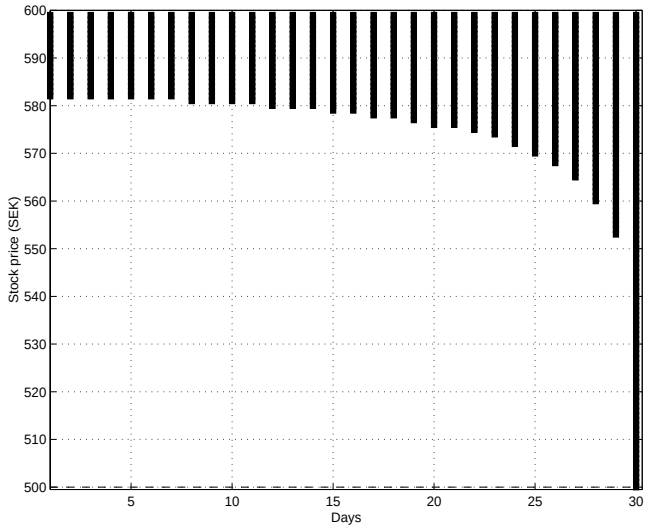


Figure 7: Stopping domain for American call option with quadratic payoff function. $M = 10^6$ for $S \geq 550$, $M = 10^5$ otherwise. The values of the parameters of the underlying pricing process is $\mu = -0.015$ and $\sigma = 0.05$.

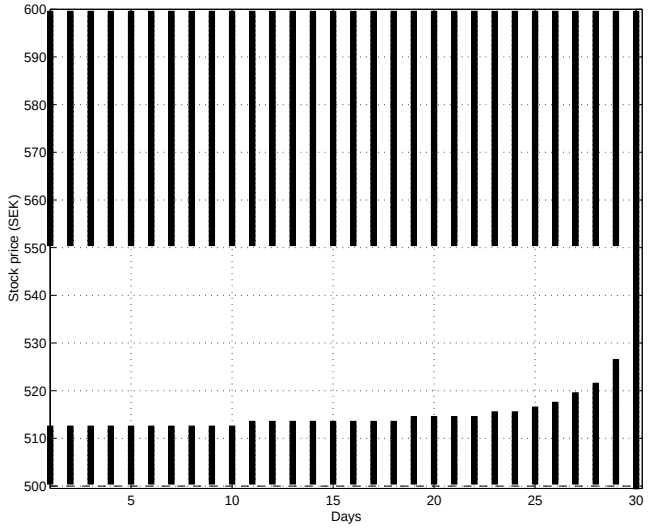


Figure 8: Stopping domain for American call option with stepwise payoff function with two steps. Strike prices $K_1 = 500$, $K_2 = 550$ and steps $L_1 = 10$, $L_2 = 30$. $M = 10^6$ for $510 < s < 530$ and $M = 10^5$ otherwise. The values of the parameters of the underlying pricing process is $\mu = -0.015$ and $\sigma = 0.05$.

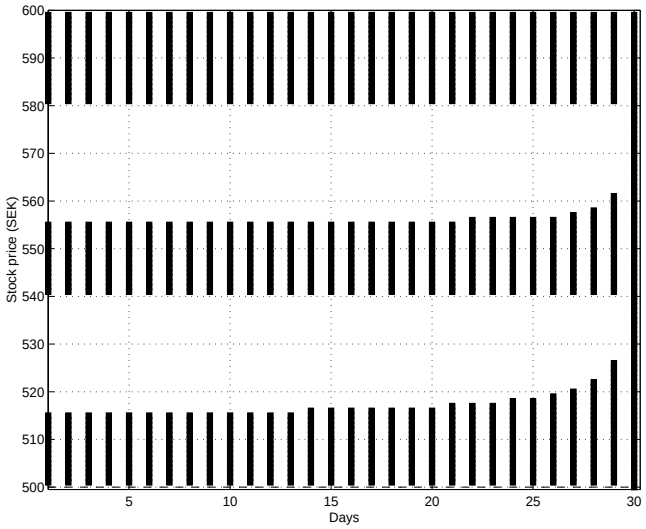


Figure 9: Stopping domain for American call option with stepwise payoff function with three steps. Strike prices $K_1 = 500$, $K_2 = 540$, $K_3 = 580$ and steps $L_1 = 10$, $L_2 = 20$, $L_3 = 40$. $M = 10^5$ for $510 < s < 525$ and $55 < s < 565$ and $M = 10^5$ otherwise. The values of the parameters of the underlying pricing process is $\mu = -0.015$ and $\sigma = 0.05$.

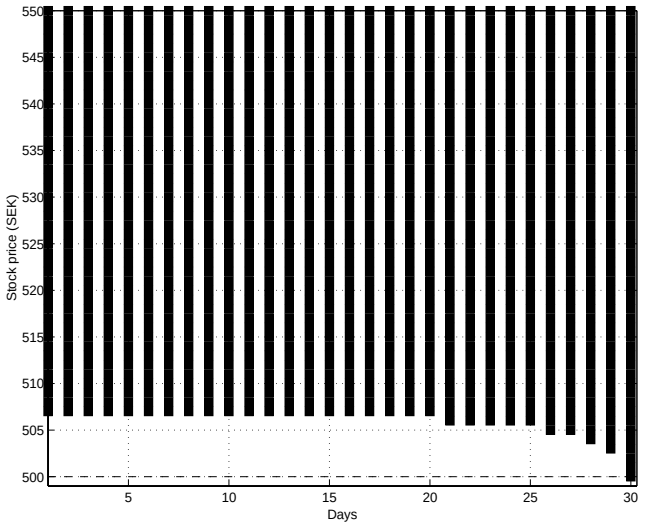


Figure 10: Stopping domain for American call option with logarithmic payoff function. Strike prices $K = 500$. $M = 10^5$. The values of the parameters of the underlying pricing process is $\mu = -0.015$ and $\sigma = 0.05$.

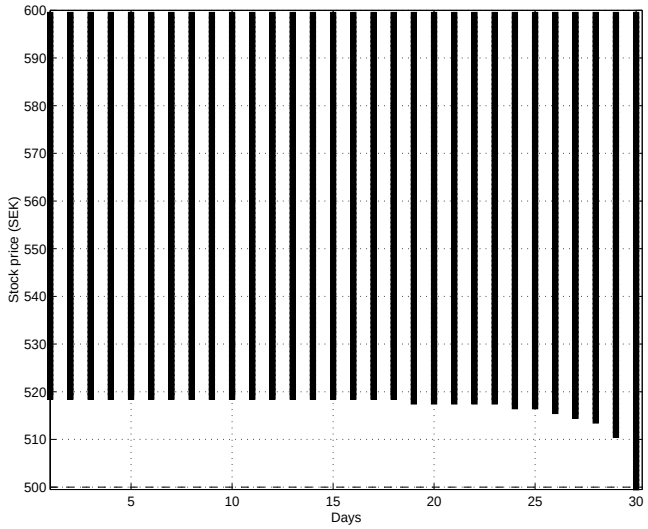


Figure 11: Stopping domain for American call option with piecewise linear payoff function with two slopes $a_1 = 1.0$ and $a_2 = 0.25$ and strike prices $K_1 = 500$ and $K_2 = 530$. $M = 10^5$. The values of the parameters of the underlying pricing process is $\mu = -0.015$ and $\sigma = 0.05$.

Bibliography

1. Boyle, P.P., Options: A Monte Carlo Approach. *Journal of Financial Economics*, **4**, (1977), 323-338.

2. Boyle, P.P., Broadie, M. and Glasserman, P., Monte Carlo Methods for Security Pricing. *Journal of Economic Dynamics and Control*, **21**, (1997), 1267-1321.

3. Broadie, M. and Detemple, Recent Advances in Numerical Methods for Pricing Derivative Securities. In *Numerical Methods In Finance*, edited by Rogers, L.C.G., and Talay, D., Cambridge University Press, (1997), p. 43-66.

4. Kuksuh, A.G. and Silvestrov, D.S., Optimal pricing of American type options with discrete time, *Research Report 2000-1*, Department of Mathematics and Physics, Mälardalen University, (2000).

5. Rogers, L.C.G. and Talay, D., *Numerical Methods In Finance*, Cambridge University Press, (1997).

6. Shiryaev, A.N., *Optimal Stopping Rules*, Springer Verlag,(1978).

7. Shiryaev, A.N., Kabanov, Yu.M., Kramkov, D.O., and Mel'nikov, A.V., Toward a theory of pricing options of European and American types. I. Discrete time. *Theory Probab. Appl.*, **39**, (1994), 14-60.

8. Westermark, P. Monte Carlo studies of optimal strategies for American Type Options. *Master's Thesis in Mathematical Statistics*, Umeå University, (1999).

Department of Mathematics and Physics, Mälardalen University, SE-721 23 Västerås, Sweden.
E-mail: henrik.jonsson@mdh.se

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.189-195

DENYS KEREKESHA

SOME GENERALIZATION OF THE RUIN PROBABILITY PROBLEM IN THE CLASSICAL RISK THEORY

The ruin probability problem in the generalization of the classical risk theory is considered.

2000 *Mathematics Subject Classifications.* 60B15.

Key words and phrases. Analytical function, risk theory, transformation of Fourier, factorization.

1. Introduction

The aim of this paper is to present some generalization of the ruin probability problem in the classical risk theory.

We shall consider

$$\frac{c(u)}{\alpha}\varphi'(u) = \varphi(u) - \int_0^u \varphi(u-t)\,dF, \quad u \geq 0. \tag{1.1}$$

Here $F(t)$ is given as distribution function of individual claim amount with density f and finite expectation m. Claims are counted according to the generalised Poisson process (with parameter a), u is the initial surplus of the insurance company,

$$\frac{c(u)}{\alpha} = \frac{\gamma_1 + \lambda_1 e^{-u}}{\gamma_2 + \lambda_2 e^{-u}} \quad (\lambda_{1,2} > 0,\ \gamma_{1,2} > 0,\ \gamma_1\lambda_2 - \gamma_2\lambda_1 \neq 0)$$

is the function of gross premium risk and $\varphi(u)$ is the probability of no ruin satisfying the following conditions:

$$0 \leq \varphi(u) \leq 1, \quad \varphi(\infty) = 1. \tag{1.2}$$

2. Reducing the integro-differential equation to the boundary value problem

For the strict exposition, we shall assume, further that the image of Fourier of the density f of the distribution F belongs to space $L_2(0,\infty)$. We shall seek the function in space satisfying the following condition

$$e^{-\nu u}\varphi(u) \in L_2(0,\infty), \quad \left(e^{-\nu u}\varphi(u)\right)' \in L_2(0,\infty), \quad \nu > 0. \tag{2.1}$$

Let us introduce the following notation

$$\varphi_+(u) = \begin{cases} \varphi(u), & u > 0 \\ 0, & u < 0; \end{cases} \qquad f_+(u) = \begin{cases} f(u), & u > 0 \\ 0, & u < 0. \end{cases} \tag{2.2}$$

The following definitions and theorems will be also necessary for us.

Definition 1. By $\{\{-m,\nu,\infty\}\}(m \geq 0)$, we denote the space of functions $\Phi(z)$ analytical in the half plane $Imz \geq \nu$ for which there exists a constant C such that for all $y \geq \nu$.

$$\int_{-\infty}^{\infty} |\Phi(x+iy)(x+iy)^m|^2 dx \leq C. \tag{2.3}$$

Definition 2. By $\{-m,\nu,\infty\}$, $(m \geq 0)$, we denote the space of original functions image which belongs to $\{\{-m,\nu,\infty\}\}$.

Theorem 1. *The function $\varphi(u) \in \{-m,\omega,\infty\}(m \geq 0)$, if and only if when $\varphi(u)$ satisfies the following conditions*

$$\frac{d}{du^k}\left(e^{-\nu u}\varphi(u)\right) \in L_2(0,\infty), \quad k = \overline{0,m}. \tag{2.4}$$

Theorem 2. *In order that a function $\varphi(u) \in \{-m,\nu,\infty\}$ it is necessary and sufficient that the Fourier transform $\Phi^+(z) \in \{\{-m,\nu,\infty\}\}$*

Validity of the theorems 1 and 2 follows from the appropriate theorems explained in [2]

Let us begin to solve equation (1.1). First, let us multiply both parts of equality (1.1) on $\gamma_2 + \lambda_2 e^{-u}$ and after that apply the Fourier transform to it. Due to the theorem 2 we shall receive the boundary value problem of the theory of analytical functions

$$\Phi^+(z+i) + \frac{\gamma_1}{\lambda_1}\left[1 + K^+(z)\right]\Phi^+(z) = G_0^+(z), \quad Imz \geq \nu. \tag{2.5}$$

Here $\Phi^+(z) = (V\varphi_+)(z)$, $\hat{f}^+(z) = (Vf_+)(z)$ (V is a Fourier operator),

$$K^+(z) = \frac{1 - \lambda_2\lambda_1^{-1} + \gamma_2\gamma_1^{-1} - \left(\gamma_2\gamma_1^{-1}\hat{f}^+(z) + \lambda_2\lambda_1^{-1}\hat{f}^+(z+i)\right)\sqrt{2\pi}}{iz - 1 + \lambda_2\lambda_1^{-1} - \lambda_2\lambda_1^{-1}\hat{f}^+(z+i)\sqrt{2\pi}}, \quad (2.6)$$

$$G_0^+(z) = \frac{\gamma_1 + \gamma_2}{\sqrt{2\pi}} \frac{\varphi_+(0)}{-iz\lambda_1 + \lambda_1 - \lambda_2 + \lambda_2\hat{f}^+(z+i)}. \quad (2.7)$$

3. Solution of the boundary value problem (2.5) by the method of factorization

At first, we shall choose the number ν so large, that the function $1 + K^+(z)$ has no zeroes and poles in the half-plane $Imz \geq \nu$. Next, we apply the method of factorization [1], which allow us to represent function $1 + K^+(z)$ in the form

$$1 + K^+(z) = \frac{X^+(z+i)}{X^+(z)}, \quad (3.1)$$

where the function $X^+(z)$ should be the bounded analytical function and have no zeroes there.

For the factorization (3.1) we shall define the logarithm

$$\ln\left(1 + K^+(z)\right) = \Omega^+(z). \quad (3.2)$$

Also we shall require, that the function

$$\Omega^+(z) \in \{\{-1, \mu, \infty\}\}. \quad (3.3)$$

Let's take the logarithm of the equality (3.1). As a result we receive the problem about " saltus" in the form

$$\Omega^+(x + i\nu) = A^+(x + i\nu + i) - A^+(x + i\nu), \quad x \in R, \quad (3.4)$$

where

$$A^+(z) = lnX^+(z). \quad (3.5)$$

The solution is the following:

$$A^+(x + i\nu) = -\frac{\Omega^+(x + i\nu)}{2} + \frac{i}{2}\int_{-\infty}^{\infty} \frac{\Omega^+(t + i\nu)\,dt}{th\pi(x - t)}. \quad (3.6)$$

Actually, applying inverse transformation of Fourier to (3.4), we shall receive the function $a_+(u) = -\frac{\omega_+(u)}{1 - e^{-u}}$, with the following properties

$$a_+(u) \equiv 0 \ \ and \ \ u < 0, \quad e^{-\nu u}a_+(u) \in L_2(0, \infty),$$

$$A^+(x) \in \{\{0,\mu,\infty\}\}\left((Va_+)(x) = A^+(x)\right).$$

The above mentioned properties follow from the theorem 2.

Applying transformation of Fourier to $a_+(u)e^{-\nu u}$, we shall receive the formula (3.6). In doing so, we have used the properties of the function $a_+(u)$ and formula [1] $\left(V\left(\frac{1}{1-e^{-u}}\right)\right)(x) = i\sqrt{\frac{\pi}{2}}cth\pi x + \sqrt{\frac{\pi}{2}}\delta(x)$ ($\delta(x)$ is the Dirac delta - function).

Note that the condition (3.3) guarantees the boundedness of the function $A^+(z)$ at $Im\ z \geq \nu$ and that in its turn also guarantees the boundedness of the function $X^+(z)$ in the same half-plane. The proof of the similar statement is present in [1]. We can easily show that if $K^+(x)$ and $(xK^+(x))$ belong to the space $L_2(-\infty,\infty)$, then condition is satisfied(3.3). In accordance with the formula (2.6) it can be possible only on condition that

$$\lambda_2\lambda_1^{-1} - \gamma_2\gamma_1^{-1} = 1. \tag{3.7}$$

Let us assume condition (3.7) in this point hold true. Then the formulae (3.5) - (3.6) give the required factorization

$$1 + K^+(z) = \frac{exp\left(A^+(z)\right)}{exp\left(A^+(z+i)\right)}, \quad Im\ z \geq \mu. \tag{3.8}$$

For convenience we introduce

$$exp\left(A^+(z)\right) = 1 + R^+(z), \quad exp\left(-A^+(z)\right) = 1 + S^+(z), \tag{3.9}$$

where

$$R^+(z) \in \{\{0,\mu,\infty\}\}, \quad S^+(z) \in \{\{0,\mu,\infty\}\}.$$

Having introduced a new unknown function

$$M^+(z) = \left[1 + S^+(z)\right]\Phi^+(z) \in \{\{0,\mu,\infty\}\}$$

$$\left(\Phi^+(z) = \left[1 + R^+(z)\right]M^+(z)\right) \tag{3.10}$$

we present the problem (2.5) in the form

$$M^+(z+i) + \frac{\gamma_1}{\lambda_1}M^+(z) = \left[1 + S^+(z)\right]G_0^+(z). \tag{3.11}$$

Applying inverse transformation of Fourier to both parts of equality (3.11) we shall receive

$$m_+(u) = g_{0+}(u)\frac{\lambda_1}{\gamma_1 + \lambda_1 e^{-u}} +$$

$$\frac{\lambda_1}{\gamma_1 + \lambda_1 e^{-u}}\frac{1}{\sqrt{2\pi}}\int_0^u s_+(u-t)\,e^{-(u-t)}g_{0+}(t)\,dt, \ u > 0. \tag{3.12}$$

Here $m_+(u) = (V^{-1})(M^+)(u)$, $s_+(u) = (V^{-1})(S^+)(u)$, $g_+(u) = (V^{-1})\left(G_0^+\right)(u)$ (V^{-1} is the inverse operator of Fourier).

At last, from (3.9) -(3.11) we shall receive the solution of equation (1.1)

$$\varphi_+(u) = \frac{g_{0+}(u)}{1+\frac{\gamma_1}{\lambda_1}e^{-u}} + \frac{1}{\sqrt{2\pi}}\int_0^u \Gamma(u,t)\, g_{0+}(t)\, dt, \quad u>0, \tag{3.14}$$

where

$$\Gamma(u,t) = \frac{\lambda_1 s_+(u-t)\, e^{-(u-t)}}{\gamma_1+\lambda_1 e^{-u}} + \frac{\lambda_1 r_+(u-t)}{\gamma_1+\lambda_1 e^{-u}} + \frac{1}{\sqrt{2\pi}}\int_t^u \frac{\lambda_1 s_+(u-t)\, e^{-(u-t)} r_+(u-\tau)}{\gamma_1+\lambda_1 e^{-\tau}} d\tau,$$

$$r_+(u) = \left(V^{-1}\right)\left(R^+\right)(u).$$

Thus we have established

Theorem 3. *Let the parameters $\lambda_{1,2}$ and $\gamma_{1,2}$ of the function $c(u)$ given in the equation (1.1) to satisfy the condition (3.7) and the image of the density of the distribution F to belong to space $L_2(0,\infty)$. Then in the space $\{-1,\nu,\infty\}$ there exists a solution of the equation (1.1) determined by the formula (3.13).*

4. Normalization of the solution. Uniqueness of the solution of equation (1.1), which satisfies to the condition (1.2)

As the equation (1.1) is homogeneous, then the unknown function is defined precisely up to the constant factor. For its fixation the probability-theoretic sense of function is used. In other words the unknown function should satisfy the conditions (1.2). In particular, for realization of the second condition (1.2) it is necessary that there existed a finite, different from zero on infinity limit of the function $\varphi_+(u)$

The immediate transition to the limit on infinity in equality (3.13) presents some difficulties. Therefore we apply another procedure to the determine the required limit.

For this purpose we shall receive the relation from equality (3.9)

$$\varphi_+(u) = m_+(u) + \frac{1}{\sqrt{2\pi}}\int_0^u r_+(u-t)m_+(t)dt. \tag{4.1}$$

Next, letting u tend to infinity in equalities (3.12), (4.1). We finally obtain the following algebraic system

$$m_+(\infty) = \frac{\lambda_1(\gamma_1+\gamma_2)}{\gamma_1} g_{0+}(\infty)\,\varphi_+(0) + \frac{\lambda_1}{\gamma_1} S^+(i)\, g_{0+}(\infty),$$

$$\varphi_+(\infty) = m_+(\infty)\left(1 + R^+(0)\right) = \frac{m_+(\infty)}{1 + S^+(0)}. \tag{4.2}$$

The system is obtained with the help the Tauberian theorem [2] and inverse transformation of Fourier function

$$G^+(z) = \frac{1}{\sqrt{2\pi}} \frac{1}{-iz\lambda_1 + \lambda_1 - \lambda_2 + \lambda_2\sqrt{2\pi}\hat{f}^+(z+i)}. \tag{4.3}$$

That is the functions determined by the equality,

$$g_+(u) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty+\nu i}^{\infty+\nu i} G^+(t+\nu i)\, e^{-(t+\nu i)u} dt, \tag{4.4}$$

where by $g_+(\infty)$ its limit on infinity is denoted.

Let us investigate the system (4.2). First of all note that if the limit of function $g_+(u)$ at $u \to \infty$ is accordingly equal to zero, infinity or does not exist, then the solution $\varphi_+(u)$ is deprived of probability-theoretic sense. It is necessary to assume, that a limit $g_+(\infty)$ different from zero exists by it is possible only on condition that the limit $g_+(\infty)$ exists and, that the function $G^+(z)$ has a pole in a point $z = 0$. It immediately brings us to the condition

$$\sqrt{2\pi}\hat{f}^+(i) = 1 - \frac{\lambda_1}{\lambda_2}. \tag{4.5}$$

Let us consider , that the limit $g_+(\infty)$ exists and that the condition (4.5) holds true. It is supposed also, that there exists a finite value $S^+(0)$. But if the function $G^+(z)$ has a pole in a point z=0, then by virtue of the structure of function $\Omega^+(z)$ a finite limit $S^+(0)$would possible not exist . It is obvious, that for its existence it is necessary, that numerator of fraction (2.6) in the point $z = 0$ converted to zero. Taking into account the condition (3.7) our last conditions will look so:

$$\hat{f}^+(i) = \frac{\gamma_2\lambda_1}{\lambda_2\gamma_1}\hat{f}^+(0). \tag{4.6}$$

Besides let us assume, that the function $\hat{f}^+(x)$ is such that there exists

$$\underset{z\to\infty}{Lim}\, K^+(z) = K^+(0) \tag{4.7}$$

and

$$1 + K^+(0) \neq 0. \tag{4.8}$$

If conditions (4.7)-(4.8)are fulfilled then condition (4.6) will be also sufficient for the existence of finite value $S^+(0)$.

Let us assume conditions (4.6) - (4.8) in this point hold true. Then taking into account the second condition (1.2) of a system (4.2) we shall receive a relation

$$\frac{\lambda_1(\gamma_1+\gamma_2)}{\gamma_1}\varphi_+(0) = \frac{1 + S^+(0) - g_+(\infty)\lambda_1\gamma_1^{-1}S^+(i)}{g_+(\infty)}. \tag{4.9}$$

At the made above suppositions we shall find out, that

$$S^+(0) = exp\left(-A^+(0)\right) - 1, \quad S^+(i) = exp\left(-A^+(i)\right) - 1, \tag{4.10}$$

where the function $A^+(z)$ is defined by the formula (3.6).

The following theorems are also true

Theorem 4. *Let the conditions of the theorem 4 be satisfied,there exists* $g_+(\infty)$*, and besides the density function of probability* $f(u)$ *and that the conditions (4.5) - (4.8) also satisfy . Then the equation (1.1) has an unique, solution which satisfies the second condition (1.2). This solution is constructed sequentially by the formulae (2.6) - (2.7), (3.2), (3.2),(3.8) - (3.9).*

Theorem 5. *Let conditions of the theorem 5 satisfy and also*

$$\frac{1 + S^+(0) - g_+(\infty)\lambda_1\gamma_1^{-1}S^+(i)}{g_+(\infty)} > 0.$$

Then the problem (1.1) - (1.2) has a unique solution. This solution determines function $\varphi(u)$ *(the probability for the company not being ruin), which is constructed sequentially by the formulae (2.6)-(2.7),(3.2),(3.7)-(3.9),(4.9)-(4.10).*

The proof is based on validity of the theorem 5 and equality (4.9)

Remark. In case the function with $c(u)$ is equal to constant a problem (1.1) - (1.2) represents a classical problem of the theory of risk [3] - [4].

Bibliography

1. Gahov F.D., Chersky Y.J. *The equation of the convolution of the type.* Moscow, (1978).

2. Titchmarsh E. *Introduction to the Theory of Fourier Integrals.* Moscow, Leningrad, (1948).

3. Leonenko M.M.,Mishura Y.S.,Parchomenko V.M.,Yadrenko M.J. *Probability-theoretic and Statistical methods in Econometrics and Financial mathematics.* Kiev Informtechnica (1995).

4. Feller B. *An Introduction to Probability Theory and its Applications .* Moscow, (1984).

Department of Probability Theory and Mathematical Statistics, Kyiv National Taras Shevchenko University, Kyiv, Ukraine
E-mail: ker@users.imem.odessa.ua

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.196-202

MACIEJ KLIMEK, GUSTAF STRANDELL AND JOHAN TYSK

TESTING WEAK STATIONARITY OF STOCK RETURNS

Geometric Brownian motion provides the most commonly used mathematical model of the behavior of stock prices. Accordingly, in the discrete approximation, the de-meaned logarithm of stock returns should represent a white noise. We examine six stocks using a test designed by Y. Okabe and Y. Nakano and based on experimental mathematics, to investigate to what extend this theoretical assumption holds in practice. In particular, we test the more general assumption of weak stationarity of the returns, where a certain dependence between returns is allowed.

2000 *Mathematics Subject Classifications.* 91B28.

Key words and phrases. Stock returns; geometric Brownian motion; weak stationarity; white noise.

1. Introduction

The claim that stock price processes can be modelled in continuous time by geometric Brownian motion has dominated modern finance for nearly half a century. Its justification is usually derived from to the Efficient Market Hypothesis. The usefulness of the geometric Brownian motion model is beyond a doubt, but as every mathematical model of reality it has limitations. In fact, numerous well documented features in the behavior of stock prices do not seem to conform to the Brownian motion approach prompting search for alternative explanations. A broad discussion of this issue can be found e.g. in Mandelbrot (1997) and (Lo and McKinlay,1999). In this situation there is a clear need for testing empirical data to ascertain boundaries of applicability of the conventional model. This note is intended as a step in this direction.

A discrete representation of a stock price process, which approximates a geometric Brownian motion can be described by the formula

$$S_{j+1} = S_j \exp(\nu \Delta t + \sigma \epsilon_j \sqrt{\Delta t}), \qquad j = 0, 1, \ldots,$$

where S_j denotes the stock price at time t_j, $\Delta t = t_{j+1} - t_j$ is the basic time interval, ν and σ are positive constants, and the ϵ_j's are independent

normally distributed random variables of mean zero and standard deviation 1. Consequently, the process

$$\mathcal{Z}(j) = \log \frac{S_{j+1}}{S_j} - \nu\Delta t, \qquad j = 0, 1, \ldots,$$

should be a Gaussian white noise. The assumption of independence of the ϵ_j's can be relaxed by requiring only weak stationarity of the sequence and thus allowing observations which are k time units apart to influence each other in a manner independent of the time when the observations are made. The purpose of this note is to check to what extent the de-meaned logarithmic return sequences represent white noise or at least are weakly stationary in the case of six major Swedish companies. In particular we will show that $\mathcal{Z}_j$ may be weakly-stationary without being a white noise or even not weakly-stationary. Our principal tool is a test devised by Yasunori Okabe and Yuji Nakano which we describe in the next section.

2. The Okabe-Nakano test of stationarity

In this section we summarize a special case of the so called **Test(S)** of weak-stationarity introduced by Y. Okabe and Y. Nakano (see (Okabe and Nakano, 1991); see also (Okabe and Inoue, 1994), Nakano (1995) and (Ohama and Yanagawa, 1997)). In its more general form the test applies also to multichannel signals. The scientific methodology of the Okabe-Nakano test is characteristic for experimental mathematics understood as a combination of a mathematical argument and computer simulation. Consequently it differs from the methodology of standard hypotheses testing used in statistics.

Recall that a finite time stochastic process (X_j) is said to be *weakly stationary* if all the random variables X_j have the same mean and if the covariance of any two of these random variables, say X_i and X_j, depends only on $j - i$. Let N be a positive integer. Suppose that the finite duration signal $\mathcal{Z} : \{0, 1, \ldots, N\} \longrightarrow \mathbf{R}$ is our experimental data. The objective is to test whether or not it represents a weakly stationary process. We calculate the sample mean $\mu^{\mathcal{Z}}$ and the sample autocovariance $R^{\mathcal{Z}}$ according to the formulas

$$\mu^{\mathcal{Z}} = \frac{1}{N+1} \sum_{m=0}^{N} \mathcal{Z}(m)$$

and

$$R^{\mathcal{Z}}(n) = \frac{1}{N+1} \sum_{m=0}^{N-n} \left(\mathcal{Z}(n-m) - \mu^{\mathcal{Z}}\right) \left(\mathcal{Z}(m) - \mu^{\mathcal{Z}}\right),$$

for $n = 0, \ldots, N$. Then we normalize the observations $\mathcal{Z}(n)$ by defining

$$\mathcal{X}(n) = \frac{\mathcal{Z}(n) - \mu^{\mathcal{Z}}}{\sqrt{R^{\mathcal{Z}}(0)}}, \qquad n = 0, \ldots, N.$$

The quantities $\mu^{\mathcal{X}}$ and $R^{\mathcal{X}}$ are calculated similarly to $\mu^{\mathcal{Z}}$ and $R^{\mathcal{Z}}$. Clearly $\mu^{\mathcal{X}} = 0$. For the test to work, we have to assume that the Toeplitz matrices

$$S_n^{\mathcal{X}} = \left[R^{\mathcal{X}}\left(|i-j|\right)\right]_{0\leq i,j\leq n}, \qquad n = 1, \ldots, N$$

are non-singular.

Define $M = \lfloor 3\sqrt{N+1} \rfloor - 1$. It can be demonstrated that only the the first $M+1$ coefficients $R^{\mathcal{X}}$ can be used reliably. Using the standard Levinson-Durbin algorithm (see e.g. Hayes (1996)), we calculate the numbers $\gamma(n,k), V(m)$ for $1 \leq n \leq M, 0 \leq k < n, 0 \leq m \leq M$ so that the following matrix equation is solved:

$$\left[\begin{array}{ccccccc} 1 & \gamma(n,n-1) & \gamma(n,n-2) & \ldots & \gamma(n,0) \end{array}\right] S_n^{\mathcal{X}} = \left[\begin{array}{ccccc} V(n) & 0 & 0 & \ldots & 0 \end{array}\right],$$

where $V(0) = R^{\mathcal{X}}(0)$.

To gain better access to the internal structure of $\mathcal{X}$ we look at shifted segments $\mathcal{X}_i : \{0, \ldots, M\} \longrightarrow \mathbf{R}$ of $\mathcal{X}$, where $\mathcal{X}_i(n) = \mathcal{X}(i+n)$ for $i \in \{0, \ldots, N-M\}$ and $n \in \{0, \ldots, M\}$. For these values of i we also define $\xi_i : \{0, \ldots, M\} \longrightarrow \mathbf{R}$ by the formulas:

$$\xi_i(0) = \mathcal{X}_i(0) V(0)^{-1/2}$$

and

$$\xi_i(n) = \left(\mathcal{X}_i(n) + \sum_{k=0}^{n-1} \gamma(n,k)\mathcal{X}_i(k)\right) V(n)^{-1/2}, \qquad n = 1, \ldots, M.$$

It can be shown (see Okabe (1988); see also e.g. (Okabe and Nakano, 1991) or (Okabe and Inoue, 1994)) that *$\mathcal{X}_i$ is a realization of a weakly stationary process (with the covariance function $R^{\mathcal{X}}$) if and only if ξ_i is a realization of a standard white noise.*

To check the latter, one uses three criteria, which will be denoted by **(M)**, **(V)** and **(O)**. Suppose that $Y \in \{\xi_0, \ldots, \xi_{N-M}\}$. Condition **(M)** requires that

$$\sqrt{M+1}|\mu^Y| < 1.96,$$

where

$$\mu^Y = \frac{1}{M+1} \sum_{k=0}^{M} Y(k).$$

Condition **(V)** requires that

$$|(v^Y - 1)\tilde{}| < 2.2414,$$

where

$$(v^Y - 1)\tilde{} = \sum_{k=0}^{M} (Y(k)^2 - 1) \left(\sum_{k=0}^{M} (Y(k)^2 - 1)^2\right)^{-1/2}.$$

Condition **(O)** has the form of the inequality

$$(M+1)|R^Y(n,m)|\left(\sqrt{L^{(1)}_{n,m}}+\sqrt{L^{(2)}_{n,m}}\right)^{-1}<1.96,$$

where

$$R^Y(n,m)=\frac{1}{M+1}\sum_{k=m}^{M-n}Y(k)Y(n+k),$$

where $1\le n\le L$, $0\le m\le L-n$, $L=\lfloor 2\sqrt{M+1}\rfloor-1$, and the numbers $L^{(l)}_{n,m}$ are defined as follows. For $l\in\{1,2\}$, $n\in\{1,\dots,L\}$ and $m\in\{0,\dots,L-n\}$ we put

$$L^{(l)}_{n,m}=n\left(q-(-1)^{s+l}\left\lceil\frac{s}{2}\right\rceil\right)-\frac{1-(-1)^{s+l}}{2}m-(-1)^l\max\left(0,(-1)^l(n-r)\right),$$

where

$$q=\left\lfloor\frac{M+1}{2n}\right\rfloor,\ r=M+1-2nq\quad\text{and}\quad s=\left\lfloor\frac{m}{n}\right\rfloor.$$

Note that $M+1-n-m=L^{(1)}_{n,m}+L^{(2)}_{n,m}$.

It is observed in (Okabe and Nakano, 1991), that if Y is obtained by sampling white noise consisting of independent random variables, then in view of the Central Limit Theorem and the Law of Large Numbers Y should satisfy Conditions **(M)**, **(V)** and **(O)** with probabilities 0.95, 0.975 and 0.90, respectively. Okabe and Nakano have used this fact in conjunction with extensive computer simulation to justify the following test of weak stationarity:

Test(S) (Okabe and Nakano, 1991): *If ξ_i satisfies Criterion* **(M)** *for over* 80% *of $i\in\{0,1,\dots,N-M\}$, Criterion* **(V)** *for over* 70% *of i and Criterion* **(O)** *for over* 80% *of i, then $\mathcal{Z}$ is a realization of a weakly stationary process.*

If $\mathcal{Z}$ passes **Test(S)**, we might ask if $\mathcal{Z}-\mu^Z$ represents a white noise. To this end we check if the data sequences $\mathcal{X}_i$ satisfy Condition **(O)**. *If $\mathcal{X}_i$ fulfills Condition* **(O)** *for over* 80% *of i, we assert that $\mathcal{Z}-\mu^{\mathcal{Z}}$ is a realization of a white noise.* In what follows we will call this statement *the white noise test*.

3. Testing stock returns

We apply the above tests as follows. Let m be either 100 or 150. Let

$$S:\{-(m-1),\dots,1000\}\longrightarrow\mathbf{R}$$

be the sequence of closing stock prices for a company over the period of $1000 + m$ trading days ending either at June 30, 1995 or at June 30, 1999. We define

$$\mathcal{Z}(n) = \log \frac{S(n)}{S(n-1)}$$

for $n = -(m-2), \ldots, 1000$ and

$$\mathcal{Z}_i(n) = \mathcal{Z}(i - m + n)$$

for $i = 1, .., 1000$ and $n = 1, \ldots, m$. Then we calculate the percentage of days j in the period $\{1, \ldots, 1000\}$ such that the data sequence $\mathcal{Z}_j$ passes **Test(S)**. If $\mathcal{Z}_j$ passes **Test(S)**, we also check if the de-meaned process $\mathcal{Z}_j$ is a representation of white noise. The results of our tests are presented in Tables 1-4. Horizontal graph bars correspond to the relevant period of 1000 trading days. The white/gray/black areas represent the days such that the prior m day period, respectively, passes the white noise test, passes the weak stationarity test and fails the weak stationarity test.

TABLE 1

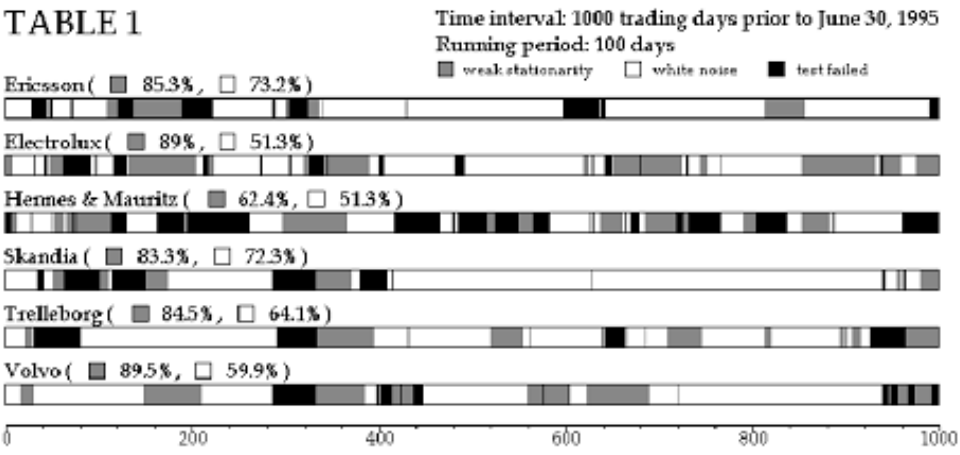


TABLE 2

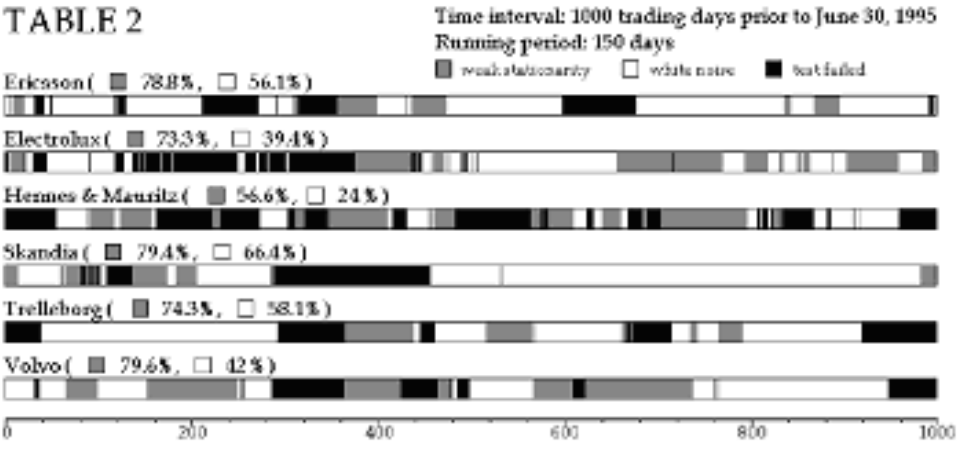


TABLE 3

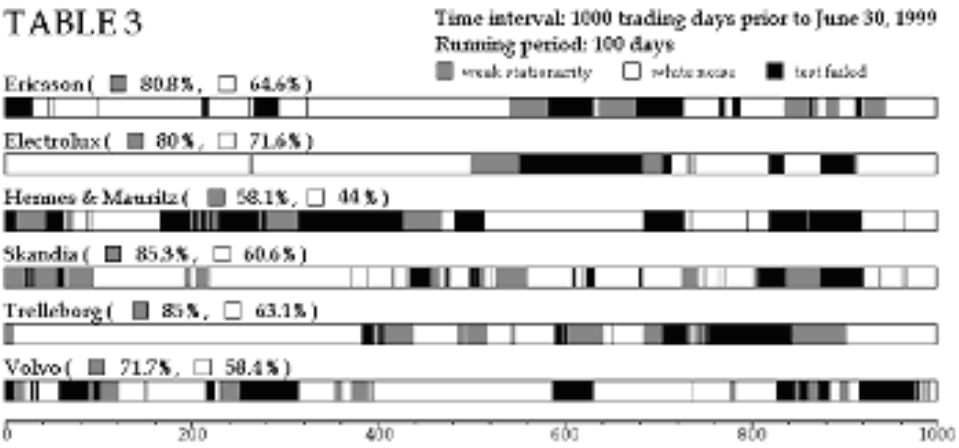
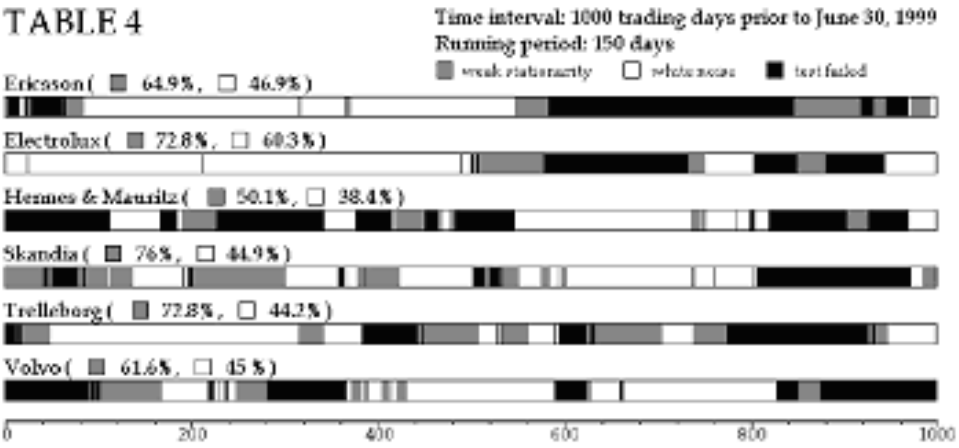


TABLE 4



4. Conclusion

The results of the tests show that the degree to which logarithmic returns of different stocks conform to the standard model varies considerably. For some stocks, the percentage of running 100-day periods (resp. 150-day periods) that pass the white noise test is as low 35.2% (resp. 24%). The highest score was 73.2% (resp. 66.4%). However a less demanding requirement of weak stationarity is satisfied more often. The lowest pass rate is 58.1% (resp. 50.1%) and the highest 89% (resp. 79.6%). Nevertheless, even in this case the rate of failure is rather high and can be linked to the fact that the volatility of the actual market data does not remain constant during the relevant period of 100 (resp. 150) days. This may be seen as further evidence in favor of models allowing stochastic volatility.

5. Acknowledgment

The authors wish to thank Öhman Fondkommission in Stockholm for supplying the market data and Professor Y. Okabe of the University of Tokyo for helpful comments.

Bibliography

1. Hayes, M.H. (1996) Statistical Digital Signal Processing and Modeling. John Wiley & Sons, Inc., New York.
2. Lo, A.W. and McKinlay, A.C. (1999) A Non-Random Walk Down Wall Street. Princeton University Press, Princeton, NJ.
3. Mandelbrot, B. (1997) Fractals and Scaling in Finance: Discontinuity, Concentration, Risk. Springer-Verlag, New York.
4. Nakano, Y. (1995) On causal analysis of economic time series. Hokkaido Mathematical Journal, 24, 1–35.
5. Ohama, G. and Yanagawa, T. (1997) Testing stationarity using residual. Bulletin of Informatics and Cybernetics, 29(1), 15–39.
6. Okabe, Y. (1988) On a stochastic difference equation for the multi-dimensional weakly stationary process with discrete time. In: Kashiwara, M., Kawai, T. (Eds.), Prospect of Algebraic Analysis, Vol. 2. Academic Press, Boston, MA, pp. 601–645.
7. Okabe, Y. and Inoue, A. (1994) The theory of KM_2O-Langevin equations and applications to data analysis (II): Causal analysis (1). Nagoya Mathematical Journal 134, 1–28.
8. Okabe, Y. and Nakano, Y. (1991) The theory of KM_2O-Langevin equations and its applications to data analysis (I): Stationary analysis. Hokkaido Mathematical Journal 20, 45–90.

Department of Mathematics, Uppsala University, P. O. Box 480, 75106 Uppsala, Sweden.
E-mail: maciej.klimek@math.uu.se

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.203-214

ALEXANDER G. KUKUSH AND YURY V. CHERNIKOV

GOODNESS-OF-FIT TESTS IN NEVZROV'S MODEL

Goodness-of-fit test is studied for Nevzorov's model.

2000*Mathematics Subject Classifications.* 62P05, 90C40, 60J25.

Key words and phrases. Goodness-of-fit test, Nevzorov's record model.

1. Introduction

We want to present an approach to the investigation of natural catastrophe claims in the presence of a trend. The so-called *Nevzorov's model* is studied. We assume that the yearly catastrophe claims are realizations of an independent sequence $\{X_i, i \geq 1\}$ of random variables with support $R^+ := [0, \infty)$ and continuous cumulative d.f. $\{F_i, i \geq 1\}$, s.t. $F_i = F^{\gamma_i}$, with $\gamma_i = \gamma^{i-1}, \gamma \geq 1$. Here F is a fixed cumulative d.f. with $F(0) = 0$. In Kukush (1999) some results concerning asymptotic properties of the MLE in the Nevzorov's record model are obtained, such as consistency, asymptotic normality and efficiency. In that paper the semi-parametric and the three-parameter model are studied. The author interprets the observed claims as a stochastically increasing sequence of Fréchet distributed random variables. This idea was first proposed in Pfeifer (1997) with some simulation study.

Here we continue that investigations. The goodness-of-fit test is obtained for the above-mentioned model.

The paper is organized as follows. In Section 2 the main results are presented, and in Section 3 proofs are given.

2. Goodness-of-fit test

Let θ_0 be an interior point of $\Theta \subset \mathbf{R}^d$, and $\hat{\theta}_n$ be a strongly consistent estimator of a parameter θ_0, i.e., $\hat{\theta}_n$ converges to θ_0 a.s., as $n \to \infty$. Consider a random functional $Q_n(\theta) \in C^1(\Theta)$ and suppose that with probability 1

$Q_n(\theta)$ converges uniformly on each compact subset of Θ to a limit functional $Q_\infty(\theta,\theta_0)$, and $Q_\infty(u,v)\in C^1(\Theta\times\Theta)$. Suppose also that $Q_n(\theta)\le Q_n(\hat\theta_n)$ a.s. for all $\theta\in\Theta$ and $Q_\infty(\theta,\theta_0)<Q_\infty(\theta_0,\theta_0)$ for all $\theta\in\Theta,\ \theta\ne\theta_0$.

Theorem 1. *Assume that:*

1) $\begin{pmatrix}\sqrt{n}(Q_n(\theta_0)-Q_\infty(\theta_0,\theta_0))\\ \sqrt{n}\ grad\ Q_n(\theta_0)\end{pmatrix}\to\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}$ *in distribution,* ξ_1 *is a random variable, and* ξ_2 *is a random vector in* $\mathbf{R}^d$*;*

2) $Q_n''(\theta_0)\to S$ *in probability,* S *is a nonsingular matrix;*

3) $\lim_{\epsilon\to0+}\limsup_n P\{\sup_{\|\theta-\theta_0\|\le\epsilon}\|Q_n''(\theta-Q_n''(\theta_0\|>\delta\}=0$ *for all* δ*.*

Then $T_n=\sqrt{n}(Q_n(\hat\theta_n)-Q_\infty(\hat\theta_n,\hat\theta_n))\to\nu=\left(1,gS^{-1}\right)\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}$ *in law.*

Here $g=\frac{\partial Q_\infty}{\partial v}(\theta_0,\theta_0)$. We regard here and further the derivative vectors as row vectors.

The model in which d.f. F_i has a form $F_i=F(x)^{\gamma^{i-1}}$, and $F(x)$ is unspecified is called *semi-parametric.*

We call *three-parameter model* to be the model in which d.f. F_i has a form $F_i=F(x;A,\alpha)^{\gamma^{i-1}}$, F is Fréchet distribution function with parameters A, α, $F(x)=\exp\left(-(Ax)^{-\alpha}\right),\ i=1,2,....$

Let L be a log-likelihood function, and $Q_n(X,\theta)$ be a normalized log-likelihood function, $Q_n=\frac{1}{n}L$. In the semiparametric model

$$Q_n(\gamma)=\frac{1}{n}\sum_{i=2}^{n}(I_i-p_i^0)\ln\frac{p_i}{1-p_i}+\frac{1}{n}\sum_{i=2}^{n}\Big[p_i^0\ln\frac{p_i}{1-p_i}+\ln(1-p_i)\Big],$$

and in the three-parameter model

$$Q_n(\theta)=\frac{1}{n}L(A,\alpha,\gamma)=\Big(1-\frac{1}{n}\Big)\frac{\psi}{2}+\ln\alpha-\alpha\ln A+(\alpha+1)\ln A_0-$$

$$-\frac{\alpha+1}{\alpha_0}\frac{\psi_0}{2}\Big(1-\frac{1}{n}\Big)-\frac{1}{n}\frac{\alpha+1}{\alpha_0}\sum_{i=1}^{n}\ln Z_i-\left(\frac{A_0}{A}\right)^{\alpha}\frac{1}{n}\sum_{i=1}^{n}e^{\frac{i-1}{n}(\psi-\psi_0\frac{\alpha}{\alpha_0})}Z_i^{-\frac{\alpha}{\alpha_0}}.$$

In the semi-parametric model for $\gamma>1,\ \gamma_0>1$ define

$$Q_\infty(\gamma,\gamma_0)=(1-\gamma_0^{-1})\ln(\gamma-1)-\ln\gamma.$$

In the three-parameter model define for $\theta=(A,\alpha,\psi)\in\Theta=(0,\infty)\times(0,\infty)\times\mathbf{R}$, $\psi=n\ln\gamma$, $\theta_0=(A_0,\alpha_0,\psi_0)\in\Theta$, $\psi_0=n\ln\gamma_0$.

$$Q_\infty(\theta,\theta_0)=\frac{1}{2}\Big(\psi-\frac{\alpha+1}{\alpha_0}\psi_0\Big)+\ln\alpha A_0+\alpha\ln\frac{A_0}{A}-\frac{\alpha+1}{\alpha_0}\gamma-$$

$$-\left(\frac{A_0}{A}\right)^{\alpha}\Gamma\Big(1+\frac{\alpha}{\alpha_0}\Big)\frac{e^{\psi-\frac{\alpha}{\alpha_0}\psi_0}-1}{\psi-\frac{\alpha}{\alpha_0}\psi_0}.$$

Applying the Theorem 1 to semi-parametric and three-parameter models we obtain the following results:

Theorem 2. *In semi-parametric model with* $\gamma_0 > 1$

$$T_n = \sqrt{n}(Q_n(\hat{\gamma}_n) - Q_\infty(\hat{\gamma}_n, \hat{\gamma}_n)) \to \ N(0, \sigma^2(\gamma_0)) \text{ in law. Here}$$

$$\sigma^2(\gamma_0) = \frac{2l_0(l_0\gamma_0^2 - 2l_0\gamma_0 + l_0 + 1)}{\gamma_0^4},\ l_0 = (\ln(\gamma_0 - 1))^2.$$

Corollary 1. *Let the conditions of Theorem 2 hold. Then* $V_n = \frac{T_n}{\sigma(\hat{\gamma}_n)} \to$ *N(0,1) in law.*

Theorem 3. *In three-parameter model with* $\gamma_0 > 1$

$$T_n = \sqrt{n}(Q_n(\hat{\theta}_n) - Q_\infty(\hat{\theta}_n, \hat{\theta}_n)) \to N(0, \sigma^2(\theta_0)) \text{ in law. Here}$$

$$\sigma^2(\theta_0) = \kappa^T BKB^T\kappa,\ \kappa^T = (1, gS^{-1}),\ g = \left(\tfrac{1}{A_0},\ \ \tfrac{1}{2}\tfrac{\psi_0+2\alpha_0+2\gamma_e}{\alpha_0^2},\ \ -\tfrac{1}{2\alpha_0}\right),$$

$$B = \begin{pmatrix} \frac{\alpha_0+1}{\alpha_0} & 1 & 0 & 0 \\ 0 & -\frac{\alpha_0}{A_0} & 0 & 0 \\ 0 & 0 & -\frac{\psi_0}{\alpha_0} & \frac{1}{\alpha_0} \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

$$K = \begin{pmatrix} \frac{\pi^2}{6} & -1 & -\frac{1}{2} & \gamma_e \\ -1 & 1 & \frac{1}{2} & 1-\gamma_e \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{3} & \frac{1-\gamma_e}{2} \\ \gamma_e & 1-\gamma_e & \frac{1-\gamma_e}{2} & 1+\frac{\pi^2}{3}+4\gamma_e+\gamma_e^2 \end{pmatrix},$$

$$S = \begin{pmatrix} -\frac{\alpha_0^2}{A_0^2} & -\frac{1}{2}\frac{\psi_0-2+2\gamma_e}{A_0} & \frac{1}{2}\frac{\alpha_0}{A_0} \\ -\frac{1}{2}\frac{\psi_0-2+2\gamma_e}{A_0} & -\frac{1}{6}\frac{6\gamma_e^2-12\gamma_e-6\psi_0+2\psi_0^2+\pi^2+6+6\psi_0\gamma_e}{\alpha_0^2} & \frac{1}{6}\frac{3\gamma_e-3+2\psi_0}{\alpha_0} \\ \frac{1}{2}\frac{\alpha_0}{A_0} & \frac{1}{6}\frac{3\gamma_e-3+2\psi_0}{\alpha_0} & -\frac{1}{3} \end{pmatrix},$$

γ_e is Euler's constant.

Corollary 2. *Let the conditions of Theorem 3 hold. Then* $V_n = \frac{T_n}{\sigma(\hat{\theta}_n)} \to$ *N(0,1) in law.*

Corollaries 1 and 2 are applied to goodness-of-fit test. In both semi-parametric and three-parameter cases we reject the hypothesis about validity of the model with $\gamma_0 > 1$ if $|V_n| > N_{\alpha/2}$, where $N_{\alpha/2}$ is $\alpha/2$-quantile of normal law, i.e., $P\{N(0,1) > N_{\alpha/2}\} = \alpha/2$.

3. Proofs

Proof of the Theorem 1. Before proving the Theorem 1 consider the following three statements:

Lemma 1. *Let $\Theta \subset \mathbf{R^d}$, θ_0 be an interior point of Θ, $\{Q_n(\theta),\ \theta \in \Theta,\ n \geq 1\}$ be sequence of random fields, which are twice differentiable in the neighborhood of θ_0. Let θ_n be a random vector defined by*

$$\theta_n \in \arg\max_{\theta\in\Theta} Q_n(\theta),$$

and suppose that $\theta_n \to \theta_0$ in probability. Assume also that:

a) $\sqrt{n}Q_n'(\theta_0)$ converges in law to a random vector γ,

b) $Q_n''(\theta_0) \to S$ in probability, where S is nonsingular nonrandom matrix,

c) For each $\delta > 0$

$$\lim_{\epsilon\to 0}\limsup_{n\to\infty} P\Big(\sup_{\|\theta-\theta_0\|\leq\epsilon} \|Q_n''(\theta) - Q_n''(\theta_0)\| > \delta\Big) = 0.$$

Then $\delta_n := \sqrt{n}(\theta_n - \theta_0) \to -S^{-1}\gamma$ in law.

Proof. $Q_n'(\theta_n) = 0$ with probability tending to 1, because $\theta_n \in \arg\max_\theta Q_n(\theta)$ and $\theta_n \to \theta_0$ in probability, θ_0 is interior point. Then

$$\frac{\partial Q_n(\theta_n)}{\partial\theta^i} = 0,\ \frac{\partial Q_n(\theta_0)}{\partial\theta^i} + \sum_{j=1}^{n} \frac{\partial^2 Q_n(\bar\theta_i)}{\partial\theta^i\partial\theta^j}(\theta_n^j - \theta_0^j) = 0,\ \bar\theta_i \in [\theta_0, \theta_n].$$

Then $\sqrt{n}Q_n'(\theta_0) + Q_n''(\theta_0)\delta_n + R_n = 0$, $R_n = \Lambda_n(\theta_n - \theta_0)\sqrt{n}$, $\Lambda_n = (\Lambda_n^{ij})_{i,j=1}^3$,

$$\Lambda_n^{ij} = \frac{\partial^2 Q_n(\bar\theta_i)}{\partial\theta^i\partial\theta^j} - \frac{\partial^2 Q_n(\theta_0)}{\partial\theta^i\partial\theta^j}.$$

So,

$$(Q_n''(\theta_0) + \Lambda_n)\delta_n = -\sqrt{n}Q_n'(\theta_0). \tag{1}$$

Now, $\Lambda_n \to 0$ in probability. Indeed,

$$P\{\|\Lambda_n\| \geq \delta\} \leq P\{\|\theta_n - \theta_0\| > \epsilon\} + P\{\sup_{\|\bar\theta_i-\theta_0\|\leq\epsilon} \|\Lambda_n\| \geq \delta\},$$

$$\limsup_{n\to\infty} P\{\|\Lambda_n\| \geq \delta\} \leq \limsup_{n\to\infty} P \sup_{\|\bar\theta_i-\theta_0\|\leq\epsilon} \{\|\Lambda_n\| \geq \delta\} \to 0,\ \epsilon \to 0+$$

by condition (c). Thus $\Lambda_n \to 0$ in probability.

Then in (1) $-\sqrt{n}Q_n'(\theta_0) \to -\gamma$ in law, $Q_n''(\theta_0) + \Lambda_n \to S$ in probability. So, $\delta_n = (Q_n''(\theta_0) + \Lambda_n)^{-1}(-\sqrt{n}Q_n'(\theta_0))$, it holds with probability tending to 1 and $\delta_n \to -S^{-1}\gamma$ in distribution, because S is nonsingular.□

Lemma 2. *Assume that:*

1) $\sqrt{n}Q_n'(\theta_0) \to \xi$ *in distribution,* ξ *is a random vector in* $\mathbf{R}^d$*;*

2) $Q_n''(\theta_0) \to S$ *in probability,* S *is a nonsingular non-random matrix;*

3) $\lim_{\epsilon\to 0+} \limsup_n P\{\sup_{\|\theta-\theta_0\|\le\epsilon} \|Q_n''(\theta - Q_n''(\theta_0\| > \delta\} = 0$ *for all* δ*.*

Then $\sqrt{n}(Q_\infty(\hat\theta_n, \hat\theta_n) - Q_\infty(\theta_0, \theta_0)) = g\delta_n + o_p(1),\ n \to \infty$*, where* $g = \frac{\partial Q_\infty}{\partial v}(\theta_0, \theta_0)$.

Proof. Expand the value $Q_\infty(\hat\theta_n, \hat\theta_n)$ into Taylor series near the point (θ_0, θ_0).

$$Q_\infty(\hat\theta_n, \hat\theta_n) = Q_\infty(\theta_0, \theta_0) + \frac{\partial Q_\infty}{\partial u}(\theta_0, \theta_0)(\hat\theta_n - \theta_0) +$$

$$+\frac{\partial Q_\infty}{\partial v}(\theta_0, \theta_0)(\hat\theta_n - \theta_0) + o(\|\hat\theta_n - \theta_0\|).$$

$\theta = \theta_0$ is a maximum point of $Q_\infty(\theta, \theta_0)$. So $\frac{\partial Q_\infty}{\partial u}(\theta_0, \theta_0) = 0$. According to lemma 1, all the conditions of which are satisfied, $\delta_n = \sqrt{n}(\hat\theta_n - \theta_0) \to -S^{-1}\xi$. So, $\sqrt{n}(\hat\theta_n - \theta_0) = O_p(1)$ and $\sqrt{n}\|\hat\theta_n - \theta_0\| = O_p(1)$. Thus $o(\|\hat\theta_n - \theta_0\|) = o(\frac{1}{\sqrt{n}})O_p(1)) = o_p(\frac{1}{\sqrt{n}})$. And finally $\sqrt{n}(Q_\infty(\hat\theta_n, \hat\theta_n) - Q_\infty(\theta_0, \theta_0)) = g\delta_n + o_p(1)$.□

Lemma 3. *Let the conditions of lemma 2 hold.*
Then $\sqrt{n}(Q_n(\hat\theta_n) - Q_n(\theta_0)) = o_p(1),\ n \to \infty$.

Proof. Expand the value $Q_n(\theta_0)$ into Taylor series near the point $\theta = \hat\theta_n$.

$$Q_n(\theta_0) = Q_n(\hat\theta_n) + \frac{dQ_n}{d\theta}(\hat\theta_n)(\theta_0 - \hat\theta_n) + \frac{d^2Q_n}{d\theta^2}(\bar\theta)(\theta_0 - \hat\theta_n)^2,\ \bar\theta \in [\theta_0, \hat\theta_n].$$

With probability 1 $\hat\theta_n \to \theta_0$, θ_0 is interior point, so $\hat\theta_n$ is interior point for $n > n_0(\omega)$. $\theta = \hat\theta_n$ is a maximum point of $Q_n(\theta)$. So $\frac{dQ_n}{d\theta}(\hat\theta_n) = 0$. Using again Lemma 1 we obtain $\delta_n = \sqrt{n}(\hat\theta_n - \theta_0) = O_p(1),\ n \to \infty$. Thus $n(\hat\theta_n - \theta_0)^2 = O_p(1),\ n \to \infty$ and $\sqrt{n}(\hat\theta_n - \theta_0)^2 = \frac{1}{\sqrt{n}}O_p(1) = o_p(1),\ n \to \infty$. Now, show that $\frac{d^2Q_n}{d\theta^2}(\bar\theta)$ is stochastically bounded.

$$P\Big\{\Big\|\frac{d^2Q_n}{d\theta^2}(\bar\theta)\Big\| \ge C\Big\} \le P\{\|\bar\theta_n - \theta_0\| > \epsilon\} + P\Big\{\sup_{\|\bar\theta-\theta_0\|\le\epsilon}\Big\|\frac{d^2Q_n}{d\theta^2}(\bar\theta)\Big\| \ge C\Big\} \le$$

$$P\{\|\hat\theta_n - \theta_0\| > \epsilon\} + P\Big\{\sup_{\|\bar\theta-\theta_0\|\le\epsilon}\Big(\Big\|\frac{d^2Q_n}{d\theta^2}(\bar\theta) - \frac{d^2Q_n}{d\theta^2}(\theta_0)\Big\| + \Big\|\frac{d^2Q_n}{d\theta^2}(\theta_0)\Big\|\Big) \ge C\Big\} \le$$

$$P\{\|\hat\theta_n - \theta_0\| > \epsilon\} + P\Big\{\sup_{\|\bar\theta-\theta_0\|\le\epsilon}\Big\|\frac{d^2Q_n}{d\theta^2}(\bar\theta) - \frac{d^2Q_n}{d\theta^2}(\theta_0)\Big\| \ge \frac{C}{2}\Big\} +$$

$$+P\Big\{\Big\|\frac{d^2Q_n}{d\theta^2}(\theta_0)\Big\| \geq \frac{C}{2}\Big\}.$$

Since $\hat{\theta}_n \to \theta_0$, a.s., $n \to \infty$, for any $\epsilon > 0$

$$\limsup_{n\to\infty} P\Big\{\Big\|\frac{d^2Q_n}{d\theta^2}(\bar{\theta})\Big\| \geq C\Big\} \leq$$

$$\limsup_{n\to\infty} P\Big\{\sup_{\|\bar{\theta}-\theta_0\|\leq\epsilon}\Big\|\frac{d^2Q_n}{d\theta^2}(\bar{\theta}) - \frac{d^2Q_n}{d\theta^2}(\theta_0)\Big\| \geq \frac{C}{2}\Big\}+$$

$$+\limsup_{n\to\infty} P\Big\{\Big\|\frac{d^2Q_n}{d\theta^2}(\theta_0)\Big\| \geq \frac{C}{2}\Big\}.$$

Now, tend ϵ to 0 and use the condition 3) of the lemma.

$$\limsup_{n\to\infty} P\Big\{\Big\|\frac{d^2Q_n}{d\theta^2}(\bar{\theta})\Big\| \geq C\Big\} \leq \limsup_{n\to\infty} P\Big\{\Big\|\frac{d^2Q_n}{d\theta^2}(\theta_0)\Big\| \geq \frac{C}{2}\Big\}.$$

According to the condition 2) of the lemma $\Big\|\frac{d^2Q_n}{d\theta^2}(\theta_0)\Big\| = O_p(1)$. By this way we have proved that $\sqrt{n}\frac{d^2Q_n}{d\theta^2}(\bar{\theta})(\theta_0 - \hat{\theta}_n)^2 = o_p(1), \bar{\theta} \in [\theta_0, \hat{\theta}_n]$.

From this the statement of the lemma follows.□

Now, prove Theorem 1. At first mention that condition 1) of Lemma 2 and conditions of Lemma 3 follow from the condition 1 of the Theorem 1.

From Lemma 1 we have: $\delta_n = \sqrt{n}(\hat{\theta}_n - \theta_0) = (Q_n''(\theta_0))^{-1}(-\sqrt{n}Q_n'(\theta_0)) + o_p(1)$, $n \to \infty$; $S\delta_n + \sqrt{n}Q_n'(\theta_0) \to 0$, $n \to \infty$ in probability. Thus

$$\begin{pmatrix}\sqrt{n}(Q_n(\theta_0) - Q_\infty(\theta_0, \theta_0))\\ S\delta_n\end{pmatrix} \to \begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix} \text{ in probability.}$$

Apply to this vector the continuous transformation

$$F\begin{pmatrix}\upsilon_1\\ \upsilon_2\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & S^{-1}\end{pmatrix}\begin{pmatrix}\upsilon_1\\ \upsilon_2\end{pmatrix}.$$

From condition 1) of Lemma 1 using Slutsky lemma we obtain:

$$\begin{pmatrix}\sqrt{n}(Q_n(\theta_0) - Q_\infty(\theta_0, \theta_0))\\ \delta_n\end{pmatrix} \to \begin{pmatrix}\xi_1\\ -S^{-1}\xi_2\end{pmatrix} \text{ in distribution.}$$

And similarly

$$\begin{pmatrix}\sqrt{n}(Q_n(\theta_0) - Q_\infty(\theta_0, \theta_0))\\ g\delta_n\end{pmatrix} \to \begin{pmatrix}\xi_1\\ -gS^{-1}\xi_2\end{pmatrix} \text{ in distribution.}$$

In Lemma 3 it was proved that

$$\sqrt{n}(Q_\infty(\hat{\theta}_n, \hat{\theta}_n) - Q_\infty(\theta_0, \theta_0)) = g\delta_n + o_p(1),\ n \to \infty.$$

So,

$$\begin{pmatrix} \sqrt{n}(Q_n(\theta_0) - Q_\infty(\theta_0,\theta_0)) \\ \sqrt{n}(Q_\infty(\hat\theta_n,\hat\theta_n) - Q_\infty(\theta_0,\theta_0)) \end{pmatrix} \to \begin{pmatrix} \xi_1 \\ -gS^{-1}\xi_2 \end{pmatrix} \text{ in distribution.}$$

Using Lemma 4 we obtain:

$$\begin{pmatrix} \sqrt{n}(Q_n(\hat\theta_n) - Q_\infty(\theta_0,\theta_0)) \\ \sqrt{n}(Q_\infty(\hat\theta_n,\hat\theta_n) - Q_\infty(\theta_0,\theta_0)) \end{pmatrix} \to \begin{pmatrix} \xi_1 \\ -gS^{-1}\xi_2 \end{pmatrix} \text{ in distribution.}$$

Finally, considering the transformation $G\begin{pmatrix} \upsilon_1 \\ \upsilon_2 \end{pmatrix} = \upsilon_1 - \upsilon_2$ we obtain:

$$\sqrt{n}(Q_n(\hat\theta_n) - Q_\infty(\hat\theta_n,\hat\theta_n)) \to \xi_1 + gS^{-1}\xi_2 = (1, gS^{-1})\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \text{ in distribution.}\square$$

Proof of the Theorem 2. We must check three conditions of the theorem 1 and prove that the vector $(1, gS^{-1})\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$ is normally distributed with parameters $(0, \sigma^2(\theta_0))$.

Check the first condition. Consider $W = \begin{pmatrix} \sqrt{n}(Q_n(\theta_0) - Q_\infty(\theta_0,\theta_0) \\ \sqrt{n}\ \text{grad}\ Q_n(\theta_0) \end{pmatrix}$. It was proved above that

$$W = \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=2}^n (I_i - p_i^0)\ln\frac{p_i^0}{1-p_i^0} \\ \frac{1}{\sqrt{n}}\sum_{i=2}^n (I_i - p_i^0)\frac{p_i'(\gamma_0)}{p_i^0(1-p_i^0)} \end{pmatrix} = \frac{1}{\sqrt{n}}\sum_{i=2}^n (I_i - p_i^0)\begin{pmatrix} \ln\frac{p_i^0}{1-p_i^0} \\ \frac{p_i'(\gamma_0)}{p_i^0(1-p_i^0)} \end{pmatrix}$$

Denote $A_i(\gamma_0) = \begin{pmatrix} \ln\frac{p_i^0}{1-p_i^0} \\ \frac{p_i'(\gamma_0)}{p_i^0(1-p_i^0)} \end{pmatrix}$. Then Var W $= \frac{1}{n}\sum_{i=1}^n A_i v A_i^T$. Here $v =$ Var $(I_i - p_i^0) = p_i^0(1-p_i^0)$. Thus

$$\text{Var } W \to \begin{pmatrix} (\ln(\gamma_0 - 1))^2\frac{(\gamma_0-1)}{\gamma_0^2} & -\frac{\ln(\gamma_0-1)}{\gamma_0^2} \\ -\frac{\ln(\gamma_0-1)}{\gamma_0^2} & \frac{1}{\gamma_0^2(\gamma_0-1)} \end{pmatrix} := K.$$

Now, the second condition of the theorem 1 holds, as $Q_n''(\gamma_0) \to S = -\frac{1}{\gamma_0^2(\gamma_0-1)}$, $n \to \infty$ in probability.

It is easy to see that $g = \frac{\partial Q_\infty}{\partial v}(\theta_0,\theta_0) = \frac{\ln\gamma_0 - 1}{\gamma_0^2}$.

The third condition of the theorem holds as well. Therefore, applying the theorem 1 we obtain:

$$T_n = \sqrt{n}(Q_n(\hat\gamma_n) - Q_\infty(\hat\gamma_n,\hat\gamma_n)) \to\ N(0,\sigma^2(\gamma_0)) \text{ in law.}$$

$$\sigma^2(\gamma_0) = \kappa K \kappa^T,\ \kappa = (1, gS^{-1})^T.$$
$$\sigma^2(\gamma_0) = \frac{2l_0(l_0\gamma_0^2 - 2l_0\gamma_0 + l_0 + 1)}{\gamma_0^4},\ l_0 = (\ln(\gamma_0 - 1))^2.\square$$

Proof of the Theorem 3. (i) Reparametrization. Consider the maximum likelihood function for the observed data set $X_1, ..., X_n$, $L(A, \alpha, \gamma) = \sum_{i=1}^n \ln f(x_i; A, \alpha, \gamma)$. Here $f_i(x; A, \alpha, \gamma)$ is density function of described above distribution, $f_i(x; A, \alpha, \gamma) = \frac{\partial F_i(x;A,\alpha,\gamma)}{\partial x}$. Let $\psi = n \ln \gamma$. Rewrite this functional using new parameter $\theta = (A, \alpha, \psi)$ (the same is for true values), $\theta \in (0, \infty) \times (0, \infty) \times \mathbf{R}$.

$$Q_n(\theta) = \frac{1}{n} L(A, \alpha, \gamma) = (1 - \frac{1}{n})\frac{\psi}{2} + \ln(\alpha A^{-\alpha}) -$$
$$\frac{1}{n}(\alpha + 1)\sum_{i=1}^n \ln X_i - \frac{1}{n}\sum_{i=1}^n e^{\frac{i-1}{n}\psi}(AX_i)^{-\alpha}$$

Rewrite the functional $Q_n(\theta)$ using the i.i.d. sequence $Z_i = \frac{(A_0 X_i)^{\alpha_0}}{(\gamma_0^{1/\alpha_0})^{i-1}}$, $i = 1, 2, ...,$

$$Q_n(\theta) = \frac{1}{n} L(A, \alpha, \gamma) = \Big(1 - \frac{1}{n}\Big)\frac{\psi}{2} + \ln\alpha - \alpha \ln A + (\alpha + 1)\ln A_0 -$$
$$-\frac{\alpha + 1}{\alpha_0}\frac{\psi_0}{2}\Big(1 - \frac{1}{n}\Big) - \frac{1}{n}\frac{\alpha + 1}{\alpha_0}\sum_{i=1}^n \ln Z_i - \left(\frac{A_0}{A}\right)^\alpha \frac{1}{n}\sum_{i=1}^n e^{\frac{i-1}{n}(\psi - \psi_0\frac{\alpha}{\alpha_0})} Z_i^{-\frac{\alpha}{\alpha_0}}.$$

We obtain

$$Q_n(\theta) = \frac{n-1}{2n}\Big(\psi - \frac{\alpha + 1}{\alpha_0}\psi_0\Big) + \ln\alpha A_0 + \alpha(\ln\frac{A_0}{A}) -$$
$$-\frac{\alpha + 1}{\alpha_0}\frac{1}{n}\sum_{i=1}^n \ln Z_i - \left(\frac{A_0}{A}\right)^\alpha \frac{1}{n}\sum_{i=1}^n e^{\frac{i-1}{n}(\psi - \psi_0\frac{\alpha}{\alpha_0})} Z_i^{-\frac{\alpha}{\alpha_0}} \qquad (2)$$

(ii) Limit functional. Consider θ belong to a compact subset of Θ. Uniformly on this set we have

$$Q_n(\theta) = Q_\infty(\theta) + R_1 + R_2 + o(1),$$

with the limit functional

$$Q_\infty(\theta, \theta_0) = \frac{1}{2}\Big(\psi - \frac{\alpha + 1}{\alpha_0}\psi_0\Big) + \ln\alpha A_0 + \alpha \ln\frac{A_0}{A} - \frac{\alpha + 1}{\alpha_0}\gamma -$$
$$-\left(\frac{A_0}{A}\right)^\alpha \Gamma\Big(1 + \frac{\alpha}{\alpha_0}\Big)\frac{e^{\psi - \frac{\alpha}{\alpha_0}\psi_0} - 1}{\psi - \frac{\alpha}{\alpha_0}\psi_0} \qquad (3)$$

and

$$R_1(\theta) = -\frac{1}{n}\frac{\alpha+1}{\alpha_0}\sum_{i=1}^{n}\Big(\ln Z_i - E\ln Z_i\Big); \tag{4}$$

$$R_2(\theta) = -\frac{1}{n}\left(\frac{A_0}{A}\right)^{\alpha}\sum_{i=1}^{n} e^{\frac{i-1}{n}(\psi-\psi_0\frac{\alpha}{\alpha_0})}\Big(Z_i^{-\frac{\alpha}{\alpha_0}} - EZ_i^{-\frac{\alpha}{\alpha_0}}\Big). \tag{5}$$

With probability 1 $Q_n(\theta)$ converges to $Q_\infty(\theta,\theta_0)$ uniformly.

(iii) We must check the conditions of Theorem 1 and prove that $(1, gS^{-1})\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}$ is normally distributed with mean 0 and variance $\sigma^2(\theta_0)$.

Check the first condition of the Theorem 1: Consider

$$W = \begin{pmatrix}\sqrt{n}(Q_n(\theta_0) - Q_\infty(\theta_0,\theta_0)\\ \sqrt{n}\ \text{grad}\ Q_n(\theta_0)\end{pmatrix}.$$

It is easy to see that $\sqrt{n}(Q_n(\theta_0) - Q_\infty(\theta_0,\theta_0) = \sqrt{n}(R_1(\theta_0) + R_2(\theta_0))$. Here

$$R_1(\theta_0) = -\frac{\alpha_0+1}{\alpha_0}\frac{1}{n}\sum_{i=1}^{n}(\ln z_i - E\ln z_i);$$

$$R_2(\theta_0) = R_2 = -\frac{1}{n}\sum_{i=1}^{n} z_i^{-1} - Ez_i^{-1}.$$

Another part of W has the form:

$$\sqrt{n}Q_n'(\theta_0) =$$

$$-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix}-\frac{\alpha_0}{A_0}(z_i^{-1} - Ez_i^{-1})\\ \frac{1}{\alpha_0}[((1-z_i^{-1})\ln z_i - \psi_0\frac{i-1}{n}z_i^{-1}) - E((1-z_i^{-1})\ln z_i - \psi_0\frac{i-1}{n}z_i^{-1})]\\ \frac{i-1}{n}(z_i^{-1} - Ez_i^{-1})\end{pmatrix}.$$

Thus W is equal to

$$-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix}(\frac{\alpha_0+1}{\alpha_0}\ln z_i + z_i^{-1}) - E(\frac{\alpha_0+1}{\alpha_0}\ln z_i + z_i^{-1})\\ -\frac{\alpha_0}{A_0}(z_i^{-1} - Ez_i^{-1})\\ \frac{1}{\alpha_0}[((1-z_i^{-1})\ln z_i - \psi_0\frac{i-1}{n}z_i^{-1}) - E((1-z_i^{-1})\ln z_i - \psi_0\frac{i-1}{n}z_i^{-1})]\\ \frac{i-1}{n}(z_i^{-1} - Ez_i^{-1})\end{pmatrix}.$$

Consider the vector

$$\zeta_i = \begin{pmatrix}\ln z_i - E\ln z_i\\ z_i^{-1} - Ez_i^{-1}\\ (1-z_i^{-1})\ln z_i - E((1-z_i^{-1})\ln z_i)\end{pmatrix}$$

with

$$\Gamma = \text{Var}\ \zeta_i = \begin{pmatrix}a & b & c\\ b & d & e\\ c & e & f\end{pmatrix}.$$

Consider also continuous linear transformation

$$A_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{i-1}{n} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i \zeta_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} \ln z_i - E \ln z_i \\ z_i^{-1} - E z_i^{-1} \\ \frac{i-1}{n}(z_i^{-1} - E z_i^{-1}) \\ (1 - z_i^{-1}) \ln z_i - E((1 - z_i^{-1}) \ln z_i) \end{pmatrix}.$$

The corresponding covariance matrix has the form

$$\operatorname{Var} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i \zeta_i \right) = \frac{1}{n} \sum_{i=1}^{n} A_i \Gamma A_i^T =$$

$$= \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} a & b & \frac{i-1}{n} b & c \\ b & d & \frac{i-1}{n} d & e \\ \frac{i-1}{n} b & \frac{i-1}{n} d & \left(\frac{i-1}{n}\right)^2 d & \frac{i-1}{n} e \\ c & e & \frac{i-1}{n} e & f \end{pmatrix} =$$

$$= \begin{pmatrix} a & b & b\frac{S_2}{n^2} & c \\ b & d & db\frac{S_2}{n^2} & e \\ b\frac{S_2}{n^2} & d\frac{S_2}{n^2} & d\frac{S_3}{n^3} & e\frac{S_2}{n^2} \\ c & e & e\frac{S_2}{n^2} & f \end{pmatrix} \longmapsto \begin{pmatrix} a & b & \frac{b}{2} & c \\ b & d & \frac{d}{2} & e \\ \frac{b}{2} & \frac{d}{2} & \frac{d}{3} & \frac{e}{2} \\ c & e & \frac{e}{2} & f \end{pmatrix} := K.$$

Here $S_k = \sum_{i=1}^{n} (i-1)^{k-1}$, $k > 1$, $\frac{S_2}{n^2} = \frac{n(n-1)}{2n^2} \to \frac{1}{2}$; $\frac{S_3}{n^3} = \frac{n(n-1)(2n-1)}{6} \to \frac{1}{3}$.

Now by multivariate CLT $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i \zeta_i \to \rho$, ρ is normal distributed with mean 0 and covariance matrix K. Indeed, check Lyapunov condition for this vector. It is sufficient to check it for the components of the vector. Check it only for third component of the vector $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i \zeta_i$. (Checking the condition for the other components is obvious.) Let

$$\xi_i = z_i^{-1} - E z_i^{-1},\ \xi_{ni} = \frac{i-1}{n\sqrt{n}} \xi_i,\ S_n = \sum_{i=1}^{n} \xi_{ni} = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i \zeta_i \right)_3.$$

Now,

$$s_n^2 = \operatorname{Var}\ S_n = \frac{1}{n} \sum_{i=1}^{n} \frac{(i-1)^2}{n^2} \to \frac{1}{3}.$$

Check Lyapunov condition:

$$\frac{1}{s_n^3} \sum_{i=1}^{n} E|\xi_{nk}|^3 \to 0,\ n \to \infty.$$

Indeed,

$$\frac{1}{(n\sqrt{n})^3}\sum_{i=1}^{n} E|\xi_i|^3 = \frac{S_4(n)}{n^4\sqrt{n}}E|\xi_1|^3 \to 0,\ n\to\infty,\ S_4(n)=\sum_{i=1}^{n}(i-1)^3.$$

Consider another transformation:

$$B = \begin{pmatrix} \frac{\alpha_0+1}{\alpha_0} & 1 & 0 & 0 \\ 0 & -\frac{\alpha_0}{A_0} & 0 & 0 \\ 0 & 0 & -\frac{\psi_0}{\alpha_0} & \frac{1}{\alpha_0} \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

It is easy to see that $W = -B\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} A_i\zeta_i\Big)$. And we obtain that $W \to N(0,T)$, where $T = \operatorname{Var} W = BKB^T$. Obviously, $(1, gS^{-1})\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}$ is normally distributed with mean 0 and variance $\sigma^2(\theta_0)$, where $\sigma^2(\theta_0) = = \kappa^T BKB^T\kappa$, $\kappa^T = (1, gS^{-1})$.

Now, check the second condition of the theorem 1.

Consider $Q_n(\theta) = Q_\infty(\theta,\theta_0) + R_1(\theta) + R_2(\theta)$. Direct calculations show that $(Q_\infty)''(\theta_0,\theta_0) =$

$$\begin{pmatrix} -\frac{\alpha_0^2}{A_0^2} & -\frac{1}{2}\frac{\psi_0-2+2\gamma_e}{A_0} & \frac{1}{2}\frac{\alpha_0}{A_0} \\ -\frac{1}{2}\frac{\psi_0-2+2\gamma_e}{A_0} & -\frac{1}{6}\frac{6\gamma_e^2-12\gamma_e-6\psi_0+2\psi_0^2+\pi^2+6+6\psi_0\gamma_e}{\alpha_0^2} & \frac{1}{6}\frac{3\gamma_e-3+2\psi_0}{\alpha_0} \\ \frac{1}{2}\frac{\alpha_0}{A_0} & \frac{1}{6}\frac{3\gamma_e-3+2\psi_0}{\alpha_0} & -\frac{1}{3} \end{pmatrix} := S.$$

It is easy to see that $R_1''(\theta) \equiv 0$ for all $\theta\in\Theta$.

Now consider $R_2''(\theta_0)$. Every element of this matrix consists of linear combination of the following expressions:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{(i-1)^k}{n^k}(z_i^{-1} - Ez_i^{-1}),\ \frac{1}{n}\sum_{i=1}^{n}\frac{(i-1)^k}{n^k}(z_i^{-1}\ln z_i - E(z_i^{-1}\ln z_i)),$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{(i-1)^k}{n^k}(z_i^{-1}\ln^2 z_i - E(z_i^{-1}\ln^2 z_i)),\ k=0,1,2.$$

Thus $R_2''(\theta_0)\to 0,\ n\to\infty$, in probability. Consider, for example, the most interesting term

$$\frac{\partial^2 R_2}{\partial\psi^2}(\theta) = -\left(\frac{A_0}{A}\right)^{\alpha}\frac{1}{n}\sum_{i=1}^{n}\frac{(i-1)^2}{n^2}e^{\frac{i-1}{n}(\psi-\psi_0\frac{\alpha}{\alpha_0})}\Big(z_i{}^{\frac{\alpha}{\alpha_0}} - Ez_i{}^{\frac{\alpha}{\alpha_0}}\Big),$$

$$\frac{\partial^2 R_2}{\partial\psi^2}(\theta_0) = \frac{1}{n}\sum_{i=1}^{n}\frac{(i-1)^2}{n^2}\Big(z_i{}^{-1} - Ez_i{}^{-1}\Big).$$

But

$$\operatorname{Var}\frac{\partial^2 R_2}{\partial\psi^2}(\theta_0)=\frac{1}{n^2}\frac{1-3n+2n^2}{n}\mathrm{Const}\asymp\frac{1}{n}\to 0,\ n\to\infty.$$

And $\frac{\partial^2 R_2}{\partial\psi^2}(\theta_0)\to 0,\ n\to\infty$ in probability.

Finally, check the last condition of the theorem 1. Consider $Q_n'''(\theta_0)$. Every element of this tensor consists of linear combination of the following expressions:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{(i-1)^k}{n^k}(z_i^{-1}-Ez_i^{-1}),\ \frac{1}{n}\sum_{i=1}^{n}\frac{(i-1)^k}{n^k}(z_i^{-1}\ln z_i-E(z_i^{-1}\ln z_i)),$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{(i-1)^k}{n^k}(z_i^{-1}\ln^2 z_i-E(z_i^{-1}\ln^2 z_i)),\ \sum_{i=1}^{n}\frac{(i-1)^k}{n^k}(z_i^{-1}\ln^3 z_i-E(z_i^{-1}\ln^3 z_i)),$$

$k=0,1,2,3$. Consider the most interesting: $\frac{1}{n}\sum_{i=1}^{n}\frac{(i-1)^3}{n^3}\Big({z_i}^{-1}-E{z_i}^{-1}\Big)$.

$$\operatorname{Var}\Big\{\frac{1}{n}\sum_{i=1}^{n}\frac{(i-1)^3}{n^3}\Big({z_i}^{-1}-E{z_i}^{-1}\Big)\Big\}=$$

$$\frac{1}{n^2}\frac{1-2n+n^2}{n}\mathrm{Const}\asymp\frac{1}{n}\to 0,\ n\to\infty.$$

Therefore there exists a constant $C=C(\omega)$ s.t. $|Q_n'''(\theta_0)|<C$. From this statement the third condition follows.□

Bibliography

1. Kukush, A.G. *On maximum likelihood estimator in a statistical model of natural catastrophe claims.* Theory Stoch. Proces., 1999, 5 (21).

2. Pfeifer, D. *A statistical model to analyse natural catastrophe claims by mean of record values.* Proceedings of the 28 International ASTIN Colloquium, Cairns, Australia 10-12 August, 1997, The institute of Actuaries of Australia, (1997).

Department of Mathematics, Kyiv National Taras Shevchenko University, Kyiv, Ukraine.
E-mail: kuog@mechmat.univ.kiev.ua and chernikov@europe.com

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.215-230

ALEXANDER G. KUKUSH AND DMITRII S. SILVESTROV

SKELETON APPROXIMATIONS OF OPTIMAL STOPPING STRATEGIES FOR AMERICAN TYPE OPTIONS WITH CONTINUOUS TIME[1][2]

American type options are studied for continuous pricing processes. The skeleton type approximations are considered. The explicit upper bounds are given for the step of discretisation for ε-optimal stopping strategies.

2000 *AMS subject classifications.* 60J05, 60H10, 91B28, 91B70.

Key words and phrases. Markov process, optimal stopping, American options, skeleton approximation.

1. Introduction

Traditional methods of option pricing are based on models of pricing processes which are various modifications of the classical model of geometrical Brownian motion. Stochastic differential equations can be written down for such pricing processes. Then partial differential equations and the corresponding variational problems can be derived for functions which represent optimal strategies, see for instance Øksendal (1992), Duffie (1996) and Karatzas and Shreve (1998). Finally various numerical algorithms can be applied to find optimal strategies for continuous time models and their discrete time approximations. The extended survey of latest results can be found in the book edited by Rogers and Talay (1998), in particular in the paper by Broadie and Detemple (1998).

We do prefer to use an alternative approach for evaluation of optimal stopping Buyer strategies for American type options. The structure of optimal stopping strategies is investigated by applying of the direct probabilistic analysis under general assumptions for underlying pricing processes.

[1]The paper represents a part of an invited lecture.

[2]This work is supported in part by the project "Stochastic modelling of insurance and finance processes and systems" funded by the Knowledge Foundation.

In the papers by Kukush and Silvestrov (2000a, 2000b) the structure of optimal stopping strategies were investigated for a general model of discrete time pricing processes and pay-off functions. The model of pricing process is a two component inhomogeneous in time Markov process with a phase space $[0,\infty) \times Y$. The first component is the corresponding pricing process and the second component (with a general measurable phase space Y) represents some stochastic index process controlling the pricing process. Pay-off functions under consideration are in sequel: (a) an inhomogeneous in time analogue of a standard one $g_n(x) = a_n[x-K_n]_+$; (b) piecewise linear convex functions, and finally (c) general convex functions.

At present paper we study skeleton type approximations for continuous time pricing processes. The explicit upper bounds are given for the step of discretisation for ε-optimal stopping strategies. These upper bounds enable us to use the results given in Kukush and Silvestrov (2000a, 2000b) for constructive description of ε-optimal stopping strategies for American type options with continuous time. The special attention is paid to the case of general model of pricing processes which are geometrical diffusion processes controlled by stochastic index processes.

We think that the main advantage of direct probabilistic approach in structural studies of optimal stopping strategies is that this approach is much more flexible and less sensitive to the modifications of models of underlying pricing processes, pay-off functions and other characteristics of the models.

The knowledge of the explicit structure of optimal stopping strategies is the base for the creation of effective optimising Monte Carlo pricing algorithms for numerical evaluation of the corresponding optimal strategies. Such algorithms and programs have been elaborated by Silvestrov, Galochkin and Sibirtsev (1999). We would like also to refer to the papers by

We would like to refer to the book by Shiryaev (1978) and the paper by Shiryaev, Kabanov, Kramkov, and Mel'nikov (1994), which stimulated the present research. We also refer to the paper by Kukush and Silvestrov (1999), where part of the current results was presented without the proofs.

2. Skeleton approximations for American type options in continuous time

Consider a two component inhomogeneous in time Markov process $Z_t = (S_t,\, I_t),\ \ t \geq 0$, with a phase space $Z = [0,\infty) \times Y$. Here $(Y, \mathcal{B}_Y)$ is a general measurable space and as usual we consider Z as a measurable space with the σ–field $\mathcal{B}_Z = \sigma(\mathcal{B}_+ \times \mathcal{B}_Y)$ where $\mathcal{B}_+$ is a Borel σ–field on $R^+ = [\,0,\infty)$.

We assume that $Z_t,\ t \geq 0$ is a measurable process ($Z_t(\omega),\ t \geq 0$ are $\mathcal{B}_Z$–measurable functions with respect to (t,ω)). Without loss of generality we assume that $Z_0 = (S_0, I_0)$ is a non-random value in Z.

We interpret the first component S_t as a pricing process and the second component I_t as a stochastic index process controlling the pricing process.

A basic example of the model described above is the pricing process given in the following form:

$$S_t = S_0 \cdot \exp\{\int_0^t (a(u, I_u) - \frac{1}{2}\,\sigma(u, I_u)^2)du + \int_0^t \sigma(u, I_u)dw(u)\},\ t \geq 0,$$

where (a) $a(t, y)$ and $\sigma(t, y) \geq 0$ are measurable real-valued functions defined on Z, (b) $I_t,\ t \geq 0$ is a measurable inhomogeneous in time Markov process such that functions $E|a(t, I_t)|$ and $E\sigma(t, I_t)^2$ are integrable at finite intervals and $w(u), u \geq 0$ is the Wiener process independent of process $I_t,\ t \geq 0$, (d) $Z_0 = (S_0, I_0)$ is a non-random value in Z.

In this case vector process $Z_t = (S_t, I_t), t \geq 0$ is an inhomogeneous Markov process with the first component $S_t, t \geq 0$ is a continuous geometrical diffusion process controlled by the index process $I_t, t \geq 0$.

Let $\mathcal{F}_t,\ t \geq 0$ be a flow of σ–fields, associated with process $Z_t,\ t \geq 0$. We shall consider Markov moments τ with respect to $\mathcal{F}_t$. It means that τ is a random value distributed in $[0, \infty]$ and with the property $\{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t, t \geq 0$.

Introduce further a pay-off function $g(x, t), x \in R^+, t \geq 0$. We assume that $g(x, t)$ is a nonnegative measurable function. Let also $R_t, t \geq 0$ be a nondecreasing function with $R_0 = 0$. Typically $R_t = \int_0^t r(s)ds$, where $r(s) \geq 0$ is a Borel function representing riskless interest rate at moment s.

The typical example of pay-off function is:

$$g(x, t) = a_t\,[x - K_t]_+ = \begin{cases} a_t\,(x - K_t), & \text{if} \quad x > K_t, \\ 0, & \text{if} \quad 0 \leq x \leq K_t, \end{cases}$$

where $a_t, t \geq 0$ and $K_t, t \geq 0$ are two nonnegative measurable functions. The case, where $a_t = a$ and $K_t = K$ do not depend on t, corresponds to the standard American call option.

We fix parameter $T > 0$ which we call an expiration date. It is convenient to operate with the transformed pricing process $S_g(t) = e^{-R_t} g(S_t, t), t \geq 0$. Let us formulate conditions which we impose on pricing processes and pay-off functions:

A: $S_g(t),\ t \geq 0$ is a.s. continuous from the right process.

B: $E \sup\limits_{0 \leq t \leq T} S_g(t) < \infty$.

Let denote $\mathcal{M}_{max,T}$ the class of all Markov moments $\tau \leq T$. Let now choose a partition $\Pi = \{0 = t_0 < t_1 < \ldots t_N = T\}$ of interval $[0, T]$. We also consider the class $\hat{\mathcal{M}}_{\Pi,T}$ of all Markov moments from $\mathcal{M}_{max,T}$ which only take the values $t_0, t_1, \ldots t_N$, and the class $\mathcal{M}_{\Pi,T}$ of all Markov moments from $\hat{\mathcal{M}}_{\Pi,T}$ such that event $\{\omega : \tau(\omega) = t_k\} \in \sigma[Z_0, \ldots Z_{t_k}]$ for $k = 0, \ldots N$. By definition

$$\mathcal{M}_{\Pi,T} \subseteq \hat{\mathcal{M}}_{\Pi,T} \subseteq \mathcal{M}_{max,T}. \tag{1}$$

The goal functional under consideration is:

$$\Phi_g(\tau) = Ee^{-R_\tau} g(S_\tau, \tau). \tag{2}$$

Denote for a class of Markov moments $\mathcal{M}_T \subseteq \mathcal{M}_{max,T}$

$$\Phi_g(\mathcal{M}_T) = \sup_{\tau \in \mathcal{M}_T} Ee^{-R_\tau} g(S_\tau, \tau). \tag{3}$$

Conditions **A**, **B** and relation (1) imply that

$$\Phi_g(\mathcal{M}_{\Pi,T}) \le \Phi_g(\hat{\mathcal{M}}_{\Pi,T}) \le \Phi_g(\mathcal{M}_{max,T}) < \infty. \tag{4}$$

Random variables $Z_{t_0}, Z_{t_1}, \dots Z_{t_N}$ are connected in an inhomogeneous Markov chain with discrete time. The optimisation problem (2)-(3) for the class $\mathcal{M}_{\Pi,T}$ is a problem of optimal pricing for American type options with discrete time.

In Kukush and Silvestrov (2000a, 2000b) the structure of optimal and ε-optimal stopping moments in the class $\mathcal{M}_{\Pi,T}$ is described for various classes of convex in x pay-off functions $g(x, t_k), k = t_0, \dots t_N$. Also, optimising Monte Carlo algorithms and programs for numerical evaluation of optimal stopping strategies, functionals $\Phi_g(\mathcal{M}_{\Pi,T})$ and other functionals for standard American options with discrete time are described in Silvestrov, Galochkin and Sibirtsev (1999).

Our goal is to show in which way the functional $\Phi_g(\mathcal{M}_{max,T})$ can be approximated by functionals $\Phi_g(\mathcal{M}_{\Pi,T})$ and to give explicit upper bounds for the accuracy of this approximation. This makes it possible to find stopping moments $\tau_\varepsilon \in \mathcal{M}_{\Pi,T}$ that are 2ε-optimal stopping moments in the class $\mathcal{M}_{max,T}$.

The next important statement is a base of skeleton approximation.

Lemma 1. *For every partition* $\Pi = \{0 = t_0 < t_1 < \dots < t_N = T\}$ *of interval* $[0, T]$ *and for the classes* $\mathcal{M}_{\Pi,T}$ *and* $\hat{\mathcal{M}}_{\Pi,T}$ *of Markov moments*

$$\Phi_g(\mathcal{M}_{\Pi,T}) = \Phi_g(\hat{\mathcal{M}}_{\Pi,T}). \tag{5}$$

Proof. Consider the optimization problem (2)-(3) for the class $\hat{\mathcal{M}}_{\Pi,T}$ as a problem of optimal pricing for American type options with discrete time. For this purpose add to the random variables Z_{t_n} additional components $\bar{Z}_{t_n} = \{Z_t, t_{n-1} < t \le t_n\}$ with the phase space $Z^{(t_{n-1}, t_n]}$ endowed by cylindrical σ- field. Consider the extended Markov chain $I_n = (Z_{t_n}, \bar{Z}_{t_n})$. As is known (Shiryaev (1978)) the optimal stopping moment τ exists in any discrete time model, and it has the form of the first hitting time $\tau = \min\{0 \le n \le N : I_n \in D_n\}$, where optimal stopping domains D_n are determined by the transition probabilities of Markov chain I_n. However, in this case the transition probabilities depend only on values of the first component Z_{t_n}. This case was considered in the papers by Kukush and

Silvestrov (2000a, 2000b). To imbed the model described above in the model introduced in these papers one should to consider the two component Markov chain (S_n, I_n) with the components $S_n = S_{t_n}, I_n = (Z_{t_n}, \bar{Z}_{t_n})$. The first component S_n is in this case completely determined by the component $Z_{t_n} = (S_{t_n}, I_{t_n})$ while, as was pointed out above, transition probabilities of Markov chain I_n do depend only of the values of the first component Z_{t_n}. As was shown in Kukush and Silvestrov (2000a, 2000b) in this case the optimal stopping moment has the the form of the first hitting times for the process (S_{t_n}, Z_{t_n}) and do not depend on the component $\bar{Z}_{t_n}$. Since S_{t_n} is determined by Z_{t_n} this moment can by represented in the form $\tau = \min\{0 \le n \le N : Z_{t_n} \in D'_n\}$, i.e. as the first hitting time for the Markov Chain Z_{t_n}.

Therefore for the optimal stopping moment $\tau \in \mathcal{M}_{\Pi,T}$. Hence $\Phi_g(\mathcal{M}_{\Pi,T}) \ge \Phi_g(\hat{\mathcal{M}}_{\Pi,T})$, and by (4) we obtain equality (5). $\bigoplus$

For any Markov moment $\tau \in \mathcal{M}_{max,T}$ and a partition $\Pi = \{0 = t_0 < t_1 < \ldots < t_N = T\}$ one can define the discretisation of this moment

$$\tau[\Pi] = \begin{cases} 0, & \text{if } \tau = 0, \\ t_k, & \text{if } t_{k-1} < \tau \le t_k,\ k = 1, \ldots N. \end{cases}$$

Now, let τ_ε be ε-optimal stopping moment in the class $\mathcal{M}_{max,T}$, i.e. $ES_g(\tau_\varepsilon) \ge \Phi_g(\mathcal{M}_{max,T}) - \varepsilon$. Since $\tau_\varepsilon[\Pi] \in \hat{\mathcal{M}}_{\Pi,T}$ the relation (5) implies

$$ES_g(\tau_\varepsilon[\Pi]) \le \Phi_g(\hat{\mathcal{M}}_{\Pi,T}) = \Phi_g(\mathcal{M}_{\Pi,T}) \le \Phi_g(\mathcal{M}_{max,T}). \tag{6}$$

Denote $d(\Pi) = \max\{t_k - t_{k-1}, k = 1, \ldots N\}$. Let also $\Pi_N = \{0 = t_{0N} < t_{1N} < \ldots t_{NN} = T\}$ be a sequence of partitions such that $d(\Pi_N) \to 0$ as $N \to \infty$.

By definition $\tau_\varepsilon \le \tau_\varepsilon[\Pi_N] \le \tau_\varepsilon + d(\Pi_N)$. That is why condition **A** implies that random variables $S_g(\tau_\varepsilon[\Pi_N]) \to S_g(\tau_\varepsilon)$ as $N \to \infty$ almost surely. This relation, condition **B** and Lebesgue theorem easily implies that $ES_g(\tau_\varepsilon[\Pi_N]) \to ES_g(\tau_\varepsilon) \ge \Phi_g(\mathcal{M}_{max,T}) - \varepsilon$ as $N \to \infty$. Since ε can be chosen arbitrary small the last relation and (6) implies in an obvious way that under conditions **A** and **B**

$$\lim_{N\to\infty} \Phi_g(\mathcal{M}_{\Pi_N,T}) = \Phi_g(\mathcal{M}_{max,T}). \tag{7}$$

Relation (7) gives the base for the use of skeleton discrete time approximation for continuous time model. This relation guarantees that for any fixed $\varepsilon > 0$ there exists $N = N_\varepsilon$ such that $\Phi_g(\mathcal{M}_{max,T}) - \Phi_g(\mathcal{M}_{\Pi_{N_\varepsilon},T}) \le \varepsilon$. Let τ'_ε be an ε-optimal stopping moment in the class $\mathcal{M}_{\Pi_{N_\varepsilon},T}$, i.e. $ES_g(\tau'_\varepsilon) \ge \Phi_g(\mathcal{M}_{\Pi_{N_\varepsilon},T}) - \varepsilon$. Obviously τ'_ε is a 2ε-optimal stopping moment in the class $\mathcal{M}_{max,T}$.

However, relation (7) does not give quantitative estimates which connect the maximal step of the partition $d(\Pi_N)$ with ε. Such estimates can be obtained with the use of inequality (6).

For a separable process $S(t), t \geq 0$ the modulus of continuity on the interval $[0, T]$ is defined in the following way:

$$\Delta_{h,T}(S(\cdot)) = \sup_{t',t'',\in[0,T],|t'-t''|\leq h} |S(t') - S(t'')|, h > 0.$$

Condition **B** implies that $E\Delta_{h,T}(S(\cdot)) < \infty$ for all $h > 0$. Note also that $E\Delta_{h,T}(S(\cdot))$ monotonically does not decrease in $h > 0$.

Let us assume the following condition:

C: $E\Delta_{h,T}(S_g(\cdot)) \to 0$ as $h \to 0$.

Under minimal assumption of separability of the process $S_g(t), t \geq 0$ condition **C** implies that this process is an a.s. continuous process. Therefore condition **A** holds.

Let τ_ε be ε-optimal stopping moment in the class $\mathcal{M}_{max,T}$. Then inequality (6) and the relation $\tau_\varepsilon \leq \tau_\varepsilon[\Pi_N] \leq \tau_\varepsilon + d(\Pi_N)$ imply that

$$\Phi_g(\mathcal{M}_{max,T}) - \Phi_g(\mathcal{M}_{\Pi_N,T}) \leq \varepsilon + ES_g(\tau_\varepsilon) - ES_g(\tau_\varepsilon[\Pi_N]) \leq \tag{8}$$

$$\leq \varepsilon + E\left|S_g(\tau_\varepsilon) - S_g\left(\tau_\varepsilon[\Pi_N]\right)\right| \leq \varepsilon + E\Delta_{d(\Pi_N),T}\big(S_g(\cdot)\big).$$

Since ε can be chosen arbitrary small relation (8) implies finally that

$$\Phi_g(\mathcal{M}_{max,T}) - \Phi_g(\mathcal{M}_{\Pi_N,T}) \leq E\,\Delta_{d(\Pi_N),T}\,\left(S_g\left(\cdot\right)\right). \tag{9}$$

Condition **C** implies that there exists $h = h_\varepsilon$ such that $E\Delta_{h_\varepsilon,T}(S_g(\cdot)) \leq \varepsilon$. Since $d(\Pi_N) \to 0$ as $N \to 0$ there exists $N = N_\varepsilon$ such that $d(\Pi_{N_\varepsilon}) \leq h_\varepsilon$. Let τ'_ε be an ε-optimal stopping moment in the class $\mathcal{M}_{\Pi_{N_\varepsilon},T}$. Then (9) implies that τ'_ε is a 2ε-optimal stopping moment in the class $\mathcal{M}_{max,T}$.

So, the problem is reduced to solving with respect to h the following inequality:

$$E\,\Delta_{h,T}\left(S_g(\cdot)\right) \leq \varepsilon. \tag{10}$$

In the next section we give explicit upper bounds for the expectation of the modulus of continuity $E\Delta_{h,T}(S_g(\cdot))$ in terms of moments of increments of the transformed pricing processes $S_g(t), t \geq 0$ and link explicitly the parameters h and ε.

3. Upper bounds for expectation of the modulus of continuity

Let $S(t), t \geq 0$ be a separable real-valued process. We assume that the following condition holds:

D: $E|S(t') - S(t'')|^m \leq H|t' - t''|^r, 0 \leq t', t'' \leq T$ for some $H > 0$ and $m, r > 1$.

We use estimates for tail probabilities for the modulus of continuity given in Gikhman and Skorokhod (1974). However, we estimate the expectation for the modulus of continuity and give the upper bounds with explicit constants due to detailed technical account at all steps of calculations.

Lemma 2. (Gikhman and Skorokhod (1974)). *Let $S(t), t \in [0, T]$ be a separable process, such that there exist nonnegative, nondecreasing function $g(h)$ and function $q(C, h)$, $C > 0$, $h > 0$, with*

$$P\{|S(t+h) - S(t)| > Cg(h)\} \leq q(C, h), \tag{11}$$

and

$$G = \sum_{n=0}^{\infty} g(T/2^n) < \infty, \quad Q(C) = \sum_{n=1}^{\infty} 2^n q(C, T/2^n) < \infty. \tag{12}$$

Then for each $\delta > 0$

$$P\{\sup_{0 \leq t', t'' \leq T} |S(t') - S(t'')| > \delta\} \leq Q(\delta/2G), \tag{13}$$

and for each $\varepsilon > 0$, $C > 0$

$$P\{\Delta_{\varepsilon, T}(S(\cdot)) > CG([\log_2 T/2\varepsilon])\} \leq Q([\log_2 T/2\varepsilon], C), \tag{14}$$

where

$$G(p) = \sum_{n=p}^{\infty} g(T/2^n), \quad Q(p, C) = \sum_{n=p}^{\infty} 2^n q(C, T/2^n). \tag{15}$$

Lemma 3. *Let condition* **D** *holds. Then $S(t)$, $t \in [0, T]$ is a.s. continuous process and for every $1 < r' < r$ and for every $0 < h \leq T$:*

$$E\,\Delta_{h,T}(S(\cdot)) \leq B_1 h^{\frac{r-1}{m}}, \tag{16}$$

where

$$B_1 = \frac{m}{m-1}\, 2^{\frac{r-1}{m}} (1 - 2^{-\frac{r'-1}{m}})^{-1}\, (1 - 2^{-(r-r')})^{-\frac{1}{m}} H^{\frac{1}{m}}\, T^{\frac{1}{m}}. \tag{17}$$

Proof. Apply Lemma 2 to the process $S(t)$. Fix a positive number $r' < r$ and set $g(h) = h^{\frac{r'-1}{m}}$. Find $G(p)$, $q(C, h)$ and $Q(p, C)$ defined in (15), (11) and (12).

$$G(p) = T^{\frac{r'-m}{m}}\, 2^{-\frac{p(r'-1)}{m}}\, (1 - 2^{-\frac{r'-1}{m}})^{-1}\,,$$

therefore

$$G\left([\log_2 \frac{T}{2\varepsilon}]\right) \le 2^{\frac{r'-1}{m}}\ \varepsilon^{\frac{r'-1}{m}}\ (1-2^{-\frac{r'-1}{m}})^{-1}.$$

We have by **D**:

$$P\{|S(t+h)-S(t)| > C\,g(h)\} \le$$

$$\frac{E\,|S(t+h)-S(t)|^m}{C^m\,g^m(h)} \le \frac{H}{C^m}\ h^{1+r-r'}\ := q(C,h),$$

and

$$Q(p,C) = \frac{H\ T^{1+r-r'}}{C^m}\ 2^{-p(r-r')}\ (1-2^{-(r-r')})^{-1}.$$

Then

$$Q\left([\log_2 \frac{T}{2\varepsilon}],\ C\right) \le C^{-m}\ T\,H\,\varepsilon^{r-r'}\ (1-2^{-(r-r')})^{-1}\cdot 2^{r-r'}.$$

By (14) we obtain

$$P\{\Delta_{\varepsilon,T}\big(S(\cdot)\big) > C\ 2^{\frac{r'-1}{m}}\ \varepsilon^{\frac{r'-1}{m}}\ (1-2^{-\frac{r'-1}{m}})^{-1}\} \le$$

$$\le C^{-m}\ T\ H\ \varepsilon^{r-r'}\ (1-2^{-(r-r')})^{-1}\cdot 2^{r-r'}.$$

Denote

$$\delta = C\ 2^{\frac{r'-1}{m}}\ \varepsilon^{\frac{r'-1}{m}}\ (1-2^{-\frac{r'-1}{m}})^{-1}.$$

Then

$$P\{\Delta_{\varepsilon,T}(S(\cdot)) > \delta\} \le \frac{TkH\varepsilon^{r-1}}{\delta^m},$$

where

$$k = 2^{r-1}\ (1-2^{-\frac{r'-1}{m}})^{-m}\ (1-2^{-(r-r')})^{-1}.$$

Next,

$$E\ \Delta_{h,T}\left(S(\cdot)\right) = \int_0^\infty P\{\Delta_{h,T}\left(S(\cdot)\right) > v\}dv \le \int_0^{(T\,k\,H)^{\frac{1}{m}}\,h^{\frac{r-1}{m}}} dv +$$

$$+ \int_{(T\,k\,H)^{\frac{1}{m}}\,h^{\frac{r-1}{m}}}^\infty \frac{T\,k\,H\,h^{r-1}}{v^m}\,dv = \frac{m}{m-1}\,(T\,k\,H)^{\frac{1}{m}}\ h^{\frac{r-1}{m}} = B_1\,h^{\frac{r-1}{m}},$$

where B_1 is given by (17). Inequality (16) is proved.

Finally, for a separable process $S(t)$ condition **D** implies continuity of the paths, see Gikhman and Skorokhod (1974). $\bigoplus$

Corollary. *Let condition* **D** *holds. Then for every* $0 < u < m,\ 0 < r' < r$

$$E(\sup_{0\le t',\,t''\le T} |\,S(t') - S(t'')\,|^u) \le \frac{m}{m-u}\ k_1\ H^{\frac{u}{m}}\ T^{\frac{ru}{m}}, \qquad (18)$$

where

$$k_1 = 2^u\,(1-2^{-\frac{r'-1}{m}})^{-u}\,(2^{r-r'}-1)^{-\frac{u}{m}}\,.$$

Proof. Use (13) for the process $S(t)$. Let again $g(h) = h^{\frac{r'-1}{m}},\ 0 < r' < r$, and $q(C,h) = \dfrac{H}{C^m}\,h^{1+r-r'}$. Then according to (12)

$$G = T^{\frac{r'-1}{m}}\,(1-2^{-\frac{r'-1}{m}})^{-1}, \quad Q(C) = \frac{H}{C^m}\,T^{1+r-r'}\,(2^{r-r'}-1)^{-1}.$$

Now,

$$Q(\delta/2G) = k_0\,\frac{H\,T^r}{\delta^m},$$

where

$$k_0 = 2^m\,(1-2^{-\frac{r'-1}{m}})^{-m}\,(2^{r-r'}-1)^{-1}.$$

By (13) we have

$$P\{\sup_{0\le t',\,t''\le T}|\,S(t')-S(t'')\,| > \delta\} \le \frac{k_0\,H\,T^r}{\delta^m},$$

and

$$E(\sup_{0\le t',\,t''\le T}|\,S(t')-S(t'')\,|^u) = \int_0^\infty P\{\sup_{0\le t',\,t''\le T}|\,S(t')-S(t'')\,| > v^{\frac{1}{u}}\}dv$$

$$\le \int_0^A dv \;+\; \int_A^\infty \frac{k_0\,H\,T^r}{v^{\frac{m}{u}}}\,dv.$$

Choose A from the condition $k_0\,H\,T^r\,A^{-\frac{m}{u}} = 1$. After straightforward calculation we obtain

$$E(\sup_{0\le t',\,t''\le T}|\,S(t')-S(t'')\,|^u) \le \frac{m}{m-u}(k_0\,H\,T^r)^{\frac{u}{m}} = \frac{m}{m-u}\,k_1\,H^{\frac{u}{m}}\,T^{\frac{ru}{m}},$$

and (18) is proved. $\bigoplus$

Lemma 3, applied to the transformed pricing processes $S_g(t)$, yields the explicit solution in (10) and links parameters h and ε.

For example we get by substituting the corresponding upper bound in (10) the stronger inequality $B_1 h^{\frac{r-1}{m}} \le \varepsilon$, which guarantees that $h_\varepsilon = (\varepsilon/B_1)^{\frac{m}{r-1}}$ is the solution of (10). In sequel, if a partition Π_{N_ε} is chosen in such a way that the maximal step $d(\Pi_{N_\varepsilon}) \le (\varepsilon/B_1)^{\frac{m}{r-1}}$ then any a ε-optimal stopping moment τ'_ε in the class $\mathcal{M}_{\Pi_{N_\varepsilon},T}$ will be a 2ε-optimal stopping moment in the class $\mathcal{M}_{max,T}$.

Sometimes it is not convenient to apply Lemma 1 to the transformed pricing process $S_g(t)$ and it would be better to have similar estimates given in terms of increments of the pricing process S_t itself. Such estimates can be obtained in the case of smoothed pricing functions.

Let again $S(t), t \geq 0$ be a separable real-valued process for which the condition **D** holds. Let also $G(x,t)$ be a measurable real-valued function defined on $R \times R^+$. We consider the transformed process $S_G(t) = G(S(t),t)$, $t \geq 0$.

In the case of pricing processes transformation function is $G(x,t) = e^{-R_t} g(x,t)$ and the transformed pricing process is $S_g(t) = e^{-R_t} g(S_t, t), t \geq 0$.

In general case we do not make any assumptions about structure of transformation function $G(x,t)$. We assume only the following smoothness condition, which links the order of smoothness for function $G(x,t)$ with the parameter r in condition **D**:

E: (a) $G(x,t)$ is absolutely continuous upon x for every fixed $t \geq 0$ and upon t for every fixed $x \in R$; (b) for every $x \in R$ function $|\frac{\partial G(x,t)}{\partial t}| \leq K_1 |x|^{p_1}$ for almost all $t \in [0,T]$ with respect to Lebesgue measure, where $K_1 > 0$ and $0 \leq p_1 < r$; (c) for every $t \in [0,T]$ function $|\frac{\partial G(x,t)}{\partial x}| \leq K_2 |x|^{p_2}$ for almost all $x \in R$ with respect to Lebesgue measure, where $K_2 > 0$ and $0 \leq p_2 < r-1$.

Condition **E** guarantees the existence of the moments of the order m for increments of the process $S(t)$. Since nonlinear character of transformation function $G(x,t)$ we need also the following condition:

F: $E\,|S(0)|^m < \infty$.

In Lemma 3 an additional parameter $1 < r' < r$ was involved. Here we need to involve another additional parameter $1 < q < r'$. Let denote:

$$B_2 = \frac{m}{m-q}\, 2^{\frac{m+r-2q}{m}} \left(1 - 2^{-\frac{r'-q}{m}}\right)^{-1} \left(1 - 2^{-\frac{r-r'}{q}}\right)^{-\frac{q}{m}} \times \tag{19}$$

$$\times T^{\frac{q}{m}} \{K_1^{\frac{m}{q}} M_1\,(p_1 m/q)\, T^{\frac{m-r}{q}} + K_2^{\frac{m}{q}} (M_2(p_2 m/(q-1)))^{\frac{q-1}{q}}\, H^{\frac{1}{q}}\}^{\frac{q}{m}},$$

where

$$M_1\,(u) = 2^{[u-1]_+} \cdot (E\,|S(0)|^u \,+\, H^{\frac{u}{m}}\, T^{\frac{ru}{m}}),$$

$$M_2\,(u) = 2^{[u-1]_+} \cdot \{E\,|S(0)|^u + \frac{m}{m-u} 2^u (1 - 2^{-\frac{r'-1}{m}})^{-u} (2^{r-r'} - 1)^{-\frac{u}{m}} H^{\frac{u}{m}} T^{\frac{ru}{m}}\}.$$

Lemma 4. *Let conditions* **D** *with* $m \geq r > 1$ *and* **E**–**F** *hold. Then for every* q *and* r' *such that* $p_2 + 1 < q < r' < r$, $p_1 \leq q$ *and every* $0 < h \leq T$:

$$E\,\Delta_{h,T}(S_G(\cdot)) \leq B_2 h^{\frac{r-q}{m}}.$$

Proof. Assumption E implies

$$|G(S_t\,,t) - G(S_s\,,s)| \,\leq\, |G(S_t\,,t) - G(S_t\,,s)| \,+\, |G(S_t\,,s) - G(S_s\,,s)| \,\leq$$

$$\leq\, K_1\,|S_t|^{p_1}\,|t-s| \,+\, K_2 \sup_{t\,\in\,[0,T]} |S_t|^{p_2}\,|S_t \,-\, S_s|.$$

Let q and r' satisfy the conditions of Lemma 4. Then

$$E|\,G(S_t\,,t)-G(S_s\,,s)|^{\frac{m}{q}}\;\le 2^{\frac{m}{q}-1}\,\{K_1^{\frac{m}{q}}\;|t-s|^{\frac{m}{q}}\;E\,|S_t|^{\frac{mp_1}{q}}+$$

$$+K_2^{\frac{m}{q}}(E\,|S_t\;-\;S_s\,|^m)^{\frac{1}{q}}\;(E\sup_{t\in[0,T]}|S_t|^{\frac{p_2m}{q-1}})^{\frac{q-1}{q}}\}. \tag{20}$$

Now, by **D** we have for $0\le u\le m$:

$$E|S_t|^u\;\le\;2^{[u-1]_+}\times\{E|S(0)|^u\;+\;(E|\,S(t)-S(0)|^m)^{\frac{u}{m}}\}\;\le\;M_1\,(u), \tag{21}$$

and by **D** and Corollary for $0\le u<m$:

$$E\sup_{t\,\in\,[0,T]}|S_t|^u\;\le\;2^{[u-1]_+}\times\{E\,\big|S(0)|^u\;+\;E\sup_{t\,\in\,[0,T]}|S(t)-S(0)\big|^u\}\;\le$$

$$\le\;2^{[u-1]_+}\times\{E|S(0)|^u\;+\;\frac{m}{m-u}\,k_1\,H^{\frac{u}{m}}\,T^{\frac{ru}{m}}\}\;=\;M_2\,(u). \tag{22}$$

¿From (20) – (22) we obtain

$$E|G(S_t\,,t)-G(S_s\,,s)|^{\frac{m}{q}}\;\le\;H_1|t-s|^{\frac{r}{q}}, \tag{23}$$

where

$$H_1=2^{\frac{m}{q}-1}\{K_1^{\frac{m}{q}}\;T^{\frac{m-r}{q}}\;M_1(mp_1/q)\;+\;K_2^{\frac{m}{q}}\;H^{\frac{1}{q}}[M_2(mp_2/(q-1))]^{\frac{q-1}{q}}\}.$$

Finally, by Lemma 3 and inequality (23) we get

$$E\,\Delta_{h,T}(S_G(\cdot))\le B_1\,(m/q,\,r/q,\,H_1)\,h^{(\frac{r}{q}-1)\,(\frac{m}{q})^{-1}}=B_1(m/q,\,r/q,\,H_1)\,h^{\frac{r-q}{m}}\,. \tag{24}$$

Here $B_1(m/q,\,r/q,\,H_1)$ is obtained from B_1, which is given in (17), by substitution m/q, r/q and H_1 instead of m, r and H, respectively; we substitute also in (17) r'/q instead of r'. We have

$$B_1(m/q,\,r/q,\,H_1)\;=\;\frac{m}{m-q}\,2^{\frac{r-q}{m}}(1-2^{-\frac{r'-q}{m}})^{-1}\times$$

$$\times(1-2^{-\frac{r-r'}{q}})^{-\frac{q}{m}}\;\times\;H_1^{\frac{q}{m}}\,T^{\frac{q}{m}}\;=\;B_2. \tag{25}$$

Now, (24) and (25) imply that

$$E\,\Delta_{h,T}\Big(S_G(\cdot)\Big)\;\le\;B_2\,h^{\frac{r-q}{m}}\,.$$

Lemma 4 is proved. $\bigoplus$

4. Skeleton approximations for the basic example

Let us illustrate the possible application of Lemmas 3 and 4 to the model where the pricing process S_t, $t \geq 0$ is given in the form:

$$S_t = S_0 \cdot \exp\{\int_0^t (a\,(u, I_u) - \frac{1}{2}\sigma(u)^2)\,du + \int_0^t \sigma(u)\,dw(u)\}, \quad t \geq 0,$$

where (a) $a(t, y)$ is a measurable real-valued functions defined on Z, (b) $\sigma(t) \geq 0$ is a measurable real-valued functions defined on R^+, (c) I_t, $t \geq 0$ is a measurable inhomogeneous in time Markov process, (d) $w(u)$, $u \geq 0$ is the Wiener process independent of process I_t, $t \geq 0$, (e) $Z_0 = (S_0, I_0)$ is a non-random value in Z.

In this case vector process $Z_t = (S_t, I_t)$, $t \geq 0$ is an inhomogeneous Markov process with the first component S_t, $t \geq 0$ is a continuous geometrical diffusion process controlled by process I_t, $t \geq 0$.

We assume the following condition:

$$\mathbf{G}\text{: (a) } A = \sup_{0 \leq t \leq T,\, y \in Y} |a(t, y) - \frac{1}{2}\,\sigma(t)^2| < \infty; \quad \text{(b) } B = \sup_{0 \leq t \leq T} \sigma(t) < \infty.$$

Lemma 5. *Let condition* **G** *holds. Then for any* $m > 2$

$$E\,|S_{t'} - S_{t''}|^m \leq H_m\,|t' - t''|^{m/2}, \quad 0 \leq t', \quad t'' \leq T,$$

where

$$H_m = \frac{1}{2}\,(2S_0\,e^{AT + \frac{1}{2}mB^2\,T}\,T^{-\frac{1}{2}})^m\,((e^{AT} - 1)^m + E\,|e^{BT^{\frac{1}{2}}N(0,1)} - 1|^m).$$

Proof. Fix $m > 2$ and denote $b(t, y) = a(t, y) - \frac{1}{2}\sigma(t)^2$, $0 \leq\ t \leq\ T, y \in Y$. We suppose that $S_0 > 0$. Then for every $t \in [0, T]$ $S_t > 0$ a.s. Fix $t \in [0, T]$ and positive h, such that $t + h \in [0, T]$. Consider the increment

$$|S_{t+h} - S_t| = S_t \cdot |\exp\{\int_t^{t+h} b(u, I_u)\,du\ + \int_t^{t+h} \sigma(u)\,dw(u)\} - 1|\ \leq$$

$$\leq\ S_t \cdot \{\exp(\int_t^{t+h} \sigma(u)\,dw(u)) \times (e^{Ah} - 1)\ +\ |\exp\Big(\int_t^{t+h} \sigma(u)\,dw(u)) - 1|\}.$$

Now, $S_t \leq S_0 e^{AT} \cdot \exp(\int_0^t \sigma(u)\,dw(u))$. Therefore

$$\frac{|S_{t+h} - S_t|}{S_0 e^{AT}}\ \leq\ \exp\{\int_0^{t+h} \sigma(u)\,dw(u)\}\ \times\ (e^{Ah} - 1)$$

$$+\exp\{\int_0^t \sigma(u)\,dw(u)\}\ \times\ |\exp\{\int_t^{t+h} \sigma(u)\,dw(u)\}-1|.$$

Then

$$E|\frac{S_{t+h}-S_t}{S_0e^{AT}}|^m\ \le\ 2^{m-1}\,(e^{Ah}-1)^m\ \cdot\ E\exp\{m\int_0^{t+h}\sigma(u)\,dw(u)\}+$$

$$+\,2^{m-1}\ \cdot\ E\exp\{m\int_0^t \sigma(u)\,dw(u)\}\ \cdot\ E|\ \exp\{\int_t^{t+h}\sigma(u)\,dw(u)\}-1\,|^m. \quad (26)$$

For each $t\in[\,0,T\,]$ we have

$$E\exp\{m\int_0^t \sigma(u)\,dw(u)\}=\exp\{\frac{m^2}{2}\cdot\int_0^t\sigma^2(u)\,du\}\ \le\ e^{\frac{m^2B^2T}{2}}. \quad (27)$$

The inequality $|e^{\alpha z}-1|\le|e^{\beta z}-1|,\ \ 0\le\alpha\le\beta,\ z\in R$, implies

$$E|\ \exp\{\int_t^{t+h}\sigma(u)\,dw(u)\}-1|^m\ =\ E|\ \exp\{(\int_t^{t+h}\sigma^2(u)\,du)^{\frac12}\times N(0,1)\}-1|^m\ \le$$

$$\le\ E|e^{B\sqrt{h}\cdot N(0,1)}-1|^m\ \le\ (\sqrt{h/T})^m\times\ E|e^{B\sqrt{T}\cdot N(0,1)}\ -\ 1|^m. \quad (28)$$

Here we used the inequality

$$|e^{hz}-1|\ \le\ \frac{h}{T}\,|e^{Tz}-1|,\quad 0<h\le T,\quad z\in R, \quad (29)$$

which follows from the convexity of the exponential function. From (26) – (29) we obtain finally

$$E|\frac{S_{t+h}-S_t}{S_0\,e^{AT}}|^m\ \le\ (\frac{h}{T})^{\frac{m}{2}}2^{m-1}\times e^{\frac12 m^2B^2T}\{\,(e^{AT}-1)^m+E\,|e^{B\sqrt{T}\cdot N(0,1)}-1|^m\,\},$$

and

$$E|S_{t+h}-S_t|^m\ \le\ H_m\,h^{\frac{m}{2}}.$$

This completes the proof. $\bigoplus$

So, condition **D** holds and Lemma 4 can be applied to the pricing process $S_t, t\ge0$ if condition **E** holds for the transformation function $G(x,t)=e^{-R_t}g(x,t)$.

Consider the case of standard American option. Here the transformation function

$$G(x,t)=e^{-\tilde{r}t}\,[x-K]_+,\ x\ge0,\ 0\le t\le T,$$

where $\tilde{r} > 0,\ K > 0$.

Let us apply Lemmas 4 and 5.

For $m > 2$ we have

$$E\,|S_{t'} - S_{t''}|^m \le H_m\,|t' - t''|^{m/2}, \quad 0 \le t', t'' \le T,$$

where H_m is given in Lemma 5. Thus **D** holds with $m > 2, r = \frac{m}{2}, H = H_m$.

Now, check the condition **E**. We have

$$|\frac{\partial G(x,t)}{\partial t}| \le \tilde{r},\ |\frac{\partial G(x,t)}{\partial x}| \le 1$$

(except the point $x = K$), therefore **E** holds with $p_1 = p_2 = 0,\ k_1 = \tilde{r},\ k_2 = 1$.

Fix q and r' such that $1 < q < r' < \frac{m}{2}$. By lemma 4 we have

$$E\,\Delta_{h,T}\,(S_G(\cdot)) \le B_2\,h^{\frac{1}{2}-\frac{q}{m}},$$

where B_2 is given by (19), with $r = \frac{m}{2},\ K_1 = \tilde{r},\ K_2 = 1,\ H = H_m,\ p_1 = p_2 = 0,\ M_1(0) = M_2(0) = 2$.

According to (9)

$$\Phi_g(\mathcal{M}_{max,T}) - \Phi_g(\mathcal{M}_{\Pi_N,T}) \le B_2\,d(\Pi_N)^{\frac{1}{2}-\frac{q}{m}} \le \varepsilon,$$

if $d(\Pi_N) \le (\varepsilon/B_2)^{\alpha}$, with $\alpha = (\frac{1}{2} - \frac{q}{m})^{-1}$.

To find $\Phi_g(\mathcal{M}_{\Pi_N,T})$ one can apply the results given papers Kukush and Silvestrov (2000a, 2000b). Let

$$\Pi_N = \{0 = t_0 < t_1 < ... < t_N = T\}\,.$$

In order to imbed the model in those considered in these papers one should consider the two component Markov chain $(S_n, I_n = (I'_n, I''_n))$, where

$$S_n = S_{t_n},\ I'_n = I_{t_n},\ I''_n = \exp\{\int\limits_{t_{n-1}}^{t_n} (a(u, I_u) - \frac{1}{2}\,\sigma(u)^2)du + \int\limits_{t_{n-1}}^{t_n} \sigma(u)dw(u)\}.$$

Let $r_k = \tilde{r}(t_{k+1} - t_k),\ k = 0, 1, ..., N-1,\ R_0 = 0,\ R_n = r_0 + r_1 + ... + r_{n-1},\ n = 1, 2, ..., N$.

The functional $\Phi_g(\tau)$ defined in (2) for $\tau \in \mathcal{M}_{\Pi_N,T}$ coincides with the functional

$$\Phi_g(\tau) = E\,e^{-R_\tau}\,[S_\tau - K]_+ \tag{30}$$

introduced in Kukush and Silvestrov (2000a, 2000b) for the discrete Markov chain (S_n, I_n).

It follows from the formulas, which define Markov chain (S_n, I_n) that the first component can be given in the following dynamical form $S_n = S_{n-1} \cdot I''_n$. Also it is obvious that component I_n is also a Markov chain and it's

transition probabilities depend only of the first component I'_n. That is why a conditions **A-C** used in Kukush and Silvestrov (2000a, 2000b) obviously hold. In particular the dynamical transition function $A(x,(y',y'')) = x \cdot y'$, which is derived from the formula $S_n = A(S_{n-1}, I_n) = S_{n-1} \cdot I''_n$, is convex and continuous in x for every (y', y'').

Assume additionally that

H: $D = \sup\limits_{0\le t\le T,\ y\in Y} a(t,y) < \tilde{r}$.

Condition condition implies that condition **D**, introduced in Kukush and Silvestrov (2000a, 2000b), holds with $a_n \equiv 1$ (recall that we consider the case of standard American option). Really, for each $x > 0$

$$\frac{1}{x} E\{S_{t_{n+1}}/S_{t_n} = x, I_{t_n} = y\} =$$

$$E\{\exp\{\int\limits_{t_n}^{t_{n+1}} (a(u,I_u) - \frac{1}{2}\,\sigma(u)^2)du + \int\limits_{t_n}^{t_{n+1}} \sigma(u)dw(u)\}/I_{t_n} = y\} =$$

$$E\{\exp\{\int\limits_{t_n}^{t_{n+1}} a(u,I_u)du\}/I_{t_n} = y\} \le e^{D\,(t_{n+1}-t_n)} < e^{r_n}.$$

Therefore Theorem 2 from Kukush and Silvestrov (2000a, 2000b) is applicable now, and the structure of $\tau_{opt} \in \mathcal{M}_{\Pi_N,T}$ for the functional (30) is given in that theorem.

Remark also that if to replace **H** by

I: $E\{a(u,I_u)/I_t = y\} \ge \tilde{r}$, for each $0 \le t \le u \le T,\ y \in Y$,

then for $x > 0, t < s$:

$$\frac{1}{x}\, E\{S_{t_{n+1}}/S_{t_n} = x, I_{t_n} = y\} =$$

$$E\{\exp\{\int\limits_{t_n}^{t_{n+1}} (a(u,I_u) - \frac{1}{2}\,\sigma(u)^2)du + \int\limits_{t_n}^{t_{n+1}} \sigma(u)dw(u)\}/I_{t_n} = y\} =$$

$$= E\{\exp\{\int\limits_{t_n}^{t_{n+1}} a(u,I_u)du\}/I_{t_n} = y\} \ge e^{\tilde{r}\,(s-t)},$$

and the process $V_t = e^{-\tilde{r}t}[S_t - K]_+,\ 0 \le t \le T$ is a submartingale (compare with the proof of Theorem 4 from Kukush and Silvestrov (2000a)). Therefore under **I** for the functional (2) in the class $\mathcal{M}_{max,T}$ we have $\tau_{opt} = T$.

The cases of American type options with linear convex pay-off functions and with general convex pay-off functions can be considered by similar way with the use of corresponding results given in Kukush and Silvestrov (2000a, 2000b).

Bibliography

1. Broadie, M. and Detemple, J., *Recent advances in numerical methods for pricing derivative securities*, In the book Numerical Methods in Finance, Ed. by L.G.G. Rogers and D. Talay, Cambridge University Press, (1998), 43-66.
2. Duffie, D., *Dynamical Asset Pricing Theory*, Princeton University Press, (1996).
3. Gikhman, I.I. and Skorokhod, A.V., *The Theory of Stochastic Processes*, Vol.1, Springer, (1974).
4. Karatzas, I. and Shreve, S.E., *Methods of Mathematical Finance*, Springer, (1998).
5. Kukush, A.G. and Silvestrov, D.S., *Optimal pricing of American type options with discrete time*, Research Report 2000-1, Department of Mathematics and Physics, Mälardalen University, (2000a).
6. Kukush, A.G. and Silvestrov, D.S., *Structure of optimal stopping strategies for American type options*, In the book Probabilistic Constrained Optimisation: Methodology and Applications, Ed. by S. Uryasev, Kluwer, (2000b), 173-185.
7. Kukush, A.G. and Silvestrov, D.S., *Optimal stopping strategies for American type options with discrete and continuous time*, In Proceedings of the Second International School on Actuarial and Financial Mathematics, Kyiv, 1999, Theory Stoch. Proces., **5(21)**, (1999), 71-79.
8. *Numerical Methods in Finance*, Ed. by L.G.G. Rogers and D.Talay, Cambridge University Press, (1998).
9. Øksendal, B., *Stochastic Differential Equations: An Introduction with Applications*, Springer, (1992).
10. Shiryaev, A.N., *Optimal Stopping Rules*, Springer, (1978).
11. Shiryaev, A.N., Kabanov, Yu.M., Kramkov, D.O., and Mel'nikov, A.V., *Toward a theory of pricing options of European and American types. I. Discrete time*, Theory Probab. Appl., **39**, (1994), 14-60.
12. Silvestrov, D.S., Galochkin, V.G. and Sibirtsev, V.G., *Algorithms and Programs for optimal Monte Carlo pricing of American type options*, In Proceedings of the Second International School on Actuarial and Financial Mathematics, Kyiv, 1999, Theory Stoch. Proces., **5(21)**, (1999), 175-187.

Department of Mechanics and Mathematics, Kyiv National Taras Shevchenko University, 252601 Kyiv, Ukraine.
E-mail: kuog@univ.kiev.ua
Department of Mathematics and Physics, Mälardalen University, SE-721 23 Västerås, Sweden.
E-mail: dmitrii.silvestrov@mdh.se

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.231-235

OLEKSANDR O. KURCHENKO

ESTIMATION FOR THE FUNCTION OF A TIME DEFORMATION IN THE MODEL OF THE STATIONARY REDUCTION

Let $\delta(t)$,$t \in R^1$ be a stationary Gaussian random process with zero mean and the known function of correlation. The non-parametric consistent in mean square estimate of the function of the time deformation $\Phi(t)$, $t \in T$ by the observations of Gaussian random process $Z(t) = \delta(\Phi(t))$, $t \in T$ is constructed by means of Baxter sums of the random process $Z(t)$, $t \in T$.

2000 *Mathematics Subject Classifications.* 60G15, 60G17, 62G05.

Key words and phrases: Estimation, Baxter sums, Gaussian process, time deformation

1. Introduction

Often the problem of the reduction of the non-stationary process $Z(t)$, $t \in T$ to the stationary process appears in most applications dealing with non-stationary processes. To this end the non-stationary random process $Z(t)$, $t \in T$ with the function of correlation $r(t,s) = EZ(t)Z(s)$ is modeled in the form

$$Z(t) = \delta(\Phi(t)),\ t \in T \tag{1}$$

where $\Phi : T \to R^1$ is a time deformation, $\delta(s),\ s \in R^1$ is a stationary random process with zero mean and the function of correlation $E\delta(s)\delta(t) = R(s-t),\quad s,\ t \in R^1$.

The representation (1) is possible if and only if

$$r(s,t) = R(\Phi(s) - \Phi(t)),\quad s,\ t \in T. \tag{2}$$

In the article by Perrin and Senoussi (1999) the criterion of the stationary reduction was established. The problem of the estimation of the time deformation $\Phi(t)$, $t \in T$ by the observations of $Z(t), t \in T$ appears in the model of stationary reduction. In this item we construct the non-parametric consistent in mean square estimate of the time deformation by means of Baxter sums of the random process $Z(t), t \in [0,1]$.

2. The statement of the problem of estimation

Let $\delta(s),\ s \in R$ be a stationary Gaussian process with zero mean and the function of correlation $R(s),\ s \in R^1$. We need the following assumptions:

(R1) for some $\gamma \in (0,2)$, and $A > 0$

$$R(\tau) = 1 - A|\tau|^{\gamma} + o(|\tau|^{\gamma}),\ \ \tau \to 0;$$

(R2) for some $C > 0$ when $\tau \neq 0$ the second derivative $R''(\tau)$ exists and

$$|R''(\tau)| \leq \frac{C}{|\tau|^{2-\gamma}}, |\tau| \leq 1,\ \tau \neq 0;$$

(R3) for some $\beta > \gamma$

$$R(\tau) - 1 + A|\tau|^{\gamma} = O(|\tau|^{\beta}),\ \tau \to 0.$$

Let we assume with respect to the function of the time deformation $\Phi : [0,b] \to R$, where $b > 1$, the following properties:

(Φ) $\Phi(0) = 0,\ \ \Phi \in C^{(2)}([0,1])$ and $\Phi'(t) > 0$ for all $t \in [0,1]$.

The non-parametric estimate of the function $\Phi(t),\ t \in [0,1]$ must be constructed from one realization of the random process $Z(t) = \delta(\Phi(t)),\ t \in [0,1]$ observed in discrete times $\frac{k}{n},\ k = 0,1,\ldots,n,\ n \geq 1$. In the item by Perrin (1999) this problem was studied in the case of more restrictive assumptions.

3. Limit of Baxter sums

Let $Z(t),\ t \in [0,1]$ be Gaussian random process with zero mean and the function of correlation $r(s,t),\ s,t \in [0,1]$. Let $\lambda_n = \{t_{nk} = \frac{k}{n} : 0 \leq k \leq n\}$ be the equidistant partitioning of the segment [0,1]. We set for $k = 1,2,\ldots,n$

$$Z_{nk} = Z\left(\frac{k}{n}\right) - Z\left(\frac{k-1}{n}\right).$$

For real number $p > 1$ and $x \in [0,1]$ we consider Baxter sums $S_n(x)$ of Z along the partitioning $\lambda_n(x) = \{0, 1/n, 2/n, \ldots, [nx]/n\}$ (here [nx] is the greatest integer smaller than or equal to nx) as follows:

$$S_n(x) = \sum_{k=1}^{[nx]} |Z_{nk}|^p + (nx - [nx])|Z_{n([nx]+1)}|^p,\ \ n \geq 1.$$

Theorem 3.1. *Let for Gaussian random process $Z(t),\ t \in [0,b]$ $(b > 1)$ the following conditions are fulfilled:*

1) for some $\gamma \in (0,2)$ *and some positive function* $\alpha : [0,1] \to (0,\infty)$

$$\frac{E(Z(t+h)-Z(t))^2}{h^\gamma} \to \alpha(t)$$

uniformly on $[0,1]$ *when* $h \to +0$;

2) it exist $L > 0$ *and* $\gamma \in (0,2)$ *such that*

$$\left|\frac{\partial^2 r(s,t)}{\partial s \partial t}\right| \le \frac{L}{|s-t|^{2-\gamma}}, t \ne s.$$

Let $p = \frac{2}{\gamma}$. *Then for each* $x \in [0,1]$

$$S_n(x) \to S(x) = \sqrt{\frac{2^p}{\pi}}\Gamma\left(\frac{p+1}{2}\right)\int_0^x \sqrt{\alpha^p(t)}dt$$

in mean square as $n \to \infty$. *Furthermore,*

$$Var\, S_n(x) = \begin{cases} O\left(\frac{1}{n}\right) & if \quad \gamma \in (0, \frac{3}{2}) \\ O\left(\frac{\log n}{n}\right) & if \quad \gamma = \frac{3}{2} \\ O\left(\frac{1}{n^{4-2\gamma}}\right) & if \quad \gamma \in (\frac{3}{2}, 2). \end{cases} \tag{3}$$

uniformly on $x \in [0,1]$ *as* $n \to \infty$.

This theorem follows from corollary 2 of the item by Kurchenko (1999).

Theorem 3.2 *Let the* $\{n(k) : k \ge 1\}$ *be the subsequence of the natural numbers,* $x \in [0,1]$ *and the series*

$$\sum_{k=1}^{\infty} VarS_{n(k)}(x) \tag{4}$$

converges. Then

$$S_{n(k)}(x) \to S(x)$$

with probability one as $n \to \infty$.

Proof. From the convergence of the series (4) by Chebyshev inequality and Borel-Kantelli lemma it follows that

$$S_{n(k)}(x) - ES_{n(k)}(x) \to 0$$

with probability one as $k \to \infty$. But $ES_n(x) \to S(x)$ as $n \to \infty$. Hence, $S_{n(k)}(x) \to S(x)$ with probability one as $k \to \infty$. The theorem is proved.

In particular, for each $x \in [0,1]$ $S_{2^k}(x) \to S(x)$ with probability one as $k \to \infty$.

In the following theorem the rate of convergence in mean square of the sequence of Baxter sums $\{S_n(x) : n \ge 1\}$ is estimated uniformly on $x \in [0,1]$.

Theorem 3.3. *Let the conditions of the theorem 3.1 are true and the following conditions are fulfilled:*

$$1)\ (r(t+h,t+h)-2r(t+h,t)+r(t,t))^{\frac{p}{2}}-\alpha^{\frac{p}{2}}(t)h=O(h^{3/2})$$

uniformly on $t\in[0,1]$ *as* $h\to 0+$*;*

2) the function $\alpha^{p/2}(t),\ t\in[0,1]$ *belongs to Lipshits space* $Lip_{1/2}([0,1])$*. Then*

$$\sqrt{E(S_n(x)-S(x))^2}=\begin{cases} O\left(\frac{1}{\sqrt{n}}\right) & if\ \ \gamma\in(0,\frac{3}{2}) \\ O\left(\sqrt{\frac{\log n}{n}}\right) & if\ \ \gamma=\frac{3}{2} \\ O\left(\frac{1}{n^{2-\gamma}}\right) & if\ \ \gamma\in(\frac{3}{2},2). \end{cases} \tag{5}$$

uniformly on $x\in[0,1]$ *as* $n\to\infty$*.*

Proof. For each $x\in[0,1]$ by virtue of the triangle inequality

$$\sqrt{E(S_n(x)-S(x))^2}\le(\mathrm{Var}S_n(x))^{1/2}+|ES_n(x)-S(x)|.$$

From the assumptions 1), 2) it follows that

$$ES_n(x)-S(x)=O\left(\frac{1}{\sqrt{n}}\right)$$

uniformly on $x\in[0,1]$ as $n\to\infty$. Taking into account (3) we obtain (5). The theorem is proved.

4. The estimation of the function of a time deformation

Now we return to the problem of the estimation of the function of a time deformation $\Phi(x)$, $x\in[0,1]$ in the model (1), (2). Let the assumptions (R1), (R2), (Φ) are true. We go to verify the conditions of the theorem 3.1:

$$\lim_{h\to 0+}\frac{r(t+h,t+h)-2r(t,t+h)+r(t,t)}{h^\gamma}=$$

$$\lim_{h\to 0+}\frac{2A|\Phi(t+h)-\Phi(t)|^\gamma+o(|\Phi(t+h)-\Phi(t)|^\gamma)}{h^\gamma}=2A(\Phi'(t))^\gamma$$

uniformly on $[0,1]$. Thus the condition 1) of the theorem 3.1 is fulfilled for

$$\alpha(t)=2A(\Phi'(t))^\gamma,\quad t\in[0,1].$$

We consider the following estimate for the time deformation $\Phi(x)$, $x\in[0,1]$ by the observations of the random process $Z(t)=\delta(\Phi(t))$ in the points $k/n, k=0,1,\dots,n;\ n\ge 1$:

$$\Phi_n(x)=\frac{\sqrt{\pi}}{(4A)^{1/\gamma}\left(\Gamma\left(\frac{1}{\gamma}+\frac{1}{2}\right)\right)}\left(\sum_{k=1}^{[nx]}|Z_{nk}|^{2/\gamma}+(nx-[nx])|Z_{n([nx]+1)}|^{2/\gamma}\right).$$

From the theorem 3.1 and theorem 3.3 it follows

Theorem 4.1. *Let the assumptions (R1), (R2), (Φ) are fulfilled. Then the random variable $\Phi_n(x)$ for each $x \in [0,1]$ is the consistent in mean square estimate of $\Phi(x)$. If additionally the assumption (R3) is true and $\beta - \gamma \geq 1/2$ then*

$$\sup_{x\in[0,1]} \sqrt{E(\Phi_n(x) - \Phi(x))^2} =$$

$$\begin{cases} O\left(\frac{1}{\sqrt{n}}\right) & \text{if } \gamma \in (0, \frac{3}{2}) \\ O\left(\sqrt{\frac{\log n}{n}}\right) & \text{if } \gamma = \frac{3}{2} \\ O\left(\frac{1}{n^{2-\gamma}}\right) & \text{if } \gamma \in (\frac{3}{2}, 2). \end{cases}$$

It should be noted that for each $x \in [0,1]$ $\Phi_{2^n}(x) \to \Phi(x)$ with probability one as $n \to \infty$. For proof it is sufficient to observe that the series $\sum\limits_{n=1}^{\infty} \mathrm{Var}\Phi_{2^n}(x)$ converges and to make use of theorem 3.2.

5. Example

Let $X(t)$, $t \in R^1$ be a stationary Gaussian random process with zero mean and the function of correlation

$$R(u) = \mathrm{ch}(Hu) - 2^{2H-1}(\mathrm{sh}(|u|/2))^{2H}, \quad u \in R^1,$$

where $H \in (0,1)$ (Example 3.1.1 in Perrin and Senoussi (1999)). For this random process the assumptions (R1), (R2), (R3) are true for $\gamma = 2H$, $A = 1/2$, $C = 7$, $\beta = 2$.

Bibliography

1. Perrin, O., Senoussi, R., *Reducing non-stationary stochastic processes to stationarity by a time deformation,* Statistics and Probabilities Letters, **43**, (1999), 393-397.

2. Perrin, O., *Quadratic variation for Gaussian processes and application to time deformation,* Stochastic Processes and their applications, **82**, (1999), 293-305.

3. Kurchenko, O., *Convergence of the $F-$variation for Gaussian random fields,* Theory Probab. and Math. Statist., **60**, (1999), 98-108.

Departement of Mecanics and Mathematics, Kyiv Taras Schevchenko University, 01033, Kyiv, Ukraine.
E-mail: akurchenko@yahoo.fr

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.236-244

OLENA LYASHENKO

NONLINEAR INPUT-OUTPUT BALANCE FOR OPEN ECONOMY WITH HAUTHAKKER-JOHANSEN PRODUCTION FUNCTIONS

We consider the case of multibranch open economy. The way of researc is construstion and analysis of optimizing problems at balance restristions. The criterion is maximization of total output. The basis of mathematical modelling is nonlinear input-output balance of open economy with Hautakker-Johansen production function.

2000 *Mathematics Subject Classifications.* 91B66

Key words and phrases. Nonlinear input-output balance, Hautakker-Johansen production function, optimization, Khun-Tucker theorem.

1. Introduction

Now as the basic way of the description of technological structure description the inter-branch scheme input - output by V.Leontiev is used [1]. For the description of structural changes in industrial system the pure branches with many technologies are considered, which intensity of use is limited to capacities [2,3]. Natural generalization of Leontiev scheme became model with continual set of technologies, that is offered in works H.Hautakker and L.Johansen [4,5]. In [6] such description was used for study of structural changes in economy. The theory of production functions given by distribution of capacities on technologies, is advanced in [6-8].

2. Hautakker-Johansen production function

We shall describe shortly Hautakker-Johansen model, following [2,3]. Let pure branch produces a homogeneous product, using n kinds of production factors (resources). In branch there are different technological processes of manufacture, each of which is set by a vector $a = (a_1,..., a_n)$ norms of expenses of production factors on output of a unit of production. Intensity

of use of technologies $u(a)$ are limited to capacities, available in branch. It is supposed, that the technology within the framework of the given capacity is chosen at a stage of construction of branch. Let's designate $\mu(\nu)$ total capacity of technologies, which vectors of expenses norms belong to some set in non-negative orthant of Euclidean space R_+^n. The measure $\mu(\nu)$ is given on R_+^n and is called the distribution of capacity on technologies.

Then capacity of branch

$$M = \int\limits_{R_+^n} \mu\,(da)\,, \tag{1}$$

flows of production factors ensuring this capacity,

$$L_k = \int\limits_{R_+^n} a_k \mu\,(da)\,,\ k = 1, ..., n,$$

or in a vector form

$$L = \int\limits_{R_+^n} a\mu\,(da)\,,$$

where $L=(L_1,...,L_n)$ is a vector of available industrial resources.

If the flows $l_k<L_k$, total output Y of branch is less than capacity M. The value Y depends on economic mechanisms of regulation of manufacture and distribution of resources. It is supposed usually, that these mechanisms are completely competitive markets ensuring effective distribution of resources.

It is supposed, that the measure $\mu(\nu)$ is non-negative function, and function of capacities loading $u\,(a)$, $0 \le u\,(a) \le 1$, is measurable on Lebesgue and is determined on orthant R_+^n.

The task of optimal distribution of industrial resources flow $l=(l_1,...,l_n)$, come in the branch with the purpose of branch output maximization looks like

$$\int\limits_{R_+^n} u\,(a)\,\mu\,(da) \to \max_{u(a)},$$

$$\int\limits_{R_+^n} au\,(a)\,\mu\,(da) \le l, \tag{2}$$

$$0 \le u\,(a) \le 1.$$

Production function $F(l)$ in [2,3] is called the function which compares to a vector l_0 the greatest possible output of branch in a problem of optimal distribution of resources (3). The appointed production function has the

basic properties, that is postulated neo-classic theory (continuity, monotony, concavity) usually, and satisfies to the law of "ultimate output" [9].

According to (3) to calculate one value of production function, it is necessary to solve a problem of optimization in functional space. With the purpose of more simple and convenient macrodescription of branch in [2,3] it is offered to use profit function

$$\Pi\left(p_0,p\right)=\int\limits_{R_+^n} max\left(0;p_0-p\cdot a\right)\mu\left(da\right), \tag{3}$$

where p_0 is a price of output production; $p=(p_1,...,\ p_n)\geq 0$ is a price of production resources; $\Pi(p_0,p)$ is a total profit of branch.

Production function $F(l)$ and profit function $\Pi(p_0,p)$ are dual and are connected by Legendre transformation [7,8]:

$$\Pi\left(p_0,p\right)=\sup_{l\geq 0}\left[p_0F\left(l\right)-p\cdot l\right],\ F\left(l\right)=\inf_{p\geq 0}\frac{1}{p_0}\left[\Pi\left(p_0,p\right)+p\cdot l\right]. \tag{4}$$

The functions of supply and demand are expressed through the profit function:

$$g_0\left(p_0,p\right)=\frac{\partial\Pi\left(p_0,p\right)}{\partial p_0},\quad g_k\left(p_0,p\right)=\frac{\partial\Pi\left(p_0,p\right)}{\partial p_k},\quad k=1,...,n. \tag{5}$$

Thus

$$g_0(p_0,p)\equiv F(g_1(p_0,p),...,g_n(p_0,p)).$$

If the measure $\mu(\cdot)$ is absolutely continuous on orthant R_+^n and $\varphi(a)$ is its density, then the problem (3) will be transformed as

$$\begin{gathered} Y=\int\limits_{R_+^n} u\left(a\right)\varphi\left(a\right)da\rightarrow\max_{u(a)},\\ \int\limits_{R_+^n} au\left(a\right)\varphi\left(a\right)da\leq l, \\ 0\leq u\left(a\right)\leq 1. \end{gathered} \tag{6}$$

In [7] the statement similar to generalized Neuman-Pirson lemma on existence and structure of the solution of problem (6) is proved. It is shown, that there are such prices $p_0>0$, $p=(p_1,...,p_n)\geq 0$, at that the solution of the problem (6) looks like

$$u(a)=1, \text{ if } p_0 - p \cdot a > 0,$$

$$u(a)=0, \text{ if } p_0 - p \cdot a < 0.$$

Prices p_0, p are dual variables of formulated optimization problem. Thus, the market mechanisms of the perfect competition provide the maximal release of branch Y at given l; under these conditions the pure branch can be described by production function $Y=F(l)$ or supply function $g_0(p_0, p)$ and demand functions $g_i(p_0, p)$, $i=1,\ldots, n$.

3. The model of production

Let's consider one more description of production function using distribution of capacities on technologies [10]. For this purpose we shall consider the model of production, which we shall write down as the following problem of mathematical (generally nonlinear) programming

$$\begin{gathered} f(x) \to max, \\ g(x) \le b, \\ x \in T_X, \end{gathered} \tag{7}$$

where x is non-negative m-dimensional vector of output (or in other interpretation: x is a vector of intensities of technologies of complex manufacture), $T_X \subseteq R_t^m$, b is n-dimensional vector of available and necessary resources for manufacture (in further it is possible to consider this vector as strictly positive), $g(x)$ is a n-dimensional vector of expenses of resources on manufacture of a vector of products x ($0 \le g_i(x) \le b_i$, $i=1,\ldots,n$), $f(x)$ is the income of realization (or release) of vector of production x ($f(x) \ge 0$).

The problem (7) (as a problem of rational conducting of economy) consists in maximization of the income at available volumes of production factors (resources) within the framework of existing technological opportunities specified by a vector $g(x)$ and set of technologically allowable releases T_X. It is supposed logically, that $g_i(0)=0$, $i=1,\ldots,n$, i.e. the resources are used only then, when manufacture functions. Also it is supposed, that the set T_X is convex, the function $f(x)$ is concave (convex upwards), functions $g_i(x)$, $i=1,\ldots,n$ are convex. Thus, the problem (7) is a problem of convex programming. As $b_i>0$ (if even one $b_i=0$, the problem (7) has a zero vector as a solution, i.e. any manufacture in such situation does not exist), the problem (7) has the regular allowable solution, for which are carried out Sleuter's conditions of a regularity: $g_i(0)< b_i$, $i=1,\ldots,n$.

If the zero vector does not belong to set T_X, all the same it is required, that this set should be regular.

The designations are entered:

$X(b) = \{x \in T_X / g(x) \le b\}$ is allowable set of problem (7);

$B = \left\{b \in R_+^m / X(b) \neq \emptyset\right\}$ is a set of vector parameters b, at which the problem (7) has the allowable decisions;

$F(b) = \max\limits_{x \in X(b)} f(x), b \in B$, is a value of problem (7);

$\Lambda^*(b) = \left\{\lambda^* \in R_+^m / F(b) \geq L(x, \lambda^*) = f(x) + \lambda^* \cdot (b - g(x)) \, \forall x \in T_X\right\}$ is a set of Khun-Tucker's vectors (optimal dual variables) for problem (7) ($L(x,\lambda)$ is a regular Lagrange function);

$\bar{\partial} F(b) = \left\{\theta \in R^m / F\left(\tilde{b}\right) \leq F(b) + \theta \cdot \left(\tilde{b} - b\right) \forall \tilde{b} \in B\right\}$ is a superdifferential of function $F(b)$ in a point $b \in B$ [11].

At the made assumptions the set $X(b)$ is not empty and compact at everyone fixed $b \geq 0$, therefore according to the known Weierschtrass theorems the problem (7) has solution, and $F(b) < +\infty$. According to the Khun-Tucker's theorem for each solution $x^*(b) \in X(b)$ there will be such Khun-Tucker's vector $\lambda^*(b) \in \Lambda^*(b)$, that the point $(x^*(b), \lambda^*(b))$ will be a saddle point for regular Lagrange function $L(x,\lambda)$, i.e. the values of the primal problem (7) and dual to it will coincide.

The problem (7) is a general model of production function. This function is described implicitly and is a mapping of the set of allowable resources to the set of effective releases at the given technologies (generally variable and nonlinear).

The following statement follows from the theory of convex programming [12].

Theorem. *Let for the problem of convex programming (7) the earlier made assumptions are right. Then the function of values F(b) has such properties:*

1) F(b) is finite, concave and monotone non-decreasing at all $b \in B$;

2) F(b) is continuous at each internal point $b \in B$;

3) $\bar{\partial} F(b) = \Lambda^*(b)$ at all $b \in B$;

4) F(b) has derivative at any direction $h \in R^n$ in each internal point $b \in B$, and

$$F'(b, h) = \min_{\lambda^*(b) \in \Lambda^*(b)} \lambda^*(b) \cdot h;$$

5) F(b) is differentiable at internal point $b \in B$ if and only if Khun-Tucker's vector of a problem (7) is unique, i.e. if $\Lambda^*(b) = \{\lambda^*(b)\}$. Thus a gradient $F'(b) = \lambda^*(b)$;

6) if $\lambda_i^*(b) = 0$ at some $\lambda^*(b) \in \Lambda^*(b)$, then $F\left(b + \alpha e^{(i)}\right) = F(b)$ at every $\alpha > 0$, where $e^{(i)}$ is the i-th single ort at R_+^n;

7) if $\lambda_i^*(b) > 0$ at every $\lambda^*(b) \in \Lambda^*(b)$, then $F\left(b + \alpha e^{(i)}\right) > F(b)$ at all $\alpha > 0$.

From the theorem follows, that in addition to finiteness, concavity and monotonous non-decrease, the function $F(b)$ will be piecewise smooth in

general case, in particular at linear functions $f(x)$, $g_i(x)$, i=1,...,m, it will be piecewise linear [13]. The last property essentially distinguishes function $F(b)$ from neoclassic production functions, which are always assumed smooth.

4. The case of multibranch open economy

Below branch production functions given by distributions of capacities on technologies, we use for the description of open economy with the help of Leontiev scheme "input-output". Let's consider N pure branches, each of which is described by production function

$$Y_i = F_i(X_1^i, ...X_N^i, l_1^i, ..., l_n^i)\,,\ i=1,...,N.$$

Here X_j^i, i,j=1,...,N, are production inputs of branch j on output of branch i (material inputs); l_k^i, i=1,...,N, k=1,...,n, are inputs of a primary resource k on output of branch i (resource inputs). Besides that, let l_k k=1,...,n, is given quantity of resources; X_i^0, i=1,...,N, is output of final production i for internal consumption; $I_i, i \in M_I$ is import of production i, p_i^I is its price, M_I is given set of imported products; $E_i, i \in M_E$ is export of production i, p_i^E is its price, M_E is given set of exported products; $M_I \cup M_E \subseteq M = \{1, ..., N\}\,,\ M_I \cap M_E = \emptyset$; K=$\{1,...,N\}$.

Balance model by scheme "input - output" for open economy with non-negative export-import balance is written in the form:

$$\begin{gathered}
Y_i + I_i \geq \sum_{j \in M} X_j^i + X_i^0 + E_i,\ i \in M, \\
\sum_{i \in M} l_k^i \leq l_k, \quad k \in K, \\
\sum_{i \in M_E} p_i^E E_i \geq \sum_{i \in M_I} p_i^I I_i, \\
Y_i = F_i(X_1^i, ..., X_N^i, l_1^i, ..., l_n^i), i \in M, \qquad (8) \\
X_i^0 \geq 0, \quad X_j^i \geq 0, \quad l_k^i \geq 0, \quad i,j \in M, \quad k \in K, \\
I_i \geq 0, \quad i \in M_I; \quad E_i \geq 0, \quad i \in M_E.
\end{gathered}$$

It is possible to name model (8) as nonlinear input-output balance for open economy by analogy with [2,3].

With the purpose of further research model (8) shall copy as:

$$F_i\left(X^i, l^i\right) - \sum_{j \in M} X_j^i - Y_i^1 \geq 0, i \in M_1; \qquad (9)$$

$$F_i\left(X^i, l^i\right) - \sum_{j \in M} X_j^i + Y_i^2 \geq 0, i \in M_2; \tag{10}$$

$$l - \sum_{j \in M} l^j \geq 0; \tag{11}$$

$$p^E \cdot E - p^I \cdot I \geq 0; \tag{12}$$

$$X_i^0 \geq 0, \quad X_j^i \geq 0, \quad l^j \geq 0, \quad i, j \in M; \tag{13}$$

$$I_i \geq 0, \quad i \in M_I; \quad E_i \geq 0, \quad i \in M_E,$$

where

$$Y_i^1 = X_i^0 + E_i - I_i > 0, \quad i \in M_1; \tag{14}$$

$$Y_i^2 = -\left(X_i^0 + E_i - I_i\right) \geq 0, \quad i \in M_2.$$

Here all set of products (branches) is divided into two disjoint subsets M_1 and M_2 by a rule (14): $M_1 \cup M_2 = M$, $M_1 \cap M_2 = \emptyset$, and

$$X^i = \left(X_1^i, ..., X_N^i\right), \quad l^i = \left(l_1^i, ..., l_N^i\right), \quad l = (l_1, ..., l_n),$$

$$I = \left(I_i, i \in M_I\right), \quad E = \left(E_i, i \in M_E\right),$$

$$p^I = \left(p_i^I, i \in M_I\right), \quad p^E = \left(p_i^E, i \in M_E\right)$$

are vectors of appropriate dimension.

Correlations (8) - (10) determine the set $E(l)$ of final outputs of products $X^0 = (X_1^0, ..., X_N^0)$. From concavity of branch production functions follows, that the set $E(l)$ is convex. Besides that if $X^0 \in E(l)$ and $0 \leq Y^0 \leq X^0$, then $Y^0 \in E(l)$.

Let's assume, that the economy of industrial branches is productive, that is to say there are such non-negative vectors $\left\{Z^1, ..., Z^N, L^1, ..., L^N\right\}$, that $F_i\left(Z^i, L^i\right) - \sum\limits_{j \in M} Z_j^i > 0$, $i = 1, ..., N$. Let's assume also, that the conditions of efficiency of open economy are satisfied, that is to say for considered X^0, I, E, which are given, the sufficient conditions of existence of non-negative solutions X^i, l^i, i=1,,N, of inequalities system (8) - (10) are satisfied. The hyperplane $p^E \cdot E = p^I \cdot I$ passes through a point of zero position of export and import (E=0, I=0), which is admissible for productive open economy.

Let's choose number t so that $0< t <1$ and $t \sum\limits_{j\in M} L^j < l$. Such t exists, if $l>0$. From concavity of production functions $F_i(X^i,\ l^i)$ and from efficiency of considered group of branches follows, that at $l>0$ will be

$$F_i\left(tZ^i, tL^i\right) - \sum_{j\in M} tZ_j^i \geq t\left[F_i\left(Z^i, L^i\right) - \sum_{j\in M} Z_j^i\right] > 0.$$

Thus, at $l>0$ restrictions (8) - (13) satisfy to Slater condition (see [14]).

The mutual deliveries of branches and distribution of primary resources between them are determined by economic mechanisms, working in industrial system. The economic activity is an aspiration of the agent in system of the usual relations to use resources with the greatest benefit, which he can dispose of. Agreeing, competing, submitting, the agents establish system of the formal and informal arrangements - balance - how to divide common benefit. If thus the agents take all possible benefit from resources, which they dispose of, their economic activity can be simulated by a problem of optimal distribution of resources [15].

Bibliography

1. Leontiev, V.V. *Interbranch economy.* M.: Publishing house "Economics", (1997), 479 p. (in Russian).

2. Petrov, A.A. and Shananin, A.A. *Economic Mechanisms and the Problem of Input-Output Model Aggregation*, Mathematical modeling, (1993), **5**, No.9, 18-42 (in Russian).

3. Petrov, A.A., Pospelov, I.G. and Shananin, A.A. *Experience of Mathematical Modeling of Economy*, M.: Energoatomizdat, (1996), 544 p. (in Russian).

4. Hautakker, H.S. *The Pareto-Distribution and the Cobb-Douglas Production Function in Activity Analysis*, Rev. Econ. Stud. (1955/56), Vol.**23** (2), No.60, 27-31.

5. Johansen, L. *Outline of an Approach to Production Studies* Memorandum from inst. of economics. Univ. of Oslo, 28 April, (1969), 68 p.

6. Petrov, A.A. and Pospelov I.G. *The System Analysis of Developing Economy: To the Theory of Production Functions*, News of Academy of Sciences of the USSR. Technical Cybernetics. (1979), No.2, 18-27 (in Russian).

7. Shananin, A.A. *One Class of Production Functions Investigation arising at Macrodescription of Economic Systems*, Zhurnal Vychislit. Matem. I Mat. Fiziki. (1984), Vol.**24**, No.12, 1799-1811 (in Russian).

8. Shananin, A.A. *Research of One Class of Profit Functions Arising at Macrodescription of Economic Systems*, Zhurnal Vychislit. Matem. I Mat. Fiziki. (1985), vol.**25**, No.1, 53-65 (in Russian).

9. Johansen, L. *Production Functions*, Amsterdam - London: North Holland Co., (1972), 274 p.

10. Grigorkiv, V.S. *Generalized Linearly Homogeneous Production Functions*, Cybernetics and system analysis, (1999), No.5, 124-132 (in Russian).

11. Dem'yanov, V.F.and Vasiliev, L.V. *Nondifferentiable Optimization*, M.: Nauka, 1981. - 384 p.(in Russian).

12. Sukharev, A.G., Timohov, A.V. and Fedorov V.V. *Course of Optimization Methods*, M.: Nauka, (1986,) - 328 p.(in Russian).

13. Ashmanov, S.A. *Linear Programming*, M.:Nauka, (1981)-340 p. (in Russian).

14. Karmanov, V.G. *Mathematical Programming*, M.: Nauka, (1975), 272 p. (in Russian).

15. Nikaido, H. *Convex Structures and Mathematical Economy*, M.: Mir, (1972), 518 p. (in Russian).

Department of Economic Cybernetics, Kiev Taras Shevchenko University, Kiev, Ukraine.
E-mail: lei@mmed.cyb.univ.kiev.ua and lyashenkoolena@hotmail.com

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.245-252

OLEKSANDR MERTENS

STOCHASTIC QUASI-GRADIENT TECHNIQUES IN VAR-BASED ALM MODELS

The paper presents VaR-based stochastic optimization model of asset-liability management with no particular assumptions about distribution of returns and other random parameters. VaR is widely accepted measure of portfolio risk and a number of research on implementation of the VaR measure in portfolio optimization models appeared recently. Proposed approach allows to include VaR constraints into the optimization model using combination of Monte-Carlo simulation and stochastic quasi-gradient techniques.

2000 *Mathematics Subject Classifications.* 91B28.

Key words and phrases. Asset-liability management, Value-at-Risk, Portfolio optimization, Stochastic quasi-gradient methods.

1. Value-at-Risk measure in the optimization framework

In the last decade Value-at-Risk (*VaR*) became industry standard as measure of risk of investment portfolios and widely used tool for risk evaluation and control. The important problem which is considered in this context is the construction of the portfolio with pre-determined constraints on *VaR* or with minimum possible *VaR*. As a consequence the task of including *VaR* measure into optimization problem appears.

Traditionally *VaR* is determined as lowest amount L such that with probability α the loss in portfolio value will not exceed L within some time interval t. Approaches for calculation of *VaR* can be divided into two groups. The first approach is based on assumption of probability distribution of risk factors (normal for returns or log-normal for values). The second approach use Monte-Carlo simulations generating scenarios and does not depend on particular distributional assumptions. The latter is very often the only possible technique for large portfolios especially when it include instruments with options properties. But including *VaR* into optimization problem is difficult due to bad mathematical properties (e.g. non-convexity).

One approach for optimizing *VaR*, which is calculated from scenarios, was proposed in Rocafellar and Uryasev (1999). It uses alternative risk measure, Conditional Value-at-Risk (*CVaR*). *CVaR* is defined as conditional expectation of losses above amount L (where L is defined as in above definition of *VaR*). *CVaR* is better than *VaR* in terms of its properties, but at the same time, optimizing *CVaR* is very close to optimizing *VaR*.

According to Rocafellar and Uryasev (2000) if we denote $f(x, r(\theta))$ - the function of losses in portfolio value which depends on decision vector x and random parameters vector $r(\theta)$, and $\Phi(x, L)$ - the probability that losses will not exceed L:

$$\Phi(x, L) = \int_{r:f(x,r)\leq L} p(r)\, dr, \tag{1}$$

VaR and *CVaR* could be defined respectively as:

$$VaR(x, \alpha) = min\{L : \Phi(x, L) \geq \alpha\}, \tag{2}$$

$$CVaR(x, \alpha) = \frac{1}{1-\alpha} \int_{r:f(x,r)\geq VaR(x,\alpha)} f(x, r)\, p(r)\, dr. \tag{3}$$

With function $F_\alpha(x, L)$ defined as:

$$F_\alpha(x, L) = L + \frac{1}{1-\alpha} \int_{r\in R^m} max\{f(x, r) - L, 0\}\, p(r)\, dr \tag{4}$$

($F_\alpha(x, L)$ is convex and continuously differentiable as function of α), *CVaR* and *VaR* could be expressed as following:

$$CVaR(x, \alpha) = \min_{L\in R} F_\alpha(x, L), \tag{5}$$

$$VaR = min\left\{ \underset{L\in R}{argmin}\, F_\alpha(x, L) \right\}. \tag{6}$$

Theorem 2 in Rocafellar and Uriasev (2000) postulates that minimizing of *CVaR* over $x \in X$ is equivalent to minimizing $F_\alpha(x, L)$ over $(x, L) \in X \times R$ so that:

$$\min_{x\in X} CVaR(x, \alpha) = \min_{(x,L)\in X\times R} F_\alpha(x, L), \tag{7}$$

while $F_\alpha(x, L)$ is convex on (x, L), and if $X$is convex set, minimization of $F_\alpha(x, L)$ is convex problem.

In practice, having the set of observations over random vector $r(\theta)$:

$$r_1, r_2, \ldots, r_N,$$

one can approximate the function $F_{\alpha}(x, L)$:

$$\bar{F}_{\alpha}(x, L) = L + \frac{1}{(1-\alpha) N} \sum_{i=1}^{N} max\{f(x, r_i) - L, 0\}. \tag{8}$$

Function $\bar{F}_{\alpha}(x, L)$ could be included into optimization problem either as objective:

$$\min_{x,L} \bar{F}_{\alpha}(x, L), \tag{9}$$

or as constraint:

$$\bar{F}_{\alpha}(x, L) \leq \bar{L}. \tag{10}$$

In order to linearize the problem, one can use auxiliary variables, so that:

$$\bar{F}_{\alpha}(x.L) = L + \frac{1}{(1-\alpha) N} \sum_{i=1}^{N} u_i, \tag{11}$$

$$f(x, r_i) - L \leq u_i, \quad u_i \geq 0, \quad i = 1, \ldots, N. \tag{12}$$

2. VaR-based optimization in the two-stage model

In many practical problems of asset-liability management some corrections of decision parameters could be made after obtaining information about realizations of random parameters. It means two-stage (or multistage in general case) setting of decision problem. Some decision variables should be chosen *before* observation of the state of nature, but some corrections are possible *after* observation. In this case, for example, the loss function could be presented as $f(x, y(x, \theta), r(\theta))$, where x is *ex ante* (strategic) decision and $y(x, \theta)$ is correction (or *adaptation*) which is dependent on the decision x and the state of nature.

The problem for minimizing *CVaR* could be written as follows:

$$\min_{x,y,L} L + \frac{1}{1-\alpha} E[u(\theta)] \tag{13}$$

s.t.:

$$f(x, y(\theta), r(\theta)) - L \leq u(\theta) \quad a.s., \tag{14}$$

$$u(\theta) \geq 0 \quad a.s., \quad x \in X, \quad y(x, \theta) \in Y(\theta) \tag{15}$$

If loss function $f(x, y(x, \theta), r(\theta))$ is linear in x and y and sets X and Y include only linear in x and y constraints, problem (13) - (14) become

well known *two-stage linear stochastic programming problem* (Dantzig and Madansky (1961)). In its general form this problem can be written as:

$$\begin{array}{l} \underset{x}{max} \quad x^T a + E\left[y\left(x,\theta\right)^T b\left(\theta\right)\right], \\ s.t.: \\ Dx \leq d, \quad x \geq 0, \\ y\left(x,\theta\right) = arg\underset{y}{max}\left\{y^T b\left(\theta\right) : B\left(\theta\right) y \leq q\left(\theta\right) - A\left(\theta\right) x, \quad y \geq 0\right\}. \end{array} \tag{16}$$

where x and $y\left(x,\theta\right)$ are first and second stage decisions respectively, a, d and D - deterministic parameters, $b\left(\theta\right)$, $q\left(\theta\right)$, $A\left(\theta\right)$ and $B\left(\theta\right)$ - random parameters, θ - state of nature.

3. Approaches for solving two-stage model: combining of Monte-Carlo simulation with optimization

The approach, which is most often used in practice to solve problem like (16), is based on Benders (1962) decomposition (see e.g. Infager (1994)). Modern techniques utilize Monte-Carlo simulation for generating scenarios and efficient algorithms, which allow solution of very large scale problems.

In this paper we propose to use alternative technique, which is well-known stochastic quasi-gradient algorithm (see e.g. Ermoliev (1976), Ermoliev and Yastremsky (1979), Ermoliev and Wets (1988)). Despite widely accepted thought about slow convergence rate of quasi-gradient methods, our experience suggests that this approach is fully operational for practical problems while considerably more flexible (in terms of class of problems which could be solved) comparing to Benders decomposition and related algorithms.

For the problem (16) so-called linearization quasi-gradient method could be used (Ermoliev (1976)). One calculate new approximation of the optimum on each iteration, which is based on random direction calculated using stochastic quasi-gradient:

$$\begin{array}{l} x^{s+1} = x^s + \rho_s\left(\bar{x}^s - x^s\right) \\ z^{s+1} = z^s + \delta_s\left(\xi^s - z^s\right) \\ \bar{x}^s = arg\underset{x}{max}\left\{x^T z^s : Dx \leq d, x \geq 0, x \in I^s\right\}, \end{array} \tag{17}$$

where x^s is approximation of the solution on the iteration s, ξ^s is stochastic quasi-gradient - random vector which satisfy following conditions:

$$\begin{array}{l} E\left[\xi_s / x^0, x^1, \ldots, x^s\right] = \nabla F\left(x^s\right) + b^s, \\ b^s \to 0 \quad a.s. \quad (s \to \infty) \end{array} \tag{18}$$

z^s is average of stochastic quasi-gradient over all iterations (it is necessary to guarantee convergence), I^s is the set of so-called induced constraints (we

need induced constraints to guarantee non-empty set for the second-stage problem).

Stochastic quasi-gradient for the problem (16) is calculated from the following conditions (actually from the solution of the dual to the second-stage problem):

$$\begin{array}{l} \xi^s = a - A\left(\theta^s\right)^T v\left(x^s, \theta^s\right) \\ v\left(x^s, \theta^s\right) = arg\, \underset{v}{min} \left\{ \left(v^s\right)^T \left(q - A\left(\theta^s\right) x^s\right) : B\left(\theta^s\right) v, v \geq 0 \right\}. \end{array} \tag{19}$$

The set of induced constraint could be written as follows:

$$I^s = \left\{ x : A\left(\theta^s\right) x \leq q\left(\theta^s\right) \right\} \tag{20}$$

The algorithm consists in generating random scenarios θ^s on each iteration, calculating quasi-gradient using (19) (it requires solution of linear programming problem) and calculating next approximation of the solution using (17) (one more linear problem).

It was proved in Ermoliev (1976) that process (17) converges to the optimal solution if step coefficients ρ_s and δ_s satisfy following conditions:

$$\begin{array}{l} \rho_s \geq 0, \quad \delta_s \geq 0, \quad \rho_s / \delta_s \to 0 \quad a.s. \quad (s \to \infty), \\ \sum\limits_{s=0}^{\infty} \rho_s = \infty, \quad \sum\limits_{s=0}^{\infty} \delta_s = \infty \quad a.s., \\ \sum\limits_{s=0}^{\infty} \rho_s \left\| b^s \right\| < \infty \quad a.s., \quad \sum\limits_{s=0}^{\infty} E\left[\rho_s^2 + \delta_s^2\right] < \infty. \end{array} \tag{21}$$

An example of such coefficients is as follows:

$$\begin{array}{l} \rho_s = K_1 / (1+s)^{\alpha}, \quad \delta_s = K_2 / (1+s)^{\beta}, \\ \alpha > \beta, \ 1/2 < \alpha \leq 1, \ 1/2 < \beta \leq 1 \\ 0 < \underline{K} \leq K_1 \leq K_2 \leq \overline{K} < \infty \end{array} \tag{22}$$

The software for implementing algorithm (17) for the problem (16) was developed at Economic Cybernetics department of the Kiev State University in 1991 (see Mertens (1991)). This software (called ROS, i.e. Risk Optimization System), which is developed using platform-independent C-code, solves general form of the two-stage stochastic linear problem (16) and utilizes number of techniques, which improve efficiency of the algorithm (for example, speed up of the solution of linear problems on each iteration and adaptive algorithms for the step size - see e.g. Uryasev (1991)). The stopping criteria, which is one of the main difficulties in quasi-gradient algorithms, is based on observations on the statistics of objective function in (16):

$$
\begin{aligned}
f_s &= \tfrac{1}{(1+s)} \sum_{i=1}^{s} f\left(x^i, \theta^i\right), \\
F_s &= \tfrac{1}{(1+L)} \sum_{i=s-L}^{s} f\left(x^i, \theta^i\right),
\end{aligned}
\tag{23}
$$

where $f\left(x^i, \theta^i\right)$ is the random value of objective on s-th iteration. The number of stopping criteria are used in the ROS software including comparison with estimates of value of objective functions in dual problem, e.g.:

$$
\begin{aligned}
&|F_s - G_s| < \varepsilon, \\
&G_s = \tfrac{1}{(1+L)} \sum_{i=s-L}^{s} \left(\left(u^s\right)^T d + v\left(x^s, \theta^s\right) \right), \\
&u^s = arg \min_u \left\{ u^T d : D^T u \geq a - A\left(\theta^s\right)^T v\left(x^s, \theta^s\right) \right\}, \\
&v\left(x^s, \theta^s\right) = arg \min_v \left\{ \left(v^s\right)^T \left(q - A\left(\theta^s\right) x^s\right) : B\left(\theta^s\right) v, v \geq 0 \right\}.
\end{aligned}
\tag{24}
$$

or:

$$
\begin{aligned}
&\left| \left(x^s\right)^T \left(a - V_s\right) \right| < \varepsilon, \\
&V_s = \tfrac{1}{s} \sum_{i=1}^{s} \left(A\left(\theta^i\right)^T v\left(x^i, \theta^i\right) \right),
\end{aligned}
\tag{25}
$$

(for additional discussion on stopping criteria see e.g. Pflug (1996)).

4. Model of optimization of inter-bank loans portfolio

Proposed approach was used for the real-life asset-liability management problem, namely management of the inter-bank loans portfolio of commercial bank. The problem consists in establishing of tomorrow limits of operation for the instruments with different maturities so that the Value-at-Risk of inter-bank loans portfolio was at pre-determined level and the expected return on entire portfolio was at maximum.

Let us define:

$W\left(\theta\right) = \bar{W} + w\left(\theta\right)$ is exogenous to the model general limit on tomorrow inter-bank operations which is defined form financing needs or excess liquidity of the bank; $t-$ time to maturity of particular instrument; $r_t\left(\theta\right) -$ tomorrow spot rate at inter-bank market (continuously compounded); $p_t\left(\theta\right) = exp\left(-r_t\left(\theta\right) \cdot t\right) -$ tomorrow price of one currency unit of inter-bank loan with maturity t; s_t- existing position in particular instrument; x_t- tomorrow limit (decision); $y_t\left(x, \theta\right) -$ tomorrow correction of the limit which is bounded in some way.

The loss function which is necessary to build VaR constraint is defined as dollar duration of total inter-bank loans position:

$$f\left(x, y\left(x,\theta\right),\theta\right) = -\sum_{t} t \cdot p_t\left(\theta\right) \cdot \left(s_t + x_t + y_t\left(\theta\right)\right).$$

The possible objectives are minimizing of the Value-at-Risk of inter-bank loans portfolio or maximizing of the total return:

$$\max_{x} E\left[\sum_{t} r_t\left(\theta\right) \cdot \left(x_t + y_t\left(\theta\right)\right)\right].$$

The problem was solved for more than one hundred instruments (inter-bank loans with 120 different maturities) for the real-life situation at the Ukrainian inter-bank loans market. The number of utilized scenarios (iterations in quasi-gradient algorithm) was up to 100,000, but for the relatively good approximation of the optimal solution it was enough about 10,000 iterations. The time of calculations at ordinary Pentium III 1 GHz processor was about 30 min for 10,000 iterations.

5. Conclusion

The paper demonstrates possibility of using quasi-gradient techniques in *VaR*-based asset-liability management optimization models. The approach was used for real-life problem of inter-bank loans portfolio management. The obtained results demonstrate that this approach is fully operational and efficient while allowing to solve more general (comparing to traditional approach) form of two-stage stochastic programming problem.

The main directions of future research are (2) development and practical implementation of wider range of asset-liability management problems (including credit risk management, etc.) using proposed approach, and (3) precise comparison in terms of efficiency of numerical algorithms between traditional (based on Benders decomposition) techniques and quasi-gradient methods.

Bibliography

1. Benders, J.R. Partitioning Procedure for Solving Mixed-Variable Programming Problems, *Numerische Mathematik* 4, (1962). 238 - 252.

2. Birge, J.R. and F. Louveaux. *Introduction to Stochastic Programming*, Springer Verlag, (1997).

3. Dantzig, G.B. and M. Madansky. On the Solution of Two-Staged Linear Programs under Uncertainty. *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability* I, (1961), 165 - 176.

4. Ermoliev, Yu. and R. J.-B. Wets (eds.) *Numerical Techniques for Stochastic Optimization*, Springer Verlag, (1988).

5. Ermoliev, Yu. and A. Yastremsky *Stochastic Models and Methods in Economic Planning*. Moscow: Nauka, (1979).

6. Ermoliev, Yu. *Methods of Stochastic Programming*. Moscow: Nauka, (1976).

7. Infanger, G. *Planning Under Uncertainty: Solving Large-Scale Stochastic Linear Programs*. Boyd & Fraser Pub. Co., (1994).

8. Jorion, Ph. *Value at Risk: A New Benchmark for Measuring Derivatives Risk*. Irwin Professional Pub., (1996).

9. Kall, P. and S. W. Walace *Stochastic Programming*. John Wiley & Sons, (1995).

10. Mertens, A.V. *The Risk Optimization Software for Economic Models*, Operations Research and Automatic Control Systems, 32, (1992), 102 - 107.

11. Pflug, G.Ch. *Optimization of Stochastic Models: The Interface Between Simulation and Optimization*. Kluwer Academic Publishers, (1996).

12. RiskMetrics. *Technical Document*. J.P.Morgan., (1996).

13. Rocafellar T. and S. Uryasev. *Optimization of Conditional Value-at-Risk*, The Journal of Risk, (2000).

14. Uryasev S. *Adaptive algorithms of Stochastic Optimization and Game Theory*. Nauka, (1990).

15. Zenios S.A. (Ed.) *Financial Optimization*. Cambridge University Press, (1996).

Department Economics, Kiev Taras Shevchenko University, 252601 Kiev, Ukraine.
E-mail: mertens@icpim-iib.com

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.253-264

MIKHAIL MOKLYACHUK

GAME THEORY AND CONVEX OPTIMIZATION METHODS IN ROBUST ESTIMATION PROBLEMS

We considere the problem of the mean square optimal estimate of the functional

$$A\xi = \int_0^\infty \langle\, a(t),\, \xi(t)\,\rangle\, dt$$

which is determined by the unknown values of the Hilbert space valued stationary stochastic process $\xi(t)$, $t \in \mathbf{R}^1$ from observations of the process $\xi(t)+\eta(t)$ for $t < 0$, where $\eta(t)$ is an uncorrelated with $\xi(t)$ Hilbert space valued stationary stochastic process. The mean square error and the spectral characteristic of the optimal estimate of the functional $A\xi$ are proposed. The minimax spectral characteristic and the least favorable spectral densities are found for various classes of spectral densities.

2000 *Mathematics Subject Classifications.* 60G25, 62M20.

Key words and phrases. Stationary stochastic process, Hilbert space, mean square optimal estimate, minimax spectral characteristic, least favorable spectral density

1. Introduction

The Hilbert space projection method of linear interpolation, extrapolation and filtering of weakly stationary stochastic processes (see Kolmogorov (1986), Rozanov (1967)) may be employed in the case where spectral densities of stochastic processes are known. In practice, however, problems of estimation of the unknown values of stochastic processes arise where the spectral densities are not known exactly. To solve the problem the parametric or nonparametric estimates of the unknown spectral densities are found. Then the traditional method is applied provided that estimates of the densities are true. This procedure can result in a significant increasing of the value of the error as Vastola and Poor (1983) have demonstrated with the help of some examples. For this reason it is necessary to search the estimate

that has the least value of the error for all densities from a certain class of spectral densities. Such an approach to the problem of interpolation, extrapolation and filtering of stationary stochastic processes have been taken into consideration by many investigators. A survey of results in minimax (robust) methods of data processing can be found in the paper Kassan and Poor (1985). See also papers by the author [8-10]. In this article it is shown that convex optimization methods can be applied to determine the least favorable spectral densities and minimax (robust) spectral characteristic of the optimal estimate of linear functional of a Hilbert space valued stationary stochastic process from observations with noise.

2. Hilbert space projection method of extrapolation

Denote by X a separable Hilbert space with the orthonormal basis $\{e_k : k = 1, 2, \ldots\}$. Stationary stochastic processes $\xi(t)$ and $\eta(t)$ with values in X have spectral densities $f(\lambda)$ and $g(\lambda)$ if the correlation functions $B_\xi(s)$ and $B_\eta(s)$ can be represented in the form

$$\langle B_\xi(s)\, e_k, e_j\rangle = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} e^{i\lambda s}\,\langle f(\lambda)\, e_k, e_j\rangle\, d\lambda, \quad k, j = 1, 2, \ldots,$$

$$\langle B_\eta(s)\, e_k, e_j\rangle = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} e^{i\lambda s}\,\langle g(\lambda)\, e_k, e_j\rangle\, d\lambda, \quad k, j = 1, 2, \ldots$$

For almost all $\lambda \in \mathbf{R}^1$ the spectral densities $f(\lambda)$ and $g(\lambda)$ are nuclear operators with integrable nuclear norms (see Kallianpur and Mandrekar (1965)).

Denote by $\mathbf{K}(f+g)$ a set of all $k \in \mathbf{N}$ such that $f_k(\lambda) + g_k(\lambda)$, $f_k(\lambda) = \langle f(\lambda)e_k, e_k\rangle$, $g_k(\lambda) = \langle f(\lambda)e_k, e_k\rangle$ satisfy the minimality condition:

$$\int_{-\infty}^{\infty} \frac{|\gamma(\lambda)|}{f_k(\lambda) + g_k(\lambda)}\, d\lambda < \infty$$

for some nontrivial function of the exponential type $\gamma(\lambda) = \int_0^\infty \alpha(t)\, e^{it\lambda}\, dt$. We will consider the estimation problem in the case where $\mathbf{K}(f+g) \neq \emptyset$. This condition is necessary and sufficient in order that the mean square error of the optimal linear estimate of the unknown value of the process $\xi(t)$ be not equal to zero (see Rozanov (1967)).

Let the sequence $a(t)$ that determine the functional

$$A\xi = \int_0^\infty \langle\, a(t),\, \xi(t)\,\rangle\, dt = \sum_{k=1}^{\infty}\int_0^\infty a_k(t)\xi_k(t)\, dt,$$

satisfy the conditions

$$\sum_{k=1}^{\infty}\int_0^{\infty}|a_k(t)|\,dt<\infty,\qquad \sum_{k=1}^{\infty}\int_0^{\infty}t\,|a_k(t)|^2\,dt<\infty. \tag{1}$$

Under these conditions $\mathbf{E}\,|A\xi|^2<\infty$. The mean square error of a linear estimate $\widehat{A}\xi$ of the functional $A\xi$ is determined by the spectral characteristic $h(\lambda)=\left\{\,h_k(\lambda):\ k=1,2,\dots\,\right\}$ of the estimate and the spectral densities $f(\lambda)$ and $g(\lambda)$ of processes $\xi(t)$ and $\eta(t)$. The value of the mean square error

$$\Delta(h;f,g)=\mathbf{E}\left|A\xi-\widehat{A}\xi\right|^2$$

$$=\sum_{k=1}^{\infty}\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[|A_k(\lambda)-h_k(\lambda)|^2\,f_k(\lambda)+|h_k(\lambda)|^2\,g_k(\lambda)\right]\,d\lambda,$$

where

$$A_k(\lambda)=\int_0^{\infty}a_k(t)e^{it\lambda}\,dt.$$

The spectral characteristic $h(\lambda)=\left\{\,h_k(\lambda):\ k=1,2,\dots\,\right\}$ of the estimate $\widehat{A}\xi$ belongs to the subspace $\mathbf{L}_2^-(f+g)$ of the space $\mathbf{L}_2(f+g)$ generated by the functions

$$h_k(\lambda)=\int_0^{\infty}h_k(t)e^{-it\lambda}\,dt,$$

$$\sum_{k=1}^{\infty}\int_{-\infty}^{\infty}|h_k(\lambda)|^2\,(f_k(\lambda)+g_k(\lambda))\,d\lambda<\infty.$$

The spectral characteristic $h(f,g)$ of the optimal linear estimate of the functional $A\xi$ minimizes the mean square error. With the help of the Hilbert space projection method (see Kolmogorov (1986), Rozanov (1967)) we can derive the following formulas for the value $\Delta(f,g)$ of the error and the spectral characteristic $h(f,g)$ of the optimal linear estimate of the functional $A\xi$ under the condition that the spectral densities $f(\lambda),\ g(\lambda)$ of the processes $\eta(t)$ and $\xi(t)$ are known. In this case

$$\Delta(f,g)=\sum_{k\in\mathbf{K}}\left\{\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{|A_k(\lambda)g_k(\lambda)+C_k(\lambda)|^2}{(f_k(\lambda)+g_k(\lambda))^2}\,f_k(\lambda)\,d\lambda+\right.$$

$$\left.+\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{|A_k(\lambda)f_k(\lambda)-C_k(\lambda)|^2}{(f_k(\lambda)+g_k(\lambda))^2}\,g_k(\lambda)\,d\lambda\right\}=$$

$$=\sum_{k\in\mathbf{K}}\left[\int_0^{\infty}\left(\mathbf{B}_k\,c_k\right)(t)\ \bar{c}_k(t)\,dt+\int_0^{\infty}\left(\mathbf{R}_k\,a_k\right)(t)\ \bar{a}_k(t)\,dt\right]=$$

$$= \sum_{k\in\mathbf{K}} \left[\langle \mathbf{B}_k\, c_k, c_k\rangle + \langle \mathbf{R}_k\ a_k, a_k\rangle\right], \tag{2}$$

and

$$h_k(f,g) = \frac{A_k(\lambda)\, f_k(\lambda) - C_k(\lambda)}{f_k(\lambda)+g_k(\lambda)} =$$

$$= A_k(\lambda) - \frac{A_k(\lambda)\, g_k(\lambda) + C_k(\lambda)}{f_k(\lambda)+g_k(\lambda)}, \tag{3}$$

where

$$C_k(\lambda) = \int_0^\infty c_k(t)\, e^{it\lambda}\, dt, \quad c_k(t) = \left(\mathbf{B}_k^{-1}\, \mathbf{D}_k\, a_k\right)(t), \quad k \in \mathbf{K}(f+g),$$

$\langle a, c\rangle$ is the inner product and $\mathbf{B}_k$, $\mathbf{D}_k$, $\mathbf{R}_k$ are operators in the space $\mathbf{L}_2[0,\infty)$ that are determined by the relations

$$\left(\mathbf{B}_k\, c\right)(t) = \frac{1}{2\pi}\int_0^\infty c(u) \int_{-\infty}^\infty e^{i(u-t)\lambda} (f_k(\lambda)+g_k(\lambda))^{-1}\, d\lambda\, du,$$

$$\left(\mathbf{D}_k\, c\right)(t) = \frac{1}{2\pi}\int_0^\infty c(u) \int_{-\infty}^\infty e^{i(u-t)\lambda} f_k(\lambda)\, (f_k(\lambda)+g_k(\lambda))^{-1}\, d\lambda\, du,$$

$$\left(\mathbf{R}_k\, c\right)(t) = \frac{1}{2\pi}\int_0^\infty c(u) \int_{-\infty}^\infty e^{i(u-t)\lambda} f_k(\lambda)\, g_k(\lambda)\, (f_k(\lambda)+g_k(\lambda))^{-1}\, d\lambda\, du.$$

Lemma 1. *Let $\xi(t)$ and $\eta(t)$ be uncorrelated stationary stochastic processes with spectral densities $f(\lambda)$ and $g(\lambda)$ such that $\mathbf{K}(f+g) \neq \emptyset$ and condition (1) is satisfied. The mean square error $\Delta(f,g)$ and the spectral characteristic $h(f,g)$ of the optimal linear estimate of the functional $A\xi$ from observations of the process $\xi(t)+\eta(t)$ for $t<0$ can be computed by (2) and (3).*

In the case of observations without noise formulas for the mean square error and spectral characteristic are the following.

$$\Delta(f) = \sum_{k\in\mathbf{K}} \frac{1}{2\pi}\int_{-\infty}^\infty |C_k(\lambda)|^2\, f_k^{-1}(\lambda)\, d\lambda = \sum_{k\in\mathbf{K}} \left\langle \mathbf{B}_k^{-1}\ a_k, a_k\right\rangle = \sum_{k\in\mathbf{K}} \|\mathbf{A}_k\ d_k\|^2, \tag{4}$$

$$h_k(f) = A_k(\lambda) - C_k(\lambda)\, f_k^{-1}(\lambda) = A_k(\lambda) - (\mathbf{A}_k\ d_k)(\lambda) d_k^{-1}(\lambda), \tag{5}$$

where the functions $d_k(t)$ are detemined by the factorization

$$f_k(\lambda) = |d_k(\lambda)|^2 = \left|\int_0^\infty d_k(t)\, e^{-it\lambda}\, dt\right|^2$$

of the density $f_k(\lambda)$,

$$C_k(\lambda) = \int_0^\infty \left(\mathbf{B}_k^{-1}\, a_k\right)(t)\, e^{it\lambda}\, dt,$$

$$(\mathbf{A}_k\, d)(t) = \int_0^\infty a_k(t+u)d(u)du,$$

$$(\mathbf{A}_k\ d_k)(\lambda) = \int_0^\infty (\mathbf{A}_k\, d_k)\,(t)\, e^{it\lambda}\, dt,$$

$$\big(\mathbf{B}_k\, c\big)(t) = \frac{1}{2\pi}\int_0^\infty c(u)\int_{-\infty}^\infty e^{i(u-t)\lambda} f(\lambda)_k^{-1}\, d\lambda\, du.$$

Lemma 2. *Let $\xi(t)$ be a stationary stochastic process with the spectral density $f(\lambda)$ such that $\mathbf{K}(f) \neq \emptyset$ and let condition (1) be satisfied. The mean square error $\Delta(f)$ and the spectral characteristic $h(f)$ of the optimal linear estimate of the functional $A\xi$ from observations of the process $\xi(t)$ for $t < 0$ can be computed by formulas (4), (5).*

3. Minimax-robust approach to extrapolation problem

Formulas (1) – (5) can be applied to compute the mean square error and the spectral characteristic of the optimal linear estimate of the functional $A\xi$ if the spectral densities $f(\lambda)$ and $g(\lambda)$ of the processes $\eta(t)$ and $\xi(t)$ are known. In the case where the spectral densities are not known exactly, but sets $\mathcal{D}_f$ and $\mathcal{D}_g$ of possible spectral densities are given, we apply the minimax (robust) approach to the problem of estimation of the unknown values of stochastic processes and functionals. With the help of this approach we can find an estimate that minimizes the mean square error for all spectral densities $f(\lambda), \quad g(\lambda)$ from a given class $\mathcal{D} = \mathcal{D}_f \times \mathcal{D}_g$ simultaneously.

Definition 1. For a given class $\mathcal{D} = \mathcal{D}_f \times \mathcal{D}_g$ of spectral densities the spectral densities $f^0(\lambda) \in \mathcal{D}_f$ and $g^0(\lambda) \in \mathcal{D}_g$ are called the least favorable in $\mathcal{D} = \mathcal{D}_f \times \mathcal{D}_g$ for the optimal linear estimation of the functional $A\xi$ if

$$\Delta(f^0, g^0) = \Delta\big(h(f^0, g^0); f^0, g^0\big) = \max_{(f,g)\in\mathcal{D}_f\times\mathcal{D}_g} \Delta\big(h(f,g); f, g\big).$$

Definition 2. For a given class $\mathcal{D} = \mathcal{D}_f \times \mathcal{D}_g$ of spectral densities the spectral characteristic $h^0(\lambda)$ is called minimax (robust) if

$$h^0(\lambda) \in H_{\mathcal{D}} = \bigcap_{(f,g)\in\mathcal{D}_f\times\mathcal{D}_g} L_2^-(f+g),$$

$$\min_{h\in H_{\mathcal{D}}} \max_{(f,g)\in\mathcal{D}_f\times\mathcal{D}_g} \Delta\big(h; f, g\big) = \max_{(f,g)\in\mathcal{D}_f\times\mathcal{D}_g} \Delta\big(h^0; f, g\big).$$

The detailed analysis of the relations (2) – (5) makes it possible to conclude that the following statement is true.

Lemma 3. *Spectral densities $f^0(\lambda)$ and $g^0(\lambda)$ are the least favorable in the class $\mathcal{D}=\mathcal{D}_f\times\mathcal{D}_g$ for the optimal linear estimation of the functional $A\xi$ if $\mathbf{K}(f^0,g^0)\neq\emptyset$ and the Fourier transform of functions*

$$\frac{1}{\big(f_k^0(\lambda)+g_k^0(\lambda)\big)},\ \frac{f_k^0(\lambda)}{\big(f_k^0(\lambda)+g_k^0(\lambda)\big)},\ \frac{f_k^0(\lambda)g_k^0(\lambda)}{\big(f_k^0(\lambda)+g_k^0(\lambda)\big)}$$

form operators $\mathbf{B}_k^0,\ \mathbf{D}_k^0,\ \mathbf{R}_k^0$ that determine a solution of the conditional extremum problem

$$\max_{(f,g)\in\mathcal{D}_f\times\mathcal{D}_g}\sum_{k\in\mathbf{K}}\left[\left\langle\mathbf{D}_k\,a_k,\,\mathbf{B}_k^{-1}\,\mathbf{D}_k\,a_k\right\rangle+\left\langle\mathbf{R}_k\,a_k,\,a_k\right\rangle\right]=$$

$$=\sum_{k\in\mathbf{K}}\left[\left\langle\mathbf{D}_k^0\,a_k,\,(\mathbf{B}_k^0)^{-1}\,\mathbf{D}_k^0\,a_k\right\rangle+\left\langle\mathbf{R}_k^0\,a_k,\,a_k\right\rangle\right].\tag{6}$$

The minimax (robust) spectral characteristic $h^0=h(f^0,g^0)$ can be computed by the formula (3) if $h(f^0,g^0)\in H_{\mathcal{D}}$.

For the case of observations without noise we have the following statement.

Lemma 4. A spectral density $f^0(\lambda)\in\mathcal{D}_f$ is the least favorable in the class $\mathcal{D}_f$ for the optimal linear estimation of the functional $A\xi$ from observations of the process $\xi(t)$ for $t<0$ if $\mathbf{K}(f^0)\neq\emptyset$ and the Fourier transformat of functions $\left(f_k^0(\lambda)\right)^{-1},k=1,2,\dots$ form operators $\mathbf{B}_k^0,k=1,2,\dots$ that determine a solution of the conditional extremum problem

$$\max_{f\in\mathcal{D}_f}\sum_{k\in\mathbf{K}}\left\langle\mathbf{B}_k^{-1}\,a_k,\,a_k\right\rangle=\sum_{k\in\mathbf{K}}\left\langle(\mathbf{B}_k^0)^{-1}\,a_k,\,a_k\right\rangle.\tag{7}$$

The minimax (robust) spectral characteristic $h^0=h(f^0)$ can be computed by the formula (5) if $h(f^0)\in H_{\mathcal{D}_f}$.

The least favorable spectral densities $f^0(\lambda)$, $g^0(\lambda)$ and the minimax (robust) spectral characteristic $h^0=h(f^0,g^0)$ form a saddle point of the function $\Delta(h;f,g)$ on the set $H_{\mathcal{D}}\times\mathcal{D}$. The saddle point inequalities

$$\Delta\big(h^0;f,g\big)\leq\Delta\big(h^0;f^0,g^0\big)\leq\Delta\big(h;f^0,g^0\big)$$

$$\forall h\in H_{\mathcal{D}}\qquad\forall f\in\mathcal{D}_f\qquad\forall g\in\mathcal{D}_g$$

hold when $h^0=h(f^0,g^0)$ and $h(f^0,g^0)\in H_{\mathcal{D}}$, and (f^0,g^0) is a solution of the conditional extremum problem

$$\sup_{(f,g)\in\mathcal{D}_f\times\mathcal{D}_g}\Delta\big(h(f^0,g^0);f,g\big)=\Delta\big(h(f^0,g^0);f^0,g^0\big),\tag{8}$$

where

$$\Delta\big(h(f^0,g^0);f,g\big)=$$

$$= \sum_{k\in\mathbf{K}} \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|A_k(\lambda) g_k^0(\lambda) + C_k^0(\lambda)|^2}{\left(f_k^0(\lambda) + g_k^0(\lambda)\right)^2} f_k(\lambda)\, d\lambda + \right.$$

$$\left. + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|A_k(\lambda) f_k^0(\lambda) - C_k^0(\lambda)|^2}{\left(f_k^0(\lambda) + g_k^0(\lambda)\right)^2} g_k(\lambda)\, d\lambda \right\}.$$

The conditional extremum problem (8) is equivalent to the unconditional extremum problem:

$$\Delta_{\mathcal{D}}(f,g) = -\Delta\big(h(f^0,g^0); f, g\big) + \delta\left((f,g) \mid \mathcal{D}_f \times \mathcal{D}_g\right) \to \inf, \qquad (9)$$

where $\delta\big((f,g) \mid \mathcal{D}_f \times \mathcal{D}_g\big)$ is the indicator function of the set $\mathcal{D} = \mathcal{D}_f \times \mathcal{D}_g$. A solution (f^0, g^0) of problem (9) is characterized by the condition $0 \in \partial\Delta_{\mathcal{D}}(f^0, g^0)$, where $\partial\Delta_{\mathcal{D}}(f^0, g^0)$ is the subdifferential of the convex functional $\Delta_{\mathcal{D}}(f,g)$ at the point (f^0, g^0) (see Pshenichnyi (1971))

This condition gives us a possibility to determine the least favorable spectral densities for concrete classes of spectral densities.

Lemma 5. Let (f^0, g^0) be a solution of the extremum problem (9). The spectral densities $f^0(\lambda)$ and $g^0(\lambda)$ are the least favorable in the class $\mathcal{D} = \mathcal{D}_f \times \mathcal{D}_g$ and the spectral characteristic $h^0 = h(f^0, g^0)$ is the minimax (robust) for the optimal linear estimate of the functional $A\xi$ if $h(f^0, g^0) \in H_{\mathcal{D}}$.

4. Least favorable spectral densities in the class $\mathcal{D}_f^0 \times \mathcal{D}_g^0$

Let us consider the problem for the set of spectral densities $\mathcal{D} = \mathcal{D}_f^0 \times \mathcal{D}_g^0$, where

$$\mathcal{D}_f^0 = \left\{ f(\lambda) \mid \frac{1}{2\pi} \sum_{k=1}^{\infty} \int_{-\infty}^{\infty} f_k(\lambda)\, d\lambda \le P_1 \right\},$$

$$\mathcal{D}_g^0 = \left\{ g(\lambda) \mid \frac{1}{2\pi} \sum_{k=1}^{\infty} \int_{-\infty}^{\infty} g_k(\lambda)\, d\lambda \le P_2 \right\}.$$

Stochastic processes which have spectral densities from the class $\mathcal{D}_f^0$ satisfy the power constraint: $\mathbf{E}|\xi(t)|^2 \le P_1$. Let the densities $f^0(\lambda) \in \mathcal{D}_f^0$ and $g^0(\lambda) \in \mathcal{D}_g^0$, $\mathbf{K}(f^0 + g^0) \ne \emptyset$ and the functions

$$h_{kf}(f^0, g^0) = \frac{\left|A_k(\lambda) g_k^0(\lambda) + C_k^0(\lambda)\right|}{f_k^0(\lambda) + g_k^0(\lambda)}, \qquad (10)$$

$$h_{kg}(f^0, g^0) = \frac{\left|A_k(\lambda) f_k^0(\lambda) - C_k^0(\lambda)\right|}{f_k^0(\lambda) + g_k^0(\lambda)}, \qquad (11)$$

$$k \in \mathbf{K}(f^0 + g^0),$$

are bounded. Under these conditions the functional $\Delta\big(h(f^0,g^0);f,g\big)$ is a continuous linear functional in the space $L_1\times L_1$ and

$$\partial\Delta_{\mathcal{D}_f^0\times\mathcal{D}_g^0}(f^0,g^0)=-\partial\Delta\big(h(f^0,g^0);f^0,g^0\big)+\partial\delta\big((f^0,g^0)\mid\mathcal{D}_f^0\times\mathcal{D}_g^0\big).$$

From the condition $0\in\partial\Delta_{\mathcal{D}}(f^0,g^0)$ for $\mathcal{D}=\mathcal{D}_f^0\times\mathcal{D}_g^0$ we find that the components of the least favorable densities $f^0\in\mathcal{D}_f^0$, $g^0\in\mathcal{D}_g^0$ satisfy the equations

$$\alpha_{k1}\big|A_k(\lambda)g_k^0(\lambda)+C_k^0(\lambda)\big|=\big(f_k^0(\lambda)+g_k^0(\lambda)\big) \tag{12}$$

$$\alpha_{k2}\big|A_k(\lambda)f_k^0(\lambda)-C_k^0(\lambda)\big|=\big(f_k^0(\lambda)+g_k^0(\lambda)\big), \tag{13}$$

where constants $\alpha_{k1}\geq 0$, $\alpha_{k2}\geq 0$.

Note, that $\alpha_{k1}\neq 0$ if

$$\sum_{k\in\mathbf{K}}\frac{1}{2\pi}\int_{-\infty}^{\infty}f_k^0(\lambda)\,d\lambda=P_1 \tag{14}$$

and $\alpha_{k2}\neq 0$ if

$$\sum_{k\in\mathbf{K}}\frac{1}{2\pi}\int_{-\infty}^{\infty}g_k^0(\lambda)\,d\lambda=P_2. \tag{15}$$

Theorem 1. Let spectral densities $f^0(\lambda)$ and $g^0(\lambda)$ belong to the set $\mathcal{D}=\mathcal{D}_f^0\times\mathcal{D}_g^0$, $\mathbf{K}(f^0+g^0)\neq\emptyset$ and the functions $h_{kf}(f^0,g^0)$ and $h_{kg}(f^0,g^0)$, $k\in\mathbf{K}(f^0+g^0)$ computed by the formulas (10) and (11) are bounded. The spectral densities $f^0(\lambda)$ and $g^0(\lambda)$ are the least favorable in the class $\mathcal{D}=\mathcal{D}_f^0\times\mathcal{D}_g^0$ for the optimal linear estimation of the functional $A\xi$ if they satisfy the relations (12) and (13) and determine a solution of the extremum problem (6). The minimax (robust) spectral characteristic of the optimal linear estimate of the functional is computed by the formula (3).

Theorem 2. Let a spectral density $f(\lambda)$ is known, a density $g^0(\lambda)$ belong to the set $\mathcal{D}_g^0$, $\mathbf{K}(f+g^0)\neq\emptyset$ and the functions $h_{kg}(f,g^0)$, $k\in\mathbf{K}(f+g^0)$ computed by the formula (11) are bounded. The spectral density $g^0(\lambda)$ is the least favorable in the class $\mathcal{D}_g^0$ for the optimal estimation of the functional $A\xi$ if the components of the density satisfy the relations

$$g_k^0(\lambda)=\max\big\{0,\alpha_{k2}\big|A_k(\lambda)f_k(\lambda)-C_k^0(\lambda)\big|-f_k(\lambda)\big\}$$

and $(f(\lambda),g^0(\lambda))$ determine a solution of the extremum problem (6). The function $h(f,g^0)$ computed by the formula (3) is the minimax (robust) spectral characteristic of the optimal linear estimate of the functional $A\xi$.

Theorem 3. Let a spectral density $f^0(\lambda)$ belong to the set $\mathcal{D}_f^0$, $\mathbf{K}(f^0)\neq\emptyset$ and let functions $h_{kf}(f^0)$, $k\in\mathbf{K}(f^0)$ computed by formula (5) be bounded. The spectral density $f^0(\lambda)$ is the least favorable in the class $\mathcal{D}_f^0$ for the

optimal estimation of the functional $A\xi$ from observations of the process $\xi(t)$ for $t < 0$ if it satisfy the equation

$$f_k^0(\lambda) = \alpha_{k1}\big|C_k^0(\lambda)\big|$$

and determine a solution of the extremum problem (7). The minimax (robust) spectral characteristic of the optimal linear estimate of the functional is computed by the formula (5).

5. Least favorable spectral densities in the class $\mathcal{D} = \mathcal{D}_\varepsilon \times \mathcal{D}_v^u$

Let us consider the problem for the set of spectral densities $\mathcal{D} = \mathcal{D}_\varepsilon \times \mathcal{D}_v^u$,

$$D_\varepsilon = \left\{ f \;\Big|\; f_k(\lambda) = (1-\varepsilon) f_k^1(\lambda) + \varepsilon w_k(\lambda), \quad \frac{1}{2\pi}\sum_{k=1}^{\infty}\int_{-\infty}^{\infty} f_k(\lambda)\, d\lambda = P_1 \right\},$$

$$\mathcal{D}_v^u = \left\{ g \;\Big|\; v_k(\lambda) \le g_k(\lambda) \le u_k(\lambda); \quad \frac{1}{2\pi}\sum_{k=1}^{\infty}\int_{-\infty}^{\infty} g_k(\lambda)\, d\lambda \le P_2 \right\},$$

where the spectral densities $v(\lambda)$, $u(\lambda)$ and $f^1(\lambda)$ are known and the densities $u_k(\lambda)$, $k = 1, 2, \dots$ are bounded. The class $\mathcal{D}_\varepsilon$ describes the "ε - pollution" model of stochastic processes. The class $\mathcal{D}_v^u$ describes the "band" model of stochastic processes.

If $f^0(\lambda) \in \mathcal{D}_\varepsilon$, $g^0(\lambda) \in \mathcal{D}_v^u$, $\mathbf{K}(f^0 + g^0) \neq \emptyset$ and the function$h_{kf}(f^0, g^0)$ and $h_{kg}(f^0, g^0)$, $k \in \mathbf{K}(f^0 + g^0)$ computed by the formulas (10) and (11) are bounded, the condition $0 \in \partial\Delta_{\mathcal{D}}(f^0, g^0)$ for $\mathcal{D} = \mathcal{D}_\varepsilon \times \mathcal{D}_v^u$ is satisfied if components of the densities $f^0(\lambda)$ and $g^0(\lambda)$ satisfy the equations

$$\big|A_k(\lambda) g_k^0(\lambda) + C_k^0(\lambda)\big| = \big(f_k^0(\lambda) + g_k^0(\lambda)\big)\big(\varphi_k(\lambda) + \alpha_{k1}^{-1}\big), \tag{16}$$

$$\big|A_k(\lambda) f_k^0(\lambda) - C_k^0(\lambda)\big| = \big(f_k^0(\lambda) + g_k^0(\lambda)\big)\big(\gamma_{k1}(\lambda) + \gamma_{k2}(\lambda) + \alpha_{k2}^{-1}\big), \tag{17}$$

where $\gamma_{k1}(\lambda) \le 0$ a. e. and $\gamma_{k1}(\lambda) = 0$ if $g_k^0(\lambda) \ge v_k(\lambda)$; $\gamma_{k2}(\lambda) \ge 0$ a. e. and $\gamma_{k2}(\lambda) = 0$ if $g_k^0(\lambda) \le u_k(\lambda)$; $\varphi_k(\lambda) \le 0$ a. e. and $\varphi_k(\lambda) = 0$ if $f_k^0(\lambda) \ge (1-\varepsilon) f_k^1(\lambda)$.

Theorem 4. Let spectral densities $f^0(\lambda)$ and $g^0(\lambda)$ belong to the set $\mathcal{D} = \mathcal{D}_\varepsilon \times \mathcal{D}_v^u$, $\mathbf{K}(f^0 + g^0) \neq \emptyset$ and the functions $h_{kf}(f^0, g^0)$ and $h_{kg}(f^0, g^0)$, $k \in \mathbf{K}(f^0 + g^0)$ computed by the formulas (10) and (11) are bounded. The spectral densities $f^0(\lambda)$ and $g^0(\lambda)$ are the least favorable in the class $\mathcal{D} = \mathcal{D}_\varepsilon \times \mathcal{D}_v^u$ for the optimal linear estimation of the functional $A\xi$ if they satisfy the relations (14) – (17) and determine a solution of the extremum problem (6). The minimax (robust) spectral characteristic of the optimal linear estimate of the functional is computed by the formula (3).

Theorem 5. Let a spectral density $f(\lambda)$ is known, a density $g^0(\lambda)$ belong to the set $\mathcal{D}_v^u$, $\mathbf{K}(f + g^0) \neq \emptyset$ and the functions $h_{kg}(f, g^0)$ and $k \in \mathbf{K}(f + g^0)$

computed by the formula (11) are bounded. The spectral density $g^0(\lambda)$ is the least favorable in the class $\mathcal{D}_v^u$ for the optimal estimation of the functional $A\xi$ if components of the density satisfy the relations

$$g_k^0(\lambda)=\max\left\{v_k(\lambda),\min\left\{u_k(\lambda),\alpha_{k2}\left|A_k(\lambda)f_k(\lambda)-C_k^0(\lambda)\right|-f_k(\lambda)\right\}\right\}$$

and $(f(\lambda),g^0(\lambda))$ determine a solution of the extremum problem (6). The function $h(f,g^0)$ computed by the formula (3) is the minimax (robust) spectral characteristic of the optimal linear estimate of the functional $A\xi$.

Theorem 6. Let a spectral density $f^0(\lambda)$ belong to the set $\mathcal{D}_\varepsilon$, $\mathbf{K}(f^0)\neq\emptyset$ and the functions $h_{kf}(f^0)$, $k\in\mathbf{K}(f^0)$ computed by the formula (5) are bounded. The spectral density $f^0(\lambda)$ is the least favorable in the class $\mathcal{D}_\varepsilon$ for the optimal estimation of the functional $A\xi$ from observations of the sequence $\xi(t)$ for $t<0$ if the components $f_k^0(\lambda)$, $k\in\mathbf{K}(f^0)$ of the density satisfy the equation

$$f_k^0(\lambda)=\max\left\{(1-\varepsilon)f_k^1(\lambda),\alpha_{k1}\left|C_k^0(\lambda)\right|\right\}$$

and determine a solution of the extremum problem (7). The minimax (robust) spectral characteristic of the optimal linear estimate of the functional is computed by the formula (5).

6. Least favorable spectral densities in the class $\mathcal{D}_{2\delta_1}\times\mathcal{D}_{1\delta_2}$

Let the set of densities is of the form $\mathcal{D}=\mathcal{D}_{2\delta_1}\times\mathcal{D}_{1\delta_2}$,

$$\mathcal{D}_{2\delta_1}=\left\{f\mid\frac{1}{2\pi}\sum_{k=1}^{\infty}\int_{-\infty}^{\infty}\left|f_k(\lambda)-f_k^1(\lambda)\right|^2\,d\lambda\leq\delta_1\right\},$$

$$\mathcal{D}_{1\delta_2}=\left\{g\mid\frac{1}{2\pi}\sum_{k=1}^{\infty}\int_{-\infty}^{\infty}\left|g_k(\lambda)-g_k^1(\lambda)\right|\,d\lambda\leq\delta_2\right\},$$

where $f_k^1(\lambda)$ and $g_k^1(\lambda)$ are known bounded spectral densities. The sets $\mathcal{D}_{2\delta_1}$ and $\mathcal{D}_{1\delta_2}$ describe the "δ – neighbourhood" models of stochastic processes in the spaces $\mathbf{L}_1$ and $\mathbf{L}_2$. If $f^0(\lambda)\in\mathcal{D}_{2\delta_1}$ and $g^0(\lambda)\in\mathcal{D}_{1\delta_2}$, $\mathbf{K}(f^0+g^0)\neq\emptyset$ and the functions $h_{kf}(f^0,g^0)$ and $h_{kg}(f^0,g^0)$, $k\in\mathbf{K}(f^0+g^0)$ computed by the formulas (10) and (11) are bounded, the condition $0\in\partial\Delta_{\mathcal{D}}(f^0,g^0)$ for $\mathcal{D}=\mathcal{D}_{2\delta_1}\times\mathcal{D}_{1\delta_2}$ is satisfied if the components of the densities $f^0(\lambda)$ and $g^0(\lambda)$ satisfy the equations

$$\left|A_k(\lambda)g_k^0(\lambda)+C_k^0(\lambda)\right|^2=\left(f_k^0(\lambda)+g_k^0(\lambda)\right)^2\left(f_k^0(\lambda)-f_k^1(\lambda)\right)\alpha_{k1},\qquad(18)$$

$$\left|A_k(\lambda)f_k^0(\lambda)-C_k^0(\lambda)\right|=\left(f_k^0(\lambda)+g_k^0(\lambda)\right)\psi_k(\lambda)\alpha_{k2},\qquad(19)$$

$$k\in\mathbf{K}(f^0+g^0),$$

where $|\psi_k(\lambda)| \le 1$ a. e. and $\psi_k(\lambda) = \operatorname{sign}\left(g_k^0(\lambda) - g_k^1(\lambda)\right)$ if $g_k^0(\lambda) \ne g_k^1(\lambda)$. The equations (18) and (19) with the extremum condition (6) and the conditions

$$\sum_{k=1}^{\infty} \frac{1}{2\pi} \int_{-\infty}^{\infty} \left|f_k(\lambda) - f_k^1(\lambda)\right|^2 d\lambda = \delta_1, \tag{20}$$

$$\sum_{k=1}^{\infty} \frac{1}{2\pi} \int_{-\infty}^{\infty} \left|g_k(\lambda) - g_k^1(\lambda)\right| d\lambda = \delta_2 \tag{21}$$

determine the least favorable spectral densities.

Theorem 7. Let spectral densities $f^0(\lambda)$ and $g^0(\lambda)$ belong to the set $\mathcal{D}_{2\delta_1} \times \mathcal{D}_{1\delta_2}$, $\mathbf{K}(f^0 + g^0) \ne \emptyset$ and the functions $h_{kf}(f^0, g^0)$ and $h_{kg}(f^0, g^0)$, $k \in \mathbf{K}(f^0 + g^0)$ computed by the formulas (10) and (11) are bounded. The spectral densities $f^0(\lambda)$ and $g^0(\lambda)$ are the least favorable in the class $\mathcal{D}_{2\delta_1} \times \mathcal{D}_{1\delta_2}$ for the optimal linear estimation of the functional $A\xi$ if they satisfy the relations (18) – (21) and determine a solution of the extremum problem (6). The minimax (robust) spectral characteristic of the optimal linear estimate of the functional is computed by the formula (3).

Theorem 8. Let a spectral density $f(\lambda)$ is known, a density $g^0(\lambda)$ belong to the set $\mathcal{D}_{1\delta_2}$, $\mathbf{K}(f + g^0) \ne \emptyset$ and the functions $h_{kg}(f, g^0)$, $k \in \mathbf{K}(f + g^0)$ computed by the formula (11) are bounded. The spectral density $g^0(\lambda)$ is the least favorable in the class $\mathcal{D}_{1\delta_2}$ for the optimal estimation of the functional $A\xi$ if components of the density satisfy the relations

$$g_k^0(\lambda) = \max\left\{g_k^1(\lambda), \alpha_{k2}\left|A_k(\lambda) f_k(\lambda) - C_k^0(\lambda)\right| - f_k(\lambda)\right\}$$

and $(f(\lambda), g^0(\lambda))$ determine a solution of the extremum problem (6). The function $h(f, g^0)$ computed by the formula (3) is the minimax (robust) spectral characteristic of the optimal linear estimate of the functional $A\xi$.

Theorem 9. Let a spectral density $f^0(\lambda)$ belong to the set $\mathcal{D}_{2\delta_1}$, $\mathbf{K}(f^0) \ne \emptyset$ and the functions $h_{kf}(f^0)$, $k \in \mathbf{K}(f^0)$ computed by the formula (5) are bounded. The spectral density $f^0(\lambda)$ is the least favorable in the class $\mathcal{D}_{2\delta_1}$ for the optimal estimation of the functional $A\xi$ from observations of the sequence $\xi(t)$ for $t < 0$ if the components $f_k^0(\lambda)$, $k \in \mathbf{K}(f^0)$ of the density satisfy the equation

$$\left|C_k^0(\lambda)\right|^2 = \left(f_k^0(\lambda)\right)^2 \left(f_k^0(\lambda) - f_k^1(\lambda)\right)\alpha_{k1},$$

and determine a solution of the extremum problem (7). The minimax (robust) spectral characteristic of the optimal linear estimate of the functional is computed by the formula (5).

Bibliography

1. Kolmogorov, A.N., *Theory of Probability and Mathematical Statistics. Selected Papers.* Nauka, Moscow. (1986).

2. Rosanov, Yu.A., *Stationary Random Processes.* Holden-Day, San Francisco. (1967).

3. Vastola, K.S. and Poor, H.V., *An analysis of the effects of spectral uncertainty on Wiener filtering* Automatica, **19**, (1983), 289-293.

4. Franke, J., *Minimax-robust prediction of discrete time series.* Z. Wahrsch. verw. Gebiete. **68**, No. 2, (1985), 337-364.

5. Franke, J. and Poor, H.V., *Minimax-robust filtering and finite length robust predictors. Robust and Nonlinear Time Series Analysis.* Lecture Notes in Statist. **26**, (1984), 87-126.

6. Kassam, S.A. and Poor, H.V., *Robust techniques for signal processing. A Survey.* Proc. IEEE. **73**, (1985), 433-481.

7. Kallianpur, G. and Mandrekar, V., *Multiplicity and representation theory of purely nondeterministic stochastic processes.* Theor. Probab. Appl. **10**, (1965), 614-644.

8. Moklyachuk, M.P., *On an antagonistic game and prediction of stationary random processes in a Hilbert space.* Theor. Probab. Math. Stat. **24**, (1982), 121-127.

9. Moklyachuk, M.P., *Interpolation of vector-valued stochastic processes.* Skorokhod, A. V. (ed.) et al., Exploring stochastic laws. Festschrift in honour of the 70th birthday of Academician Vladimir Semenovich Korolyuk. Utrecht: VSP. (1995), 329-341.

10. Moklyachuk, M.P., *Estimates of stochastic processes from observations with noise.* Theory Stoch. Process. **3(19)**, No.3-4, (1997), 330-338.

11. Pshenichnyi, B.N., *Necessary Conditions for an Extremum.* Marcel Dekker, New York, (1971).

Department of Probability Theory and Mathematical Statistics,
Kyiv National Taras Shevchenko University, Kyiv, Ukraine
E-mail: mmp@mechmat.univ.kiev.ua

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.265-272

ANDRIY Ya. OLENKO AND DMYTRO V. PAVLOV

ON SOME PROPERTIES OF PROBABILITY METRICS

Some estimates of the closeness in different metrics of spectral and correlation functions of random fields are obtained.

2000 *Mathematics Subject Classifications*: 60G60, 60E99.

Key words and phrases. Probability metric, random field, correlation function, spectral function, closeness.

1. Introduction

In many cases mathematical models for spatial phenomenon or images are obtained as particular instances of random fields. Models of this type are often characterized reasonably well by their correlation or spectral functions. There are a lot of numerical methods for estimating the values of correlation or spectral functions. Since we usually have a finite number of observations, it is clear that these methods build estimates only for finite area. For this reason it is important to obtain estimates of the closeness in different metrics of the spectral and correlation functions of random fields. Some results on multidimensional probability metrics were obtained in the paper by Malyarenko (1979), book by Rachev and Rüschendorf (1998) and papers by Olenko (1993, 1996, 1997, 2000) and Pavlov (1999, 2000, 2001).

Denote by $F_\xi(x), F_\eta(x)(x \in \mathbf{R})$ the distribution functions of random variables ξ, η and by $\varphi_\xi(t), \varphi_\eta(t)(t \in \mathbf{R})$ their characteristic functions.

Let us introduce the following probability metrics (see book by Zolotarev (1986)) for one-dimensional distribution functions $F_\xi(x), F_\eta(x)$:

a) Kolmogorov's (uniform) metric:

$$\rho(\xi,\eta) \equiv \rho(F_\xi, F_\eta) = \sup_{x\in\mathbf{R}} |F_\xi(x) - F_\eta(x)|;$$

b) average metric: $\kappa_1(\xi,\eta) \equiv \kappa_1(F_\xi, F_\eta) = \int\limits_{\mathbf{R}} |F_\xi(x) - F_\eta(x)|dx$;

c) weighted uniform metric for characteristic functions:

$$\chi_s(\xi,\eta) \equiv \chi_s(\varphi_\xi, \varphi_\eta) = \sup_{t\in\mathbf{R}} \frac{|\varphi_\xi(t) - \varphi_\eta(t)|}{|t|^s}$$

and

d) modified weighted semi-metric

$$\tilde{\kappa}_{k;a}(\xi,\eta) \equiv \tilde{\kappa}_{k;a}(F_\xi, F_\eta) = \left| \int_0^\infty (F_\xi(\lambda) - F_\eta(\lambda)) \lambda^k e^{-a\lambda} d\lambda \right|.$$

Let $\gamma_1(x), \gamma_2(x)$ $(x \in \mathbf{R}^n)$ be real, measurable, mean-square continuous, homogeneous isotropic random fields (see book by Yadrenko (1983)).

Denote by $B_{n,1}(t)$, $B_{n,2}(t)$ their correlation functions and by $\Phi_{n,1}(\lambda)$, $\Phi_{n,2}(\lambda)$ corresponding spectral functions.

Suppose that $\mathbf{E}\gamma_1(x) = \mathbf{E}\gamma_2(x) = 0, \mathbf{E}\gamma_1^2(x) = \mathbf{E}\gamma_2^2(x) = 1$ for all $x \in \mathbf{R}^n$ and, therefore, $B_{n,1}(0) = B_{n,2}(0) = 1$.

Through entire article we will suppose that $\Phi_{n,1}(\lambda) - \Phi_{n,2}(\lambda)$ is not a singular function.

The following conditions will be considered:

$$(i) \quad \forall r \in [0;H] \quad B_{n,1}(r) = B_{n,2}(r);$$

$$(ii) \quad \forall \lambda \in [0;K] \quad \Phi_{n,1}(\lambda) = \Phi_{n,2}(\lambda);$$

$$(iii) \quad \forall \lambda \geq c \quad \Phi_{n,1}(\lambda) = \Phi_{n,2}(\lambda).$$

Denote $b_\nu = \sup\limits_{z \geq 0} |J_\nu(z)|$, $r_\nu = \inf\{z > 0 \,|\, J_\nu(z) = 0\}$, where J_ν is the first kind Bessel function of the order ν.

Denote for $n > 1, H > 0, y > 0$ and one-dimensional distribution functions $F_1(x), F_2(x)$

$$T_n(H, y, F_1, F_2) =$$
$$= \min\left\{ \tfrac{\pi}{H}; \int_y^{+\infty} |F_1(u) - F_2(u)|\, du + 3(y+2)\sqrt{2}\left(\tfrac{\pi\sqrt{n}}{H}\right)^{\frac{n}{n+1}} \right\}$$

and for $n > 1, H > 0$ and probability density $p(\lambda)$

$$U_n(H,p) = \min\left\{ \frac{48}{\pi H} \cdot \sup_{\lambda \geq 0} |p(\lambda)|; \left(1 + \sup_{\lambda \geq 0} p(\lambda)\right) \cdot \left(\frac{\pi\sqrt{n}}{H}\right)^{\frac{n}{n+1}} \right\}.$$

Denote by $BV([a;b])$ the set of functions of bounded variation on $[a;b]$ and by $V(f;[a;b])$ the variation of the function $f \in BV([a;b])$.

2. Main results

Theorem 2.1. *If $\kappa_1(\Phi_{n,1}, \Phi_{n,2}) < +\infty$ and there exists such $H > 0$ that (i) holds, then for all $y > 0$*

$$\chi_1(B_{n,1}, B_{n,2}) \leq \frac{\Gamma\left(\frac{n}{2}\right) T_n(H, y, \Phi_{n,1}, \Phi_{n,2})}{\Gamma\left(\frac{n+1}{2}\right)\sqrt{\pi}}.$$

Proof. See corollary 3 and theorem 3 in papers by Olenko (1993, 1997) respectively and theorem 1 in the paper by Pavlov (1999).

Theorem 2.2. *Let $\Phi_{n,1} - \Phi_{n,2}$ be a piecewise continuously differentiable function, $n \geq 5$ and there exist $H > 0, K > 0$ such that $(i), (ii)$ hold.*

If $\Phi_{n,1}(\lambda)$ has a bounded density $\Phi'_{n,1}(\lambda)$, then

$$\chi_{\frac{4-n}{2}}\left(B_{n,1}, B_{n,2}\right) \leq \frac{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right) b_{\frac{n}{2}}}{(n-4)\cdot K^{\frac{n-4}{2}}} \cdot U_n(H, \Phi'_{n,1}).$$

Proof. See theorem 2 with remark in the paper by Pavlov (1999) and corollary 2 and theorem 5 in papers by Olenko (1993, 1997) respectively.

Theorem 2.3. *Let η_n be a random variable with the distribution function*

$$F_{\eta_n}(x) = \begin{cases} 1 & \forall x \geq 1 \\ \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\cdot\Gamma\left(\frac{n-1}{2}\right)} \cdot \int\limits_{-1}^{x} (1-u^2)^{\frac{n-3}{2}} du & \forall x \in [-1;1] \\ 0 & \forall x \leq -1 \end{cases},$$

the random variables ξ_1, ξ_2 be independent of η_n, $F_{\xi_1}(0) = F_{\xi_2}(0) = 0$, $\int\limits_0^{+\infty} |F_{\xi_1}(\lambda) - F_{\xi_2}(\lambda)| d\lambda < +\infty$ and there exists such $H > 0$ that for all $t \in [0; H]$ $\varphi_{\xi_1\eta_n}(t) = \varphi_{\xi_2\eta_n}(t)$.

Then for all $y > 0$

$$\frac{\Gamma\left(\frac{n}{2}\right)\tilde{\kappa}_{0;0}(\xi_1, \xi_2)}{\Gamma\left(\frac{n+1}{2}\right)\sqrt{\pi}} \leq \kappa_1\left(\xi_1 \cdot \eta_n, \xi_2 \cdot \eta_n\right) \leq$$

$$\leq \frac{\Gamma\left(\frac{n}{2}\right)\kappa_1\left(\xi_1, \xi_2\right)}{\Gamma\left(\frac{n+1}{2}\right)\sqrt{\pi}} \leq \frac{\Gamma\left(\frac{n}{2}\right) T_n(H, y, F_{\xi_1}, F_{\xi_2})}{\Gamma\left(\frac{n+1}{2}\right)\sqrt{\pi}}.$$

Proof. See theorem 3 and corollary 3 in papers by Olenko (1997, 1993) respectively and corollary in the paper by Pavlov (1999).

Theorem 2.4. *If (iii) holds with $c \in \left[0; r_{\frac{n}{2}}\right]$, then*

$$|B_{n,1}(t) - B_{n,2}(t)| \leq \left| 2^{\frac{n-2}{2}}\Gamma\left(\frac{n}{2}\right) \frac{J_{\frac{n-2}{2}}(ct)}{(ct)^{\frac{n-2}{2}}} - 1 \right| \rho\left(\Phi_{n,1}, \Phi_{n,2}\right)$$

for all $t \in [0;1]$.

Proof. Let us denote $h_n(\lambda) = \Phi_{n,1}(\lambda) - \Phi_{n,2}(\lambda)$, $g_{n,t}(\lambda) = \frac{J_{\frac{n-2}{2}}(\lambda t)}{(\lambda t)^{\frac{n-2}{2}}}$.

Since $c \in \left[0; r_{\frac{n}{2}}\right], t \in [0;1], \lambda \in [0;c], t\lambda \in [0;c]$ and $\frac{J_{\frac{n}{2}}(z)}{z^{\frac{n-2}{2}}}$ is of constant sign on $[0;c]$, then $g_{n+2,t}$ is of constant sign on $[0;c]$ too.

Hence, $g_{n,t}(\lambda)$ is monotonic on $[0;c]$.

The theorem of mean gives us that $\exists \Lambda_{n,t} \in \left[\inf_{[0;c]} h_n; \sup_{[0;c]} h_n\right]$ (i.e. $\exists \Lambda_{n,t} \in [-\rho(\Phi_{n,1}, \Phi_{n,2}); \rho(\Phi_{n,1}, \Phi_{n,2})]$) such that

$$\int_0^c h_n(\lambda) dg_{n,t}(\lambda) = \Lambda_{n,t} \cdot (g_{n,t}(c) - g_{n,t}(0)).$$

If h_n is continuous (though it is not necessary), then $\exists \theta_{n,t} \in [0;c]$ such that $h_n(\theta_{n,t}) = \Lambda_{n,t}$.

$$|B_{n,1}(t) - B_{n,2}(t)| = \left| -2^{\frac{n-2}{2}} \Gamma\left(\frac{n}{2}\right) \int_0^c h_n(\lambda) dg_{n,t}(\lambda) \right| =$$

$$= \left| -2^{\frac{n-2}{2}} \Gamma\left(\frac{n}{2}\right) \cdot \Lambda_{n,t} \cdot (g_{n,t}(c) - g_{n,t}(0)) \right| \leq$$

$$\leq \left| 2^{\frac{n-2}{2}} \Gamma\left(\frac{n}{2}\right) \frac{J_{\frac{n-2}{2}}(ct)}{(ct)^{\frac{n-2}{2}}} - 1 \right| \rho(\Phi_{n,1}, \Phi_{n,2}).$$

Theorem 2.5. *Let there exist* $H > 0$ *and* $c > 0$ *such that* $(i), (iii)$ *hold. Then*

a) for small $t > 0$ *and for all* $y > 0$

$$|B_{n,1}(t) - B_{n,2}(t)| \leq 2^{\frac{n-2}{2}} \Gamma\left(\frac{n}{2}\right) \cdot t \cdot \sup_{\lambda \in [0;c]} \left| \frac{J_{\frac{n}{2}}(\lambda t)}{(\lambda t)^{\frac{n-2}{2}}} \right| \cdot T_n(H, y, \Phi_{n,1}, \Phi_{n,2});$$

b) if $\Phi_{n,1}(\lambda)$ *has a bounded density* $\Phi'_{n,1}(\lambda)$*, then for all* $t \geq 0$

$$|B_{n,1}(t) - B_{n,2}(t)| \leq 2^{\frac{n-2}{2}} \cdot \Gamma\left(\frac{n}{2}\right) \cdot V(g_{n,t}; [0;c]) \cdot U_n(H, \Phi'_{n,1}).$$

Proof. See corollaries 2 and 3 in the paper by Olenko (1993), theorems 3 and 4 in the paper by Pavlov (2001), and theorems 3 and 5 in the paper by Olenko (1997).

Theorem 2.6. *Let* $a > 0, k \in \{0; 1\}, \chi_0 (B_{n,1}, B_{n,2}) < +\infty$ *and*

$$\int_0^{+\infty} \left| (B_{n,1}(r) - B_{n,2}(r)) J_{\frac{n}{2}}(\lambda r) \right| r^{\frac{n}{2}-1} dr < +\infty$$

for all $\lambda > 0$.

Then

$$\tilde{\kappa}_{k;a} (\Phi_{n,1}, \Phi_{n,2}) \leq \frac{\chi_0 (B_{n,1}, B_{n,2})}{a^{k+1}}$$

and

$$\tilde{\kappa}_{k;a} (\Phi_{n,1}, \Phi_{n,2}) \leq \frac{a^k \Gamma\left(\frac{n+1}{2} + k\right)}{2^{\frac{n-1}{2}} \Gamma\left(\frac{n}{2}\right) \sqrt{\pi}} \cdot \int_0^a \frac{|B_{n,1}(r) - B_{n,2}(r)| \, dr}{r^{2k+2}} +$$

$$+ \frac{\Gamma\left(\frac{n+1}{2} + k\right)}{2^{\frac{n-1}{2}} a^{n+1+k} \Gamma\left(\frac{n}{2}\right) \sqrt{\pi}} \cdot \int_a^{+\infty} r^{n-1} |B_{n,1}(r) - B_{n,2}(r)| \, dr$$

provided both integrals exist.

Proof.

$$\tilde{\kappa}_{k;a} (\Phi_{n,1}, \Phi_{n,2}) =$$

$$= \frac{1}{2^{\frac{n-2}{2}} \Gamma\left(\frac{n}{2}\right)} \cdot \left| \int_0^{\infty} \int_0^{\infty} \frac{B_{n,1}(r) - B_{n,2}(r)}{r} J_{\frac{n}{2}}(\lambda r) (\lambda r)^{\frac{n}{2}} \lambda^k e^{-a\lambda} dr d\lambda \right|.$$

Under the theorem assumption it is possible to exchange the order of integration.

Since for $\mathrm{Re}\nu > -\frac{1}{2}, a > 0, b > 0, k \in \{0; 1\}$

$$\int_0^{+\infty} J_\nu (bt) e^{-at} t^{\nu+k} dt = \frac{(2a)^k (2b)^\nu \Gamma\left(\nu + k + \frac{1}{2}\right)}{(a^2 + b^2)^{\nu+k+\frac{1}{2}} \sqrt{\pi}}$$

(see, for example, the book by Watson (1945)), then

$$\tilde{\kappa}_{k;a} (\Phi_{n,1}, \Phi_{n,2}) = \frac{(2a)^k \cdot 2\Gamma\left(\frac{n+1}{2} + k\right)}{\Gamma\left(\frac{n}{2}\right) \sqrt{\pi}} \cdot \left| \int_0^{+\infty} \frac{(B_{n,1}(r) - B_{n,2}(r)) \, r^{n-1} dr}{(a^2 + r^2)^{\frac{n+1}{2}+k}} \right| \leq$$

$$\leq \frac{(2a)^k \cdot 2\Gamma\left(\frac{n+1}{2} + k\right)}{\Gamma\left(\frac{n}{2}\right) \sqrt{\pi}} \cdot \int_0^{+\infty} \frac{r^{n-1} dr}{(a^2 + r^2)^{\frac{n+1}{2}+k}} \cdot \chi_0 (B_{n,1}, B_{n,2}).$$

It is easy to show that for $a>0, n\in\mathbf{N}, k\in\{0;1\}$

$$\int\limits_0^{+\infty}\frac{r^{n-1}dr}{\left(a^2+r^2\right)^{\frac{n+1}{2}+k}}=\frac{1}{a^{2k+1}}\int\limits_0^{\frac{\pi}{2}}\sin^{n-1}(t)\cos^{2k}(t)\,dt=$$

$$=\frac{1}{a^{2k+1}}\cdot\frac{(n-2)!!}{(n-1+2k)!!}\cdot\left(1_{2\mathbf{N}}(n)+\frac{\pi}{2}\cdot 1_{2\mathbf{N}-1}(n)\right),$$

$$\frac{\Gamma\left(\frac{n+1}{2}+k\right)}{\Gamma\left(\frac{n}{2}\right)}\cdot\frac{(n-2)!!}{(n-1+2k)!!}=\frac{\frac{\sqrt{\pi}}{2}\cdot 1_{2\mathbf{N}}(n)+\frac{1}{\sqrt{\pi}}\cdot 1_{2\mathbf{N}-1}(n)}{2^k},$$

where $\mathbf{N}$ is the set of all positive integer numbers and $1_A(x)$ is the indicator function of the set A.

Hence, we get the first inequality of the theorem:

$$\tilde{\kappa}_{k;a}\left(\Phi_{n,1},\Phi_{n,2}\right)\le\frac{\chi_0\left(B_{n,1},B_{n,2}\right)}{a^{k+1}}.$$

Finally,

$$\tilde{\kappa}_{k;a}\left(\Phi_{n,1},\Phi_{n,2}\right)=\frac{(2a)^k\cdot 2\Gamma\left(\frac{n+1}{2}+k\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi}}\cdot\left|\int\limits_0^{+\infty}\frac{\left(B_{n,1}(r)-B_{n,2}(r)\right)r^{n-1}dr}{\left(a^2+r^2\right)^{\frac{n+1}{2}+k}}\right|\le$$

$$\le\frac{(2a)^k\cdot 2\Gamma\left(\frac{n+1}{2}+k\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi}}\left(\frac{1}{2^{\frac{n+1}{2}+k}}\int\limits_0^{a}\frac{\left|B_{n,1}(r)-B_{n,2}(r)\right|dr}{r^{2k+2}}+\right.$$

$$\left.+\frac{1}{(2a^2)^{\frac{n+1}{2}+k}}\int\limits_a^{+\infty}\left|B_{n,1}(r)-B_{n,2}(r)\right|r^{n-1}dr\right)$$

implies the second inequality of theorem.

3. Remarks

1. The following example illustrates that the result of theorem 2.4 cannot be improved by the constant reduction.

Let $\Phi_{n,1}(c,\lambda)=1_{\lambda>c}, \Phi_{n,2}(\varepsilon,\lambda)=1_{\lambda>\varepsilon}$, where $\varepsilon\in(0;c]$.

Then $B_{n,1}(c,t)=2^{\frac{n-2}{2}}\Gamma\left(\frac{n}{2}\right)\frac{J_{\frac{n-2}{2}}(ct)}{(ct)^{\frac{n-2}{2}}}$, $B_{n,2}(\varepsilon,t)=2^{\frac{n-2}{2}}\Gamma\left(\frac{n}{2}\right)\frac{J_{\frac{n-2}{2}}(\varepsilon t)}{(\varepsilon t)^{\frac{n-2}{2}}}$.

By theorem 2.4

$$2^{\frac{n-2}{2}}\Gamma\left(\frac{n}{2}\right)\left|\frac{J_{\frac{n-2}{2}}(ct)}{(ct)^{\frac{n-2}{2}}}-\frac{J_{\frac{n-2}{2}}(\varepsilon t)}{(\varepsilon t)^{\frac{n-2}{2}}}\right|\le\left|2^{\frac{n-2}{2}}\Gamma\left(\frac{n}{2}\right)\frac{J_{\frac{n-2}{2}}(ct)}{(ct)^{\frac{n-2}{2}}}-1\right|$$

for all $t \in [0;1]$.

The left side is continuous with respect to ε. Let us make ε going to 0. We get then the identity of the right and left sides, in other words, the inequality becomes an equality. We cannot get $\varepsilon = 0$ because the theorem requires the condition $\Phi_{n,1}(c, 0+) = \Phi_{n,2}(\varepsilon, 0+)$. Finally, it is obvious that the only $a \in \mathbf{R}$, for which for all $\delta > 0$ there exists $\varepsilon > 0$ such that

$$0 < \left| 2^{\frac{n-2}{2}} \Gamma\left(\frac{n}{2}\right) \frac{J_{\frac{n-2}{2}}(ct)}{(ct)^{\frac{n-2}{2}}} - 1 \right| - a \cdot |B_{n,1}(c,t) - B_{n,2}(\varepsilon,t)| < \delta,$$

is $a = 1$.

Therefore, it is clearly seen that it is impossible to improve the result of the theorem 2.4 by the reduction of the constant.

2. Let us make two remarks to the theorem 2.6.

a) *If there exists such $H > a$ that (i) holds, then*

$$0 \leq \tilde{\kappa}_{k;a}(\Phi_{n,1}, \Phi_{n,2}) \leq$$

$$\leq \frac{\Gamma\left(\frac{n+1}{2} + k\right)}{2^{\frac{n-1}{2}} a^{n+1+k} \Gamma\left(\frac{n}{2}\right) \sqrt{\pi}} \cdot \int_H^{+\infty} r^{n-1} |B_{n,1}(r) - B_{n,2}(r)| \, dr \to 0,$$

if $H \to +\infty$.

b) *By the Stirling formula for all n, k there exist $\theta_1(n,k), \theta_2(n,k) \in (0;1)$ such that*

$$\frac{\Gamma\left(\frac{n+1}{2} + k\right)}{2^{\frac{n-1}{2}} \Gamma\left(\frac{n}{2}\right)} = \frac{\sqrt{2\pi}\left(\frac{n+1}{2} + k\right)^{\frac{n}{2}+k} e^{-\frac{n+1}{2}-k} e^{\frac{\theta_1(n,k)}{6(n+1)+12k}}}{2^{\frac{n-1}{2}} \sqrt{2\pi}\left(\frac{n}{2}\right)^{\frac{n-1}{2}} e^{-\frac{n}{2}} e^{\frac{\theta_2(n,k)}{6n}}} =$$

$$= \frac{\left(\frac{n+1}{2} + k\right)^{k+\frac{1}{2}}}{2^{\frac{n-1}{2}}} \cdot \frac{\left(1 + \frac{2k+1}{n}\right)^{\frac{n-1}{2}}}{e^{k+\frac{1}{2}}} \cdot e^{\frac{\theta_1(n,k)}{6(n+1)+12k} - \frac{\theta_2(n,k)}{6n}} \to 0,$$

if $n \to +\infty$.

Bibliography

1. Zolotarev, V.M., *Contemporary Theory of Summation of Independent Random Variables*, Nauka, Moscow, (1986). (in Russian)

2. Yadrenko, M.I., *Spectral Theory of Random Fields*, Optimization Software Inc., New York (distributed by Springer-Verlag), (1983).

3. Olenko, A.Ya., *On properties of spectral and correlation functions*, 4th World congress of the Bernoulli Society, Abstracts, Vienna, (1996), 363-364.

4. Olenko, A.Ya., *On proximity of the spectral functions of homogeneous isotropic fields*, Theory of Probability and Mathematical Statistics, **46**, (1993), 117-119.

5. Olenko, A.Ya., *On properties of correlation functions with common distribution functions*, Theory of Stochastic Processes, **3(19)**, No. 3-4, (1997), 347-354.

6. Pavlov, D.V., *On some relations for probability metrics in spectral theory of random fields*, Visnyk Kyiv. Univ., Ser. Phys.-Math., **2**, (1999), 135-141. (in Ukrainian)

7. Pavlov, D.V., *On the closeness of correlation functions of homogeneous isotropic random fields which spectral functions coincide on some set*, Ukrainian Mathematical Journal, **53**, No. 1, (2001), 130-134. (in Ukrainian)

8. Olenko, A.Ya. and Pavlov, D.V., *On some estimates for the closeness of correlation and spectral functions of random fields*, National University "Kyiv-Mohyla Academy", Naukovi Zapysky, **18**, Phys.-Math., (2000), 17-19. (in Ukrainian)

9. Malyarenko, A.A., *A generalization of a theorem of Esseen*, Visnyk Kyiv. Univ., Ser. Mech.-Math., **21**, (1979), 84-88. (in Ukrainian)

10. Watson, G.N., *A Treatise of the Theory of Bessel Functions*, (1945).

11. Rachev, S.T. and Rüschendorf, L., *Mass Transportation Problems*, Vol. 1, Springer Verlag, (1998).

Department of Mechanics and Mathematics,
Kyiv National Taras Shevchenko University, 01033, Kyiv, Ukraine.
E-mail: olenk@mechmat.univ.kiev.ua
Department of Mechanics and Mathematics,
Kyiv National Taras Shevchenko University, 01033, Kyiv, Ukraine.
E-mail: dmytrop@yahoo.com

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.273-277

OLEKSANDR PONOMARENKO

THE MASTER EDUCATIONAL PROGRAMME IN THE MATHEMATICAL ECONOMICS AND ECONOMETRICS AT Kyiv NATIONAL TARAS SHEVCHENKO UNIVERSITY

The outline of the Master's Education Programme for specialization "Mathematical Economics and Econometrics@ of new educational direction "Statistics" at Kyiv National University is described.

2000 *Mathematics Subject Classifications.* 97C90

Key words and phrases. Master educational programme, mathematical economics, econometrics.

1. Introduction

In the beginning of the nineties years in connection with the transformation processes in social, political and economical life in Ukraine there were arisen the problems of corresponding transformation in educational system.

The Department of Probability and Mathematical Statistics of Kyiv National Taras Shevchenko University (head of Department prof.M.Yadrenko), Economic and Mathematical Center (president O.Ponomarenko), the Faculty of Mechanics and Mathematics of Kyiv National University (dean prof. M.Perestyuk) made the big work for implementation and development a new educational direction for Higher Educational System of Ukraine "Statistics" together with other mathematical directions as Pure Mathematics and Applied Mathematics. This new direction is oriented mainly on Economics, Financial and Actuarial Applications of statistical, stochastic and mathematical methods and models.

In April 1996 the Council of Kyiv National University decided to introduce since 1996/1997 academic year at Faculty of Mechanics and Mathematics new educational direction "Statistics" as third mathematical direction in education and to begin training the specialists (on three levels of Bachelors, Specialists and Masters) in framework of such educational direction which includes three new economic-statistical specializations:

(i) Mathematical Economics and Econometrics;

(ii) Financial and Actuarial Mathematics;

(iii) Applied Statistics.

The first version of educational programmes for these specializations was worked out by Economic and Mathematical Center in 1994 and the last (second) version (only for of so-called normative courses which are the same for all statistical specializations) was created in 1998. This last version has some non-principal changes and improvements.

The important role in creation of new statistical specializations at Kyiv National University played the project "Mathematical Economics" (1994-1995) which was supported by US AID through "The EuroAsia Foundation". Realization of this project gave the possibilities to write and publish first Ukrainian textbooks on Modern Microeconomic and Macroeconomic Theories, Mathematical Economics, Theory of Social Choice, Applied Statistics, Econometrics, Decision Making in Economics and Management, Actuarial and Financial Mathematics ([1]-[3]). This project also gave possibility to begin organizing of the library of economic and financial literature at the Department of Probability and Statistics of Kyiv National University.

The very important role in further development and updating of curricula for new statistical specializations in Kyiv National University (especially in respect of creation of systems of special professionally- oriented courses) plays the project JEP-10353-97 "Statistical Aspect of Economics" in framework of the TEMPUS-TACIS Programme of European Union.

2. Outline of Programme

The programme presented below follows the general structure of training for higher school in Ukraine. It is supposed that the Master education programme for specialization "Mathematical Economics and Econometrics" is a special part of a general programme for speciality "Statistics". The last programme include three levels of training. The first level is a level of Junior Specialist of Statistics (first two years studies). The second level is a level of Bachelor of Statistics (else two years of studies). Such level include system of obligatory normative courses for all statistical specializations and system of special courses, which is different for different specializations. The third level is a level of Master of Statistics or Specialist of Statistics (else one year of study). The Master Level deals with narrow specialization of students with high educational rating. The other students with Bachelor degree deals with Specialist Level with respect to narrow specialization also.

The structure of first educational concentre (level of Junior Specialist of Statistics) is the following.

I. Basic Mathematical Courses (which are the same as for speciality "Pure Mathematics"): Mathematical Analysis I and II (840 hours); Analytical Geometry (108 hours); Linear Algebra (289 hours); Common Algebra and Number Theory (157 hours); Differential Equations (210 hours); Informatics and Programming (420 hours);

II. Standard Cycle of Humanitarian Disciplines (for all Ukrainian Universities): History of Ukraine (108 hours); History of Ukrainian and World Culture (156 hours); Foreign Language (420 hours); Sport training (244 hours).

III. Professionally-oriented normative courses (which are the same for all statistical specializations): Basic Financial Theory (102 hours); Foundation of Microeconomics and Macroeconomics (108 hours); Financial Analysis and Calculations (108 hours); Discrete Models in Probability Theory and its Applications (102 hours).

The structure of second educational concentre (level of Bachelor of Statistics, 5-8 semesters) is the following.

I. Basic Mathematical Courses: (the same as for speciality "Pure Mathematics"): Theory of Measure and Integral (108 hours); Theory of Probability (135 hours); Mathematical Statistics (102 hours); Mathematical Logic (54 hours); Complex Analysis (210 hours); Functional Analysis (210 hours); Equations of Mathematical Physics (210 hours).

II. The Cycle of Social and Humanitarian Branches: Essentials of State and Law (51 hours); Politology (81 hours); Management and Marketing (48 hours); Psychology (77 hours); History of Mathematics (54 hours); Basic Ecology (48 hours).

III. Normative professionally-oriented courses: Additional Chapters of Probability Theory (77 hours); Mathematical Statistics II (108 hours); Theory of Stochastic Processes (102 hours); Mathematical Economics (153 hours); Methods of Economic and Financial Computations (108 hours); Basic Actuarial Mathematics (158 hours); Stochastic Financial Mathematics (102 hours).

IV. Professionally-oriented special courses of specialization "Mathematical Economics and Econometrics": (Year 3) Methods of Economic and Financial Statistics (54 hours); Essentials of Econometrics (51 hours); Principles of Book-keeping, Financial and Management Accounting (51 hours); (Year 4) History of Economic and Statistical Thoughts (54 hours); Basic Statistics of Stochastic Processes (54 hours); Nonstationary and Nonlinear Time Series (51 hours); Computer Statistics (51 hours).

Structure of Master's Level training of specialization "Mathematical Economics and Econometrics" (fifth year of study) is the following.

I. Cycle of Humanitarian Disciplines: Philosophic Problems of Natural Sciences (48 hours); Foreign Language (96 hours); Sociology (96 hours).

II. Normative professionally-oriented courses: Sampling Survey

(48 hours); Stochastic Models in Management (96 hours); Nonsmooth Analysis and Optimization (96 hours).

III. Special professionally-oriented courses: Modern Statistical Microeconomics (48 hours); Modern Macroeconomic Theory (48 hours); Nonlinear Economic Dynamics (48 hours); Bayesian Methods in Econometrics (48 hours).

IV. Scientific seminars (96 hours).

V. Pedagogical and assistant practice (216 hours).

VI. Qualification thesis (324 hours).

At last time were created some new textbook and manuals for statistical specializations and specialization on "Mathematical Economics and Econometrics" [4] - [7]. Part of them are connected with the project of Economic and Mathematical Center "Analytical and Statistical Models and Methods for Social and Humanitarian Sciences", which was supported by "Renaissance International Foundation" [8]- [10].

In conclusion we give brief sketch of syllabuses for some main courses of specialization in "Mathematical Economics and Econometrics". The detailed syllabuses of all normative and special courses for all statistical specializations were published in special collection books [11] [12] and methodological investigation [13] edited by the author.

1. Mathematical Economics. The course intends to introduce basic concepts and models of mathematical economics: Consumption Models, Theory of Production Function, Theory of firm, Modeling of Markets and Market Economy, Theory of General Economic Equilibrium, Models of Behavior of Economic Agents under Uncertainty, Financial and Insurance Markets, Theory of Economic Welfare and Cooperative Economic Decision, Statistical Leontief's input-output Models, Dynamical Multi-branch Models, Models of Economic Growth, Economic-Ecological Models.

2. Essentials of Econometrics. The course intends to introduce basic concepts and models of econometrics: Two-variable linear regression models (estimation, hypothesis testing, applications), Multiple regression (estimation and hypothesis testing); Functional Forms of Regression Models; Regression of Dummy Explanatory Variables; Regression Analysis in Practice (Multi-collinearity, heteroscedasticity), Autocorrelation, Model Selections: Criteria and Tests; Elements of Nonlinear Regression and its Applications.

3. Modern Statistical Microeconomics. The course intends to introduce basic practical problems of statistical microeconomics: Methodology of Applications of Informational and Statistical Methods in Microeconomics, Express Estimators for Microeconomics characteristics and Indexes, Modern Statistical Methods of Analysis for Market Demand, Forecasting of Microe-

conomic Indexes, Statistical Methods of Marketing Investigations, Analysis of Dynamical Processes in Microeconomics, Statistical Estimation of Microeconomics Parameters based on Bounded Data, Simulation of Microeconomic Indexes and Processes.

Bibliography

1. Ponomarenko O.I., Perestyuk M.O., Burym V.M. *Foundations of Mathematical Economics*, Informtechnika, Kyiv (1995) (in Ukrainian).
2. Leonenko N.N., Mishura Yu.S., Parhomenko V.M., Yadrenko M.I. *Theoretical, Probabilistic and Statistical Methods in Econometrics and Financial Mathematics,* Informtechnika, Kyiv (1995) (in Ukrainian).
3. Ponomarenko O.I., Ponomarenko V.O. *System Methods in Economics, Management and Business,* Lybid', Kyiv (1995) (in Ukrainian).
4. Ponomarenko O.I. *Modern Economic Analysis: Microeconomics,* Vyscha Schkola, Kyiv (2001) (in Ukrainian).
5. Perestyuk M.O., Burym V.M., *Modern Economic Analysis: Macroeconomics,* Vyscha Schkola, Kyiv (2001) (in Ukrainian).
6. Borisenko A.D., Ponomarenko A.I. *Collection of Tests on Microeconomics and Macroeconomics,* EMC, Kyiv (1999) (in Russian).
7. Ponomarenko O.I. *Modern Microeconomic Analysis,* Kyiv National University (2000)
8. Ponomarenko O.I. *Modern Analytical Politology,* Publish Center "Kyiv University", Kyiv (2000) (in Ukrainian).
9. Zinchenko N.M., Olenko A.Ya. *Analytical Models and Methods of Sociology,* Publish Center "Kyiv University", Kyiv (2000) (in Ukrainian).
10. Borisenko O.D., Maiboroda R.E. *Analytical and Statistical Methods and Models of Psychology and Pedagogic,* Publish Center "Kyiv University", Kyiv (2000) (in Ukrainian).
11. *Collection of the programmes of normative courses for higher education with professional direction 6.080102 Statistics.Bachelor level.* (Ed. Ponomarenko O.I.), EMC, Kyiv (1999) (in Ukrainian).
12. *Collection of the programmes of special courses for higher education with professional direction Statistics (6.080102, 7.080102, 8.080102). Bachelor's, Specialist's and Master's Levels.* (Ed. Ponomarenko O.I.), Kyiv University (2000) (in Ukrainian).
13. Ponomarenko O.I. *Master Educational Programmes of Educational direction "Statistics",* EMC, Kyiv (2000) (in Ukrainian).
14. Ponomarenko O.I. *Foundation of the Theory of Finance,* Kyiv University (2001) (in Ukrainian).

Department of Probability Theory and Mathematical Statistics,
Kyiv National Taras Shevchenko University, Kyiv, Ukraine
E-mail: ponomar@mechmat.univ.Kyiv.ua

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.278-290

VICTOR SHPYRKO

THE APPROXIMATIONS OF THE RUIN PROBABILITY IN CLASSICAL RISK MODEL

The problem is devoted to approximated evaluations of the ruin probability for insurance companies. In Classic Risk Model the ruin probability can be directly calculated only for exponential distribution of the insurance claims. Otherwise it is advisable to use approximated estimations of this probability. In this article three of such approximations were used: these are the De Vylder, the Beekman-Bowers and the diffusion approximations.

2000 *Mathematics Subject Classifications.* 91B30

Key words and phrases. Ruin probability, approximation, insurance company.

1. Introduction

The problem is devoted to approximated evaluations of the ruin probability for insurance companies. In Classic Risk Model the ruin probability can be directly calculated only for exponential distribution of the insurance claims. Otherwise it is advisable to use approximated estimations of this probability. In my work three of such approximations were used: these are the De Vylder, the Beekman-Bowers and the diffusion approximations. I estimated the ruin probability of 30 largest Ukrainian insurance companies applying mentioned approximations depending on initial funds of these companies. For such evaluation it is also necessary to know the mean of the insurance claims and relative safety loading of specific company. Comparison of these estimations makes it possible to draw a conclusion about reliability of each insurance company.

2. Classical risk model

The values of payments, which are carried out by insurance company, form the sequence of independent and identically distributed random variables $(Y_k,\ k \geq 1)$ with the distribution function $F(x)$. We assume that

$F(0) = 0$, i.e. risk sums are positive ($Y_k > 0$). The expectation $EY_k = \mu$, the variance $DY_k = \sigma^2$.

The insurance payment takes place when a company receives a claim. Let us make assumptions about the coming of insurance claims:

1) the entrances of insurance claims on nonoverlapping time intervals are independent random events;

2) the number of claims received on interval $[t,\ t+h)$ depends on h and is independent of t;

3) the coming of at the least one claim on interval $[t,\ t+h)$ happens with probability $\alpha h + o(h)$, where $\alpha = const$ and $\lim\limits_{h\to\infty} \frac{o(h)}{h} = 0$;

4) the coming of more than one claim on interval $[t,\ t+h)$ happens with probability $o(h)$.

Let N_t be the number of insurance claims received on $[0,\ t)$ and

$$P_m(t) = P\{N_t = m\}. \tag{1}$$

Under these assumptions the next theorem takes place.

Theorem 1. *Random variable N_t has a Poisson distribution with intensity αt, i.e.*

$$P_m(t) = e^{-\alpha t}\frac{(\alpha t)^m}{m!}, \quad m = 0,\ 1,\ \ldots. \tag{2}$$

The proof of this theorem see for instance in [1].

Since N_t is a Poisson process with intensity αt it follows that $EN_t = \alpha t$.

Random process

$$S_t = \sum_{k=1}^{N_t} Y_k \tag{3}$$

expresses the sum of payments which are carried out by insuranse company on interval $[0,\ t]$ (we consider that $\sum\limits_{k=1}^{0} Y_k = 0$). So we have that the expectation

$$ES_t = EN_t EY_k = \alpha t\mu = \alpha\mu t. \tag{4}$$

The profit of the company on interval $[0,\ t]$ equals to

$$Q_t = ct - S_t, \tag{5}$$

where c is the constant which defines intensity of entrance of insurance premiums. The expectation of this profit is equal to

$$EQ_t = ct - \alpha\mu t = (c - \alpha\mu)t. \tag{6}$$

The relative safety loading ρ is defined as the ratio of company's average profit and average value of payments:

$$\rho = \frac{EQ_t}{ES_t} = \frac{c - \alpha\mu}{\alpha\mu} = \frac{c}{\alpha\mu} - 1. \tag{7}$$

The total capital of the company at moment t equals to

$$U_t = u + ct - S_t, \tag{8}$$

where u is the initial capital of the company.

Let us consider the question about the ruin probability $\psi(u)$ for the insurance company which has the initial capital u on time interval $[0,\ +\infty)$. Than $\psi(u) = P\{U_t < 0 \text{ for some value } t > 0\}$.

Here we have the important result which will be used later on.

Theorem 2. *Let the insurance payments be exponentially distributed variables with the average μ, i.e. density function is*

$$f(x,a) = \begin{cases} ae^{-ax}, & x \geq 0, \\ 0, & x < 0, \end{cases} \tag{9}$$

at that $a = \frac{1}{\mu}$. Then the ruin probability $\psi(u)$ for initial capital u is equal to

$$\psi(u) = \begin{cases} \frac{1}{1+\rho} e^{-\frac{\rho u}{(1+\rho)\mu}}, & if\ c > \alpha\mu, \\ 1, & if\ c \leq \alpha\mu. \end{cases} \tag{10}$$

3. Asymptotic behavior of the ruin probability when $u \to +\infty$

Let us analyse asymptotic behavior of the ruin probability $\psi(u)$ on interval $[0,\ +\infty)$ for initial capital u when $u \to +\infty$**.**

Put

$$\overline{\mu} = \frac{\alpha}{c} \int_0^{+\infty} ye^{Ry} \left[1 - F(y)\right] dy. \tag{11}$$

Then the next theorem exists.

Theorem 3. *Let $\frac{\alpha\mu}{c} < 1$, equation*

$$\frac{\alpha}{c} \int_0^{+\infty} e^{Ry} \left[1 - F(y)\right] dy = 1 \tag{12}$$

has a root R and $\overline{\mu} < +\infty$. Then if $u \to +\infty$

$$\psi(u) \sim \frac{\rho}{(1+\rho) R\overline{\mu}} e^{-Ru}. \tag{13}$$

Now we will examine the existence conditions for the root of equation (12). Let

$$h(r) = \int_0^{+\infty} e^{rz} dF(z) - 1. \tag{14}$$

Let us make a **supposition**. There exists such value of $r_\infty > 0$ that $h(r) \uparrow +\infty$, when $r \uparrow r_\infty$ (it is allowed that $r_\infty = +\infty$).

Under this supposition the equation (14) may be written as

$$h(R) = \frac{c}{\alpha} R. \tag{15}$$

Lemma 1. *On these assumptions the equation (15) has the single root R, at that* $R < r_\infty$.

Thus, the theorem 3 may be formulated in such a way.

Theorem 4. *(Cramer-Lundberg Theorem). Under the assumptions about h(r) and if* $u \to +\infty$

$$\psi(u) \sim \frac{\rho\mu}{h'(R) - \frac{c}{\alpha}} e^{-Ru}, \tag{16}$$

where R is the root of equation (15).

The right-hand member of (16) is called the Cramer-Lundberg approximation.

Notice that in the case of exponentially distributed claims the Cramer-Lundberg approximation is exact, see [1].

Let us assume that $c > \alpha\mu$ (if $c \leq \alpha\mu$, then company's ruin takes place with probability 1).

We can point at upper estimate for probability $\psi(u)$ which is valid for any $u > 0$.

Theorem 5. *Let equation (12) have a positive root R. Then for any* $u > 0$ *the inequation (17) holds true:*

$$\psi(u) \leq e^{-Ry}. \tag{17}$$

The inequality (17) is called *the Cramer-Lundberg inequality*, and number R is *the Lundberg coefficient.*

4. Some approximations for the ruin probability in classical risk model

In practice it is rather difficult to calculate the exact values of the ruin probability, unless the payments are exponentially distributed (in this case accurate formula (10) for $\psi(u)$ applies).

So it is natural to use approximations for $\psi(u)$. Three of such approximations are given below.

4.1. The Beekman-Bowers approximation for $\psi(u)$.

Let

$$H(u) = P\left\{\inf_{t\geq 0} Q(t) < -u / \inf_{t\geq 0} Q(t) < 0\right\} \tag{18}$$

Then

$$H(u) = 1 - (1+\rho)\,\psi(u), \tag{19}$$

whence

$$\psi(u) = \frac{1}{1+\rho}\left[1 - H(u)\right]. \tag{20}$$

Let μ_H and σ_H^2 are the expectation and the variance of the distribution $H(u)$. The idea of the approximate formula lies in replacement of $H(u)$ in (20) for Gamma distribution $G(u)$, the first two moments of which coincide with the moments of $H(u)$.

Then the approximated formula for $\psi(u)$ is

$$\psi_{BB}(u) = \frac{1}{1+\rho}\left[1 - G(u)\right]. \tag{21}$$

Denote the k-th moment of the distribution function $F(y)$ of payments as μ_k, i.e.

$$\mu_k = EY_1^k, \quad k = 1,\ 2,\ 3. \tag{22}$$

Using the Laplace-Stieltjes transformation of function $F(y)$ we can define variables μ_H and σ_H^2 with the moments of function $F(y)$, see [1]:

$$\mu_H = \frac{\mu_2(1+\rho)}{2\rho\mu_1} \tag{23}$$

$$\sigma_H^2 = \frac{\mu_2(1+\rho)}{2\rho\mu_1}\left[\frac{2}{3}\frac{\mu_3}{\mu_2} + \frac{\mu_2(1+\rho)}{2\rho\mu_1}\right]. \tag{24}$$

An algorithm to apply formula (21) follows below:

1) we find first three moments $\mu_1,\ \mu_2,\ \mu_3$ of the distribution function $F(y)$;

2) we calculate the value of the ruin probability $\psi(u)$ by formula (21) using the Gamma-distribution table for the expectation μ_H and the variance σ_H^2.

When the payments are exponentially distributed, we can determine that the approximate formula (21) is accurate.

4.2. The De Vylder approximation. We approximate the process $Q(t)$ in general Classical Risk Model by the process $\tilde{Q}(t)$ so that the payments of process $\tilde{Q}(t)$ have exponential distribution and

$$EQ^k(t) = E\tilde{Q}^k(t)\, for k = 1, 2, 3. \tag{25}$$

Knowing the exact formula for the ruin probability $\psi_{DV}(u)$, we take $\psi_{DV}(u)$ in process $\tilde{Q}(t)$ for the ruin probability $\psi(u)$.

The risk process $\tilde{Q}(t)$ is determined by the three parameters $(\tilde{\alpha}, \tilde{c}, \tilde{\mu})$ or $(\tilde{\alpha}, \tilde{\rho}, \tilde{\mu})$, taking into account that $\tilde{\rho} = \frac{\tilde{c}}{\tilde{\alpha}\tilde{\mu}} - 1$.

Then the values of

$$\tilde{\mu} = \frac{\mu_3}{3\mu_2}, \tilde{\rho} = \frac{2\mu_1\mu_3}{3\mu_2^2}\rho, \tilde{\alpha} = \frac{9\mu_2^3}{2\mu_2^2}\alpha. \tag{26}$$

Thus the value of the ruin probability

$$\psi(u) \equiv \psi_{DV}(u) = \frac{1}{1+\tilde{\rho}} a^{-\frac{u\tilde{\rho}}{\tilde{\mu}(1+\tilde{\rho})}}. \tag{27}$$

By construction of the approximation it follows that $\psi_{DV}(u) = \psi(u)$ in the case of exponentially distributed claims.

4.3. The diffusion approximation. Let D be space of right continuous functions on $[0, \infty)$ and limit on left existing (space of functions without nonremovable discontinuities).

Definition. Consecution X_n converges distributionwise to the random process X (we will write: $X_n \xrightarrow{d} X$) if for any bounded and continuous function f on space D

$$Ef(X_n) \to Ef(X). \tag{28}$$

Using the idea of convergence distributionwise, such diffusion approximation for $\psi(u)$ can be determined as:

$$\psi(u) \tilde{} \psi_D(u) = e^{-u\rho\frac{2\mu}{\mu^2+\sigma^2}}. \tag{29}$$

5. Comparison of the approximations

As it was shown above, under certain assumptions there exists number R (the Lundberg coefficient) such that $\psi(u) \leq e^{-Ru}$.

In the case of the diffusion approximation (29) it is natural to consider

$$R_D = \frac{2\rho\mu}{\mu^2 + \sigma^2} = \frac{2\mu_1}{\mu_2}\rho, \tag{30}$$

as the diffusion approximation of R.

In the same way for the De Vylder approximation (27) we put

$$R_{DV} = \frac{\tilde{\rho}}{\tilde{\mu}(1+\tilde{\rho})} = \frac{2\mu_1}{\mu_2 + \frac{2\mu_1\mu_3}{3\mu_2}}\rho \tag{31}$$

and also consider R_{DV} as the approximation of the Lundberg coefficient.

The Beekman-Bowers approximation (21) is not exponential but it is possible to define

$$R_{BB} = \frac{\mu_B}{\sigma_B^2} = \frac{2\mu_1}{\mu_2 + \left(\frac{4\mu_1\mu_3}{3\mu_2} - 1\right)\rho}\rho \tag{32}$$

For each of the three ruin probability approximations we consider the relative error of the approximation. For example,

$$\varepsilon_D(u) = \frac{\psi_D(u) - \psi(u)}{\psi(u)} \tag{33}$$

is the relative error of the diffusion approximation.

Let us consider an example.

Example 1.
The payments are Gamma-distributed with the mean $\mu = 1$ and the variance σ^2 =100. We assume that $\rho = 10\%$. Then $\mu_1 = 1$, $\mu_2 = 101$, $\mu_3 = 20301$. Exact values of $\psi(u)$ are given in [4]. The values of $\psi(u)$ for different values of u and the relative errors for all three approximations are listed below. Also notice that R = 0,0017450, R_{BB} = 0,0016992, R_D = 0,0019802, R_{DV} = 0,0017483.

u	$\psi(u)$	ε_D	ε_{BB}	ε_{DV}
300	0,52114	5,9%	-0,1%	0,3%
600	0,30867	-1,3%	-0,8%	0,2%
900	0,18287	-8,0%	-0,9%	0,1%
1200	0,10834	-14,3%	-0,7%	0,0%
1500	0,06418	-20,1%	-0,2%	-0,01%
1800	0,03803	-25,5%	0,3%	-0,2%

2100	0,02253	-30,6%	1,0%	-0,3%
2400	0,01335	-35,4%	1,8%	-0,4%
2700	0,00791	-39,8%	2,7%	-0,5%
3000	0,00468	-43,8%	3,6%	-0,5%

From the table we can see that the De Vylder approximation is the best of the three approximations for the given example. The examination of other practical cases also confirms advantages of the De Vylder approximation.

6. Practical application of the ruin probability approximations for Ukrainian insurance market

In applications of classical risk model the intensity α and the distribution function $F(y)$ (or its moments μ_1, μ_2, μ_3) are the parameters of the model, ρ and u being the variables that influence the degree of risk. If the ruin probability p is determined, it is natural to choose ρ and u so that $\psi(u) = p$. In practice $\psi(u)$ is replaced by one of the approximations. Using the diffusion approximation we get

$$u = \frac{-lnp}{R_D}, \tag{34}$$

and applying the De Vylder approximation we have

$$u = \frac{-lnp - ln(1+\tilde{\rho})}{R_{DV}}. \tag{35}$$

It is slightly more difficult to define the necessary value of the initial capital u, using the Beekman-Bowers approximation. In that case it is equal to

$$u = G_{inv}(1 - p(1+\rho)), \tag{36}$$

where G_{inv} is the inverse function of Gamma distribution with the parameters μ_H and σ_H^2, which are defined from the equations (23) and (24). The expression $(1 - p(1+\rho))$ is the argument of this function.

As it was shown above, for the estimation of the ruin probability it is necessary to know the distribution function of the payments or its first three moments. Certainly, for each insurance company we can construct only the sum polygon, thereto we should have a sample of the insurance payments.

In contrast to commercial banks the information about insurance companies' activities is not available in Ukraine. Granting this, during the calculation of the ruin probability approximations we made certain assumptions about the distribution function and its parameters.

6.1. The exponential distribution. First we assume that the insurance payments are exponentially distributed with the density function (9). Then if random variable ς has the exponential distribution, its expectation $E\varsigma = \frac{1}{\alpha}$ and the variance $D\varsigma = \frac{1}{\alpha^2}$.

As it was shown in item 1, in the case of exponentially distributed payments we can define explicit function (10) for $\psi(u)$. For calculation of $\psi(u)$ we should know only the mean value of payments μ and relative safety loading ρ. The results of calculations of the ruin probability $\psi(u)$ for 30 largest Ukrainian insurance companies are given in table 1. At that the value of the initial capital was defined by the next formula:

$$u = initialfund + insurancefunds. \tag{37}$$

Starting from the rough mean of an insurance payment and the value of relative safety loading, the ruin probability was calculated for the mean of payment $\mu = 500$, 1000, 1500 and 2000 UAH (Ukrainian hryvna) and the relative safety loading $\rho = 30\%$, 40% and 50% (12 combinations of values in all).

Obviously, on increase of the mean of payments μ the ruin probability $\psi(u)$ increases and on increase of the relative safety loading ρ it decreases. For example, the initial capital u of the insurance company "UASK ASKA" makes up 10975,2 thousand of UAH. Then for $\rho = 30\%$ and $\mu = 1000$ UAH the ruin probability $\psi(u) = 6{,}11\%$. If μ rises to 1500 UAH, $\psi(u)$ increases to 14,22%; if $\mu = 1000$ UAH and ρ rises to 40%, then the ruin probability $\psi(u)$ reduces to 3,10%.

If an insurance company wants to get the value of the ruin probability for values μ and ρ, which are not stated in the table (e.g. $\mu = 1200$ UAH and $\rho = 35\%$), the electronic table developed in Microsoft Excel gives the possibility to obtain the necessary results.

Under the values of $\psi(u)$ for 30 companies for each of the 12 combinations of average payments and relative safety loadings there is the minimum initial capital (in thousands of UAH), which is necessary for insurance company in order that the ruin probability does not exceed 5% or 1%. These items are calculated using the next formula:

$$u = -\frac{(1+\rho)\,\mu}{\rho} ln\left(p\,(1+\rho)\right), \tag{38}$$

where p is the required ruin probability (5% or 1%). This formula is inverse to the formula (10). Knowing the value of the minimum initial capital defined by (38) the insurance company can determine the lowest level of this item necessary to hold the ruin probability on the safe level if $\psi(u)$ was less than the required value. Also the company can define the necessary increase of the initial capital or the relative safety loading in order to guarantee the safe level of $\psi(u)$ if the value of $\psi(u)$ was more than this safe level. For

example, the initial capital u of the insurance company "Avante" amounts 4844,3 thousand of UAH. Then for $\rho = 30\%$ and $\mu = 500$ UAH the ruin probability $\psi(u) = 8{,}22\%$. In order to obtain the ruin probability $\psi(u) = 5\%$ the company should increase its initial capital to 5922,3 thousand of UAH or it should rise the safety loading ρ to 40% (in this case the value of $\psi(u)$ becomes equal to 4,48%).

We should note that using the electronic table a company can compute the minimum initial capital for any other level of the ruin probability (for instance, 3%, 10% etc.).

6.2. Gamma distribution

Let us consider a situation of non-exponentially distributed insurance payments. Gamma distribution is one of the most popular distributions.

The density function of Gamma distribution is

$$f(x,\alpha,\beta) = \begin{cases} \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}, & x \geq 0, \\ 0, & x < 0, \end{cases} \tag{39}$$

at that the expectation of the random variable ς which has Gamma distribution (39) is equal to $E\varsigma = \alpha\beta$ and its variance $D\varsigma = \alpha\beta^2$.

Notice that when $\alpha = 1$ the distribution (39) is exponential with parameter $\frac{1}{\beta}$ and when $\beta = 1$ the distribution (39) is called the standard Gamma distribution. The density function of the standard Gamma distribution is

$$f(x,\alpha) = \begin{cases} \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}, & x \geq 0, \\ 0, & x < 0, \end{cases} \tag{40}$$

In this case $E\varsigma = D\varsigma = \alpha$.

As it was stated above, in practice it is impossible to get the exact value of the ruin probability for non-exponentially distributed payments. That is why we use three ruin probability approximations: the Beekman-Bowers, the De Vylder and the diffusion approximation. The algorithm of calculating these approximations and formulas (21), (27), (29) applied for calculation of estimates are given in item 4.

The values of the three ruin probability approximations for 30 largest Ukrainian insurance companies in case of standard Gamma distributed payments are listed in table 2. The relative safety loading equals 30%, and the mean of payments $\mu = 500$, 1000, 1500, 2000 UAH. As it was shown above the standard Gamma distribution has only one parameter α, which is equal to the mean of payments, so the data about the safety loading and the mean of payments are enough to compute all the three approximations.

The same way as in the situation of exponential distribution under the values of $\psi(u)$ there is the minimum initial capital (in thousands of UAH), which is necessary for insurance company in order that the ruin probability

does not exceed 5% or 1%. These items were calculated with the use of all the three approximations by the formulas (34), (35), (36).

Analysing the received results we can note that in all cases three approximations give almost the same results if the ruin probability is close to 20%. As it was stated above the De Vylder approximation works the best as a rule. Also the company may take into consideration the approximation which gives the largest value of the ruin probability (i.e. the worst situation for the insurance company).

At last we should note that if it is necessary to use some other (not exponential and not Gamma) distribution function of payments while calculating the ruin probabilities, then it is enough to define the first three moments of this function and to change corresponding formulas in the electronic table. All ruin probabilities will be automatically recalculated according to the new data.

Bibliography

1. Leonenko M., Mishura Y., Parkhomenko V., Yadrenko M. *Probability-theoretic methods in econometrics and financial mathematics*, Kyiv, 1995.
2. Beekman J. *A ruin function approximation*, Trans. of the Soc. of Actuaries, **21** (1969), 41-48 and 275-279.
3. De Vylder F. *A practical solution to the problem of ultimate ruin probability*, Scand. Actuarial Journ, (1978), 114-119.
4. Grandell J., Segerdahl C.-O. *A comparison of some approximations of ruin probabilities*, Skand. AktuarTidskr., (1971), 144-158.
5. A.Straub. *Non-Life Insurance Mathematics.* Springer Verlag, 1988.

Department of economic cybernetics, economics faculty, Kiev National Taras Shevchenko University, Kiev, Ukraine.
E-mail: viksh@ukr.net

Table 1. Estimations of insurance companies' ruin probability (exponential distribution)

Num	Insurance Company	Initial fund, thousands of UAH	Insurance funds, thousands of UAH	Initial capital u, thousands of UAH	Relative safety loading 30%				Relative safety loading 40%				Relative safety loading 50%			
					Average payments, UAH				Average payments, UAH				Average payments, UAH			
					500	1000	1500	2000	500	1000	1500	2000	500	1000	1500	2000
1	DASK	1120,2	100700,2	101820,4	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
2	NASK "Oranta"	22323,6	42741,4	65065,0	0,00%	0,00%	0,00%	0,04%	0,00%	0,00%	0,00%	0,01%	0,00%	0,00%	0,00%	0,00%
3	"AKV Garant"	1500,0	41507,1	43007,1	0,00%	0,00%	0,10%	0,54%	0,00%	0,00%	0,02%	0,15%	0,00%	0,00%	0,00%	0,05%
4	"Omega"	2000,0	28911,7	30911,7	0,00%	0,06%	0,66%	2,17%	0,00%	0,01%	0,20%	0,86%	0,00%	0,00%	0,07%	0,39%
5	"Zahid Reserv"	2250,0	23802,4	26052,4	0,00%	0,19%	1,40%	3,81%	0,00%	0,04%	0,50%	1,73%	0,00%	0,01%	0,20%	0,87%
6	"Ukr. Neftegas."	1550,0	22401,0	23951,0	0,00%	0,31%	1,93%	4,85%	0,00%	0,08%	0,75%	2,33%	0,00%	0,02%	0,33%	1,23%
7	"Lemma"	10530,0	6595,3	17125,3	0,03%	1,48%	5,52%	10,66%	0,00%	0,54%	2,74%	6,19%	0,00%	0,22%	1,48%	3,84%
8	"Kredo-Klassic"	1591,1	14177,1	15768,2	0,05%	2,02%	6,80%	12,47%	0,01%	0,79%	3,54%	7,51%	0,00%	0,35%	2,01%	4,81%
9	"Garant-Avto"	2433,5	12071,5	14505,0	0,10%	2,71%	8,26%	14,43%	0,02%	1,13%	4,51%	8,99%	0,00%	0,53%	2,65%	5,94%
10	"Ostra-Kyiv"	3600,0	7453,1	11053,1	0,47%	6,00%	14,05%	21,49%	0,13%	3,04%	8,70%	14,73%	0,04%	1,67%	5,72%	10,56%
11	"UASK ASKA"	6802,6	4172,6	10975,2	0,49%	6,11%	14,22%	21,68%	0,13%	3,10%	8,83%	14,89%	0,04%	1,72%	5,82%	10,70%
12	"Rostok"	1112,0	6217,1	7329,1	2,61%	14,17%	24,91%	33,02%	1,08%	8,80%	17,68%	25,07%	0,50%	5,79%	13,08%	19,65%
13	"Ukrgasprompolys"	2000,0	5145,6	7145,6	2,84%	14,79%	25,62%	33,73%	1,20%	9,27%	18,31%	25,74%	0,57%	6,16%	13,62%	20,26%
14	"Alkona"	1198,9	4974,8	6173,7	4,45%	18,51%	29,76%	37,73%	2,10%	12,24%	22,04%	29,57%	1,09%	8,51%	16,91%	23,83%
15	"Ukr. Transportnaya"	1262,7	4336,0	5598,7	5,81%	21,13%	32,51%	40,32%	2,91%	14,43%	24,59%	32,10%	1,60%	10,31%	19,21%	26,22%
16	INTO	1300,0	3697,0	4997,0	7,66%	24,28%	35,66%	43,22%	4,11%	17,13%	27,57%	34,98%	2,38%	12,60%	21,96%	28,99%
17	"Ukrinmedstrah"	1150,0	3843,3	4993,3	7,68%	24,30%	35,68%	43,24%	4,12%	17,15%	27,59%	35,00%	2,39%	12,62%	21,98%	29,01%
18	"Energopolys"	3360,0	1529,0	4889,0	8,06%	24,89%	36,26%	43,76%	4,37%	17,67%	28,15%	35,53%	2,56%	13,07%	22,49%	29,51%
19	"Avante"	1998,0	2846,3	4844,3	8,22%	25,15%	36,51%	43,98%	4,48%	17,90%	28,39%	35,75%	2,64%	13,26%	22,72%	29,73%
20	"Ukr. pozharno-strah."	1500,0	3208,1	4708,1	8,76%	25,95%	37,28%	44,68%	4,85%	18,61%	29,13%	36,46%	2,89%	13,88%	23,42%	30,42%
21	"Euroreserv"	1669,1	2513,3	4182,4	11,16%	29,30%	40,42%	47,48%	6,55%	21,62%	32,20%	39,30%	4,10%	16,54%	26,32%	33,20%
22	"Galinstrah"	1300,0	2621,0	4121,0	11,48%	29,72%	40,81%	47,81%	6,78%	22,00%	32,58%	39,65%	4,27%	16,88%	26,68%	33,54%
23	"Velta"	1330,5	2640,2	3970,7	12,31%	30,77%	41,76%	48,65%	7,39%	22,97%	33,53%	40,51%	4,72%	17,75%	27,59%	34,40%
24	"Kuby insurance"	2140,1	1742,2	3882,3	12,82%	31,40%	42,33%	49,15%	7,77%	23,56%	34,10%	41,02%	5,01%	18,28%	28,13%	34,91%
25	"Inter-polys"	1207,0	2463,0	3670,0	14,14%	32,98%	43,74%	50,37%	8,77%	25,03%	35,50%	42,28%	5,77%	19,62%	29,49%	36,16%
26	"Tekom"	1057,8	2495,1	3552,9	14,92%	33,88%	44,53%	51,05%	9,38%	25,88%	36,30%	43,00%	6,24%	20,40%	30,27%	36,88%
27	"Oranta-Donbass"	1009,0	2361,0	3370,0	16,24%	35,34%	45,80%	52,14%	10,41%	27,27%	37,59%	44,14%	7,05%	21,68%	31,53%	38,02%
28	"Krym-Oranta"	1280,0	1762,3	3042,3	18,89%	38,12%	48,17%	54,15%	12,56%	29,95%	40,01%	46,25%	8,77%	24,18%	33,91%	40,15%
29	"Dysko"	1143,3	1492,1	2635,4	22,79%	41,87%	51,28%	56,75%	15,84%	33,64%	43,24%	49,02%	11,50%	27,69%	37,12%	42,97%
30	"Busyn"	1100,0	1406,4	2506,4	24,19%	43,14%	52,31%	57,60%	17,06%	34,90%	44,31%	49,93%	12,54%	28,91%	38,20%	43,90%
The minimum initial capital (thousands of UAH), which is necessary for insurance company in order to reach the value of ruin probability				5%	5922,3	11844,6	17766,9	23689,2	4653,7	9307,4	13961,1	18614,8	3885,4	7770,8	11656,2	15541,6
				1%	9409,4	18818,8	28228,2	37637,7	7470,2	14940,4	22410,7	29880,9	6299,6	12599,1	18898,7	25198,2

Table 2. Estimations of insurance companies' ruin probability (standard gamma distribution)

Relative safety loading 30%

Num	Insurance Company	Initial fund, thousands of UAH	Insurance funds, thousands of UAH	Initial capital u, thousands of UAH	Average payments 500			Average payments 1000			Average payments 1500			Average payments 2000		
					Approximations			Approximations			Approximations			Approximations		
					BB	DV	dif	BB	DV	dif	BB	DV	dif	BB	DV	dif
1	DASK	1120,2	100700,2	101820,4	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
2	NASK "Oranta"	22323,6	42741,4	65065,0	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,01%	0,00%	0,00%	0,05%	0,01%	0,00%
3	"AKV Garant"	1500,0	41507,1	43007,1	0,00%	0,00%	0,00%	0,01%	0,00%	0,00%	0,10%	0,02%	0,00%	0,36%	0,13%	0,00%
4	"Omega"	2000,0	28911,7	30911,7	0,00%	0,00%	0,00%	0,06%	0,01%	0,00%	0,42%	0,16%	0,00%	1,13%	0,67%	0,01%
5	"Zahid Reserv"	2250,0	23802,4	26052,4	0,00%	0,00%	0,00%	0,15%	0,04%	0,00%	0,78%	0,40%	0,00%	1,83%	1,30%	0,04%
6	"Ukr. Nefregas."	1550,0	22401,0	23951,0	0,00%	0,00%	0,00%	0,23%	0,07%	0,00%	1,03%	0,58%	0,01%	2,26%	1,74%	0,08%
7	"Lemma"	10530,0	6595,3	17125,3	0,04%	0,00%	0,00%	0,82%	0,43%	0,00%	2,54%	2,02%	0,11%	4,61%	4,40%	0,59%
8	"Kredo-Klassic"	1591,1	14177,1	15768,2	0,06%	0,01%	0,00%	1,07%	0,62%	0,01%	3,06%	2,59%	0,18%	5,34%	5,30%	0,88%
9	"Garant-Avto"	2433,5	12071,5	14505,0	0,09%	0,02%	0,00%	1,37%	0,87%	0,02%	3,64%	3,25%	0,30%	6,14%	6,29%	1,29%
10	"Ostra-Kyiv"	3600,0	7453,1	11053,1	0,32%	0,11%	0,00%	2,73%	2,23%	0,13%	5,98%	6,10%	1,21%	9,12%	10,07%	3,64%
11	"UASK ASKA"	6802,6	4172,6	10975,2	0,33%	0,11%	0,00%	2,78%	2,28%	0,14%	6,05%	6,18%	1,24%	9,20%	10,18%	3,72%
12	"Rostok"	1112,0	6217,1	7329,1	1,33%	0,84%	0,02%	6,04%	6,16%	1,24%	10,63%	12,00%	5,34%	14,48%	16,74%	11,11%
13	"Ukrgasprompolys"	2000,0	5145,6	7145,6	1,43%	0,92%	0,02%	6,29%	6,48%	1,38%	10,96%	12,41%	5,75%	14,83%	17,16%	11,73%
14	"Alkona"	1198,9	4974,8	6173,7	2,10%	1,57%	0,06%	7,84%	8,45%	2,47%	12,88%	14,80%	8,48%	16,91%	19,60%	15,71%
15	"Ukr. Transportnaya"	1262,7	4336,0	5598,7	2,66%	2,15%	0,12%	8,97%	9,88%	3,49%	14,22%	16,44%	10,67%	18,33%	21,19%	18,66%
16	INTO	1300,0	3697,0	4997,0	3,41%	2,98%	0,25%	10,36%	11,64%	5,00%	15,82%	18,34%	13,57%	20,00%	23,01%	22,35%
17	"Ukrinmedstrah"	1150,0	3843,3	4993,3	3,41%	2,99%	0,25%	10,36%	11,66%	5,01%	15,83%	18,35%	13,59%	20,01%	23,02%	22,37%
18	"Energopolys"	3360,0	1529,0	4889,0	3,57%	3,16%	0,29%	10,63%	11,99%	5,34%	16,13%	18,70%	14,17%	20,32%	23,35%	23,09%
19	"Avante"	1998,0	2846,3	4844,3	3,63%	3,24%	0,30%	10,75%	12,14%	5,48%	16,27%	18,85%	14,42%	20,46%	23,49%	23,40%
20	"Ukr. pozharno-strah."	1500,0	3208,1	4708,1	3,85%	3,49%	0,36%	11,11%	12,60%	5,95%	16,68%	19,32%	15,23%	20,88%	23,93%	24,37%
21	"Euroreserv"	1669,1	2513,3	4182,4	4,82%	4,65%	0,67%	12,67%	14,54%	8,15%	18,39%	21,26%	18,79%	22,63%	25,71%	28,53%
22	"Galinstrah"	1500,0	2621,0	4121,0	4,95%	4,81%	0,72%	12,87%	14,79%	8,46%	18,61%	21,50%	19,26%	22,85%	25,93%	29,06%
23	"Velta"	1330,5	2640,2	3970,7	5,28%	5,22%	0,86%	13,38%	15,41%	9,25%	19,15%	22,10%	20,45%	23,40%	26,46%	30,40%
24	"Kuby insurance"	2140,1	1742,2	3882,3	5,49%	5,48%	0,96%	13,68%	15,78%	9,76%	19,49%	22,46%	21,18%	23,73%	26,78%	31,22%
25	"Inter-polys"	1207,0	2463,0	3670,0	6,03%	6,15%	1,23%	14,46%	16,72%	11,08%	20,31%	23,34%	23,06%	24,56%	27,57%	33,27%
26	"Tekom"	1057,8	2495,1	3552,9	6,36%	6,56%	1,42%	14,92%	17,27%	11,89%	20,79%	23,84%	24,17%	25,04%	28,02%	34,46%
27	"Oranta-Donbass"	1009,0	2361,0	3370,0	6,90%	7,25%	1,77%	15,66%	18,15%	13,27%	21,57%	24,65%	26,00%	25,81%	28,72%	36,40%
28	"Krym-Oranta"	1280,0	1762,3	3042,3	8,02%	8,66%	2,62%	17,13%	19,85%	16,15%	23,09%	26,16%	29,64%	27,30%	30,04%	40,16%
29	"Dysko"	1143,3	1492,1	2635,4	9,71%	10,82%	4,26%	19,23%	22,18%	20,60%	25,20%	28,17%	34,87%	29,37%	31,75%	45,37%
30	"Busyn"	1100,0	1406,4	2506,4	10,33%	11,61%	4,97%	19,97%	22,97%	22,26%	25,94%	28,84%	36,72%	30,08%	32,31%	47,16%
The minimum initial capital (thousands of UAH), which is necessary for insurance																
company in order to reach the value of ruin probability				**5%**	4097,4	4049,9	2501,4	8189,6	8096,6	4997,9	12281,8	12143,3	7494,3	16374,1	16189,9	9990,8
				1%	8055,1	7000,0	3845,3	16105,2	13997,3	7683,0	24155,4	20994,7	11520,6	32205,3	27992,0	15358,2

The value of average payments is indicated in UAH. BB - the Beekman-Bowers approximation, DV - the De Vylder approximation, dif - the diffusion approximation

Theory of Stochastic Processes
Vol.7 (23), no.1-2 ,2001, pp.291-300

DMITRII S. SILVESTROV, VIKTOR G. GALOCHKIN,
AND ANATOLIY A. MALYARENKO

OPTAN — A PILOT PROGRAM SYSTEM FOR ANALYSIS OF OPTIONS[1][2]

We present a project, which intend the elaboration of a pilot program system OptAn (OPTion ANalyzer) for analysis of options of different types. The system is based on advanced optimizing Monte Carlo methods. It works under Microsoft Windows 95/98/ME/NT/2000. The structure of the system let one flexibly modify and extend its algorithmic contents. The Option Analyzer intends to provide a flexible tool for research studies related to options market and analysis of options: (a) solving of wide range of problems of option analysis, including optimal execution, re-selling, evaluation and forecast of American, European and some exotic options; (b) both automatic and manual choice of models of the pricing process, including classical geometrical Brownian motion, geometrical fractional Brownian motion, parametric and nonparametric dynamical models and stochastic automata; (c) comparison of several different options; (d) extended presentation of output information including optimal stopping strategies, dynamical histograms of expected profit and other profit-risk parameters.

2000 *AMS Mathematics Subject Classifications.* 62P05.

Key words and phrases. Option, Monte-Carlo method, Program system.

1. Introduction

The elaboration of a pilot program system OptAn (OPTion ANalyzer) is a project focused on the development of new optimizing simulation algorithms and programs for analysis of options.

[1]The paper represents a part of an invited lecture.

[2]This work is supported in part by the project "Stochastic modelling of insurance and finance processes and systems" funded by the Knowledge Foundation.

The principle problem, which exists in the area of optimal option pricing, is that mathematical models of pricing processes used nowadays in options calculations, are still too simple in comparison with the corresponding real-life pricing processes. The present project is based on the alternative (to the traditional analytical one) optimising Monte Carlo (simulation) approach based on extensive multi-step simulation-optimisation computer procedures. The principle advantage of this approach is the possibility to use more complicated and advanced models of pricing processes in the option analysis. Our aim is also to demonstrate a great potential of Monte Carlo optimising methods in problems of optimal option pricing as universal methods of global statistical analysis including selection of models, estimation of parameters, simulation, optimisation and visual presentation of data.

The program system OptAn is a pilot program system oriented to two categories of users, first researchers and second students of specialities in mathematical economics, analytical finance and similar specialities. It can also be useful for the individual investors, which would like to perform a sophisticated analysis of real or potential options.

At this stage we do not intend to develop a commercial program system. However, we search for partners from financial industry that would be interested to support the development of the project in this direction.

In Section 2 we describe the principles of design of OptAn program system. Section 3 is devoted to the description of three main problems, which can be solved by the existing variant of the system. In Section 4 we describe in brief the components, which are not ready now, but are planned to be added to the system in the nearest future.

We would like also to refer to papers by Boyle, Broadie and Glasserman (1997), Broadie and Glasserman (1997), Broadie, Detemple, Ghysels and Torres (2000a, b) and Broadie and Glasserman (2000), where different aspects of applications of Monte Carlo methods to the problem of option pricing are considered.

2. Principles of design of the OptAn program system

We follow to the approach, that the interface of the program should be similar, as the most popular programs like Microsoft Word have, that is: multiple document interface (Figure 12).

One can see, that our program has all the elements of the standard Windows interface, i.e., toolbar, status bar, menu bar, common controls etc. The multi-document interface allows the user to work with one window while the system makes long calculations in another one.

The first group box contains wide range of problems of option analysis, which can be solved with the help of the program. They will be described in details later.

The second group box contains possible types of options, with which the program can work. We plan to investigate not only standard options of American and European type, but also some exotic options.

The third group box simulates the future Internet interface of our program. With the help of it the user will have a possibility to choose a company, a type of historical data, i.e. share prices or options prices related to this company. Then the program should generate a query to the database located somewhere in the Internet, obtain the corresponding data and represent it to the user in both table and graphical forms. Later these data will be used as an input for advanced optimising Monte Carlo methods calculating options.

The fourth and fifth group boxes are intended for more experienced users. The fourth group box allows them to choose a model of the pricing process, with the help of which the options will be calculated. In future this list will be essentially extended.

The fifth group box will allow the users to select the model of a pricing process, to check goodness of a fit test for the chosen model and to estimate parameters of a model.

3. Main problems

Three main problems can be solved by the existing variant of the system. These task are: (a) access to the current option's data through the Internet; (b) solving the problem of optimal execution of the option of American type; (c) forecast of the option and share prices.

The first problem, the decision of which is already realised in the program, is the access to the current option's data through the Internet. The user should check the radio button called "Viewing" (Figure 12) and press Go button. The result is shown on Figure 13.

The access to the current option's data is realised as the very first problem in the list. In this case the user can choose an Internet site from the list (Figure 13) and analyze options data containing inside.

The second problem is the optimal execution of the option of American type. The user should check the radio button called "Execution" (Figure 14) and press Go button. In this case the previously written program "Optimal pricing" is called.

The detailed description of this program was given in Silvestrov, Galochkin and Sibirtsev (1999). In particular, this program has a module for estimation of threshold levels for optimal stopping strategies. The output data of this module are represented in a graphical form. The typical example of output picture is given in Figure 15 (upper window, the differences between the values of optimal threshold levels and the striking price are displayed).

The module for estimation of basic option characteristics let estimate and output the following characteristics of options (a) expected profit for optimal stopping strategy (buyer's income); (b) standard deviation, minimum and maximum of profit for optimal stopping strategy, quotient of expected profit and standard deviation, probability of realizing of the option; (c) profit histogram for optimal stopping strategies, (d) probabilities that the profit for optimal stopping strategies belongs to some given interval. The typical example of output picture is also given in Figure 16 (lower window).

The third problem is the forecast of option and share prices. The user should check the radio button called "Forecast" (Figure 16). After that the first list of the group box "Shares data" becomes enabled. It simulates the future access to the Internet database, which should contain historical data about options up to today.

After choosing of a database the second list of the group box "Shares data" becomes enabled. It simulates the choice of company, for which the forecast of option and share prices should be done. After choosing of a company the Monte Carlo simulation will start, and the progress control "Calculation progress" will indicate the process of calculations.

After finishing of the Monte Carlo simulations the user should press Go button. A window containing the results of simulation will appear (Figure 16). The line denoting historical data is placed on the left hand side from the vertical axis. The digits on the horizontal axis denote day numbers (today's number is equal to 0). The lines denoting quantiles of the price distribution are placed on the right hand side from the vertical axis. The denote levels corresponding to 10%, 25%, 50%, 75%, and 90% of the distribution's content. There exists also a line representing the average. Note that "today" can be any fixed date in historical data array. In this case a real realization of pricing process can be also drawn and compared with dynamic histogram.

The user can right-click in the area of quantiles. In this case a local menu will appear (Figure 16). One of the items in the menu is: building the histogram of the distribution of a price. The number of a day depends on the current position of a mouse cursor. In our case the user calls a histogram of the seventh day.

Calling the histogram can be repeated. For example, Figure 17 shows the case, when the user called histograms for both seventh and thirtieth days. The columns corresponds to the histogram, the line corresponds to the theoretical log-normal distribution of prices. One can see, that the more late histogram is wider than more early one.

4. The development of the program system

One of the important task from the programmer's point of view is to

realise a connection through the Internet with database, which contains historical share and option prices as well as provide daily updating of such prices. At the moment, this problem is solved in principle, and is in the process of practical realisation.

The historical data are used in the system for estimation of parameters of the corresponding dynamical models of pricing processes. After that artificial trajectories of the corresponding pricing processes are generated and used for the prediction of prices, finding optimal strategies, etc.

At present time, the following models of discrete time pricing processes are implemented in the OptAn program system: geometrical random walk with multiplicative log-normal increments, which is a discrete time analogue of classical model of geometrical Brownian motion, similar model with Bernoullian increments and additive analogues of these models.

Also a couple of automaton models of a pricing process is realised. These models show very promising results, when appliyng to some real pricing data. The automaton models are based on approximation of pricing processes by a geometrical random walk controlled by a Markov chain. This automaton use historical price data by accumulating (as a current state of the controlling Markov chain) information about appearance of special serial events in the realisation of pricing processes. As far as distributions of multiplicative increments are concerned, they are smoothed empirical distributions of increments (constructed with the use of sub-samples of increments corresponding to different states of the automaton).

In the simplest case of a trivial automaton with one state the distribution of the increment is a smoothed variant of the empirical distribution constructed from the sample of historical multiplicative increments of the corresponding pricing process.

The list of models is planned to be essentially extended. Some parametrical variants of automaton models will be included. Another interesting model for a pricing process, which is planned to be implemented is a fractional Brownian motion (see Cavler, Fisher and Mandelbrot (1997)).

One of the most interesting task for the further development of the OptAn program system is to implement algorithms of evaluation and reselling of the option of European type. The reselling option problem do requires the development of the model which describes dynamics of deviations of market option prices of theoretical Black-Sholes prices. We have been elaborating such models.

Also the list of option types will be essentially extended. It will include variants of American type options with various payoff functions and various exotic options. We would like to refer here to Jönsson (2001), where results related to the development of the project in this direction, are presented.

The possibility to use different models of pricing processes for option analysis causes appearance of an additional problem of a model selection

(the fifth group box on Figure 12). Statistical tests connected to the choice of a model and various goodness of fit tests will be implemented in the system. Any good Windows program should contain print subsystem, which prints the results of calculations. Such a subsystem must be added to the program system.

A lot of users never read documentation, but prefer to read a help file. The system already contains help subsystem, but it must be extended. A written documentation in PDF format with a detailed description of a program should also be prepared.

The Monte Carlo optimising methods, which is the scientific base for the development of OptAn program system can be also effectively used for portfolio analysis. The transformation of the system in the effective tool for analysis of multi-variate share-option portfolios will be the next step in the development of the project.

Bibliography

1. Boyle, P., Broadie, M. and Glasserman, P. *Monte Carlo methods for security pricing*, Journal of Economic Dynamics and Control, **21** (1997), 1267–1321.
2. Broadie, M. and Glasserman, P. *Pricing American-style securities using simulation*, Journal of Economic Dynamics and Control, **21** (1997), 1323–1352.
3. Broadie, M., Detemple, J., Ghysels, E. and Torres, O. *American options with stochastic dividends and volatility: A nonparametric investigation*, Journal of Econometrics, **94** (2000a), 53–92.
4. Broadie, M., Detemple, J., Ghysels, E. and Torres, O. *Nonparametric estimation of American options' exercise boundaries and call prices*, Journal of Economic Dynamics and Control, **24** (2000b), 1829–1857.
5. Broadie, M., Glasserman, P. and Ha, Z. *Pricing American options by simulations using a stochastic mesh with optimized weights*, in S. P. Uryasev (ed.), *Probabilistic Constraint Optimization*, Kluwer, Dordrecht, 2000, 26–44.
6. Cavler, L., Fisher, A. and Mandelbrot, B. B. *A multifractal model of asset returns*, Discussion paper of the Cowles Foundation for Economics, No. 1164, 1997.
7. Jönsson, H. *Monte Carlo studies of American type call options with discrete time* In Proceedings of the International School on Mathematical and Statistical Applications in Economics, Västerås, Sweden, 2001, Theory Stoch. Proces. **7(23)** (2001),
8. Silvestrov, D. S., Galochkin, V. G and Sibirtsev, V. G. *Algorithms and programs for optimal Monte Carlo pricing of American options*, In Proceedings of the Second International School on Actuarial and Financial Mathematics, Kiev, 1999, Theory Stoch. Proces. **5(21)** (1999), 175-187

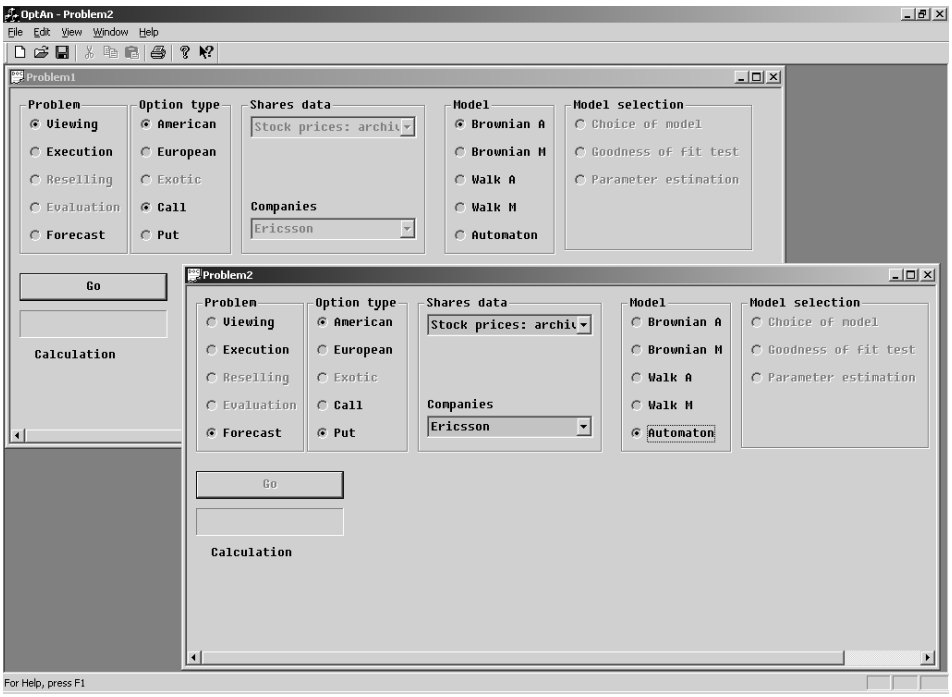


Figure 12: Multiple document interface

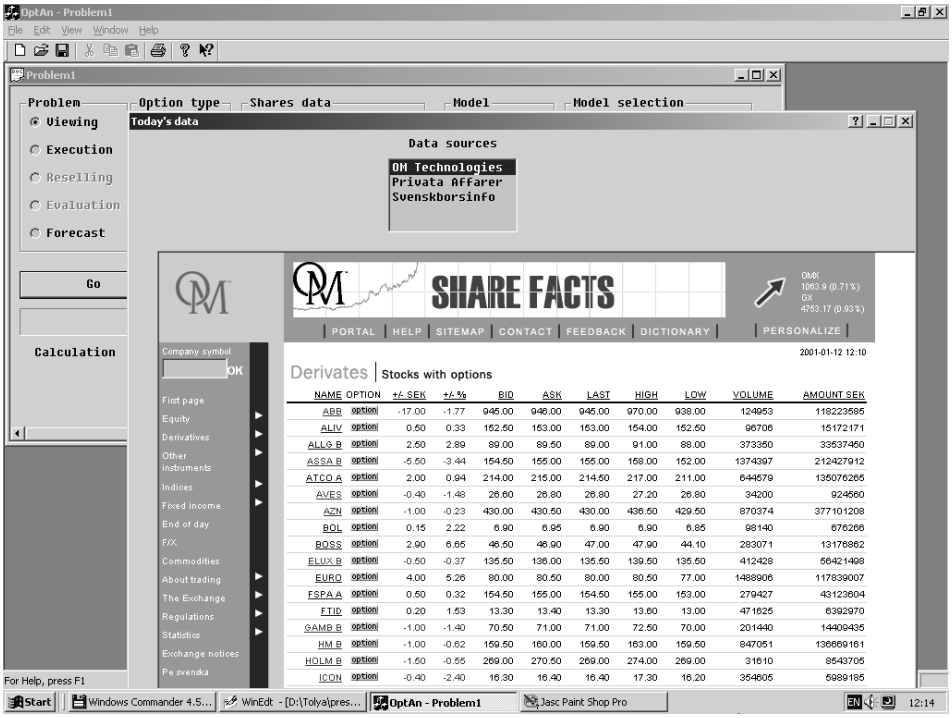


Figure 13: Current option's data

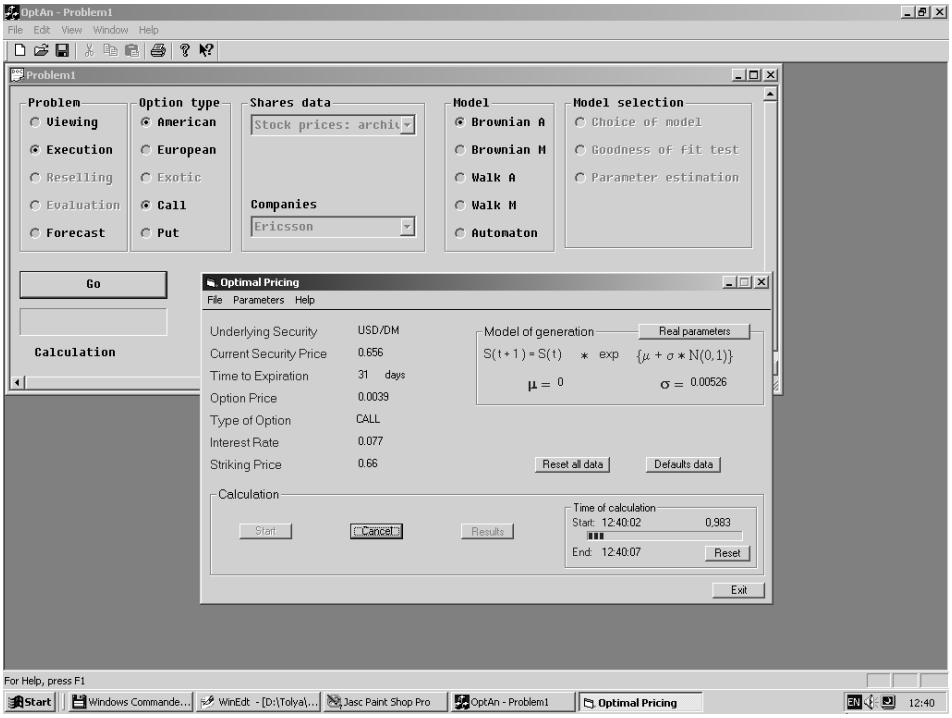


Figure 14: Estimation of threshold levels

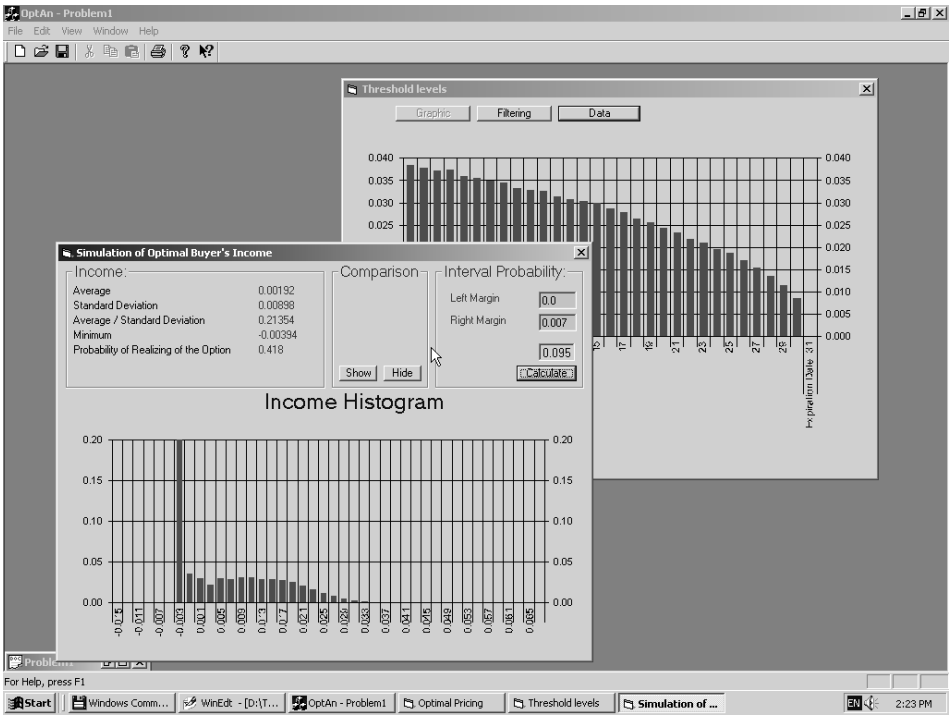


Figure 15: Optimal execution of the option

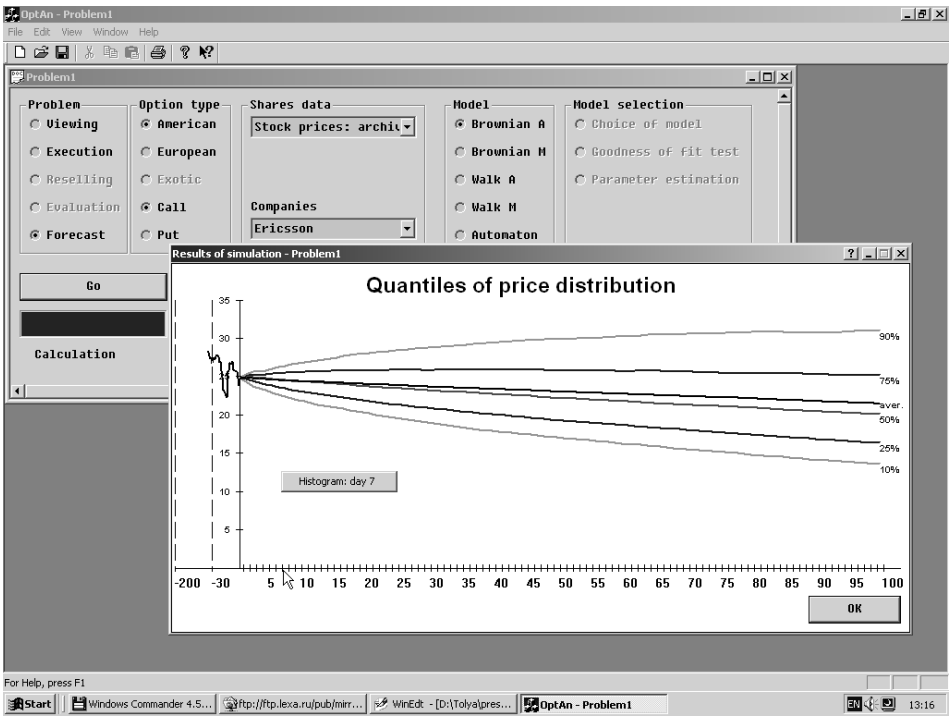


Figure 16: Forecast of option and share prices

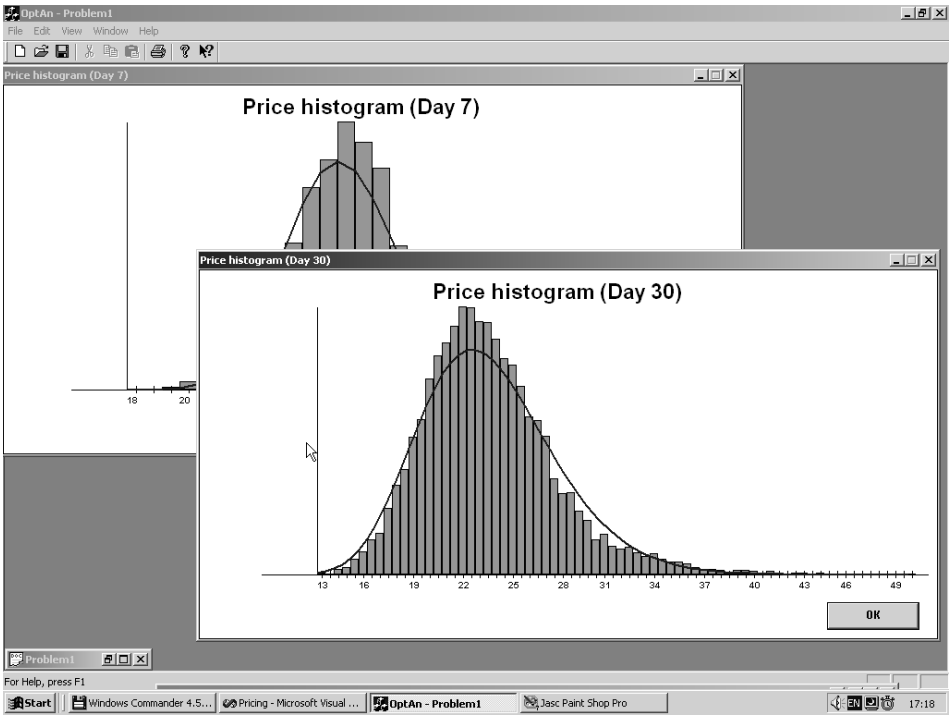


Figure 17: Comparison of histograms

Department of Mathematics and Physics, Mälardalen University, SE-72 123 Västerås, Sweden
E-mail: dmitrii.silvestrov@mdh.se

Center for Practical Informatics of the National Academy of Sciences of Ukraine and Department of Mathematics and Physics, Mälardalen University, SE-72 123 Västerås, Sweden.
E-mail: gal@nas.gov.ua

International Mathematical Center of the National Academy of Sciences of Ukraine and Department of Mathematics and Physics, Mälardalen University, SE-72 123 Västerås, Sweden.
E-mail: malyarenko@themail.com

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.301-316

IMBI TRAAT, KADRI MEISTER, KAJA SÕSTRA

STATISTICAL INFERENCE IN SAMPLING THEORY [1][2]

In this paper a stochastic representation of survey data is presented allowing to handle them along classical statistical traditions. The man-made randomization law, the sampling design, is given by a discrete multivariate disribution. The design-based, model-based and model-design-based finite population inference under with- and without-replacement sampling designs are covered. A very general variance formula of the estimator of the sum of population means is derived. The framework is developed to handle sampling and resampling methods jointly. The keypoint in this is the two-phase sampling design expressed by multivariate distributions. The random selection variables of the first phase are explicitly given in the second-phase design.

2000 *Mathematics Subject Classifications.* 62D05, 62E15, 62F10, 62H10.

Keywords and phrases. Finite population, random survey data, model-design-based inference, multivariate Bernoulli and multinomial designs, two-phase sampling, resampling, bootstrap.

1. Introduction

Classical inference theory is based on the distributional assumptions of the sample. Usually, it is assumed that elements of the sample are i.i.d. random variables − generated independently from the population distribution of the study variable. If the independence assumption is violated then, often, a random process (time series, Markov chain etc.) can be assumed for the observations. Thus, the observations are viewed as random variables with their randomness described by the assumptions on the data generation

[1]Invited lecture.

[2]This work was supported by the grant No. 3939 of the Estonian Science Foundation and the grant No. 2461/2000(381/N20) of the Visby Programme of the Swedish Institute.

process. These assumptions are the building blocks for theoretical derivations of various inference matters in classical statistics.

In the finite population sampling theory a special kind of data generation process has appeared − the man-made randomization which according to some probability mechanism, called *sampling design*, selects observations from the population. Until recently, the man-made sampling process was not expressed in the observations in a way which would have allowed to handle survey data along classical statistical traditions. Therefore, its own inference theory has been developed for the finite population statistics. Through the history it has been developing separately in several different branches (forming e.g. the design-based, model-based and model-design-based inference, the with- and without-replacement inference, etc.), without general unifying approach being available. Different approaches in the finite population inference have been systematically covered by Cassel, Särndal and Wretman (1977).

In the first part of this paper a stochastic representation of survey data is presented in which two random processes − the process generating values of the study variable(s) in the population and the process randomly selecting some of these values into a sample − are explicitly given in the observations. The idea was put forward in Traat (1997). In Traat and Meister (1998) it was shown that the elements of the classical estimation theory (the methods of moments and least squares) can be easily applied to the random survey data defined in the proposed way. The material of the first part of the paper was delivered in a lecture during the 52nd Session of the ISI in Helsinki, summarized in Traat and Meister (1999). The same idea was independently used, but only for without-replacement sampling designs, in Molina, Smith and Sugden (1999) who have made further developments in the estimation theory on survey data. Using the unified approach on the sampling designs as multivariate discrete distributions (Traat 1997, 2000, Traat, Bondesson and Meister 2000) we are able to incorporate also with-replacement sampling designs here. In this way, the presented stochastic representation of data, and the inference built on data, will have wider statistical applicability, satisfying also needs of the resampling theories.

In sections 3-4 the stochastic representation of survey data is presented in a very general form which includes the design-based, model-based and model-design-based special cases. The randomization process may be either with or without replacement. Based on the distributional assumptions of data generation processes the random feature of sampled data is described and applied in the inference. An unbiased estimator of the sum of the population means is derived. It includes the classical Horvitz-Thompson (1952) and Hansen-Hurwitz (1943) estimators as special cases. A very general variance formula of this estimator is derived and numerous important special cases are drawn from it.

In section 5 a wider look on the sampling concept in statistics is given by considering the finite population sampling and resampling like bootstrap,

jacknife and subsampling (see Politis, Romano and Wolf 1999) in the same framework. In fact, the probability mechanisms of the classical resampling methods are the simpliest sampling designs of the sampling theory − simple random sampling with replacement for nonparametric bootstrap, simple random sampling without replacement for jacknife and for what is called subsampling procedure. In this section the two-phase sampling design is defined in the language of multivariate distribution. A crucial component in it is a probability function of the second-phase sampling design which explicitly includes the first-phase selection variables. In the examples different sampling designs for the phases are combined with each other and the probability law of the resulting selection mechanism is expressed.

2. Sampling design

Let us have a population U with labelled elements $U = \{1, 2, \ldots, N\}$. Let a random vector $\boldsymbol{I} = (I_1, I_2, \ldots, I_N)$ describe the sampling process in U, so that its outcome $\boldsymbol{k} = (k_1, k_2, \ldots, k_N)$ identifies a sample by $k_i = 0$ meaning that the unit i is not sampled, and $k_i > 0$ meaning that the unit i is sampled k_i times. The multivariate distribution $p(\boldsymbol{k}) = \Pr(\boldsymbol{I} = \boldsymbol{k})$, with $\boldsymbol{k}$ belonging to the space of nonnegative integers, is called *sampling design*. This concept of the sampling design (Traat 1997, 2000) differs from the traditional one, the latter being defined as a distribution on the sets or ordered sets (Cassel, Särndal and Wretman 1977). The advantages of the present definition appear in the possibility to consider both the with- and without-replacement sampling designs jointly in a more general level, and in the availability of the knowledge and tools worked out for the multivariate distributions in mathematical statistics and probability theory.

In our approach all without-replacement sampling designs are certain multivariate Bernoulli distributions with probability function in the most general form being simply a list of probabilities on all possible points $\boldsymbol{k}$,

$$p(\boldsymbol{k}) = \Pr(\boldsymbol{I} = \boldsymbol{k}),\ k_i \in \{0, 1\},\ \textstyle\sum_{\mathbf{k}} p(\boldsymbol{k}) = 1. \tag{1}$$

Any first order marginal distribution of (1) is a Bernoulli distribution

$$I_i \sim B(1, \pi_i), \tag{2}$$

where $\pi_i = E(I_i) = \Pr(I_i = 1)$ is the inclusion probability of the unit i. The random sample size is $\sum I_i$ and the realized sample size is $\sum k_i$. The fixed-sample-size-n-design has $\sum I_i \equiv n$. In this paper the summation index, if not specified, runs from 1 to N.

The with-replacement sampling design with predetermined selection probabilities p_i, $\sum p_i = 1$, and n independent draws from U, is a multinomial distribution

$$p(\boldsymbol{k}) = \Pr(\boldsymbol{I} = \boldsymbol{k}) = \frac{n!}{\prod_{i=1}^{N} k_i!} \prod_{i=1}^{N} p_i^{k_i},\ \text{if } |\boldsymbol{k}| = n, \tag{3}$$

where $|\boldsymbol{k}| = \sum k_i$. Marginal distributions of (3) are binomial

$$I_i \sim B(n, p_i). \tag{4}$$

The most important special case of the multivariate Bernoulli design is simple random sampling without-replacement or simple multivariate Bernoulli design (SI-design)

$$p(\boldsymbol{k}) = \binom{N}{n}^{-1}, \text{ if } |\boldsymbol{k}| = n, \tag{5}$$

and, of the multinomial design, simple random sampling with-replacement or simple multinomial design (SIR-design)

$$p(\boldsymbol{k}) = \frac{n!}{N^n \prod_{i=1}^{N} k_i!}, \text{ if } |\boldsymbol{k}| = n. \tag{6}$$

Many other probability functions of sampling designs are derived in Traat (2000) and Traat, Bondesson and Meister(2000).

3. Stochastic representation of survey data

Survey data are generated by at least two random processes − the process generating finite population values, denoted by the random N-vector $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_N)$, and the process selecting a sample from the population, described by the design vector $\boldsymbol{I}$. These processes can be simultaneously expressed by the random vector $\boldsymbol{Y}_s$:

$$\boldsymbol{Y}_s = (Y_{si}) = (I_1 Y_1, I_2 Y_2, \ldots, I_N Y_N), \; Y_{si} = I_i Y_i. \tag{7}$$

The vector $\boldsymbol{Y}_s$ displays in a simple manner the random variables Y_i selected by the design vector $\boldsymbol{I}$. When random $\boldsymbol{Y}$ generates the observations in the population then $\boldsymbol{Y}_s$ generates them in the sample. The vector $\boldsymbol{Y}_s$ expresses the random character of survey data, and, therefore, plays a crucial role while making inference from survey data.

Let us denote the outcome of $\boldsymbol{Y}_s$ by $\boldsymbol{y}_s$:

$$\boldsymbol{y}_s = (y_{si}), \; y_{si} = k_i y_i. \tag{8}$$

It is clear that not all the components of $\boldsymbol{y}_s$ are observations of the study variable. If $y_{si} = 0$ due to $k_i = 0$ then it is not an observed value. If $y_{si} = k_i y_i$ with $k_i > 1$ then it is a multiple of the observed value y_i. Of course, real observations can be extracted from $\boldsymbol{y}_s$ by using the outcome $\boldsymbol{k}$ of the design vector $\boldsymbol{I}$. Therefore the full description of survey data is given by the pair $(\boldsymbol{y}_s, \boldsymbol{k})$ which is an outcome of the random survey data

$$(\boldsymbol{Y}_s, \boldsymbol{I}). \tag{9}$$

The pair of random vectors (9) is *a stochastic representation of survey data*. The inference can be built upon this pair and on the assumptions on the

design vector $\boldsymbol{I}$ and population vector $\boldsymbol{Y}$. Note, that (9) with $\boldsymbol{Y}_s$ given in (7) describes also the situation in the classical resampling context. In the case of bootstrap $\boldsymbol{I}$ has the SIR-design (6) with $n = N$, in the case of jacknife $\boldsymbol{I}$ has the SI-design (5) with $n = N - 1$, and in case of subsampling (in the sense of Politis, Romano and Wolf 1999) $\boldsymbol{I}$ has the SI-design (5) with $n < N - 1$. In survey situation, it is straightforward to include vector of auxiliary variables known for the whole population in the representation (9). It is also easy to think about Y_i as a vector of response and explanatory variables associated with the unit i.

In the earlier sampling literature also other possibilities for describing survey data are considered. For example, the element-wise definition of survey data has been used for the inference theory by Cassel, Särndal and Wretman (1977). They present realized data as a collection of values $d = \{(i,\ y_i) :\ i \in s\}$, produced by the random analog $\mathcal{D} = \{(i,\ Y_i) :\ i \in S\}$, where s and S are the realized and random sample from U, respectively. We would say that the representation (9) is more traditional for the classical statistical derivations than the representation $\mathcal{D}$.

In some other works it has been customary to devide the population into sampled and unsampled parts $\boldsymbol{y} = (\boldsymbol{y}_{sa}, \boldsymbol{y}_{\bar{sa}})$, and to call $(\boldsymbol{y}_{sa}, \boldsymbol{k})$ the survey data with the corresponding random analog $(\boldsymbol{Y}_{sa}, \boldsymbol{I})$. The pair $(\boldsymbol{Y}_{sa}, \boldsymbol{I})$ itself can not be used as a description of random survey data. But, for example, Pfeffermann, Krieger, Rinott (1998) have expressed the density function of $\boldsymbol{Y}_{sa}$, and develop classical maximum-likelihood inference on survey data under general conditions including informative designs. The restrictions of their approach stand in the assumption of the independent or asymptotically independent components of the design vector $\boldsymbol{I}$. Also, the approach is not meant for the purely design-based case with fixed values y_i.

The advantages when working with the representation (9) of survey data are listed below:

- $\boldsymbol{Y}_s$ displays naturally the sampled and unsampled parts of the population letting the design vector $\boldsymbol{I}$ to do it.

- $(\boldsymbol{Y}_s,\ \boldsymbol{I})$ includes simultaneously information on the labels i, observed variables Y_i and the design variables I_i.

- $(\boldsymbol{Y}_s,\ \boldsymbol{I})$ describes random nature of data both for with- and without-replacement sampling designs (including those used for the classical resampling procedures).

- $\boldsymbol{Y}_s$ includes the information about the selection dependencies imposed by the sampling design, by the multivariate distribution of $\boldsymbol{I}$.

- In the special case, when no random feature of the finite population is assumed, i.e. $\boldsymbol{Y} \equiv \boldsymbol{y}$, the vector $\boldsymbol{Y}_s$ still remains random due to the sampling design $\boldsymbol{I}$ involved in it, and, thus, represents random observation vector for the fixed finite population case.

- In the special case of fixed $\boldsymbol{I} \equiv \boldsymbol{k}$, the distribution of $\boldsymbol{Y}_s = (k_i Y_i)$ is determined by the marginal distribution of $\boldsymbol{Y}$, by the one corresponding to the components Y_i for which $k_i > 0$. If $k_i \equiv 1$, i.e. no random selection is involved, then $\boldsymbol{Y}_s \equiv \boldsymbol{Y}$ which is the classical statistical situation without man-made randomization.

- It is easy to find distributional characteristics (moments) of the vector $\boldsymbol{Y}_s$.

4. Estimation

Let us have random survey data $(\boldsymbol{Y}_s,\ \boldsymbol{I})$. The inference is directed to the population vector $\boldsymbol{Y}$ and its characteristics or, in case of the fixed finite population, to the vector $\boldsymbol{y}$. For the inference a suitably defined statistic to some extent known random behaviour is needed. According to Cassel, Särndal and Wretman (1977), in our notations, a statistic T is a function of survey data such that for any given selection result $\boldsymbol{k}$ of $\boldsymbol{I}$, the statistic T depends on $\boldsymbol{Y}$ only through those Y_i for which $k_i > 0$ (i.e. through the selected elements). It is possible to construct many functions of $(\boldsymbol{Y}_s, \boldsymbol{I})$ which do not depend on the non-sampled elements. For example, $T(\boldsymbol{Y}_s, \boldsymbol{I}) = \sum c_i Y_{si} = \sum c_i (I_i Y_i)$ and $T(\boldsymbol{Y}_s, \boldsymbol{I}) = \frac{\sum c_i I_i Y_i^{a_i}}{\sum I_i}$, with any constants $c_i,\ a_i$ are the statistics.

4.1 Distributional characteristics of survey data

Let us call the probabilistic assumptions on $(\boldsymbol{Y},\ \boldsymbol{I})$ *the survey model*. Survey model determines the probabilistic behaviour of survey data $(\boldsymbol{Y}_s,\ \boldsymbol{I})$, and hence, the behaviour of statistics built on data. Let the survey model be given as follows:

$$E_\xi(Y_i) = \mu_i,\ V_\xi(Y_i) \quad = \sigma_{ii},\ Cov_\xi(Y_i, Y_j) = \sigma_{ij}, \tag{10}$$

$$E_p(I_i) = m_i,\ V_p(I_i) \quad = \Delta_{ii},\ Cov_p(I_i, I_j) = \Delta_{ij}, \tag{11}$$

$$\boldsymbol{I},\ \boldsymbol{Y} \text{ independent.} \tag{12}$$

Here ξ marks the distribution of $\boldsymbol{Y}$. The condition (12) expresses the ignorable or noninformative feature of the sampling design.

For the probability sampling design $m_i > 0,\ \forall i$. If $p(\boldsymbol{k})$ is the multivariate Bernoulli design then the moments in (11) take the form,

$$m_i = \pi_i,\ \Delta_{ii} = \pi_i(1 - \pi_i),\ \Delta_{ij} = \pi_{ij} - \pi_i \pi_j, \tag{13}$$

where $\pi_{ij} = E_p(I_i I_j)$. If the sampling design is multinomial, $p(\boldsymbol{k}) \sim M(n, p_1, p_2, \ldots, p_N)$, then

$$m_i = np_i,\ \Delta_{ii} = np_i(1 - p_i),\ \Delta_{ij} = -np_i p_j. \tag{14}$$

The fixed finite population case, $\boldsymbol{Y} \equiv \boldsymbol{y}$, is an important special case of the survey model above, in which case $E_\xi(Y_i) = y_i,\ \sigma_{ii} = 0,\ \sigma_{ij} = 0$. Another important special case is the model with fixed selection result, $\boldsymbol{I} \equiv \boldsymbol{k}$, in which case $E_p(I_i) = k_i,\ \Delta_{ii} = 0,\ \Delta_{ij} = 0$. Note that the ignorability condition (12) looses its sense for the special models. The model $\boldsymbol{I} \equiv \boldsymbol{k}$ needs more care in the inference stage since here $E_p(I_i) = k_i$ can equal to zero.

It is easy to derive the first moments of the observations under the survey model (10)-(12):

$$\begin{aligned} E_{p\xi}(Y_{si}) &= E_{p\xi}(I_i Y_i) = m_i \mu_i, && (15)\\ V_{p\xi}(Y_{si}) &= (m_i^2 + \Delta_{ii})\sigma_{ii} + \Delta_{ii}\mu_i^2, && (16)\\ Cov_{p\xi}(Y_{si}, Y_{sj}) &= (m_i m_j + \Delta_{ij})\sigma_{ij} + \Delta_{ij}\mu_i\mu_j,\ i \neq j. && (17) \end{aligned}$$

It can be seen from (17) that for the observed data to be uncorrelated, it has to hold $\sigma_{ij} = 0$, and either $\Delta_{ij} = 0$ or $\mu_i = 0,\ \forall i, j$. It is useful to know that observations from the i.i.d. populations with $\mu_i = 0$ can be uncorrelated in spite of the correlated design. The uncorrelated designs are rare, Poisson sampling design being an example here.

Moments of the observations in (15)-(17) take the following forms under special survey models:

$$E_{p\xi}(Y_{si}) = m_i y_i,\ V_{p\xi}(Y_{si}) = \Delta_{ii} y_i^2,\ Cov_{p\xi}(Y_{si}, Y_{sj}) = \Delta_{ij} y_i y_j, \quad \text{if } \boldsymbol{Y} \equiv \boldsymbol{y} \quad (18)$$

$$E_{p\xi}(Y_{si}) = k_i \mu_i,\ V_{p\xi}(Y_{si}) = k_i^2 \sigma_{ii},\ Cov_{p\xi}(Y_{si}, Y_{sj}) = k_i k_j \sigma_{ij}, \quad \text{if } \boldsymbol{I} \equiv \boldsymbol{k} \quad (19)$$

Note that if no selection process is present ($\boldsymbol{k} = \boldsymbol{1}$, the vector consisting of 1's) then the moments of the observations in (19) coincide with the population moments in (10).

4.2 Unbiased estimation

Let us assume the survey model (10)-(12). Let us look at the statistic $\sum c_i Y_{si}$ where $\boldsymbol{c} = (c_i)$ is a vector of constants. Due to the expectations in (15)-(17) it follows that

$$E_{p\xi}(\sum c_i Y_{si}) = \sum c_i m_i \mu_i. \quad (20)$$

From (20) it is clear that by taking $c_i = 1/m_i$ we receive an unbiased estimator $\hat{t}$ for the parameter $t = \sum \mu_i$ under the assumed model:

$$\hat{t} = \sum Y_{si}/m_i = \sum I_i Y_i / m_i. \quad (21)$$

By inserting the outcome $y_{si} = k_i y_i$ of the random observation Y_{si} in (21) we get the following point estimate for t:

$$\hat{t} = \sum y_{si}/m_i = \sum k_i y_i / m_i = \sum_{k_i > 0} k_i y_i / m_i. \quad (22)$$

In the special case $\boldsymbol{Y} \equiv \boldsymbol{y}$, the estimated parameter transforms to the fixed population total $t = \sum y_i$, and the estimator in (21) takes the form

$$\hat{t} = \sum I_i y_i / m_i. \tag{23}$$

The corresponding estimate, $\hat{t} = \sum_{k_i>0} k_i y_i / m_i$, has the same form as the estimate under general survey model, given in (22).

For a multivariate Bernoulli design with inclusion probabilities π_i we have $k_i \in \{0, 1\},\ m_i = \pi_i,\ \forall i$, and the estimate in (22) turns out to be the famous π-estimate, $\hat{t} = \sum_{k_i>0} y_i/\pi_i$, invented by Horvitz and Thompson (1952). For a multinomial design with parameters (n, p_i) we have $m_i = np_i$, and the estimate in (22) turns out to be another famous estimate (*pwr*-estimate), $\hat{t} = \sum_{k_i>0} k_i y_i/(np_i)$, invented by Hansen and Hurwitz (1943).

Under the special case of the model where $\boldsymbol{I} \equiv \boldsymbol{k}$ we can not derive the special form of the estimator from (21) since some m_i in the denominator may be equal to zero. Writing the initial equality (20) down for this special case we have

$$E_{p\xi}(\sum c_i k_i Y_i) = \sum c_i k_i \mu_i. \tag{24}$$

The relationship in (24) is equivalent to

$$E_{p\xi}(\textstyle\sum_{k_i>0} c_i k_i Y_i) = \sum_{k_i>0} c_i k_i \mu_i, \tag{25}$$

from which we get the following unbiased estimator for the parameter $t = \sum_{k_i>0} \mu_i$:

$$\hat{t} = \textstyle\sum_{k_i>0} Y_i. \tag{26}$$

The inference under the model $\boldsymbol{I} \equiv \boldsymbol{k}$ is not directed to the parameters of the full vector $\boldsymbol{Y}$ but to those of the selected components alone. Note, that the repeatedly selected elements appear only once in the estimator.

The variance of the estimator in (21) is

$$V_{p\xi}(\hat{t}) = \textstyle\sum_i V_{p\xi}(Y_{si})/m_i^2 + \sum \sum_{i\neq j} Cov_{p\xi}(Y_{si}, Y_{sj})/(m_i m_j), \tag{27}$$

which can be developed with moments in (16)-(17) to the following form:

$$V_{p\xi}(\hat{t}) = \sum \sum ((m_i m_j + \Delta_{ij})\sigma_{ij} + \Delta_{ij}\mu_i\mu_j)/(m_i m_j). \tag{28}$$

The double sum with unspecified summation indeces means that both indeces run from 1 to N. In fact, the variance in (28) depends on the sampling design only through the quantity $c_{ij} = \Delta_{ij}/m_i m_j$ which may be viewed as a coefficient of co-variation (it is a square of the well-known coefficient of variation if $i = j$):

$$V_{p\xi}(\hat{t}) = \sum \sum ((1 + c_{ij})\sigma_{ij} + c_{ij}\mu_i\mu_j). \tag{29}$$

The variance formula in (29) is very general. Some of its special cases are listed below:

$$\begin{aligned}
V_{p\xi}(\hat{t}) &= \sum(1+c_{ii})\sigma_{ii} + \sum\sum c_{ij}\mu_i\mu_j, \quad \text{for uncorrelated } Y_i,\ (\sigma_{ij}=0,\ i\neq j);\\
&= \sum\sum\sigma_{ij} + \sum c_{ii}(\sigma_{ii}+\mu_i^2), \quad \text{for uncorrelated } I_i,\ (\Delta_{ij}=0,\ i\neq j);\\
&= \sigma^2(N+\sum c_{ii}) + \mu^2\sum\sum c_{ij}, \quad \text{for i.i.d. } Y_i,\ (\sigma_{ii}=\sigma^2,\ \mu_i=\mu,\ \forall i);\\
&= \sigma^2(N+\sum c_{ii}), \quad \text{for i.i.d. } Y_i, \text{ self-weighting fixed size } n \text{ design} \quad (30)\\
&\qquad (m_i = n/N\ \forall i, \sum\sum c_{ij}=0, \quad \text{due to } \sum I_i = n);\\
&= N^2\sigma^2/n, \quad \text{for conditions in (30) and multiv. Bernoulli design;} \quad (31)\\
&= N^2\sigma^2/n(1+\frac{n-1}{N}), \quad \text{for conditions in (30) and multinomial} \quad (32)\\
&\qquad \text{design (SIR-design).}
\end{aligned}$$

Under the fixed finite population case, $\boldsymbol{Y} \equiv \boldsymbol{y}$, the parameter transforms to the finite population total $t = \sum y_i$, and the variance of the estimator $\hat{t}$ takes the following well-known forms of the pure design-based variance (Särndal, Swensson, Wretman 1992):

$$\begin{aligned}
V_{p\xi}(\hat{t}) &= \sum\sum c_{ij}y_iy_j;\\
&= \sum\sum(\pi_{ij}-\pi_i\pi_j)y_iy_j/(\pi_i\pi_j), \quad \text{multivariate Bernoulli design;}\\
&= \sum(y_i/p_i - t)^2p_i/n, \ \text{multinomial } M(n,p_1,p_2,\ldots,p_N) \text{ design.}
\end{aligned}$$

Under the special case of the fixed selection result, $\boldsymbol{I} \equiv \boldsymbol{k}$, the parameter to be estimated is $t = \sum_{k_i>0}\mu_i$ with the estimator given in (26). The general variance formula in (28) obtains the following form for this special case:

$$V_{p\xi}(\hat{t}) = \sum\sum k_ik_j\sigma_{ij} = \sum\sum_{k_i>0,k_j>0} k_ik_j\sigma_{ij}. \quad (33)$$

Comment 1. It can be seen from the general variance formula (28) that the cosistency statements are not so straightforward. It has to be formulated how the parameters in (28) behave when the sample and the population mutually increase. These conditions are much easier to establish for i.i.d. Y_i and fixed size sampling designs. The formulae (31) and (32) tell that in this case they can be formulated in terms of the sample size n and the population size N alone. If we look the average population mean $\sum\mu_i/N$ and the corresponding estimator $\hat{t}/N$ then the variance formulae in (31) and (32) for i.i.d. Y_i and self-weighting fixed size sampling designs take the forms:

$$\begin{aligned}
V_{p\xi}(\hat{t}/N) &= \sigma^2/n, \quad \text{multivariate Bernoulli design;} \quad (34)\\
&= \sigma^2(1+\frac{n-1}{N})/n, \quad \text{multinomial design.} \quad (35)
\end{aligned}$$

Now, if N is fixed, we have that $\lim_{n\to N} V_{p\xi}(\hat{t}/N) = \frac{\sigma^2}{N}$ under multivariate Bernoulli design, and $\lim_{n\to N} V_{p\xi}(\hat{t}/N) = \frac{\sigma^2}{N}(2-\frac{1}{N})$ under multinomial design. Letting also N free ($n < N$) we have $\lim_{n,N\to\infty} V_{p\xi}(\hat{t}/N) = 0$. Note

that if N is fixed then under multinomial sampling n can increase the value N, and then $\lim_{n\to\infty} V_{p\xi}(\hat{t}/N) = \sigma^2/N$.

Comment 2. The stochastic representation (9) of survey data enables to develop inference for informative sampling designs too. Let $\boldsymbol{I}$ be an informative sampling design. This means that $\boldsymbol{I}$ and $\boldsymbol{Y}$ are dependent. Now $E_{p\xi}(I_iY_i) \neq E_\xi(I_i)E_p(Y_i)$ and instead $E_\xi E_p(I_iY_i \mid Y_i)$ should be evaluated. Since the inner expectation can be developed for the fixed y_i as

$$E_p(I_iY_i \mid Y_i = y_i) = y_iE_p(I_i \mid Y_i = y_i) = y_im_i(y_i), \tag{36}$$

where $m_i(y_i)$ is the expected selection frequency of the element i conditional on y_i. In sample surveys this conditional expected selection frequency is known either for each i or for i in the sample. Since the expected selection frequency is expressed either by the inclusion or by the selection probability ($m_i(y_i) = \pi_i(y_i)$ for the multivariate Bernoulli design, and $m_i(y_i) = np_i(y_i)$ for the multinomial design) then it is necessary to model dependence of the probability on the study variable. Pfeffermann, Krieger and Rinott (1998) offer several models to describe $\pi_i(y)$. The same models can be applied to describe the selection probability $p_i(y)$.

Comment 3. Molina, Smith and Sugden (1999) consider the survey model (10)-(12) in the form of a linear model where the vector of population means is expressed by $\boldsymbol{\mu} = (\mu_i) = \boldsymbol{X}\boldsymbol{\beta}$ with $\boldsymbol{\beta}$ being a vector of regression coefficients, $\boldsymbol{X}$ being a design matrix, and $\boldsymbol{I}$ being an without-replacement sampling design. Using the framework similar to ours (in matrix form) and the theory of estimating functions they have elegantly derived a very general estimator for $\boldsymbol{\beta}$ and corresponding efficiency. Using our definition of the sampling design their results can also be written down for the with-replacement designs by using covariance matrix of the multinomial distribution.

5. Sampling and resampling

There are many common features between finite population sampling and resampling theories. For example, sampling from the initial sample is a common element both in the two-phase sampling and resampling procedures. Substantial difference is in the reasons for sampling from a sample. In the two-phase sampling the estimator can be calculated only in the second-phase sample and its variability is created by the designs of both phases. In the resampling situation the estimator is calculated in the first-phase sample and its variability is formed by the first-phase sampling design (and model distribution, if present). In the classical resampling situation the first-phase sample is a sample with a model distribution. Due to the difficulty to find an analytical variance formula, the second-phase sampling is started. With suitable choice of the second-phase design, and some other manipulations, it is sometimes possible to achieve that

the second-phase variance approximates the first-phase variance. The second-phase variance can be calculated exactly, if not analytically then by evaluating the estimator on all possible second-phase samples and calculating its variance with the help of design probabilities. Although, usually it is estimated from a smaller number of samples generated from the second-phase design. Generally, arbitrary sampling designs can be used in the phases of a two-phase sampling procedure. We benefit from the fact that approach of the present paper enables to consider with- and without-replacement sampling designs in the same framework. Yet, in this paper the variability of an estimator in the stages of the two-phase sampling design is not considered (remains for future work). The main stress has been put to the description of the two-phase sampling design in the language of multivariate distributions.

5.1 Two-phase sampling

Here the two-phase sampling design is considered in the framework of multivariate discrete distributions. The presentation is more general than the one in Särndal, Swensson, Wretman (1992) since it allows joint consideration of both the with- and without-replacement sampling designs. The latter makes it possible to study resampling methods (involving also with-replacement sampling) for samples drawn with more or less complex designs. Below the probability function of the two-phase sampling design is presented and interpreted. The randomness due to the sampling in the first phase is explicitly incorporated to the design expression of the second phase, making theoretical operations in deriving design properties simpler. Special cases where with- and without-replacement sampling designs are combined in the phases are considered in the examples. The first moments of corresponding two-phase designs are presented.

Two-phase sampling procedure in the finite population $U = \{1, 2, \ldots, N\}$ is a subsequent sampling where in the first phase a random sample is drawn, and in the second phase a sample from it is drawn. The sampling designs in the phases may be arbitrary.

Let $\boldsymbol{I}_a \sim p_a(\boldsymbol{k}_a)$ be sampling design in the first phase: $p_a(\boldsymbol{k}_a) = \Pr(\boldsymbol{I}_a = \boldsymbol{k}_a)$. Let $\boldsymbol{I}|\boldsymbol{k}_a \sim p(\boldsymbol{k}|\boldsymbol{k}_a)$ be sampling design in the second phase conditional on the first phase outcome: $p(\boldsymbol{k}|\boldsymbol{k}_a) = \Pr(\boldsymbol{I} = \boldsymbol{k}|\boldsymbol{I}_a = \boldsymbol{k}_a)$. Denote $\boldsymbol{I} \sim p(\boldsymbol{k})$ the two-phase sampling design expressing probability to get the sample $\boldsymbol{k}$ through the two phases. Then

$$p(\boldsymbol{k}) = \Pr(\boldsymbol{I} = \boldsymbol{k}) = \sum_{\boldsymbol{k}_a} p(\boldsymbol{k}|\boldsymbol{k}_a)\ p_a(\boldsymbol{k}_a). \tag{37}$$

Note, that $p(\boldsymbol{k}|\boldsymbol{k}_a) = 0$ for some $\boldsymbol{k}_a$ since the point $\boldsymbol{k}$ on the left-hand side of (37) can not be received from each $\boldsymbol{k}_a$. For example $p(\boldsymbol{k}|\boldsymbol{k}_a) = 0$ for such $\boldsymbol{k}$ in which $k_i > 0$, but $k_{ai} = 0$ in $\boldsymbol{k}_a$, meaning that unit i is to be sampled in the second phase, though it is not sampled in the first phase. Denoting

by $\boldsymbol{k}_a(\boldsymbol{k})$ such $\boldsymbol{k}_a$ for which sample $\boldsymbol{k}$ is possible in the second phase we can write (37) alternatively

$$p(\boldsymbol{k}) = \sum_{\boldsymbol{k}_a(\boldsymbol{k})} p(\boldsymbol{k}|\boldsymbol{k}_a)\ p_a(\boldsymbol{k}_a). \tag{38}$$

It is confirmed by (38) that if the entire population is sampled in the first phase $p_a(\boldsymbol{k}_a) = 1$ for $\boldsymbol{k}_a = \mathbf{1}$, the two-phase sampling design equals the second phase design, $p(\boldsymbol{k}) = p(\boldsymbol{k}|\mathbf{1})$. Thus, the classical resampling set-up (the i.i.d. random variables $Y_1, Y_2, \ldots, Y_N$ are the first-phase sample and subsamples are drawn from it) is a special case of the two-phase sampling procedure.

The formulae (37) or (38) can be alternatively read as expectation with respect to the first-phase design of the conditional probability to get sample $\boldsymbol{k}$ in the second phase:

$$p(\boldsymbol{k}) = E_a\ [p(\boldsymbol{k}|\boldsymbol{I}_a)]. \tag{39}$$

Note that the design probabilities $p(\boldsymbol{k}|\boldsymbol{I}_a)$ are random. If the functional form of $p(\boldsymbol{k}|\boldsymbol{I}_a)$ is known, and even more important, if the functional forms of corresponding marginal distributions are known, they can be conveniently utilized (as shown below) for evaluating characteristics of the two-phase sampling design. Functional forms of many sampling designs are presented in Traat, Bondesson, Meister (2000). Moments of the two-phase sampling design can be evaluated by the following standard formulae:

$$E(I_i) = E_a\ E(I_i|\boldsymbol{I}_a), \tag{40}$$

$$V(I_i) = E_a V(I_i|\boldsymbol{I}_a) + V_a E(I_i|\boldsymbol{I}_a), \tag{41}$$

$$Cov(I_i, I_j) = E_a Cov(I_i, I_j|\boldsymbol{I}_a) + Cov_a(E(I_i|\boldsymbol{I}_a), E(I_j|\boldsymbol{I}_a)). \tag{42}$$

From (39) it also follows that the second-phase design $p(\boldsymbol{k}|\boldsymbol{I}_a)$ is an unbiased estimator of the two-phase design $p(\boldsymbol{k})$. The same holds for many second-phase design characteristics. For example, second-phase design moments $E(I_i^r|\boldsymbol{I}_a)$ are unbiased estimators of the two-phase design moments EI_i^r (see formula (40) for $r = 1$). On the other hand, since a two-phase sampling design is an averaged second-phase design with respect to the first-phase design, this averaged measure can be used to assess the performance of a resampling procedure.

Example 1. It is easy to see that with SI-sampling in both phases the final design is a SI-design. Let the sample size be n in the first phase. Then $p_a(\boldsymbol{k}_a) = \binom{N}{n}^{-1}$, if $|\boldsymbol{k}_a| = n$. The sample size m in the second phase gives $p(\boldsymbol{k}|\boldsymbol{k}_a) = \binom{n}{m}^{-1}$, if $|\boldsymbol{k}| = m$. The final outcome $\boldsymbol{k}$ has m components being equal to 1. To find the probability (38) we note that there are $\binom{N-m}{n-m}$ such vectors $\boldsymbol{k}_a$ in which the above mentioned m 1's are fixed, and the remaining $n - m$ 1's stand in the $N - m$ places. Finally the two-phase design is

$$p(\boldsymbol{k}) = \tbinom{N-m}{n-m}\tbinom{n}{m}^{-1}\tbinom{N}{n}^{-1} = \tbinom{N}{m}^{-1},\ \text{if } |\boldsymbol{k}| = m, \tag{43}$$

which is a SI-design with sample size m and with moments

$$E(I_i) = \frac{m}{N}, \quad V(I_i) = \frac{m}{N}(1 - \frac{m}{N}), \tag{44}$$

$$Cov(I_i, I_j) = -\frac{1}{N-1}\frac{m}{N}(1 - \frac{m}{N}). \tag{45}$$

Classical jacknife situation is a special case of the present example with $n = N$ and $m = N - 1$.

Let us now assume multinomial sampling with sample size m in the second phase

$$p(\boldsymbol{k}|\boldsymbol{I}_a) = \frac{m!}{\prod_{i=1}^{N} k_i!} \prod_{i=1}^{N} p_i^{k_i}, \text{ if } |\boldsymbol{k}| = m, \tag{46}$$

where $k_i \in \{0, 1, \ldots, m\}$ is a realized selection count of the element i in the second phase, and p_i is the probability to sample unit i in one trial. The p_is are random depending on the first-phase design:

$$p_i = \frac{u_i I_{ai}}{\sum u_l I_{al}}, \tag{47}$$

where u_is are certain positive unit-specific quantities in the population level (size measure, for example). Note, that index a marks the first phase, and the second index, if in the sum, runs from 1 to N. The first order marginal distributions of (46) are binomial

$$I_i|\boldsymbol{I}_a \sim B(m, p_i). \tag{48}$$

The terms in (46) depending on the first-phase outcome can have the following values:

$$p_i^{k_i} = (\frac{u_i I_{ai}}{\sum u_i I_{ai}})^{k_i} = \begin{cases} 0, & \text{if } I_{ai} = 0,\ k_i \neq 0; \\ 1, & \text{if } I_{ai} = 0,\ k_i = 0; \\ 1, & \text{if } I_{ai} \neq 0,\ k_i = 0; \\ (\frac{u_i k_{ai}}{\sum u_l k_{al}})^{k_i}, & \text{if } I_{ai} = k_{ai} \neq 0,\ k_i \neq 0. \end{cases} \tag{49}$$

One can see from (49) that if the unit i is not selected in the first phase $I_{ai} = 0$ then it can not be selected also in the second phase (the probability of corresponding sample becomes zero), and if the unit is selected in the first phase then it can be either unselected or selected in the second phase.

Expectations, variances and covariances of the components of $\boldsymbol{I}|\boldsymbol{I}_a$ in the second phase are the following multinomial moments:

$$E(I_i|\boldsymbol{I}_a) = mp_i, \tag{50}$$

$$V(I_i|\boldsymbol{I}_a) = mp_i(1 - p_i), \tag{51}$$

$$Cov(I_i, I_j|\boldsymbol{I}_a) = -mp_i p_j, \tag{52}$$

where p_i is given by (47).

Example 2. Let us assume SI-sampling in the first phase and SIR-sampling in the second phase ($\sum I_{ai} = n,\ \sum I_i = m,\ u_i \equiv c$ (a constant), i.e. $p_i = \frac{I_{ai}}{n}$). Then the two-phase sampling design has the form

$$p(\boldsymbol{k}) \;=\; \frac{m!}{n^m \prod_{i=1}^N k_i!} \sum_{\boldsymbol{k}_a(\boldsymbol{k})} p_a(\boldsymbol{k}_a),\ \text{if } |\boldsymbol{k_i}| = m. \tag{53}$$

The design (53) is not a SIR-design, except for the special case $\boldsymbol{k}_a \equiv \mathbf{1}$. Its moments are

$$E(I_i) \;=\; \frac{m}{N}, \tag{54}$$

$$V(I_i) \;=\; \frac{m}{N}(1 - \frac{1}{n} + \frac{m}{n}(1 - \frac{n}{N})), \tag{55}$$

$$Cov(I_i, I_j) \;=\; -\frac{m}{nN}(\frac{n-1}{N-1} + \frac{m}{N}\frac{N-n}{N-1}). \tag{56}$$

In the classical bootstrap situation we have $m = n = N$, and consequently, $E(I_i) = 1,\ V(I_i) = 1 - 1/N,\ Cov(I_i, I_j) = -1/N$.

Example 3. Let us assume a multinomial sampling with sample size $\sum I_{ai} = n$ and selection probabilities q_i in the first phase, and a multinomial sampling with sample size m and $p_i = \frac{I_{ai}}{n}$, in the second phase, i.e. $u_i \equiv c$. Then applying expressions of the multinomial moments for the first phase we get from (40)-(42) and (50)-(52)

$$E(I_i) \;=\; mq_i, \tag{57}$$

$$V(I_i) \;=\; \frac{m}{n}(n + m - 1)q_i(1 - q_i), \tag{58}$$

$$Cov(I_i, I_j) \;=\; -\frac{m}{n}(n + m - 1)q_i q_j. \tag{59}$$

We see the interesting phenomena that if the first-phase sample size $n \to \infty$ then two-phase sampling design tends to multinomial having sample size m and first-phase selection probabilities q_i.

Let us further assume hypergeometric sampling in the second phase. An hypergeometric sampling design is described by the multivariate hypergeometric distribution. It is the with-replacement sampling design where for each unit a predetermined upper bound of selection multiplicity is given. The hypergeometric second-phase design with sample size m can be presented in the following form:

$$p(\boldsymbol{k}|\boldsymbol{I}_a) = \frac{\prod_{i=1}^N \binom{u_i I_{ai}}{k_i}}{\binom{\sum u_l I_{al}}{m}},\ \text{if } |\boldsymbol{k}| = m, \tag{60}$$

where $u_i I_{ai}$ is the upper bound of the selection multiplicity for the unit i in the second phase and u_is are some predetermined counts. The first-order marginal distributions of this design are univariate hypergeometric distributions

$$\Pr(I_i = k_i | \boldsymbol{I}_a) = \frac{\binom{u_i I_{ai}}{k_i}\binom{\sum u_l I_{al} - u_i I_{ai}}{m - k_i}}{\binom{\sum u_l I_{al}}{m}}. \tag{61}$$

Using relationship (47) for p_i we have the following expressions for the first moments of the second-phase design (about hypergeometric distribution see Johnson, Kotz and Balakrishnan 1997):

$$E(I_i|\boldsymbol{I}_a) = mp_i, \tag{62}$$

$$V(I_i|\boldsymbol{I}_a) = m\frac{\sum u_l I_{al} - m}{\sum u_l I_{al} - 1} p_i(1 - p_i), \tag{63}$$

$$Cov(I_i, I_j|\boldsymbol{I}_a) = -m\frac{\sum u_l I_{al} - m}{\sum u_l I_{al} - 1} p_i p_j. \tag{64}$$

The variances and covariances differ from the ones of multinomial distribution by the factors less (or equal) than one.

Example 4. Let us have SIR-design with size n in the first phase ($\sum I_{ai} = n$, $EI_{ai} = \frac{n}{N}$, $V(I_{ai}) = \frac{n}{N}(1 - \frac{1}{N})$, $Cov(I_{ai}, I_{aj}) = -\frac{n}{N^2}$). Let us have hypergeometric design in the second phase with $u_i \equiv 1$, meaning that the upper bound of the multiplicity for each unit in the second phase is just the first-phase result of this unit. Using (62)-(64) and (40)-(42) we get the following moments for the two-phase design:

$$E(I_i) = \frac{m}{N}, \tag{65}$$

$$V(I_i) = \frac{m(n-m)}{N(n-1)}(1 - \frac{N+n-1}{nN}) + \frac{m^2}{nN}(1 - \frac{1}{N}), \tag{66}$$

$$Cov(I_i, I_j) = -\frac{m}{N^2}. \tag{67}$$

In the special case of $m = n$ (all units sampled in the first phase are sampled also in the second phase), the formulae (65)-(67) express the moments of the initial SIR-design.

Bibliography

1. Cassel, C.-M., Särndal, C.-E. and Wretman, J.H., *Foundations of Inference in Survey Sampling*, New York: John Wiley, (1977).

2. Hansen, M.H. and Hurwitz, W.N., On the theory of sampling from finite populations, *Annals of Mathematical Statistics*, **14**, (1943), 333-362.

3. Horvitz, D.G. and Thompson, D.J., A generalisation of sampling without replacement from a finite universe, *Journal of the American Statistical Association*, **47**, (1952), 663-685.
4. Johnson, N. L., Kotz, S. and Balakrishnan, N., *Discrete Multivariate Distributions*, New York: John Wiley, (1997).
5. Molina, E. A., Smith, T. M. F. and Sugden, R. A., Analytical inferences from finite population: a new perspective, *Preprint Series No. 330, Faculty of Mathematical Studies, University of Southampton*, (1999).
6. Pfefferman, D., Krieger A.M. and Rinott, Y., Parametric distributions of complex survey data under informative probability sampling, *Statistica Sinica*, **8**, (1998), 1087-1114.
7. Politis, D. N., Romano, J. P. and Wolf, M., *Subsampling*, New York: Springer-Verlag, (1999).
8. Särndal, C.-E., Swensson, B. and Wretman, J., *Model Assisted Survey Sampling*, New York: Springer-Verlag, (1992).
9. Traat, I., Sampling design as a multivariate distribution. *Research Report No. 17, Department of Mathematical Statistics, Umeå University*, Sweden, (1997).
10. Traat, I., Meister, K., Unified inference in survey sampling, In *Probability Theory and Mathematical Statistics, Proceedings of the* 7^{th} *Vilnius Conference*, Vilnius, Utrecht: VSP/TEV, (1998), 697-700.
11. Traat, I., Meister, K., Distributional assumptions for the inference in survey sampling, In *Bulletin of the International Statistical Institute, ISI 99, 52nd Session, Contributed Papers, Book 3*, (1999), 365-366.
12. Traat, I., Sampling design as a multivariate distribution. In *New trends in Probability and Statistics* **5***, Multivariate Statistics.* Vilnius, Utrecht: VSP/TEV, (2000), 195-208.
13. Traat, I., Bondesson, L., Meister, K., Distribution theory for sampling designs. *Research Report No. 2, Department of Mathematical Statistics, Umeå University*, Sweden, (2000).

Department of Mathematical Statistics
University of Tartu, 2 Liivi Str., 50409 Tartu, Estonia
E-mail: imbi@ut.ee

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.317-320

JOHAN TYSK

ABSENCE OF ARBITRAGE IN MARKETS WITH INFINITELY MANY ASSETS

We consider two different conditions for the absence of arbitrage in one-period models of markets with infinitely many assets. The notions of arbitrage under study we refer to as approximate and strict arbitrage, respectively. Strict arbitrage corresponds to making a profit without risk, whereas approximate arbitrage is making a profit with arbitrarily small risk. The condition of absence of approximate arbitrage is used, in the case of one-period models, to study the mean-variance efficient set. We show that for markets with infinitely many assets the absence of approximate arbitrage is equivalent to the conditions of absence of strict arbitrage and the cost functional being continuous.

2000 *Mathematics Subject Classification*: 91B28

Key words and phrases. Arbitrage, One-period model.

1. Introduction

The interest in different definitions of the notion of arbitrage has increased over the last few years where the main objective has been to find the right notion of arbitrage that establishes the equivalence between the absence of arbitrage and the existence of a martingale measure for the price process. In this direction various notions of approximate arbitrage have been introduced, notably the condition of "no free lunch with vanishing risk" of F. Delbaen and W. Schachermeyer in [1].

In the present paper we consider two different notions of absence of arbitrage for one-period models of markets with infinitely many assets. We state the conditions presented by G. Chamberlain and M. Rothschild in [2]. We refer to their definition of arbitrage as approximate arbitrage, since it involves making a profit with arbitrarily small risk. In their paper it is shown, as a consequence of the absence of this approximate arbitrage, that the cost functional is continuous, which in turn is used to study the mean-variance efficient set. In the present paper we study the difference between the stronger assumption of absence of approximate arbitrage and the weaker one of absence of strict arbitrage. Strict arbitrage should here be

interpreted as the possibility of a risk free profit with no initial investment. We show that assumption of absence of approximate arbitrage is the same as assumptions of absence of strict arbitrage and of the continuity of the cost functional.

2. A model of a market with infinitely many assets

We consider a market with countably many assets where the return of one currency unit invested in the ith asset is a random variable $x_i, i = 1, 2, 3, ...$, on some underlying probability space with probability measure P. Assuming that the returns x_i have finite variances, the returns belong to a Hilbert space $L_2(P)$ with an inner product

$$(p, q) = E(pq) = Cov(p, q) + E(p)E(q)$$

where $E(\cdot)$ denotes expected value and $Cov(\cdot, \cdot)$ is the covariance. The variance of a random variable x is denoted by $V(x)$ so that $V(x) = Cov(x, x)$. In this framework, portfolios p are just linear combinations $p = \sum_i \alpha_i x_i$ of the random variables x_i. The coefficients α_i are the amounts invested in the corresponding assets so that the cost of the portfolio p is $C(p) = \sum_i \alpha_i$. Short-selling is allowed so the coefficients α_i may be negative. To ensure that the cost functional is well-defined on the space of portfolios we may assume the $x_i's$ are linearly independent or more generally that the price of the zero portfolio, having zero variance and return, also has the price zero. We denote by M, the closure of the space of portfolios in $L_2(P)$, with respect to norm given above and refer to the elements of this space as limit portfolios. In this way M becomes a closed subspace of $L_2(P)$ and a Hilbert space in itself. We say that the market M permits no approximate arbitrage opportunities if the following conditions hold for sequences of finite portfolios p_n in M:

Condition AAi): If $V(p_n) \to 0$ **and** $C(p_n) \to 0$**, then** $E(p_n) \to 0$**.**

Condition AAii): If $V(p_n) \to 0$**,** $C(p_n) \to 1$**, and** $E(p_n) \to \alpha$**, then** $\alpha > 0$**.**

As a comparison we note that the absence of a risk free profit in M can be stated as

Condition A): If $V(p) = 0$ **and** $C(p) \leq 0$**, then** $E(p) \leq 0$**,**

for any portfolio p which we refer to as the condition of absence of strict arbitrage.

3. Equivalence between absence of arbitrage and continuity of the cost functional

We can now state the result on the equivalence between absence of approximate arbitrage and the conditions of continuity of the cost functional and absence of strict arbitrage.

Theorem. *There is an absence of approximate arbitrage if and only if the cost functional is continuous and there is an absence of strict arbitrage.*

Proof. We first assume an absence of approximate arbitrage. Let p_n be a sequence of portfolios with $\|p_n\| \to 0$. Following the argument in [2] we asssume, to reach a contradiction, that $C(p_n)$ does not tend to zero. Note that since C is a linear functional it is enough to reach this contradiction to verify that C is continuous. Since $C(p_n)$ does not tend to zero there is an $\epsilon > 0$ and a subsequence $\{p'_n\}$ with $|C(p'_n)| \geq \epsilon$. Defining $q_n = p'_n/C(p'_n)$ we have $C(q_n) = 1$ and

$$\|q_n\| = \|p'_n\|/|C(p'_n)| \leq \|p'_n\|/\epsilon \to 0.$$

Hence $E(q_n)$ tends to zero and we have arrived at the desired contradiction to AAii). Then C can be extended by continuity to M. By choosing p_n to be constant sequences in Condition AAi) and AAii) we see that these conditions imply Condition A, the absence of strict arbitrage.

To prove the converse we first assume that there are no non-zero limit portfolios with variance equal to zero in M. Then the variance is a norm in M equivalent to the given norm. This is a well-known fact, see for instance [2], but since we do not need explicit constants in the inequalites between the norms we present a simple proof here. Assume to reach a contradiction that there is a sequence p_n of portfolios with $V(p_n)$ tending to zero, but that (p_n, p_n) does not tend to zero. By choosing a subsequence and multiplying by suitable scalars we can always assume that $E(p_n) = 1$. This means that $(p_n - 1, p_n - 1)$ tends to zero, i.e. p_n tends to a non-zero asset with zero variance which is the desired contradiction. Thus Condition AAi) holds and AAii) is vacously satisfied.

Next we assume that there are assets with variance equal to zero. From condition A we see that an asset with zero variance with cost normalized to one is unique. Let us call this asset s and refer to it as the riskless asset. The expected return of s has to be positive by condition A. We note that s is a limit portfolio and Condition A is formulated only for portfolios, but of course this condition can be extended to limit portfolios by the continuity of the cost functional. Now, we can write our space of portfolios F as the direct sum $F = F' + F_s$, where F' contains no non-zero elements of variance zero and F_s is the one-dimensional space being the linear span of s. We

can for instance simply choose F' to be the orthogonal complement of F_s. Now, let us first verify Condition AAi). Thus we take a sequence p_n with variance and cost tending to zero with n. We write $p_n = p'_n + \lambda_n s$, where p'_n belongs to F' and λ_n is a real number. Then

$$V(p_n) = V(p'_n) + V(\lambda_n s) = V(p'_n),$$

where the equalities hold since s has variance zero. In F' the covariance is an inner product so we can conclude that p'_n tends to zero. Since the cost functional is assumed to be continuous we then know that

$$C(p_n) \to \lambda_n C(s) = \lambda_n.$$

Since $C(p_n)$ is assumed to tend to zero with n this means that λ_n tends to zero, i.e. p_n tends to zero and thus the expected return tends to zero since it is continuous with respect to the given norm.

Finally, to verify condition Aii), we consider a sequence p_n of portfolios and make the same decompostion as above. Also in this case we conclude that p_n tends to $\lambda_n s$. The condition that the cost tends to 1 with n means that λ_n tends to 1. Thus

$$E(p_n) \to E(s) > 0,$$

thus completing our proof.

Bibliography

1. F. Delbaen, W. Schachermayer, A general verions of the fundamental theorem of asset pricing, *Math. Ann.* **300** (1993) 463–520.
2. G. Chamberlain and M. Rothschild, Arbitrage, factor structure, and mean variance analysis on large asset markets, *Econometrica* **51** (1983) 1281–1305.

Department of Mathematics, Uppsala University, 751 06 Uppsala, Sweden.
E-mail: johan.tysk@math.uu.se

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.321-332

MIKHAIL I. YADRENKO, NADIYA M. ZINCHENKO,
DMITRII S. SILVESTROV

NEW EDUCATIONAL PROGRAMME ON APPLIED STATISTICS, FINANCIAL AND ACTUARIAL MATHEMATICS AND TEMPUS TACIS JEP "STATISTICAL ASPECTS OF ECONOMICS"

Historical remarks about the development of statistical studies at Kyiv National Taras Shevchenko University are given. New educational programmes on Applied Statistics, Financial and Actuarial Mathematics and the results of realization of the Tempus Tacis Joint Europen Project "Statistical Aspects of Economics" are presented.

2000*AMS Mathematics Subject Classifications.* 50A10, 50B20.

Key words and phrases. Educational programme, applied statistics, financial and actuarial mathematics, Tempus Tacis Joint European Project.

1. Historical remarks

Kyiv National Taras Shevchenko University (founded in 1834) is a leading higher educational institution in Ukraine. Traditionally it is a base for implementation of the best educational innovations and its teaching and training experience is used intensively by other Ukrainian universities. The lectures in mathematics were delivered at Kyiv University since its foundation; at first at the Faculty of Natural Science, then at the Faculty of Physics and Mathematics. The outstanding specialists in algebra, theory of functions, geometry, differential equations, mechanics, mathematical physics worked at the faculty, among them: D. Grave, B. Delone, O. Shmidt, M. Chebotaryov, M. Kravchuk, E. Slutskii, M. Krylov, M. Bogolyubov.

It worth to be mentioned that the first lecture in probability theory was delivered by M. Vashchenko-Zakharchenko in 1863 and the first text-book in probability theory was published by V. Ermakov in 1878 in the "University Notes".

Professor D. Grave was the first who delivered courses in actuarial mathematics and published text-books on insurance mathematics (1912), theory of pension funds (1917) and mathematics of social insurance (1924). In 1911 E. Slutskii represented the work "The Theory of Limiting Utility" to graduate competition and worn gold medal. This work was not published and hand-written variant is now in the Ukrainian National Scientific Library.

In 1940, the Faculty of Physics and Mathematics was divided into two Faculties; so, the Mechanics and Mathematics Faculty was founded.

The Department of Probability Theory and Mathematical Statistics of Kyiv University was founded in 1949 by the prominent scientist Academician of the Ukrainian Academy of Sciences, Professor B. Gnedenko, who was a head of the department from 1949 till 1959. His first students were the future well-known mathematicians V. Korolyuk, V. Mykhalyevich, A. Skorokhod. At the same time I. Gikhman began intensive investigations in the theory of random processes and stochastic analysis. Professor I. Gikhman was a head of the department from 1962 till 1966 followed by Professor M. Yadrenko, who held this position from 1966 till 1998. Academicians V. Korolyuk in 1956 - 1999 and A. Skorokhod in 1956 - 1992 lectured at the department sharing their professorships with the work in the Institute of Mathematics of Ukrainian Academy of Science. Also, I. Ezhov in 1962 - 1975 and D. Silvestrov in 1974 - 1992 were Professors at the department.

Most of representatives of the world-wide recognized Ukrainian school of probability theory and stochastic processes are the department's graduates, among them seven Full Members and eight Corresponding Members of National Academy of Science, 40 Doctors of Science and 135 Candidates (Ph.D.) of Science in Mathematics.

At present, there are six Professors at the department: Yu. Kozachenko (the head of the department), M. Yadrenko, M. Kartashov, M. Leonenko, M. Moklyachuk and R. Maiboroda. Also Associate Professors O. Borysenko, A. Olenko, O. Ponomarenko, the secretary T. Lapida and engineer O. Vasylyk work at the department. There is also a research group at the department, which staff is: Dr.Sc. N. Zinchenko and Dr.Sc. V. Masol (leading researchers), Dr. V. Parkhomenko (senior researcher), Dr. L. Sakhno and Dr. Z. Vyzhwa (researchers), G. Bagro , A. Revenko and I. Didkovsky (engineers).

Graduates of the Department of Probability Theory and Mathematical Statistics work at research and educational institutions in 27 countries. Nowadays the department has 145 students and 23 Ph.D. students on study programmes.

The Department of Probability Theory and Mathematical Statistics is responsible for the edition of two scientific journals. The first one is *Theory of Probability and Mathematical Statistics* (Editor-in-Chief is Professor A. Skorokhod). This journal has a well established international reputation and

it is translated in English by the American Mathematical Society beginning from the first issue appeared in 1970. Another journal *In the World of Mathematics* (Editor-in-Chief is Professor M.Yadrenko) is the scientific-popular mathematical journal for scholars and students. The journal represents current discoveries and achievements in mathematics, unsolved problems, typical examination tasks, gives information about mathematical olympiads of different levels. Besides this, the department takes an active part in edition of other two scientific journals: *Theory of Stochastic Processes* and *Random Operators and Equations*.

2. Studies in Probability and Mathematical Statistics

During 50 years the Department of Probability Theory and Mathematical Statistics provides the studies in probability theory and mathematical statistics in the framework of educational direction "Mathematics". The corresponding curricula include basic courses in mathematical and functional analysis, algebra, geometry, differential equations as well as general and special courses in probability theory, theory of stochastic processes, various aspects of mathematical statistics, stochastic analysis, computer statistics. In 1956 the Department initiated the course "Programming for High-Speed Computers". The first in USSR text-book in programming by Gnedenko, Korolyuk and Yushchenko (1962) also was prepared at the department.

In the 70th the department introduced new specializations connected with applied statistics and methods of optimization.

During 50 years more than 60 research monographs and text books were published by the staff of the department. We point out here only some of the most well known text books: the text book on probability theory by Gnedenko (1950) re-printed in many editions and translated to several languages (8 Russian, 11 German, 6 English editions), the text book on stochastic processes by Gikhman and Skorokhod (1965 and 1977), the text book by Skorokhod (1990) on stochastic processes, based on his well known course delivered at Kyiv University, as well as the book by Dorogovtsev, Silvestrov, Skorokhod and Yadrenko (1976), which is possibly the largest (about 1500 problems) collection of problems in probability and stochastic processes, have Ukrainian, Russian and English editions.

In the middle of 90th the transformation processes in Ukrainian economics caused the needs in qualified specialists in applied statistics skillful in financial and actuarial mathematics, financial analysis and risk theory, econometrics and sample survey. In new economic conditions such specialist are in high demand with banks, financial institutions and insurance companies and governmental organizations.

The Department of Probability Theory and Mathematical Statistics initiated the preparation new courses in economic-statistical area, implemen-

tation the new Educational Programme "Statistics" in Kyiv University and its approval by national educational authorities.

Professor M. Yadrenko, who was a head of the department at that time, Associate Professor O. Ponomarenko and the dean of the Mechanics and Mathematics Faculty , Professor M. Perestyuk initiated the preparation of the first in Ukraine text books in this area. The text-books on mathematical economics by Ponomarenko, Perestyuk and Burym (1995), statistical methods in econometrics and financial mathematics by Leonenko, Mishura, Parhomenko, and Yadrenko (1995) and system methods in economics, management and business by O.Ponomarenko and V.Ponomarenko (1995) were published under support of US AID grants.

In 1996 the new educational direction "Statistics" was approved by the Ukrainian Ministry of Education and included in the official state list of education specialities (it is the highest level of official recognition in Ukraine).

In the same teaching year 1996/97 the Mechanics and Mathematics Faculty started three level (bachelor, spesialist, master) Educational Programme in "Statistics".

In order to support the realization of this educational programme Professors G. Kulldorff and D. Silvestrov from Umeå University (Sweden) and Professor M. Yadrenko initiated an application for the Tempus Tacis Pre-JEP. This pre-project coordinated by Professor G. Kulldorff was realiszed in 1997 and has been following by a full scale tree years Tempus Tacis Joint European Project "Statistical Aspects of Economics" coordinating by Professor D. Silvestrov.

3. Curricula for three level educational programme "Statistics"

The Ukrainian university system can be characterized as a teacher-oriented system. The teachers play leading role in the studies. The main part of work is moved to classrooms. In total it can be 30-32 such hours per week. The education process is shifted towards lectures as a main element of the process.

This differs very much of Swedish university system which can be characterized as a student-oriented system, where the education process is shifted towards self-work of students and the role of teacher is in some sense to assist students in their studies. Here, a regular self-work of students with textbooks and other teaching materials is an important element of studies while an amount of classroom can actually be only 14-16 hours per week.

One can say that Ukrainian system is more concentrated on maximisation of information content of the education programme while Swedish system is more oriented towards the development self-education skill of students.

It should be also pointed out that the secondary education in Sweden is based on 12 years programme, while now in Ukraine it is 10 years programme. The studies of English, other languages and some humanitarian disciplines are realized at the secondary education programme. That is why language and humanitarian courses as a rule are not included in the university education programme at the natural science and engineering specialities in Sweden. In contrary education programme in Ukrainian universities include language courses and humanitarian disciplines. Also a military service, which is included in the programme of studies at Ukrainian universities, is not included in the programme of studies at Swedish universities. We refer here to the paper by Silverstrov and Silvestrova (1999), where one can find discussion concerning these questions.

The three level Educational Programme "Statistics" gives possibility to students to choose one of three level of education and graduate from university with of one of the following degrees: "Bachelor" (four years of studies), "Specialist" (five years of studies) and "Master" (six years of studies).

Total duration of the basic Bachelor programme is four years. It consists of two parts: (i) block of humanitarian disciplines including history of Ukraine, Ukrainian culture, world culture and history of Mathematics, philosophy, psychology, pedagogics, politology, foreign language, ecology, foundations of the law, additional military training and physical culture training (1776 hours in total), (ii) block of fundamental and professionally oriented discipline. In total there is 7568 teaching hours in the programme including 2862 lectures and 5010 other types of auditorium studies. The list of basic obligatory fundamental courses for Bachelor programme is given below in the Table 1.

1.	Management and marketing	48 hours
2.	Mathematical Analysis I	420 hours
3.	Analytical Geometry	108 hours
4.	Linear Algebra	289 hours
5.	Discrete Mathematics	130 hours
6.	Mathematical Analysis II	420 hours
7.	Algebra and Number Theory	210 hours
8.	Differential Equations	235 hours
9.	Financial analysis and calculations	108 hours
10.	Informatics	420 hours
11.	Foundations of micro- and macroeconomics	108 hours
12.	Foundations of the theory of finance	102 hours
13.	Discrete models in the Theory of Probability	102 hours
14.	Theory of Probability	135 hours
15.	Mathematical Statistics	102 hours
16.	Functional Analysis	210 hours
17.	Theory of measure and integration	108 hours

18.	Complex analysis	211 hours
19.	Decision making	158 hours
20.	Method of economical calculations	108 hours
21.	PC practical studies	210 hours
22.	Mathematical Economics	153 hours
23.	Equations of Mathematical Physics	204 hours
24.	Foundations of Actuarial Mathematics	158 hours
25.	Variational calculus and methods of optimization	96 hours
26.	Mathematical logics	81 hours

Students can also choose between specialities inside of Educational Programme "Statistics". These specialities are (1) Financial and Actuarial Mathematics, (2) Econometrics and Mathematical Economics, (3) Applied Statistics.

The Bachelor programme includes a large number of special courses, which can be obligatory or facultative (740 hours in total) and can variate depending of the Speciality.

The list of *special courses for Bachelor programme* include such courses as (number of hours is shown in the brackets): advanced course of probability theory (51), discrete probability in martingale representation (52), advanced course of mathematical statistics (68), stochastic processes (68), applied statistics (50), stochastic differential equations (32), statistical inference for random processes (50), time series analysis (52), non-stationary non-linear time series (36), limit theorems for risk processes (36), discrete models for random evolutions (70), partial differential equations with random initial conditions (34), computer statistics (68), non-negative matrices and their applications (70), statistical analysis of psychological data (34), ergodic theory (36), methods of data classification (32).

The Specialist and Master programmes do include some additional obligatory basic courses and also a large number of special courses, which can be obligatory or facultative. The curriculum for *Master* programme differs from those for *Specialist* programme by emphasis on individual activity plans for each student. The *basic courses for Specialist and Master* studies are (number of hours is shown in the brackets): optimization methods (64), stochastic methods in management (64), financial mathematics of funds market (64), theory of decision making (48), statistical methods in socio-humanitarian discipline (24).

The list of *special courses* for Specialist and Master programmes includes such courses as: statistical inference for random processes (32), multivariate statistical analysis (32), random evolutions and their applications in financial and actuarial mathematics (32), spectral theory of random fields (32), correlation and spectral theory of random series in economics (32) , mathematical models in life insurance (32), probabilistic methods of information defense (32), models with stable distributions in finance and insurance

(32), stochastic methods in economics and management (36), probabilistic methods in combinatorics (48), simulation of random processes (36), robust estimation of random processes (48), wavelet analysis in mathematical economics (48).

In June 2000 the first group (25 students) finished the four year studies and graduated from Kyiv University with degree "Bachelor in Statistics". In 2000/2001 academic year first two groups started training according to programme "Specialist in Statistics" and "Master in Statistics".

4. Tempus Tacis Joint European Project "Statistical Aspects of Economics"

The Tempus Tacis EU Programme (approved by European Council in 1990 and prolonged for six years from 1 July 2000) is directed to stimulate co-operation between higher education institutions of European Union and New Independent States in the development and restructuring of their education system.

According to this Programme the Tempus Tacis Joint European Project "Statistical Aspects of Economics" is carrying out in 1998-2001. The EU Grant supports it for the amount of 558.000 Euro. Kyiv National Taras Shevchenko University, Umeå University (co-ordinating institution), Stockholm University, Mälardalen University and University of Helsinki take part in the project.

The *project objective* is to promote the creation of the three level educational system (Bachelor, Specialist and Master degrees) at Kyiv University for the new educational direction "Statistics" including new economic-statistical specialities: (1) Financial and Actuarial Mathematics, (2) Econometrics and Mathematical Economics and (3) Applied Statistics.

The programme of the project includes creating the new curricula, new courses with the corresponding teaching materials in economic-statistical area, development of the cooperation with Faculty of Economics; improving of the academic staff qualification, visits of professors and teachers from Kyiv University to EU partner universities for collection of information, the development of new courses and training; visits of professors and teachers from EU partner universities to Kyiv University for consulting and lecturing; participation of teachers and postgraduate students from Kyiv University in conferences and workshops related to the subject of project; purchase of computer equipment, software and teaching literature to Kyiv University; organization of workshops in Ukraine, Sweden and Finland, development of co-operation with state and non-state Ukrainian institutions and dissemination of project results.

At Kyiv National Taras Shevchenko University, the project is carried out by the Department of Probability Theory and Mathematical Statistics

(Mechanics and Mathematics Faculty) in cooperation with the Department of Economic Cybernetics (Faculty of Economics).

Coordinator of the Project is Professor D. Silvestrov (Umeå University and Mälardalen University (from 1999)). Local Coordinators are Professor A. Martin-Löf (Stockholm University), Professor E. Valkeila (University of Helsinki) and Professor M. Yadrenko (Kyiv University).

The first step in achievement the project objective was the promotion of the creation the new curricula for the three level education system for Educational Programme "Statistics" adopted to the EU standards for higher education. According to new curricula 25 new courses were prepared at the Department of Probability Theory and several existing basic courses were changed and updated. Also more than 10 new courses were prepared at the Department of Economic Cybernetics.

All new and reconstructed courses are provided with relevant teaching materials (collections of syllables with detailed lists of recommended literature, lecture notes, manuals, text books, tasks for practical studies, control tests, etc.). The list of main teaching materials and text books prepared and published within the framework of the Project by the staff of the Department of Probability Theory and Mathematical Statistics is given in the bibliography under the numbers [14]-[29]. Also more than 15 works were prepared by the staff of the Department of Economical Cybernetics, the main of them are [27]-[34].

To promote these activities an intensive mobility programme involving about 35 professors and teachers from Kyiv University was realised within the Project. Professors and teachers from Kiev University visited EU partner universities for collection of information, the development of new courses and training, they also participated in conferences, schools, workshops and short intensive courses related to the subject of the Project and held in Sweden and Finland. In total the mobility programme of the Project included more than 120 such trips.

Professors G. Kulldorff, A. Martin-Löf, E. Valkeila, D. Silvestrov and other staff members in Umeå University, Stockholm University, University of Helsinki and Mälardalen University provided lecturers from Kyiv University by books, periodicals, software, advices and consultations in financial mathematics, risk theory, econometrics, survey sampling and other subjects. They gave all necessary support for efficient and productive cooperation within the Project activities.

As was mentioned above professors and teachers from Kyiv University participated in the short intensive courses, schools and workshops in Sweden and Finland. For example, seven Ukrainian lecturers took part in the work of the 52nd Session of the International Statistical Institute in 1999. This activity gave them possibility to collect the latest information in the area and to update the courses in various statistical topics.

During the second and the third Project year a student mobility programme was also realized within the project for Ph.D. students from Kyiv University. For example a special Ph.D. studies in mathematical economics were organized at the Mälardalen University in January 2001.

At the same time, a large group of Swedish and Finish professors and lecturers, among them G. Kulldorff, A. Martin-Löf, E. Valkeila, D. Silvestrov, K. Bränäs, M. Gyllenberg, H. Niemi, P. Salminen visited Kyiv University for consulting and lecturing as well as for lecturing at three schools organised in Ukraine within the Project.

The technical base of the Mechanics and Mathematics Faculty was essentially improved due to equipment purchased via Project. The Faculty of Mechanics and Mathematics have got access to Internet and created the base for further computerization. The PC class for practical studies in applied statistics was organized. This give the possibility to introduce new courses, such as *Methods of Economical Calculus*, *Mathematical Economics*, *Computer Statistics*, *Time Series Analysis* based on applied software MATEMATICA and SPSS purchased via the Project. New active methods of teaching and control testing were implemented.

Also a library of recent statistical and economics literature (including more that 150 books) was created that is of a great importance for the development of the new educational programme.

The dissemination is an essential part of Project activities. Three International Schools in economic-statistical area were successfully organized within framework of the Project: the First International School on Financial Mathematics and Mathematical Economics (Kyiv, 7-12 September 1998), the Second International School on Actuarial and Financial Mathematics (Kyiv, 8-12 June 1999) and the Third International School on Applied Statistics, Actuarial and Financial Mathematics (Feodosiya, 4-13 September 2000).

Representatives (lecturers, researchers, managers, Ph.D. and MA Students) from Kyiv University and other Ukrainian Universities, National Academy of Sciences, Foreign Exchange, Pension Fund, private and state insurance companies took part in the work of these schools. Prominent specialists from Finland, Sweden, Ukraine and also from Denmark, Estonia, Russia, USA delivered lectures in economic-statistical areas. The programmes of these schools covered the following topics: mathematical models in finance and insurance; time series in economics, finance and insurance; analytical, simulation and statistical methods in the risk theory; optimization and financial software, mathematical models in micro- and macroeconomics, survey sampling in economics and the teaching methodology and programmes in mathematical economics and related subjects. Teaching-methodological seminars and round tables were carried out within framework of the first and third International Schools for dissemination on

national level of results, teaching know-how and experience achieved within the Project in Kyiv University .

The International School on Mathematical and Statistical Application in Ecinomics (Västerås, 15-19 January 2000) was also organized in the framework of the Project and with the support of the universities participating in the project.

The Proceedings of these schools were published (See, [26]-[28] and the current volume) and have been distributing to Ukrainian universities, financial and insurance institutions, libraries that contributes to the dissemination programme realizing within the Project.

Such activities not only provides a good training/updating for a wide educational and scientific community but supports the development of the links with external environment. Due to realization of the Project the Department of Probability Theory and Mathematical Statistics (Mechanics and Mathematics Faculty) and the Department of Economic Cybernetics (Faculty of Economics) at Kyiv University become centers of implementation and dissemination of know-how, modern teaching methodology and good practices in new economic-statistical specialities.

The Tempus Tacis Joint European Project "Statistical Aspects of Economics" is a very good example of useful co-operation between Scandinavian and Ukrainian universities. The realization of the Project effectively promotes the development of new type of economic-statistical specialities answering the requirement of the transition market economics in Ukraine and adapted to the EU standards for higher education as well as the development of international co-operation in the area of higher education and science.

Bibliography

1. Grave, D. *Insurance Mathematics*, Kyiv (1912).

2. Grave, D. *Theory of Pension Funds*, Kyiv (1917).

3. Grave, D. *Mathematics of Social Insurance*, Leninggrad, Kyiv (1924).

4. Slutskii, E. *The Theory of Limiting Utility*, Unpublished manuscript (400 pages), Ukrainian National Scientific Library, Kyiv (1911).

5. Gnedenko, B., Korolyuk, V. and Yushchenko, E. *Elements of Programming*, "Fizmatgiz", Moscow (1962).

6. Gnedenko, B.V. *Course on Probability Theory*, "Fizmatgiz", Moscow (1950).

7. Gikhman, I.I., Skorokhod A.V. *Introduction in Theory of Stochastic Processes*, Nauka, Moscow, (1965, 1977), Saunders Co., Philadelphia (1969).

8. Skorokhod A.V. *Lectures on Stochastic Processes*, "Lybid", Kyiv (1990).

9. Dorogovtsev, A.Ya., Silvestrov, D.S., Skorokhod, A.V. Yadrenko, M.I. *Probability Theory. A Collection of Problems*, "Vyshcha Shcola" , Kyiv (1976, 1980), Translations of Mathematical Monographs, **163**, Amer. Math. Society (1997).

10. Ponomarenko, O.I., Perestyuk, M.O. and Burym, V.M. *Foundations of Mathematical Economics*, "Informtechnika", Kyiv (1995).

11. Leonenko,M.M., Mishura, Yu.S., Parhomenko, V.M. and Yadrenko, M.I. *Probabilistic and Statistical Methods in Econometrics and Financial Mathematics*, "Informtechnika", Kyiv (1995).

12. Ponomarenko, O.I. and Ponomarenko, V.O. *System Methods in Economics, Management and Business*, "Informtechnika", Kyiv (1995).

13. Silvestrov, D. and Silvestrova, E. *An outline of Master programme in actuarial and financial mathematics*. Theory Stoch. Proces. **5(21)**, 1-2 (1999), 188-195 (In Proceedings of the Second International School on Actuarial and Financial Mathematics, Kyiv, 1999).

14. *Collection of Programmes for Special Courses of Higher Education with Professional Direction Statistics* (6.080102, 7.080102, 8.080102). *Bachelor's, Specialist's and Master's degree*, Kyiv University (2000).

15. *Collection of Programmes for Normative Courses of Higher Education by Professional Direction* 6.080102 - *Statistics. Bachelor's degree*, Kyiv University (1999).

16. Maiboroda, R.E. *Computer Statistics*, Kyiv University (1999).

17. Zinchenko, N.M., Yadrenko, M.I., *Lecture Notes on Mathematical Models in Risk Theory. Risk Processes in the Presence of Large Claims*, Kyiv University (2000).

18. Zinchenko, N.M., Olenko, A.Ya. *Analytical Models and Methods of Sociology*, Kyiv University Publishing Center (2000).

19. Ponomarenko, O.I. *Modern Analytical Politology*, Kyiv University Publishing Center (2000).

20. Borisenko, O.D., Ponomarenko, O.I. *Collection of Tests on Foundation of Microeconomics and Macroeconomics*, Kyiv University (1999).

21. Borisenko, O.D., Maiboroda, R.E. *Analytical and Statistical Methods and Models of Psychology and Pedagogics*, Kyiv University Publishing Center (2000).

22. Ponomarenko, O.I. *Lecture Notes on Financial Analysis. Technical and Fundamental Analysis of Financial Markets*, Kyiv University (2000).

23. Moklyachuk, M.P. *Theory of Choice and Decision Making*, Kyiv University (2000).

24. Moklyachuk, M.P. *Lecture Notes on Robust Methods of Estimation*, Kyiv University (2000).

25. Kozachenko, Yu.V., Pashko, A.I. *Modeling of the Random Processes*, Kyiv University Publishing Center (2000).

26. Parhomenko, V.M. *Methods of Survey Sampling*, Kyiv (2001).

27. *Proceedings of the First International School on Financial Mathematics and Mathematical Economics*, Kyiv University Publishing Center (1998).

28. *Proceedings of the Second International School on Actuarial and Financial Mathematics*, Kyiv, 1999, Theory of Stochastic Processes, **5 (21)**, No. 1-2 (1999).

29. *Proceedings of the Third International School on Applied Statistics, Actuarial and Financial Mathematics*, Feodosiya, 2000, Theory of Stochastic Processes, **6 (22)**, No. 3-4 (2000).

30. Karagodova, O., Mertens, O. and Fedorenko, I. *Mathematical Methods in decision Theory*, Kyiv University Publishing Center (1999).

31. Kaminski, A., Shostak, N. *Methods of Mesurement of Risk of Investment Projects*, "Kozaky", Kyiv (1999).

32. Chernyak, O., Obushna, O. and Stavitsky, A. *Collection of the Problems on Theory of Probability and Mathematical Statistics*, "Znannja", Kyiv (2001).

33. Kaminski, A., Kiyak, A., Shepeleva, O. *Optimization of Investment Portfolio*, "RAPT", Kyiv (2000).

34. Chernyak, O. *Technique of Sampling Survey*, Publishing Centre MIVVC, Kyiv (2000).

Department of Probability Theory and Mathematical Statistics,
Kyiv Taras Shevchenko University, 252601 Kyiv, Ukraine.
E-mails: znm@mechmat.univ.kiev.ua and ymi@mechmat.univ.kiev.ua

Department of Mathematics and Physics, Mälardalen University,
SE-721 23 Västerås, Sweden.
E-mail: dsv@mdh.se

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.333-345

TETYANA ZATONATSKA

USING STATISTICAL METHODS FOR MARKETING RESEARCH

One of the problems in statistical processing of the marketing investigations (for example, interrogation) is the separation of the groups of objects with similar characteristics that is to say segments. The correct customers segmenting is possible with the use of mathematical instrument based on the concept of "distance" between objects. In mathematical statistics the process of division of the investigated objects population into groups, which are uniform according to definite characteristics, is called as classification. One of the instruments of the classification is cluster analysis. In the calculation of the distance between clusters the data of definite informational fields of questionnaires are used. These fields are called as classifying parameters. The parameters have different physical measurement units.

2000 *Mathematics Subject Classifications.* 62P20

Key words and phrases. Marketing investigations, customers, objects, units.

Clustering techniques try to find similar patterns within sets of data. Clustering groups elements that are similar on their dimensions into homogenous segments (clusters). And usually computer software is used to carry out the analysis.

In this arcticle we are going to consider clustering technique that uses Euclid's formula - mathimatic formula for measurenig distance in between multidimensional objects. Close consideration of this technique helps deeper comprehend clustering analysis procedure and results when it is carried out by the computer.

The technique is going to be illustrated with the example where we will group different cigarette brands, present at the Ukrainian market, by 2 parameters into 5 homogenous segments.

Let's start with the Euclid's formula as it is:

$$R_{A,B} = \sum_{j=1}^{n} ((X_{Aj} - X_{Bj})^2)^{1/2} \tag{2}$$

where,

n number of dimensions;

X_{Aj} , X_{Bj} values of j dimension of objects A and B;

$R_{A,B}$ distance between objects A and B measured by n dimensions.

Using this formula the way it is implies 2 major problems that can pervert results of the analysis:

1. Values of dimensions can be measured in different units (years, minetary units, custom scale rates, etc.).

2. Values of dimensions can substantially differ by arithmetic extent.

These problems are avoided by mean of rate setting procedure. It implies computation of fixed values of dimensions using the folloing formula:

$$Z_{ij} = (X_{ij} - X_{jm})/\sigma_j \tag{3}$$

where,

Z_{ij} fixed value of j dimension for i object;

X_{ij} initial value of j dimension for i object;

X_{jm} mean value of j dimension;

σ_j standard deviation of j dimention.

X_{jm} and σ_j are computed using the following formulas:

$$X_{jm} = \sum_{i=1}^{m} (X_{ij})/m \tag{4}$$

$$\sigma_j = ((\sum_{i=1}^{m} (X_{ij} - X_{jm})^2)/m)^{1/2} \tag{5}$$

where,

m - number of objetcs in the data array;

X_{ij} initial value of j dimension for i object.

Rate setting procedure is effective due to one important assuamtion. It is that values of most dimensions have normal probability distribution. That means that fixed values Z_{ij} appears in the range [-3σ; 3σ] with the probability of 99,7%. That allows to compare all kinds of dimensions.

For more precise clusterng weights of different dimensions are added to the formula. Weight (W_j) shows how much a specific dimenion is critical for the object of the research. Note that $0< W_j \leq 1$.

So the final formula for computation of distance in between multidimentional objects is the following:

$$R_{A,B} = (\sum_{j=1}^{n} W_j (Z_{Aj} Z_{Bj})^2)^{1/2} \tag{6}$$

Now let's illustrate the clustering technique with the example.

Data array for the example was formed of 30 (number of objects) cigarette brands that all together are about 95% of all the brands present at the Ukrane cigarette market. Two dimansions price and image rate are taken for omputation of distances. The dimensions weights are equal W_p=W_{ir}=1.

Prices are given the way they appeared at places of observation (small reatil points of sale) in national monitary units (uhr). Image rates were set based on author's perception of it (usually image rates are determined by group of experts).

The table of initial data and fixed values is given below.

Standart deviations and mean values are shown below the table. Values in columns **(P/c.-Aver.)**^{2} and **(Im.-Aver.)**^{2} are interim and were computed to simplify the computation of standard deviations.

The following table is the matrix of distances in between objects in the data array computed by formula (5).

The dimansion of the matrix is 30 x 30. The right upper part of the matrix is empty because values of distances here are the same that in the left lower part. Note that filling in the right upper part and leaving empty left lower woudn't be a mistake and is optional.

The next step after we had built the matrix of distances would be to choose the smallest value in the matrix. For our example it is 0 (see table 2). In our example we have several values that equal 0 in the matrix and we can choose whichever we want. After that we put 2 objects the distance between which is smallest togethher and reconstruct the matrix considering these binded objects as one. Note that the initial matrix of distances consists of clusters only with one element and by the end of the analysis we are going to have group clusters.

To reconstruct the matrix of distances we have to figure out distances from the new object to all ather objects in the data array. For that we compare values of distances form parts of the new object to other objects and chose the smaller ones. These would be values of distances from the new object to other objects.

The above procedure beginning with choosing the smallest value of distance in the matrix is repeted (n-n_w) times (n is number of objects in the data array; n_w is number of clusters at the end of the analysis). For our example it is (30 - 5) = 25 times.

The following table is the final 5-cluster matrix of distances:

Also discription of final clusters is given in tables below.

As it is shown in these tables as the result of clustering analysis in our example we have recieved 5 homogeneus segments:

; **segment #1** 60% of all brands, avarage price per pack of 20 pcs. is 2,25 (uhr), image rate 2,39 points;

; **segment #2** 16,7% of all brands, avarage price per pack of 20 pcs. is 4,22 (uhr), image rate 5 points;

; **segment #3** 3,3% of all brands, avarage price per pack of 20 pcs. is 7 (uhr), image rate 10 points;

; **segment #4** 13,3% of all brands, avarage price per pack of 20 pcs. is 6,5 (uhr), image rate 9 points;

; **segment #5** - 6,7% of all brands, avarage price per pack of 20 pcs. is 5,3 (uhr), image rate 6,5 points.

Considering the results of clustering analysis we can see how it helped us see major segments at the Ukrainian cigarette market.

Economic Cybernetics Department, Faculty of Economics, Kyiv National Taras Shevchenko University, Kyiv, Ukraine.
E-mail: chernyak@mail.univ.kiev.ua

pack	Image	Price/cigarette uh	(P/c.-Aver.)^2	(Im.-Aver.)^2	P/c. Norm.	Im. Norm.
20	3	0,135	0,0016	1,5211	-0,4422	-0,4561
20	2	0,085	0,0080	4,9878	-1,0020	-0,8259
20	1	0,07	0,0109	10,4544	-1,1699	-1,1957
20	5	0,225	0,0026	0,5878	0,5654	0,2835
20	3	0,135	0,0016	1,5211	-0,4422	-0,4561
20	4	0,15	0,0006	0,0544	-0,2743	-0,0863
20	2	0,085	0,0080	4,9878	-1,0020	-0,8259
20	10	0,35	0,0308	33,2544	1,9647	2,1326
20	9	0,325	0,0227	22,7211	1,6849	1,7627
20	5	0,215	0,0016	0,5878	0,4534	0,2835
20	3	0,12	0,0030	1,5211	-0,6101	-0,4561
20	6	0,255	0,0065	3,1211	0,9012	0,6533
20	5	0,215	0,0016	0,5878	0,4534	0,2835
20	2	0,085	0,0080	4,9878	-1,0020	-0,8259
20	4	0,17	0,0000	0,0544	-0,0504	-0,0863
20	5	0,2	0,0007	0,5878	0,2855	0,2835
20	4	0,16	0,0002	0,0544	-0,1623	-0,0863
20	1	0,075	0,0099	10,4544	-1,1139	-1,1957
20	1	0,05	0,0155	10,4544	-1,3938	-1,1957
20	2	0,11	0,0042	4,9878	-0,7221	-0,8259
20	9	0,325	0,0227	22,7211	1,6849	1,7627
20	9	0,325	0,0227	22,7211	1,6849	1,7627
20	2	0,11	0,0042	4,9878	-0,7221	-0,8259

11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
,87																			
,30	0,58																		
,54	2,41	1,83																	
,67	1,21	0,62	1,21																
,16	0,72	0,17	1,70	0,50															
,58	1,30	0,72	1,12	0,11	0,58														
,89	2,73	2,16	0,39	1,54	2,04	1,46													
,08	2,95	2,37	0,54	1,74	2,24	1,66	0,28												
,39	2,20	1,62	0,28	1,00	1,50	0,93	0,54	0,77											
,19	1,36	1,92	3,73	2,54	2,04	2,61	4,07	4,27	3,53										
,19	1,36	1,92	3,73	2,54	2,04	2,61	4,07	4,27	3,53	0,00									
,39	2,20	1,62	0,28	1,00	1,50	0,93	0,54	0,77	0,00	3,53	3,53								
,86	2,69	2,11	0,37	1,50	2,00	1,43	0,06	0,34	0,50	4,03	4,03	0,50							

	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1																			
2	0,58																		
3	2,41	1,83																	
4	1,21	0,62	1,21																
5	0,72	0,17	1,70	0,50															
6	1,30	0,72	1,12	0,11	0,58														
7	2,73	2,16	0,39	1,54	2,04	1,46													
8	2,95	2,37	0,54	1,74	2,24	1,66	0,28												
9	2,20	1,62	0,28	1,00	1,50	0,93	0,54	0,77											
10	1,36	1,92	3,73	2,54	2,04	2,61	4,07	4,27	3,53										
11	1,36	1,92	3,73	2,54	2,04	2,61	4,07	4,27	3,53	0,00									
12	2,20	1,62	0,28	1,00	1,50	0,93	0,54	0,77	0,00	3,53	3,53								
13	2,69	2,11	0,37	1,50	2,00	1,43	0,06	0,34	0,50	4,03	4,03	0,50							
14	1,21	0,62	1,21	0,00	0,50	0,11	1,54	1,74	1,00	2,54	2,54	1,00	1,50						
15	1,36	1,92	3,73	2,54	2,04	2,61	4,07	4,27	3,53	0,00	0,00	3,53	4,03	2,54					

11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
0,58	0,58																		
2,41	2,41	1,83																	
1,21	1,21	0,62	1,21																
0,72	0,72	0,17	1,70	0,50															
1,30	1,30	0,72	1,12	0,11	0,58														
2,73	2,73	2,16	0,39	1,54	2,04	1,46													
2,95	2,95	2,37	0,54	1,74	2,24	1,66	0,28												
2,20	2,20	1,62	0,28	1,00	1,50	0,93	0,54	0,77											
1,36	1,36	1,92	3,73	2,54	2,04	2,61	4,07	4,27	3,53										
1,36	1,36	1,92	3,73	2,54	2,04	2,61	4,07	4,27	3,53	0,00									
2,20	2,20	1,62	0,28	1,00	1,50	0,93	0,54	0,77	0,00	3,53	3,53								
2,69	2,69	2,11	0,37	1,50	2,00	1,43	0,06	0,34	0,50	4,03	4,03	0,50							
1,21	1,21	0,62	1,21	0,00	0,50	0,11	1,54	1,74	1,00	2,54	2,54	1,00	1,50						

11	12	13	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
0,87																		
0,80	0,58																	
0,57	1,21	0,62																
0,16	0,72	0,17	0,50															
0,58	1,30	0,72	0,11	0,58														
0,89	2,73	2,16	1,54	2,04	1,46													
0,08	2,95	2,37	1,74	2,24	1,66	0,28												
0,89	2,20	1,62	1,00	1,50	0,93	0,54	0,77											
0,19	1,36	1,92	2,54	2,04	2,61	4,07	4,27	3,53										
0,19	1,36	1,92	2,54	2,04	2,61	4,07	4,27	3,53	0,00									
0,89	2,20	1,62	1,00	1,50	0,93	0,54	0,77	0,00	3,53	3,53								
0,36	2,69	2,11	1,50	2,00	1,43	0,06	0,34	0,50	4,03	4,03	0,50							
0,67	1,21	0,62	0,00	0,50	0,11	1,54	1,74	1,00	2,54	2,54	1,00	1,50						
0,19	1,36	1,92	2,54	2,04	2,61	4,07	4,27	3,53	0,00	0,00	3,53	4,03	2,54					
0,16	0,72	0,17	0,50	0,00	0,58	2,04	2,24	1,50	2,04	2,04	1,50	2,00	0,50	2,04				
0,89	2,20	1,62	1,00	1,50	0,93	0,54	0,77	0,00	3,53	3,53	0,00	0,50	1,00	3,53	1,50			

10	11	12	13	15	16	17	18	19	20	22	23	24	25	26	27	28	29	30
30																		
58	1,87																	
00	1,30	0,58																
62	0,67	1,21	0,62															
17	1,16	0,72	0,17	0,50														
72	0,58	1,30	0,72	0,11	0,58													
16	0,89	2,73	2,16	1,54	2,04	1,46												
87	1,08	2,95	2,37	1,74	2,24	1,66	0,28											
62	0,39	2,20	1,62	1,00	1,50	0,93	0,54	0,77										
92	3,19	1,36	1,92	2,54	2,04	2,61	4,07	4,27	3,53									
62	0,39	2,20	1,62	1,00	1,50	0,93	0,54	0,77	0,00	3,53								
11	0,86	2,69	2,11	1,50	2,00	1,43	0,06	0,34	0,50	4,03	0,50							
62	0,67	1,21	0,62	0,00	0,50	0,11	1,54	1,74	1,00	2,54	1,00	1,50						
92	3,19	1,36	1,92	2,54	2,04	2,61	4,07	4,27	3,53	0,00	3,53	4,03	2,54					
17	1,16	0,72	0,17	0,50	0,00	0,58	2,04	2,24	1,50	2,04	1,50	2,00	0,50	2,04				
62	0,39	2,20	1,62	1,00	1,50	0,93	0,54	0,77	0,00	3,53	0,00	0,50	1,00	3,53	1,50			
92	3,19	1,36	1,92	2,54	2,04	2,61	4,07	4,27	3,53	0,00	3,53	4,03	2,54					
17	1,16	0,72	0,17	0,50	0,00	0,58	2,04	2,24	1,50	2,04	1,50	2,00	0,50	2,04				
62	0,39	2,20	1,62	1,00	1,50	0,93	0,54	0,77	0,00	3,53	0,00	0,50	1,00	3,53	1,50			

2	10	11	12	13	15	16	17	18	19	20	23	24	25	26	27	28	29	30
02																		
9	1,30																	
6	0,58	1,87																
2	0,00	1,30	0,58															
4	0,62	0,67	1,21	0,62														
4	0,17	1,16	0,72	0,17	0,50													
1	0,72	0,58	1,30	0,72	0,11	0,58												
7	2,16	0,89	2,73	2,16	1,54	2,04	1,46											
7	2,37	1,08	2,95	2,37	1,74	2,24	1,66	0,28										
3	1,62	0,39	2,20	1,62	1,00	1,50	0,93	0,54	0,77									
3	1,62	0,39	2,20	1,62	1,00	1,50	0,93	0,54	0,77	0,00								
3	2,11	0,86	2,69	2,11	1,50	2,00	1,43	0,06	0,34	0,50	0,50							
4	0,62	0,67	1,21	0,62	0,00	0,50	0,11	1,54	1,74	1,00	1,00	1,50						
0	1,92	3,19	1,36	1,92	2,54	2,04	2,61	4,07	4,27	3,53	3,53	4,03	2,54					
4	0,17	1,16	0,72	0,17	0,50	0,00	0,58	2,04	2,24	1,50	1,50	2,00	0,50	2,04				
3	1,62	0,39	2,20	1,62	1,00	1,50	0,93	0,54	0,77	0,00	0,00	0,50	1,00	3,53	1,50			
1	1,70	0,43	2,28	1,70	1,08	1,58	1,00	0,46	0,67	0,11	0,11	0,43	1,08	3,61	1,58	0,11		
3	1,00	2,28	0,43	1,00	1,62	1,12	1,70	3,15	3,36	2,61	2,61	3,11	1,62	0,93	1,12	2,61	2,69	

21222	10	11	12	13	15	16	17	18	19	20	23	24	25	27	28	29	30
1,92																	
3,19	1,30																
1,36	0,58	1,87															
1,92	0,00	1,30	0,58														
2,54	0,62	0,67	1,21	0,62													
2,04	0,17	1,16	0,72	0,17	0,50												
2,61	0,72	0,58	1,30	0,72	0,11	0,58											
4,07	2,16	0,89	2,73	2,16	1,54	2,04	1,46										
4,27	2,37	1,08	2,95	2,37	1,74	2,24	1,66	0,28									
3,53	1,62	0,39	2,20	1,62	1,00	1,50	0,93	0,54	0,77								
3,53	1,62	0,39	2,20	1,62	1,00	1,50	0,93	0,54	0,77	0,00							
4,03	2,11	0,86	2,69	2,11	1,50	2,00	1,43	0,06	0,34	0,50	0,50						
2,54	0,62	0,67	1,21	0,62	0,00	0,50	0,11	1,54	1,74	1,00	1,00	1,50					
2,04	0,17	1,16	0,72	0,17	0,50	0,00	0,58	2,04	2,24	1,50	1,50	2,00	0,50				
3,53	1,62	0,39	2,20	1,62	1,00	1,50	0,93	0,54	0,77	0,00	0,00	0,50	1,00	1,50			
3,61	1,70	0,43	2,28	1,70	1,08	1,58	1,00	0,46	0,67	0,11	0,11	0,43	1,08	1,58	0,11		
0,93	1,00	2,28	0,43	1,00	1,62	1,12	1,70	3,15	3,36	2,61	2,61	3,11	1,62	1,12	2,61	2,69	

8	9	21	22	26	10	13	11	12	15	16	17	18	19	20	23	24	25	27	28	29	30
46																					
39			1,92																		
65			3,19		1,30																
82			1,36		0,58	1,87															
00			2,54		0,62	0,67	1,21														
50			2,04		0,17	1,16	0,72	0,50													
07			2,61		0,72	0,58	1,30	0,11	0,58												
53			4,07		2,16	0,89	2,73	1,54	2,04	1,46											
73			4,27		2,37	1,08	2,95	1,74	2,24	1,66	0,28										
00			3,53		1,62	0,39	2,20	1,00	1,50	0,93	0,54	0,77									
00			3,53		1,62	0,39	2,20	1,00	1,50	0,93	0,54	0,77	0,00								
50			4,03		2,11	0,86	2,69	1,50	2,00	1,43	0,06	0,34	0,50	0,50							
00			2,54		0,62	0,67	1,21	0,00	0,50	0,11	1,54	1,74	1,00	1,00	1,50						
50			2,04		0,17	1,16	0,72	0,50	0,00	0,58	2,04	2,24	1,50	1,50	2,00	0,50					
00			3,53		1,62	0,39	2,20	1,00	1,50	0,93	0,54	0,77	0,00	0,00	0,50	1,00	1,50				
07			3,61		1,70	0,43	2,28	1,08	1,58	1,00	0,46	0,67	0,11	0,11	0,43	1,08	1,58	0,11			
39			0,93		1,00	2,28	0,43	1,62	1,12	1,70	3,15	3,36	2,61	2,61	3,11	1,62	1,12	2,61	2,69		

Theory of Stochastic Processes
Vol.7 (23), no.1-2, 2001, pp.346-362

NADIYA M. ZINCHENKO

HEAVY-TAILED MODELS IN FINANCE AND INSURANCE: A SURVEY

Heavy-tailed distributions and processes have long been studied by probabilistics, but the major advances in the practical applications have been made in last 10-15 years. We give the survey of the main trends in theoretical investigations and practical applications of heavy-tailed models with emphasis on subexponencial, Pareto-type and stable distributions. Certain problems connected with limit theorems, approximation, estimation, numerical simulation for heavy tails are treated as well as the connection with the risk theory.

2000 *Mathematics Subject Classifications*: 62P05, 60E07, 60F15.

Key words and phrases. Stable distributions, weak convergence, domains of attraction, extreme value distributions, stable Lévy processes,parameter estimation, numerical simulation, ruin probabilities, subexponencial distributions.

1. Introduction

Statistical applications in various research fields often deals with stochastic models based on Gaussian distributions or Gaussian processes (Brownian motion, geometrical Brownian motion, fractional Brownian motion, etc).

But real data often show the deviations from Gaussian hypothesis, for instance, they have heavy-tailed asymmetric histograms. Economists and financial analysts, for example, have been aware for almost 40 years that much economic data are of such category. In fact, the presence of the heavy tails in financial asset return distributions became obvious after the Mandelbrot's work (1963) on cotton prices, see also Fama (1965). Although the discussion about the adequacy of stable models for assets return is still going on in the financial research community, the presence of heavy-tailed distributions in this set-up is doubtless.

The notion of *heavy tails* needs more rigorous mathematical formulation.

It is often said that a random variable ξ follow a heavy tailed distribution if

$$P(\xi > x) \sim x^{-\alpha} \text{ as } x \to \infty, 0 < \alpha < 2. \quad (1)$$

The simpliest example is the *Pareto* distribution with density function $p(x) = \alpha k^{\alpha} x^{-\alpha-1}$, $\alpha, k > 0$, $x \geq k$ and distribution function (d.f.) $F(x) = 1 - (k/x)^{\alpha}$.

Another example of the distributions satisfying (1) is the class of *stable distributions*.

The fact that α-stable distribution decay like a power function means that a stable r.v. exhibits much more variability than a Gaussian one: it is much more likely to take large values, i.e. values far from the median. The high variety of the stable distributions is one of the reasons of their important role in modeling not only economic phenomena (stock market prices, foreign rate returns, large claims in insurance, etc.), but also in other areas such as: physics and astronomy (gravitation fields of stars, temperature distributions in nuclear reactors, stresses in crystalline lattices), telecommunications, World Wide Web and Network traffic, climatology.

Stable distributions have several desirable theoretical properties; the main of them are: stability under summation and deep relations with limit theorems for the sums $S_n = \sum_1^n \xi_i$ of independent identically distributed random variables (i.i.d.r.v.). More precise, only stable laws can appear as limit distributions for the linear normalized sums of i.i.d.r.v.

Unlike Gaussian distribution with two parameters (scale and location), stable laws constitute the four parameter family of distributions and their density functions can have a various shape: symmetric and asymmetric, more or less curtozis, with various tail decay depending on α.

The wider class of heavy-tailed distributions is formed by distributions attracted to the α-stable laws. These distributions have *regularly varying tails* with exponent -α, and, thus, their tail behaviour generalise (1).

So, the models with α-stable distributions or distributions in the domain of attraction of an α-stable law can be considered as a good alternative to the Gaussian models. The main reason which made them less popular was absence of well-developed statistical technique and good numerical procedures. This situation changed during last 10-15 years or so , when appearing of new generation of PC and corresponding software essentially simplified the computational problems connected with applications of the models based on stable distributions.

The other class of heavy-tailed laws appears in the risk theory for a formal definition of the large claim sizes. It is the class of *subexponential* distributions which is rather rich and includes the set of distributions on $(0, \infty)$ with regularly varying tails.

In the Section 2 we give the various equivalent definitions and basic properties of the stable distributions; Section 3 clearifies the links with the

limit theorems for i.i.d.r.v., describes the domains of attraction to the stable law. The results on limit behavior of the maximum of a sequence of iidrv is also in this section. Section 4 presents the results on the fluctuation of the sums of totally asymmetric stable variables. Section 5 is devoted to the classical problem of statistical inference - parameter estimation in the models including α-stable distributions or distributions attracted to α-stable law, p.4.3. presents the brief overview of the methods of numerical simulation. Section 5 deals with the ruin probability calculating in the case of large claims. It is demonstrated how classical risk theory has to be developed to take into account such claims. The elements of the theory of subexponential distributions are discussed with emphasis on the properties applicable within risk theory.

Finally, I want to remark that the list of references is far from complete: except a few original pioneer works, the main attention being paid to the recent works, especially monographs, guides and surveys where the reader can find additional references.

2. One-dimensional stable distributions. Basic properties.

Stable laws were introduced and studied by Paul Lévy in 1920s, the remarkable results were also obtained by Aleksander Khintchine in 1930s. The classical results which underline the links with the weak convergence of the sums of iidrv are presented in the famous monograph by Gnedenko and Kolmogorov (1949), see also Ibragimov & Linnik (1965), Feller (1971). The analytical properties of the univariate (one-dimensional) stable distributions and various fields of their applications are presented in the book by Zolotarev (1983). The basic theoretical facts as well as contemporary achievements in the theory of stable processes are summarised in monograph by Samorodnitsky and Taqqu (1994); for those who are interested in this topic and more general theory of Lévy processes the books by Gikhman and Skorokhod (1974), Skorokhod (1963,1986) are yet of great interest as well as a recent Bertoin (1996). A monograph by Yanicki and Weron (1994) is devoted to the simulation problems and "A Practical Guide to Heavy Tails" under edition of Adler, Feldman and Taqqu (1998) is directed more on applications than on theory and is primary concerning with statistical techniques appropriated for heavy-tailed data. It also worth to mention the fundamental work by Embrehts et al. (1997), which includes theoretical material, statistical methodology and a lot of interesting examples of using heavy-tailed models in finance and insurance.

2.1. Definitions. Below we give several equivalent definitions of a stable distribution:

Definition 1. *A non-degenerate variable ξ is said to have a stable distribution if for any positive numbers a and b ,there are $c > 0$ and $d \in R^1$ such*

that

$$a\xi_1 + b\xi_2 \stackrel{d}{=} c\xi + d, \tag{2}$$

where ξ_1 *and* ξ_2 *are independent copies of* ξ *and notation* $\stackrel{d}{=}$ *stands for equality in distribution.*

For any stable ξ there exist a number $\alpha \in (0, 2]$ such that c in (2) satisfies

$$c^\alpha = a^\alpha + b^\alpha. \tag{3}$$

A random variable ξ is called *strictly stable* if (2) is true with $d = 0$.

Definition 2. *A non-degenerate variable* ξ *is said to have a stable distribution if for any* $n \geq 2$ *there are* $c = c_n > 0$ *and real* $d = d_n$ *such that*

$$\xi_1 + \cdots + \xi_n \stackrel{d}{=} c_n\xi + d_n, \tag{4}$$

in the other words, the n-th convolution of the identical stable distributions is again of the same type.

In (4), necessarily, $c_n = n^{1/\alpha}$ for some $\alpha \in (0, 2]$, where α is the same as in (3).

Stable laws constitute the subclass of the infinitely divisible distributions with canonical Levy-Khintchine representation for characteristic function of α-stable distribution

$$\log g_\alpha(t) = it\gamma - \frac{\sigma^2}{2} + \int_{-\infty}^{\infty} (e^{itx} - 1 - \frac{itx}{1+x^2}) dH(x), \tag{5}$$

where $\alpha \in (0, 2]$, $\gamma \in R^1$, $\sigma^2 \geq 0$, $\int_{-\infty}^{\infty} x^2 dH(x)$, and include the Gaussian distribution ($\alpha = 2, \sigma^2 > 0$). For $0 < \alpha < 2$ there is no Gaussian component and Lévy spectral function $H(x) = C_1|x|^{-\alpha}$ for $x < 0$ and $H(x) = -C_2x^{-\alpha}$ for $x > 0$, where $C_1, C_2 \geq 0$, $C_1 + C_2 > 0$.

The stable characteristic function (ch.f.) admits the simpler representation, which also can be considered as a definition of a stable distribution

$$g_{\alpha,\beta}(t) = g(t; \alpha, \beta, \gamma, \lambda) = \exp\{it\gamma - \lambda|t|^\alpha\{1 - i\beta\text{sign}(t)\omega(t, \alpha)\}\}, \tag{6}$$

where the main parameters $\alpha \in (0, 2]$ and $\beta \in [-1, 1]$ characterize the shape and symmetry properties of the distribution, $\lambda \geq 0$ is a scale parameter and $\gamma \in R^1$ is a location parameter. Here $\omega(t, \alpha) = \tan(\pi\alpha/2)$ as $\alpha \neq 1$ and $\omega(t, \alpha) = -(2/\pi)\beta \ln t$ as $\alpha = 1$, parameter α is often called *index of stability or characteristic exponent* and β is a *skewness* parameter.

There are other than (6) parametrizations of the α-stable ch.f. more or less convenient for various purposes (for details see Zolotarev (1986)).

The support of stable distributions is the real line except if $\alpha < 1$ and $|\beta| = 1$, in which case it is $(-\infty, 0)$ for $\beta = -1$ and $(0, \infty)$ for $\beta = 1$. The distributions with $\beta = 0$ are symmetric about γ, while those with $\beta < 0$

are left skewed and those with $\beta > 0$ are right skewed. The distributions with $|\beta| = 1$ are called "totally skewed" or "extreme/ totally asymmetric" stable distributions. All symmetric stable distributions are strictly stable, but inverse assertion is not true. In fact, the α-stable distribution with $\alpha \neq 1$ is strictly stable iff $\gamma = 0$.

Denote by $S_{\alpha,\beta}(x) = S_{\alpha,\beta}(x;\lambda,\gamma)$ the distribution function (d.f.) of the stable law. We omit the index β if its value is not essential and write $\xi \sim S_{\alpha,\beta}(\lambda,\gamma)$ to indicate that r.v. ξ has the stable d.f. $S_{\alpha,\beta}(x;\lambda,\gamma)$.

Zolotarev (1983) gives an expression for the stable d.f. at $x = 0$:

$$S_{\alpha,\beta}(x;1,0) = \frac{1}{2}(1 - \beta^* K(\alpha)/\alpha), \alpha \neq 1, \text{ where } K(\alpha) = \alpha - 1 + \text{sign}(1-\alpha)$$

and β^* satisfies: $\tan(\beta^* \pi K(\alpha)/2) = \beta \tan(\pi\alpha/2)$.

Stable distributions are continuous with smooth unimodal densities, but explicit form of the stable densities is known only in 4 exceptional cases: Gaussian distribution ($\alpha = 2$), Caushy distribution ($\alpha = 1$, $\beta = 0$), Levy distribution ($\alpha = 1/2$, $\beta = 1$), and its symmetric contrapart ($\alpha = 1/2$, $\beta = -1$). Nevertheless in general case the stable densities or d.f. can be represented (Zolotarev (1986)) as the integrals or as the sums of infinite series, which however, can be numerically approximated. These facts as well as using the fast Fourier transform (FFT) algorithms for inverting the characteristic functions give the possibility to calculate numerically the densities and d.f. of the stable laws.

2.2. Tail behaviour and moments. If $\xi \sim S_{\alpha,\beta}(\lambda,\gamma)$ with $0 < \alpha < 2$, then

$$\lim_{x\to\infty} x^\alpha P(\xi > x) = C_\alpha \frac{1+\beta}{2}\lambda^\alpha, \lim_{x\to\infty} x^\alpha P(\xi < -x) = C_\alpha \frac{1-\beta}{2}\lambda^\alpha, \quad (6)$$

where

$$C_\alpha = (1-\alpha)/\Gamma(2-\alpha)cos(\pi\alpha/2), \alpha \neq 1 \text{ and } C_\alpha = 2/\pi, \alpha = 1. \quad (7)$$

This fact shows that the tails of α-stable d.f.(as well as the tails of the α-stable densities) decay like a power function (unlike the exponential decay of the tails of Gaussian distributions).

An interesting consequence followers from (6) : α-stable r.v. ξ possesses the finite moments of order $\delta < \alpha$, but all moments of order $\delta > \alpha$ do not exist. So, in the set of stable laws the Gausian distribution is only one possessing all moments (particularly, finite variance). Besides this, the finite mathematical expectation exists only for α-stable distributions with $1 < \alpha \leq 2$ and for $\alpha < 1$, $E|\xi| = \infty$. Thus, such distributions are useful when some very large observations may be expected and may dominate sums of other observations. It is not valid to treat such values as outliers,

because excluding them takes away much of specifity of the data and leads to reduce the useful information. Precisely these observations may be of main interest in understanding the data generating mechanism. The fact that α-stable variables with $\alpha < 2$ have infinite variance also means that techniques valid for the Gaussian case do not apply and special methods must be developed.

Remark . Formula (6) for $\beta = -1$ established that $P(\xi > x) \to 0$ faster than $x^{-\alpha}$. In the case $\beta = -1$ and $1 < \alpha < 2$ the true rate of decay is exponential. This fact was proved by Skorokhod (1953), who also obtained the asymptotic expression for $P(\xi > x)$ (see also Zolotarev (1983), Samorodnitsky & Taqqu (1994)). For $\beta = -1$, $\alpha < 1$, $P(\xi > x) = 0$ for all $x > 0$ because ξ is totally skewed to the left.

2.3. Le Page series representation. Let $\{\varepsilon_i, i \geq 1\}$, $\{W_i, i \geq 1\}$, $\{\Gamma_i, i \geq 1\}$ be three independent sequences of r.v., where $\{\varepsilon_i, i \geq 1\}$ is an iid sequence of Rademacher variables, $\{W_i, i \geq 1\}$ are iidrv with finite $E|W_1|^\alpha$, $\{\Gamma_i, i \geq 1\}$ is a sequence of arrival times of a Poisson process with unite arrival rate, i.e. Γ_i has gamma distribution with parameter i. Any symmetric α-stable r.v. $\xi \sim S_\alpha(\lambda, 0)$ admits the series representation (i.e. has the same distribution as)

$$\lambda(C_\alpha / E|W_1|^\alpha)^{1/\alpha} \sum_{i=1}^{\infty} \varepsilon_i W_i {\Gamma_i}^{-1/\alpha}, \tag{9}$$

where the series in (9) converges a.s. and C_α is determined by (7).

The series representation of skewed α-stable r.v. can be found in Samorodnitsky and Taqqu (1994).

3. Limit theorems. Domains of attraction.

3.1. Limit theorems for sums of iidrv. Let $\{\xi_i, i \geq 1\}$ be i.i.d.r.v. with common d.f. $F(x)$, $S_n = \sum_1^n \xi_i$, $S_0 = 0$. As it was mentioned above stable laws and only they appear as limit distributions for the linear normalized sums $S_n^* = b_n^{-1}(S_n - A_n)$. Two questions are of a great interest:

a) What properties of d.f. $F(x)$ determine the parameters of $S_{\alpha,\beta}(x)$?

b) What is the form of norming coefficients A_n and b_n and how they are connected with d.f. F(x) and $S_{\alpha,\beta}(x)$?

The investigation of domains of attraction gave the answer to these questions.

Definition 3. *D.f. F(x) belongs to the domain of attraction of the stable distribution $S_{\alpha,\beta}(x)$ (notation $F \in DAS_{\alpha,\beta}$) if there exist sequences of norming constants $a_n \in R^1$ and $b_n \geq 0$ such that for i.i.d.r.v. $\{\xi_i, i \geq 1\}$ with common distribution F(x) linear normalized sums*

$$S_n^* = b_n^{-1}(S_n - A_n) = b_n^{-1}(\xi_1 + \cdots + \xi_n - A_n) \Rightarrow S_{\alpha,\beta}. \tag{10}$$

The set of all d.f. satisfying (10), forms the domain of attraction of $S_{\alpha,\beta}(x)$, and only stable distributions possess their domains of attraction.

Theorem 1. *D.f.* $F \in DAS_{\alpha,\beta}$ *if and only if*

$$\frac{F(-x)}{1-F(x)+F(-x)} \to p, \frac{1-F(-x)}{1-F(x)+F(-x)} \to q, \tag{11}$$

where $p+q=1$, $\beta = p-q$, *and for any* $t>0$

$$\lim_{x\to\infty} \frac{1-F(x)+F(-x)}{1-F(tx)+F(-tx)} = t^{\alpha}. \tag{12}$$

Condition (11) is formulated in the form convenient for calculation of the parameter's β value, while condition (12) determines the value of parameter α. In the other terminology it means, that the tail distribution $P\{|\xi_1| > x\} = 1-F(x)+F(-x)$ is regularly varying function of order $-\alpha$, i.e. for large x

$$1-F(x)+F(-x) = x^{-\alpha}L(x), \tag{13}$$

where $L(x)$ is slowly varying at infinity.

Norming coefficients b_n are also regularly varying, i.e. $b_n = n^{1/\alpha}L_1(n)$, $L_1(x)$ is slowly varying, and centering coefficients can be determined as $A_n = 0$ for $0<\alpha<1$, $A_n = nE\xi_1$ for $1<\alpha<2$ and $A_n = n\int_{-b_n}^{b_n} xdF(x)$ for $\alpha = 1$.

The distributions attracted to the stable law with parameter $0<\alpha<2$ have the same moment properties as the α-stable distributions.

Every stable law S_α belongs to his own domain of attraction with norming constants $b_n = n^{1/\alpha}$. This fact was a good reason for Gnedenko to introduce the notion of *normal attraction*.

Definition 4. *D.f.* $F(x)$ *belongs to the domain of normal attraction (*$F \in DNS_{\alpha,\beta}$ *), if* $b_n = an^{1/\alpha}$, $a>0$.

Following statement completely characterized such domains:

Theorem 2. $F \in DNS_{\alpha,\beta}$ *for* $\alpha \in (0,2)$, $|\beta| \le 1$ *iff for some* $x_0 > 0$

$$F(-x) = (c_1a^{\alpha}+\varepsilon_1(x))|x|^{-\alpha},\ 1-F(x) = (c_2a^{\alpha}+\varepsilon_2(x))x^{-\alpha} \quad as\ x > x_0,$$

where $a>0$, $\lim_{x\to-\infty}\varepsilon_1(x) = \lim_{x\to\infty}\varepsilon_2(x) = 0$, $\beta = (c_1-c_2)/(c_1+c_2)$.

In this case $A_n = 0$ for $0<\alpha<1$, $A_n = nE\xi_1$ for $1<\alpha<2$ and $A_n = (2/\pi)\beta n\ln n$ for $\alpha = 1$.

3.2. Invariance principles. Let $\{\xi_i, i \ge 1\}$ be i.i.d.r.v. in domain of attraction of α-stable law $S_{\alpha,\beta}(x;1,0)$, i.e. $(n^{1/\alpha}L_1(n))^{-1}(S_n - A_n) \Rightarrow S_{\alpha,\beta}$ for appropriate slowly varying function L_1. Then the process

$$(n^{1/\alpha}L_1(n))^{-1}(S_{[nt]} - A_n), 0<t<1, \tag{14}$$

converges weakly to an α-stable Lévy process $Y_\alpha(t)$, $0 < t < 1$. Here weak convergence is convergence in $D[0,1]$ equipped with Skorokhod J_1 metric; $Y_\alpha(t) = Y_{\alpha,\beta}(t)$ is a separable process started from 0 with independent stationary increments distributed according to the stable law and $Y_\alpha(1) \sim S_{\alpha,\beta}(x;1,0)$. Such processes are cad-lag and for $|\beta| = 1$ have jumps of one sign: only positive if $\beta = 1$ and negative if $\beta = -1$.

The other type of limit theorems (which deal with a.s. convergence) is so-called *strong invariance principle*. In this context the strong invariance principle means the possibility to construct on a single probability space the sequence of partial sums S_n of i.i.d.r.v. $\{\xi_i, i \geq 1\} \in DNS_{\alpha,\beta}$ and the sequence of sums T_n of the $S_{\alpha,\beta}$-stable i.i.d.r.v. $\{\eta_i, i \geq 1\}$ in such a way that a.s.

$$\max_{m \leq n} |S_m - T_m| = o(n^{1/d}) \text{ for some } d > \alpha. \tag{15}$$

Sufficient conditions, which ensure (15), as well as possibility of a.s. approximation of S_n with error term $o(n^{1/\alpha-\rho})$, $\rho > 0$, by α-stable Lévy process $Y_{\alpha,\beta}(t)$, $t > 0$, were obtained by Zinchenko (1987, 1998). For example, such condition is boundedness of pseudomoments of order $l > \alpha$.

3.3. Weak convergence of max under affine transformations. Let $\{\xi_i, i \geq 1\}$ be i.i.d.r.v. with common d.f. $F(x)$. If there exists constant $c_n > 0$ and $d_n \in R^1$ and non-degenerate d.f. G such that

$$c_n(\max_{1 \leq i \leq n} \xi_i - d_n) \Rightarrow G, \tag{16}$$

then G has one of three types:

a) Fréchet with $G = \Phi_\alpha(x) = \exp(-x^\alpha)$, $x > 0$, $\alpha > 0$;

b) Weibull with $G = \Psi_\alpha(x) = \exp(-(-x)^\alpha)$, $x \leq 0$, $\alpha > 0$;

c) Gumbel with $G = \Lambda(x) = \exp(-\exp(-x))$,$x \in R^1$.

In many cases the following von-Mises general representation for extreme value distributions is rather useful

$$G(x) = G_h(x) = \exp(-(1+hx)^{-1/h}), h \neq 0,$$

and

$$G(x) = G_h(x) = \exp(-\exp(-x)), h = 0.$$

The case $h = 1/\alpha > 0$ corresponds to Fréchet distribution, $h = -1/\alpha < 0$ corresponds to Weibull distribution and $h = 0$ is the Gumbel case.

Definition 5. *We say that r.v. ξ (d.f. F) belongs to the maximum domain of attraction to the extreme value distribution G if there exist constants c_n and d_n such that (16) holds. We write $\xi \in MDA(G)$ ($F \in MDA(G)$) in this case.*

There is a simple characterization of maximum domain of attraction to the Fréchet distribution.

Theorem 3. *The d.f.* $F \in MDA(\Phi_\alpha)$ *if and only if* $\bar{F}$ *is regularly varying at infinity with index* $-\alpha$.

Thus, for $0 < \alpha < 2$ there is a deep inter-relation between the domains of attraction of α-stable law and maximum domains of attraction to the Fréchet distribution.

For an excellent introduction in this topic see Embrechts et al. (1997).

4. Fluctuation of the sums of heavy-tailed r.v.

Let $\{\xi_i, i \geq 1\}$ be i.i.d.r.v. with the common d.f.$F(x)$;

$$S_n = \sum_{i=1}^{n} \xi_i,\ S_0 = 0,\quad S_m^n = \sum_{i=m+1}^{n} \xi_i,\quad L(N,k) = \max_{0 \leq n \leq N-k} S_n^{n+k}.$$

What is the asymptotic behaviour of S_N^{N+k}, S_{N-k}^N and $L(N,k)$, when $1 < k = k_N < N$ and $N \to \infty$?

In particular case of i.i.d. Bernoulli r.v.'s with success probability $p \in (0,1)$ this question about the magnitude of the increments of a random walk S_n is closely related to the problem of the longest success-run. Also statistics that depend only on the last k observations are used in time series analysis and analysis of censored data. Particularly, it is interesting to study the growth rate of S_{n-k}^n and to obtain the explicit form of normalizing sequence $\gamma_n = \gamma(n, k_n)$, which provides the a.s. convergence

$$\limsup_{n \to \infty} \frac{1}{\gamma_n} S_{n-k}^n = c = \text{const} \neq 0.$$

Erdös and Rényi were first who studied the Bernoulli case as well as the case of Gaussian summands.

In this Section we consider i.i.d.r.v. with "heavy tails", and focus on i.i.d.r.v. distributed according to the stable law $S_{\alpha,\beta}$ with parameters $1 < \alpha < 2$, $\beta = -1$, $E\xi_1 = 0$.

Zinchenko (1992) proposed to use integral tests to investigate the the growth rate of increments of Levy stable process $Y_{\alpha,\beta}(t)$ with parameters $1 < \alpha < 2$, $\beta = -1$. As a consequence of such approach it is possible to obtain a number of asymptotic results for the lag sums of r.v. distributed according to the stable law with given above parameter values.

Theorem 4. *Let* $\{k_N, n \geq 1\}$ *be non-decreasing sequence of positive numbers, such that: (i)* $0 < k_N < N$, *(ii)* N/k_N *does not decrease in N. Then with the probability 1*

$$\limsup_{N \to \infty} S_N^{N+k_N} / k_N^{1/\alpha} d^*(N) = \limsup_{N \to \infty} L(N, k_N) / k_N^{1/\alpha} d^*(N) = 1, \tag{17}$$

$$\limsup_{N \to \infty} \max_{1 \leq m \leq N-k_N} \max_{1 \leq l \leq k_N} S_m^{m+l} / k_N^{1/\alpha} d^*(N) = 1, \tag{18}$$

$$\textit{where}\ \ d^*(N) = \{B^{-1}\ln\ln N + \ln N/k_N)\}^{1/\lambda}.$$

Corollary 1. For $k_N = [c\ln N]$, $c > 0$ we have a.s.

$$\limsup_{N\to\infty} L(N, [c\ln N)/c^{1/\alpha}\ln N = B^{-1/\lambda},$$

this is a straightforward generalization of Erdös-Rényi law for sums of random variables with asymmetric stable distribution.

Corollary 2. With the probability 1

$$\limsup_{N\to\infty} \frac{L(N,[cN])}{N^{1/\alpha}(\ln\ln N)^{1/\lambda}} = c^{1/\alpha}B^{-1/\lambda}, 0 < c < 1,$$

$$\limsup_{N\to\infty} \frac{L(N,[N^\rho])}{N^{\rho/\alpha}(\ln N)^{1/\lambda}} = (\frac{1-\rho}{B})^{1/\lambda}, 0 < \rho < 1;$$

$$\limsup_{N\to\infty} \frac{L(N,[\ln^\tau N])}{(\ln N)^{\tau+\alpha-1}} = B^{-1/\lambda}, \tau > 0;$$

$$\limsup_{N\to\infty} \frac{L(N,1)}{(\ln N)^{1/\lambda}} = \limsup_{N\to\infty}(\max_{1\le k\le N-1}\xi_k)/(\ln N)^{1/\lambda} = B^{-1/\lambda}.$$

The last relation in terms of order statistics $\xi_i^{(n)}$ for n independent $S_{\alpha,-1}$-distributed observations means that a.s.

$$\limsup_{N\to\infty} \frac{\xi_n^{(n)}}{(\ln N)^{1/\lambda}} = \limsup_{N\to\infty}(\max_{1\le k\le N-1}\xi_k)/(\ln N)^{1/\lambda} = B^{-1/\lambda}.$$

The statements of Theorem 3 and Corollary 1 via strong invariance principle can be easily extended on i.i.d.r.v. belonging to the domain of normal attraction of the stable law (Zinchenko (1999)).

5. Statistical inference for heavy-tailed models

5.1. Estimation of the index α of regular varying. We start with a very simple but not exact method.

5.1.1. Log-log CD (complementary distribution) plots.

CD plots present the complementary distribution $\overline{F}(x)$ on log-log axes. Ploted in this way d.f. $F \in DNG_{\alpha,\beta}$ has a property $\log(1-F(x)) \sim -\alpha\log x$ for large x. In practice one obtain an estimate for α by ploting CD plot of the data set $(x_1,\dots,x_n)$ selecting a minimal value x_0 above which the plot appears to be approximately linear. Then it is necessary to take points $x_i > x_0$ and estimate the slope using least-square regression.

5.1.2. Hill's-type estimators. This approach is based on relation between domains of attraction to α-stable law and maximal domains of attraction to extreme value law (see Theorem 3). The basic idea consists of

finding conditions equivalent to $F \in MDA(G_h)$, which involves parameter $h = 1/\alpha$ in a simple way, and applying the properties of order statistics. Note that this approach provides estimators which valid for rather wide classes of distributions.

Let $(x_1, \ldots, x_n)$ be a sample of iidrv with common d.f. F. We use notation $x^*_{(n)} \le \cdots \le x^*_{(1)}$ for the ordered sample and called $x^*_{(k)}$ the k-th upper statistics.

Hill estimator (1975) of index α for d.f. $F \in MDA(\Phi_\alpha)$ depends on the k largest observations and has a form $\hat{\alpha} = 1/\hat{H}$, where

$$\hat{H} = \hat{H}_{k,n} = (1/k) \sum_{i=1}^{k} (\log x^*_{(i)} - \log x^*_{(k+1)}).$$

Pickands estimator (1975) is applicable for estimating a shape parameter h of d.f. F in domain of maximal attraction of any extreme value distribution

$$\hat{h} = \hat{p}_k = (1/\ln 2) \frac{\log x^*_{(k)} - \log x^*(2k)}{\log x^*_{(2k)} - \log x^*_{(4k)}}.$$

So, the estimator for α is $\hat{\alpha} = 1/\hat{h}$.

De Haan-Resnick estimator (1980) has a form $\hat{\alpha} = 1/\hat{R}$, where

$$\hat{R} = \hat{R}_k = (1/\ln k)(\log x^*_{(1)} - \log x^*_{(k)}).$$

Dekker-Einmahl-de Haan estimator (1990) was proposed as modification of Hill estimator extended to cover the whole class $MDA(G_h)$, $h \in R^1$. It has a form

$$D = H + 1 - \frac{1}{2}\{1 - (H)^2/H^*\}^{-1},$$

where H is Hill estimator,

$$H^* = (1/k) \sum_{i=1}^{k} (\log x^*_{(i)} - \log x^*_{(k+1)})$$

and can be interpreted as moment estimator.

If $k \to \infty$, $k/n \to 0$ for $n \to \infty$ then Hill's estimators are consistent; they are strong consistent if, additionally, $k/\ln\ln n \to \infty$.

The main disadvantages of Hill's type estimations are presence of bias and absence of formal procedures for optimal choice of $k = k(n)$; Hill estimator is very sensitive with respect to dependence. The numerical investigation of these estimators carried out by Pitet, Dacorogna and Muller (1998)demonstrated that Dekker-Einmahl-de Haan estimator is generally downward bias, de Haan-Resnick estimator systematically has upward bias, Pickand's and Hill's estimators show fluctuations around true value.

5.2. Estimation of stable law parameters. McCulloch (1984,1996) proposed estimators based on the functions of the sample quantiles; his approach seems to be optimal in the exact stable case, but situation changes when in data there are deviations from assumption of stability. In this case Hill estimators may be preferable. An alternative approach based on properties of d.f. in $DAS_{\alpha,\beta}$ (analogous to Theorem 1) and empirical d.f. is discussed in de Haan & Pereira (1997), while Koutrouvelis (1980) suggested estimators based on empirical ch.f.

5.3. Numerical simulation. LePage series representation can be taken as a base for computer simulation, but a very slow convergence of LePage sums makes this method practically usefulles.

The best method of computer simulation of an important class of symmetric α-stable variables $X \sim S_\alpha(x;1,0)$ includes following steps:

i) generate a r.v. V uniformly distributed on $(-\pi/2, \pi/2)$ and an exponent r.v. W with mean 1; ii) compute

$$X = \frac{\sin(\alpha V)}{\{\cos(\alpha V)\}^{1/\alpha}} \{ \frac{cos(V - \alpha V)}{W} \}^{(1-\alpha)/\alpha}.$$

The analogous algorithm providing simulation of skewed stable r.v. with $\alpha \neq 1$ is presented in Yanicki and Weron (1994).

5.4. Comments. Stable distributions were successfully used to fit data sets that includes interest rate changes, stock returns and foreign exchange rates (Mitinik and Rachev (1995), McCuloch (1996), McDonalds (1996), Embrechts et al (1998), see also articles by Mitinik , Rachev & Peolella and Mullar, Dacorogna & Pictet in collection [1] with further references); portfolio optimization and Capital Asset Pricing Model as wel as the ways of extension of Black-Sholes option pricing model for α-stable Paretian case ($\alpha < 2$) are also discussed in these works. Chapter 7 of Embrechts et al. (1998) and Rachev, Kim & Mitinik (1999) can serve as a good introduction in time series analysis under stable non-Gaussian hypothesis.

6. Ruin probabilities in the presence of large claims.

6.1. Ruin probabilities in classical Cramer-Lundberg set-up. We start with Cramer-Lundberg set-up, when the resulting risk process $U(t)$, $t \geq 0$, is defined as

$$U(t) = u + ct - S(t) = u + ct - \sum_{i=1}^{N(t)} X_i, \tag{19}$$

where the claim arrival process $N(t)$, $t > 0$ is a Poisson process with rate λ; the claim sizes $\{X_i, i \geq 1\}$ are positive iidrv with common non-lattice d.f. F and finite mean $\mu = \mathrm{E}X_1$, u - initial capital, c - loaded premium rate .

Denote by $M_X(\nu) = E(\exp \nu X_1)$, $\overline{F}(x) = 1 - F(x)$ stands for tail of F, F^{*n} denotes the n-th convolution of F and

$$F_I(x) = \frac{1}{\mu} \int_0^x \overline{F}(y)dy, \quad x > 0,$$

is called *integrated tail distribution.*

The classical problem of calculating ruin probabilities deals with

(a) $\psi(u, T) = P\{u(t) < 0 \quad \text{for some} \quad 0 < t \leq T\}$, $0 < T < \infty$, $u > 0$ - the ruin probability in finite time;

(b) $\psi(u) = \psi(u, \infty) = P\{u(t) < 0 \text{ for some } t > 0\}$ - the ruin probability in infinite time.

We focused on asymptotic behaviour of $\psi(u)$ when initial capital u grows.

Let net profit condition $\rho = (c/\lambda\mu) - 1 > 0$ be true and assume that there exists a solution $\nu > 0$ (called *adjustment coefficient*) of the equation $M_X(\nu) = (c\nu + \lambda)/\lambda$. Then the selebrated Cramer-Lundberg's inequality states that for any initial capital $u \geq 0$

$$\psi(u) \leq e^{-\nu u}.$$

If, moreover, $\int_0^\infty xe^{\nu x}\overline{F}(x)\,dx < \infty$, then $\psi(u) \sim Ce^{-\nu x}$ for certain $C > 0$.

Condition on $M_X(\nu)$ means that claim distribution has exponentially bounded tails. The class of distributions with "light tails" , which satisfy this condition, includes, for instance, exponential, gamma, truncated normal and Weibull distribution with df $F(x) = 1 - \exp(-cx^\tau), c > 0, \tau \geq 1$, as well as all distributions with bounded support; for all of them Cramer-Lundberg estimate holds true.

Pareto distribution is often used to model the claim sizes in property insurance, but unfortunately, we cannot use Cramer-Lundberg estimate in this case because an adjustment coefficient does not exists. Two questions arise from this simple example:

1)Is it possible to obtain simple estimates for $\psi(u)$ for concrete heavy-tailed distribution?

2) Is there alternative methodology and rich classes of heavy-tailed distributions which admit general approach to ruin probability evaluation ?

In 1970s the positive answers on the first question were given by von Bahr for Pareto distributions and by Thorin and Wikstad for lognormal claim sizes.

On the other hand, Embrechts and Veraverbeke (1982) pointed out the fundamental role of class of subexponential distributions for ruin theory in the case of heavy tails.

6.2. Ruin probabilities in subexponential case. In this section assume that df F has support $(a, \infty), a \geq 0$.

Definition 6. *A d.f. F is subexponential if for all $n \geq 2$*

$$\lim_{x\to\infty} \frac{\overline{F^{n*}(x)}}{\overline{F}(x)} = \lim_{x\to\infty} \frac{1-F^{n*}(x)}{1-F(x)} = n. \tag{20}$$

The class of subexponential df is denoted by $\mathcal{S}$. A few words about probabilistic sense of (20): if $X_1, \ldots, X_n$ are iidrv with df F, $S_n = \sum_{i=1}^n X_i$, then $1 - F^{n*}(x) = P\{S_n \geq x\}$ and $P\{\max_{1\leq i\leq n} X_i \geq x\} = \overline{F}^n(x) \sim nF(x)$, as $x \to \infty$. Thus,

$$P\{S_n \geq x\} \sim P\{\max_{1\leq i\leq n} X_i \geq x\}.$$

So, asymptotically, behaviour of the total claim amount is governed by one very big claim. This is one of the intuitive notions of heavy-tailed distributions, which explained why S can be applied to model big claim amounts and clarifies the links with extremal value theory.

Condition (20) can be simplified, so that one needs to check (20) only for $n = 2$.

Class $\mathcal{S}$ is rather rich, it includes, besides other, the class $R_{-\alpha}$ of the distributions with regular varying tails with index $-\alpha$, $\alpha > 1$. Examples of subexponential distributions are: lognormal,Pareto, Burr, log-gamma, Weibull with parameter $0 < \tau < 1$, Benktander type I, II; truncated stable distribution $1 - F(x) = P\{|\xi| > x\}$ where ξ is α-stable r.v.with $|\beta| \leq 1, 1 < \alpha < 2$. The assumption $1 < \alpha < 2$ provides the existence of the finite mean $E\xi < \infty$. The typical example of d.f. not belonging to $\mathcal{S}$ is an exponential distribution itself.

In the case of claims with subexponential integrated tail distributions the ruin probability has rather simple asymptotics:

Theorem 5. *Consider the Cramer-Lundberg model with net profit condition $\rho > 0$ and $F_I(x) \in \mathcal{S}$. Then*

$$\psi(u) \sim \rho^{-1}\overline{F}_I(u), u \to \infty. \tag{21}$$

The next theorem (see Embrehts et al(1997)) states that formula (21) is only possible under condition $F_I \in \mathcal{S}$.

Theorem 6. *In Cramer-Lundberg model with net profit $\rho > 0$ following assertions are equivalent:*

$$\text{(a)} \qquad F_I \in S,$$

$$\text{(b)} \lim_{u\to\infty} \psi(u)/F_I(u) = \rho^{-1}.$$

This fact ones more underline the natural choice (from analytic point of view) of class $\mathcal{S}$ when classical Cramer-Lundberg condition is violated.

The above results can be generalized to more general risk models (Grandell (1991), Rolski et al. (1999), Schmidli (1999), Asmussen et al. (1999)).

A number of sufficient conditions for $F \in \mathcal{S}$ and $F_I \in \mathcal{S}$ in terms of the hazard rate $g(x) = f(x)/\overline{F}(x)$ and hazard function $Q(x) = -\ln \overline{F}(x)$ are given in Kluppelberg (1989), Embrechts and Veraverbeke (1982), Embrechts et al(1997)). This conditions help to check that all distributions, mentioned above, i.e. Pareto, Weibull ($\tau < 1$), lognormal, loggamma, Burr, Benktander-type-I and II are subexponential as well as their integrated tails F_I. Thus, for these distributions theorem 5 is true.

6.3. Ruin probabilities in the presence of heavy tails and interest rates. Consider the classical Cramer-Lundberg model when insurance company in addition to the linear premium income also receives interest on its reserves with a constant force of interest $\delta > 0$, then the risk process

$$U_\delta(t) = ue^{\delta t} + c\int_0^t e^{\delta v}dv - \int_0^t e^{\delta(t-v)}dS(v), \quad t \geq 0,$$

where $S(t) = \sum_{n=1}^{N(t)} X_n$, premium rate $c > 0$, u is initial capital.

It occurs that for $\delta > 0$ relation (21) is no longer valid. Instead Kluppelberg and Stadmular (1998) obtained that under assumption that size distribution F has a regular varying tail ($F \in R_{-\alpha}, \alpha > 1$):

$$\psi_\delta(u) \sim k_\delta \overline{F}(u), u \to \infty. \tag{22}$$

Hence this result is applied, for instance, to Pareto, loggamma, certain Benktander and stable claim sizes. By different methods Assmussen (1996) derived analogous results for the whole class of subexponential claims.

6.4. Perturbed processes. There are other possibilities of appearing heavy tails in risk models. For instance, we can consider the model

$$U(t) = u + ct - S(t) + Y_\alpha(t) = u + ct - \sum_{i=1}^{N(t)} X_i + Y_\alpha(t),$$

perturbed by α-stable process $Y_\alpha(t)$, $t > 0$, $1 < \alpha < 2$, $\beta = 1$. The perturbation can be interpretered as a fluctuation of the premium income, of the return from investment income, or as a fluctuation of the environment. Case $\alpha = 2$ corresponds to perturbation by Brownian motion (Grandell(1970)). When $\alpha < 2$ the perturbation leads to heavy tails even in the case of light-tailed claims. Models which includes heavy tails both in claim sizes and perturbation process are also rather interesting; for details see Furrer (1998) and Schmidli (1999) with a rich bibliography.

Bibliography

1. Adler, J., Feldman, R.and Taqqu, M.(ed.) *A Practical Guides to Heavy Tails*, Birkhauser, Boston (1998).
2. Asmussen, S., Schmidli, H. and Schmidt, V. *Tail probabilities for non-standard risk and queuering processes with subexponential jumps*, Adv.in Appl.Prob. **31** (1999).
3. Asmussen S. *Subexponential asymptotics for stochastic processes: extremal behaviour, stationary, distributions and first passage time*, (1996), Preprint.
4. Bertoin J. *An Introduction to Levy Processes*, Cambridge Univ.Press, (1996)
5. Christoph, G., Wolf, W. *Convergence theorems with a stable limit law*, Math. Reseach., **70**, Akademie-Verlag, Berlin,(1992).
6. Embrechts, P., Veraverbeke, N., *Estimates for the probability of ruin with special emphasis on the possibility of large claims*, Insurance: Math.Econ. (1982) **1**, N1, 55-72.
7. Embrechts, P., Kluppelberg, C. and Mikosh, T. *Modelling extremal events for insurance and finance*, Springer-Verlag, Berlin, (1997).
8. Fama, E.F. *Portfolio analysis in a stable Paretian market*, Management Sci., **11**, (1965), 404-419.
9. Furrer, H.J. *Risk processes perturbed by α-stable Levy motion*, Scand. Actuarial J., (1998),59-74.
10. Feller, W. *An Introduction to Probability Theory and its Applications, II*, Wiley, New York, (1971).
11. Gnedenko, B., Kolmogorov, A. *Limit Distributions for the Sums of Independent Random Variables*, Gostechizdat, M.-L, (1949).
12. Gikhman, I.I. and Skorokhod, A.V. *The Theory of Stochastic Processes, 2*, Springer-Verlag, New York, (1974) .
13. Grandell, J. *Aspects of Risk Theory*, Springer-Verlag, Berlin,(1991).
14. Grandell, J. *Mixed Poisson Processes*, Chapman and Hall, London, New York, Tokyo,(1997).
15. de Haan, L.,Pereira, T. *Estimating the index of α-stable distribution*, Statistics & Prob.Letters 41(1999), 39-55.
16. Janicki A., Weron A. *Simulation and Chaotic Behaviour of α-stable Stochastic Processes*, M.Dekker, New York, (1994).
17. Kluppelberg, C., Stadtmuller, U. *Ruin probability in the presence of heavy-tails and interest rates*, Scand. Actuarial J., N1, (1998), 49-58.
18. Kluppelberg, C. *Subexponential distributions and characterizations of related classes*, Prob. Theor. Related Fields. **82**, (1989), 259-269.
19. Lévy, P. *Calcul des Probabilites*, Paris, (1925).
20. Lévy, P. *Theorie de l'addition des variables aleatoires.* Gathier Villars. Paris, (1937 and 1954).
21. Mandelbrot B. *The variation of certain speculative prices*, J.Business, **36**, (1963), 394-419.
22. Mandelbrot B. *Fractal Geometry of Nature*, Freeman, New York, (1983).
23. McDonald J. *Probability distributions for Financial Models*, in Handbook of Statistics, **14**, Elsevier Science, (1996), 427-440.
24. McCulloch J.H. *Simple consistent estimates of stable distribution parameters*, Comm.Statist.Sim.& Comp., **15**,(1984), 1099-1136.

25. McCulloch J.H. *Financial applications of stable distributions*, in Handbook of Statistics, **14**, Elsevier Science, (1996), 393-425.
26. Mijnheer, J. *Limit theorems for sums of independent random variables in the domain of attraction of a stable law: a survey*, Theor. Probab. and Math. Statist., **53**, (1995), 109-115.
27. Mitinik, S. and Rachev, S. *Modeling Financial Assets with Alternative Stable Models*, J.Willey, New York, (1995).
28. Rachev, S., Kim, J. and Mitinik, S. (1995) *Stable Paretian models in econometrics.I.*, Math.Scientist. 24 (1999), 24-55.
29. Rolski, T., Schmidli, H., Schmidt, V. and Teugels, J.L. *Stochastic Processes for Insurance and Finance*, Wiley & Sons, Chichester, (1999).
30. Samorodnitsky, G. and Taqqu, M. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapman and Hall, London, (1994).
31. Schmidli, H. *Perturbed risk processes: a review*, Theory of Stoch.Proc., **5(21)**,(1999), 145-165.
32. Schmidli, H. *Lecture Notes on Risk Theory* , Univ. of Aarhus,(1999).
33. Skorokhod, A. *Asymptotic formulas for stable laws*, Dokl. AN USSR, **98**, (1954), 732-754.
34. Skorokhod, A. *Random Processes with Independent Increments* , Nauka, Moscow,(1963 and 1986).
35. Teugels, J.L. *The class of subexponential distributions*, Ann.Prob., 3, (1975), 1000-1011.
36. Zinchenko, N. *On the asymptotic behaviour of increments of asymmetric stable processes*, Theor.Prob. and Math.Stat., **46**, (1992), 53-60.
37. Zinchenko, N. *Generalization of strong invariance principle for multiple sums of r.v. in domain of attraction of a stable law*, Theor.Prob. and Math.Stat., **53**, (1998), 53-57.
38. Zinchenko, N.*Erdos-Renyi laws and the strong invariance principle* , Theor. Prob. and Math.Stat., **58**, (1999), 27-33.
39. Zolotarev, V. *One-dimensional Stable Distributions*, Nauka, M, (1983).

Department of Mechanics and Mathematics, Kyiv Taras Shevchenko University, 01033, Kyiv, Ukraine.
E:mail: znm@mechmat.univ.kiev.ua