Term Structure Estimation
An Implied Norm Approach

Negative Option Prices – A Puzzle or Just Noise?

Ioulia D. Ioffe

Carlson School of Management, University of Minnesota

321 19th Avenue South

Minneapolis, MN, USA 55455

Telephone: (612) 624-5869

Fax: (612) 626-1335

E-mail: iioffe@umn.edu

Alexandra E. MacKay

Joseph L. Rotman School of Management, University of Toronto

105 St. George Street

Toronto, ON Canada M5S 3E6

Telephone: (416) 978-0740

Fax: (416) 978-5433

E-mail: amackay@mgmt.utoronto.ca

Eliezer Z. Prisman

Schulich School of Business, York University

4700 Keele Street

North York, ON Canada M3J 1P3

Telephone: (416) 736-2100 ext 77948

Fax: (905) 882-9414

E-mail: eprisman@yorku.ca

Ioffe and MacKay would like to thank SSHRC of Canada, and Prisman would like to thank the York University Research Authority for financial support. We would also like to thank MathWorks Inc., in particular Sean Curry, for providing the software. The usual caveat applies.

Term Structure Estimation: An Implied Norm Approach
Negative Option Prices – A Puzzle or Just Noise?

ABSTRACT:

Knowledge of the term structure of interest rates is undeniably important. It enables the price of a stream of cash flows to be determined and is a key element for pricing certain types of derivative instruments including swaps and forward contracts. It is useful in the formation of economic policy. It is used to calibrate the parameters of the stochastic process assumed to govern the evolution of interest rates. Notwithstanding its importance, estimating the term structure of interest rates remains fraught with difficulty.

It is usual to produce an estimate of the term structure that is known to result in arbitrage opportunities. This paper suggests a new way that results in a no-arbitrage estimate of the term structure. Our method stems from the implied valuation philosophy that has its origins in the derivative market and has gained popularity there. We show how the spirit of that philosophy can be applied to the estimation of the term structure.

The goal in this paper is to uncover the discount function implicit in the prices and cash flows of bonds, and to infer the implicit selection criterion employed. It is common practice to use only straight bonds to estimate the term structure and to throw out of the sample the bonds with embedded options, resulting in the loss of information and an estimate that is quite likely inaccurate. We will show that discarding this sort of information when estimating the term structure creates the illusion of a puzzle – the illusion of arbitrage opportunities. This puzzle is apparently negatively valued options implicit in the government bonds of three countries. The approach used for term structure estimation in this paper allows the researcher to harness the information contained in government bonds with embedded options and results in an arbitrage-free estimate of the term structure.

JEL Classification: G1, G12, G13

1 Introduction

Knowledge of the term structure of interest rates is undeniably important. It enables the price of a stream of cash flows to be determined and is a key element for pricing certain types of derivative instruments including swaps and forward contracts. It is useful in the formation of economic policy. It is used to calibrate the parameters of the stochastic process assumed to govern the evolution of interest rates. Notwithstanding its importance, estimating the term structure of interest rates remains fraught with difficulty.

It is usual to produce an estimate of the term structure that is known to result in arbitrage opportunities. This paper suggests a new way that results in a no-arbitrage estimate of the term structure. Our method stems from the implied valuation philosophy that has its origins in the derivative market and has gained popularity there. We show how the spirit of that philosophy can be applied to the estimation of the term structure.

There is a relatively new trend in the finance literature. Rather than assume the risk-neutral probability distribution of stock prices, the distribution instead is imputed from the prices of derivative securities available in the market. The goal is to uncover a certain probability density function. Under this density the expected discounted values of future payoffs from a set of equity options, on the same asset and with the same maturity, are equal to their observed market prices. The density is inferred from the data rather than assumed *a priori*.

In practice, this amounts to solving a system of integral equations and so uncovering the risk-neutral probability density function of asset prices. Each equity option price and its corresponding set of future payoffs is represented by one equation. Note, that though there are infinitely many possible states of nature that could occur at each point in time, there are only a finite number of options traded on each security. This implies that there are infinitely many solutions to the system of integral equations, i.e., infinitely many risk-neutral probability density functions. The probability density function actually used for pricing depends on the criterion that is used to select it from the set of feasible density functions. In this approach to derivative pricing, the data are allowed to "speak" for themselves in the determination of the set of feasible density functions.

This paper follows the trend of allowing market data to have a voice. Our goal is to estimate the term structure of interest rates. The search is for a set of discount factors that satisfies a system of equations. Each equation in the system corresponds to the present-value-price relation for a particular bond. It is usually the case in the debt market that the number of bond payment periods exceeds the number of outstanding bonds in the market.[1] Consequently, from a theoretical standpoint the problem is underdetermined: there are infinitely many vectors of discount factors that satisfy the given system of equations.

[1] This inequality is even more pronounced if we consider only "on-the run" bonds.

Moving from the theoretical jurisdiction to empirical application, we turn from knowing there are infinitely many feasible term structures in theory to finding none empirically. The system of present-value-price equations is inconsistent and hence it is not possible to find even a single vector of discount factors satisfying the system. The inconsistency is due to the fact that it is impossible to observe synchronous prices for all bonds because the debt market is illiquid. Researchers have resorted to second best solutions. The "second best" techniques require a selection criterion[2] that accounts for the noise in observed prices and allows a discount function to be estimated from the available data. Any such estimate of the discount function depends on the criterion chosen to define "second best". Similar to the approach in derivatives markets, instead of deciding on this criterion *ad hoc*, we allow the data to "tell" us what this criterion should be.

The goal in this paper is to uncover the discount function[3] implicit in the prices and cash flows of bonds, and to infer the implicit selection criterion employed. It is common practice to use only straight bonds to estimate the term structure and to throw out of the sample the bonds with embedded options. This results in the loss of information and an estimate that is quite likely inaccurate. We will show that discarding this sort of information when estimating the term structure creates the illusion of a puzzle. This puzzle is apparently negatively valued options implicit in the government bonds of three countries. The approach used for term structure estimation in this paper allows the researcher to harness the information contained in government bonds with embedded options and results in an arbitrage-free estimate of the term structure.

Longstaff (1992) first detected the puzzling result of negatively valued call options of US Treasury bonds. Similarly puzzling values have been found in the Canadian and Israeli government bond markets. These debt markets differ from each other in many respects. Their trading rules are different. They are organized differently. They are radically different in size and liquidity. There have been many attempts to explain the results found by Longstaff. None is entirely satisfying. Not all explanations are even applicable to the other national bond markets in which the negative option pricing puzzle exists. For example, the tax-effect explanation of Jordan, Jordan and Jorgensen (1995) cannot eliminate the puzzle in the Israeli market since there is no tax on capital gains in Israel. The tax-effect argument is not useful in the Canadian government debt market either since capital gains taxation was introduced only in 1972, but the puzzle has been documented in that market as early as 1968. Jordan and Kuipers (1997) attribute the puzzle to the impact of the futures market on the bond market. However, the coexistence of a futures market with the apparent puzzle cannot be held responsible for the puzzle (at least not for the whole time period in which it is documented) in either Canada or Israel. In each of those two countries the puzzle was found to exist even before there was

[2]Examples of familiar selection criteria include the minimization of the sum of squared errors and the minimization of the sum of absolute values of errors. For the discussion of both criteria see Tuckman (1996).

[3]"Vector of discount factors" and "discount function" can be considered synonymous terms, for now.

a futures market for the corresponding instruments in either of these countries.

The documented puzzle of negative option prices is indeed a curiosity, worthy of the attention devoted to it and of the attempts to explain it away. However, it has yet to be explained away in an all-encompassing manner. In fact, though, it may not be so very curious. Despite all the differences that exist between US, Canadian and Israeli bond markets there is one thread they have in common – illiquidity. Consequently, the bond prices used in any analysis are noisy and do not represent the true market prices.[4] Bond price inaccuracies may be so severe that the puzzle is a by-product of noise, and nothing more. The results in Ioffe (1998) indicate that the puzzle exists even when the incompleteness of the bond market is taken into account. Her findings support the view that the puzzle is simply a result of the noise in the data and there are, as one might reasonably assume, no arbitrage opportunities in the market.

The evolution of the term structure can be modeled either in an arbitrage-based or an equilibrium-based framework. Ironically, researchers estimate the parameters of these models from a term structure that implies that the values of options embedded in bonds are negative. Yet, when the same options are priced based on the estimated parameters, their prices are positive. After all, what researchers do is pretend that the term structure is a realization of the "estimated" stochastic process. It is clear that this estimation mechanism is internally inconsistent. In this paper we propose a methodology that eliminates this inconsistency and produces an estimate of the term structure that does not induce the puzzling results.

The remainder of the paper is organized as follows. The next section discusses the theoretical background of the ingredients required for the recipe: the usual general term structure estimation approach, the implied valuation approach in the derivatives market and the puzzle of negative option prices. The proposed methodology (the recipe itself) is described in the third section. The numerical implementation of our methodology is presented in the fourth section and in the fifth section the empirical results are tabled and discussed. The paper concludes with a section summarizing our findings and the implications of using the methodology we propose.

2 Theoretical Setup

2.1 Term Structure Estimation

Let us now present the issue in a more formal manner. It is usually the case in the debt market that the number of bond payment periods exceeds the number of outstanding bonds in the market. Consequently, there are infinitely many discount functions, d, that satisfy the present value-price

[4] Recognizing that "true" prices fall in a range between the bid and ask prices does not alter the substance of the argument.

relations represented by the system of equations

$$Ad = p, \tag{1}$$

where p is the vector of bond prices and $A = [a_{ij}]$ is the matrix of cash flows from bonds with a_{ij} being a payoff from bond i on date j, $i = 1, ..., I$ and $j = 1, ..., J$.

Empirically, it is not possible to find a discount function that satisfies (1). Debt market prices available to researchers are sufficiently error-ridden that the system of equations (1) is inconsistent. Researchers have resorted to second best solutions. The following structure of the estimation problem is familiar:

$$\min_{\epsilon,d}\{\ \epsilon'\epsilon \mid Ad + \epsilon = p,\ d \geq \mathbf{0}\ \}, \tag{2}$$

where ϵ is the vector of errors in observed market prices.[5]

The usual practice when estimating a term structure of interest rates is to use only straight bonds that are default free. Any bonds with similar default risk outstanding in the market that have special features are not included in the sample at this stage. In particular, this means that bonds with embedded options are thrown out and the matrix A is a matrix of payoffs from straight bonds only. Once an estimate of the term structure, d^*, is obtained from (2), cash flows associated with bonds with embedded options are priced. Priced this way, some of the values of the implicit options in these bonds are negative. The existence of the negative option pricing puzzle debated in the literature is already less surprising. After all, the information contained in the bonds with embedded options was not used in the estimation of the term structure.

There have been attempts to make use of the information contained in bonds with embedded options in term structure estimation. Ho, Lee and Son (1992), for example, make use of the information contained in bonds with embedded options in their method of estimating the term structure. The prices of embedded options are estimated. Then the term structure is estimated using a sample that includes the callable Treasury bonds, their prices having been adjusted according to the call prices estimated just obtained. The estimation technique then iterates between estimating the term structure, estimating call option prices, re-adjusting the prices of callable bonds and re-estimating the term structure with new adjusted prices. The iterative procedure stops when the estimate of the term structure does not induce the negative option pricing puzzle in bonds and an extra iteration does not change the estimation.

Ho, Lee and Son admit that their estimation procedure is very sensitive to the systematic errors in pricing the call options. One should also recognize that the term structure estimate is calibrated to misleading data, i.e., negatively priced implicit calls. We propose a methodology that does not require any assumption regarding option pricing models. Our methods allows noise to be separated

[5] Notice that here we do not impose monotonicity of the discount function. The exposition is thus simplified. The continuous approximation employed in a later section will ensure monotonicity.

from the data in a manner that is imputed from the data rather than assumed *a priori*. Though the literature regarding the implied valuation operators is vast and well known, the following review of the approach serves as an introduction to our methodology.

2.2 Implied Valuation Operator in Derivatives Markets

There is a recent trend in derivatives markets to price securities and cash flows based on an implied valuation operator. Instead of assuming a lognormal risk-neutral probability distribution of the asset prices, the distribution of asset prices is estimated from the market prices of derivatives. Implicit in the market prices is the price of one dollar contingent on the value of the state of nature. This value is sometimes referred to as a stochastic discount factor or risk-neutral density.[6] Herein lies the key.

This trend amounts to "letting the data speak". The stochastic discount factors are imputed by searching for a function under which the expected discounted values of future payoffs from the derivatives at maturity are equal to their market prices. In other words, we are looking for a risk-neutral density function $f_t(\cdot)$, for which

$$p_t = e^{-r_f(T-t)} \int_0^{\infty} f_t(s) G_T(s) ds,$$

where p_t is the market price of the particular derivative security today, and $G_T(\cdot)$ is the payoff of that derivative at maturity. $G_T(\cdot)$ is a known deterministic function that depends on the price of the underlying asset. Papers in this literature include, for example, Aït-Sahalia and Lo (1998) and Rubinstein (1994).

The trend has yet to catch on and be used in the context of the debt market. This estimation philosophy has a natural application to the estimation of the (realized) term structure of interest rates. In the debt market, however, we propose to use the implied valuation philosophy to impute the norm that is used as the optimization criterion in the term structure estimation problem. Each norm induces a set of discount functions, and hence the data-induced norm together with an implied discount function is similar in nature to the data-induced risk-neutral density. We will search for a norm that separates the noise from the data and induces a set of discount factors satisfying a specified condition. The nature of this condition should reflect sound economic rationale brought to bear on the problem. The central idea is to make use of information that is often ignored in estimating the term structure. For the purposes of this paper, this information is contained in bonds with embedded options. However, a debt market with both straight bonds and bonds with imbedded options is not a requirement in order to make use of the technique we propose. All that is necessary is the existence

[6] These are actually prices of Arrow-Debreu securities and are referred to as the state-price density. In an equilibrium framework, the state-price density is expressed in terms of stochastic discount factors (Hansen and Jagannathan (1991), Hansen and Richard (1987)) and in an arbitrage-based framework it is expressed in terms of risk-neutral density (Ross (1976), Cox and Ross (1976), Black and Scholes (1973), Merton (1973)).

of securities for which a pricing hierarchy with respect to the straight bonds can be obtained. Some alternative criteria will be discussed in more detail in the next section.

2.3 The Puzzle of Negative Option Prices Implicit in Bonds

The negative option pricing puzzle was first uncovered by Longstaff (1992) in the context of callable US Treasury bonds. The essence of the problem can be understood as follows. Consider two otherwise equivalent bonds: one callable and one not. It is intuitively obvious that the market price of the callable bond should be below that of the non-callable bond. Longstaff's discovery was that this pricing hierarchy was not satisfied in some cases in the US Treasury market. This pricing hierarchy can be used to impose sensible constraints on the norm. The constraints will ensure that the term structure estimate results in prices for bonds and implicit options such that the pricing hierarchy is respected.

The puzzle of negative option prices implicit in bonds with embedded options has been uncovered in American, Canadian (Athannasakos, Caryanopoulos and Tian, 1994) and Israeli (Itzikowitz, MacKay and Prisman, 1996) bond markets. The puzzle is said to exist if the inequalities

$$Bd \geq p_c \tag{3}$$

for callable bonds and

$$Bd \leq p_e \tag{4}$$

for extendible bonds are violated for all feasible vectors of discount factors, d, obtained from solving the term structure estimation problem.[7] Here $B = [b_{mj}]$ is a matrix of payoffs from extendible/callable bonds, where b_{mj} is a payment from bond m on day j, and p_c and p_e are vectors of observed prices of callable and extendible bonds, respectively. In other words, the puzzle exists when the observed price of a callable (extendible) bond is greater (smaller) than the upper (lower) bound assigned to the value of the cash stream generated by this bond.

Since the discovery of the puzzle in the US Treasury market, attempts have been made to refute its existence or at least explain it away. The first such attempt came in 1995 by Jordon, Jordon and Jorgensen. The central argument of this paper hinges on a detailed examination of the tax treatment of the bonds making up the replicating portfolio for a callable US Treasury bond. The existence of

[7]Though this description of the puzzle holds in the US market, in fact, the method of estimating the price of the implicit option in the seminal negative option puzzle paper (Longstaff, 1992) does not rely on an estimate of the term structure of interest rates in that market. The US bond market is "thick" enough that the pricing was done with bond triplets, i.e., a linear combination of two other bonds of the same maturity and risk is used to replicate the cash flows of the bond being priced. In Ho, Lee and Son (1992), though, the puzzle in the US market is documented via term structure estimation. Also, if the triplets indicate the existence of arbitrage opportunities, so too, will the term structure. After all, the same data are used to estimate the term structure.

taxation of capital gains allows the investor to tax-time the option, and hence the relation between the bonds in the replicating portfolio relative to the bond being replicated is convex rather than linear. The puzzle's apparent existence was explained away in the majority of cases. The argument, while intriguing, cannot explain the existence of the puzzle in the Israeli market. In the Israeli economy there is no tax on capital gains and there are no strip bonds. Hence the line of reasoning proposed by Jordan *et al.* can not be applied. In the Canadian fixed income market, their argument is also troublesome since capital gains taxation was only introduced in Canada in 1972. From 1972 until 1987, only fifty percent of capital gains were taxable. This was raised to two-thirds in 1988 and then to seventy-five percent in 1990. Their argument also suffers in the Canadian context because research has shown that there is a representative investor belonging to a particular tax clientele.[8]

Jordan and Kuipers (1997) refute the existence of the negative option pricing puzzle using an argument that relies on the coexistence of the futures market and the bond market. Though the reasoning holds in most cases in the US bond market on which their research is based, it is flawed in the Israeli and Canadian debt markets. In Israel, for example, trading in futures on Treasury bonds did not begin until March 1991. Itzikowitz, MacKay and Prisman (1996) document puzzle occurrences in late 1990 and the early months of 1991, before futures trading began. In Canada, the puzzle has been documented for extendible Government of Canada bonds as early as 1968. However, trading in futures on Government of Canada bonds did not begin before the mid 1980s.

It is usual when estimating the term structure to remove all bonds with embedded options from the sample. There is a consequent loss of the information contained in the prices of these bonds. Given that the researcher is working with reduced information for the very bonds whose options are being priced, it is perhaps not entirely surprising that there are puzzling results. Moreover, while the existence of the puzzle is troubling, it has yet to be considered with full recognition of the noisy reality of the bond prices used in estimation. As a result of the illiquidity of the debt market, the bond prices available to researchers and practitioners are not "true" market prices. One might then contemplate that the puzzling results are only a consequence of noise in the data, and that there are no arbitrage opportunities in the market after all.

We accommodate the information contained in the bonds with imbedded options and we account for the noise in observed market prices in our estimation, thus correcting flaws in other estimations of the term structure. Anchored in the philosophy of implied valuation techniques, we propose a methodology that allows us to separate the noise from the true prices while accounting for the information contained in the prices of bonds with embedded options.

[8] The representative investor is one for whom all bonds are correctly priced. For further details, the interested reader is referred to Prisman and Tian (1994) and to MacKay, Prisman and Tian (2000).

3 Implied Term Structure

The primary goal in this paper is to impute from the data a norm used for the minimization criterion of the optimization problem from which the term structure is estimated. This is done making use of the information contained in bonds, even those that possess special features. For this paper, we focus on the information contained in bonds with implicit options. This information indicates to the researcher an appropriate separation of the available price into the true price and noise around that price. Our implied norm will produce an estimate of the term structure that does not result in implicit option prices that are negative.

The regression approach treats all bonds equally, minimizing the sum of squared errors. Our techniques allow us to let the data tell us which observations are more relevant than others and hence should receive more weight in the estimation process. For example, if the bond traded two weeks prior to the recorded end-of-month observation date, its recorded price is not the "true" market price for the bond on the observation day. Hence a large pricing error for this bond should not have a lot of weight. If however, the bond traded on the month-end observation day and the recorded price includes error, then this error should have a higher weight, i.e., should be penalized more severely.

In order to make the problem of uncovering the norm used in the optimization criterion computationally accessible, we assume that the norm is induced by a positive definite matrix[9] denoted by Γ. The usual minimization criterion for term structure estimation problem[10] is $\epsilon' \, \mathbb{I} \, \epsilon$: we use $\epsilon' \, \Gamma \epsilon$ instead. We further assume that the matrix Γ is diagonal[11] and denote the vector of its diagonal elements by γ. For a given[12] $\gamma > \mathbf{0}$, the term structure estimation problem can now be written as

$$\Psi(\gamma) = \inf_{\epsilon, d} \left\{ \epsilon' \, \Gamma \, \epsilon \mid \; Ad + \epsilon = P, \quad d \geq \mathbf{0} \right\}. \tag{5}$$

Notice that this problem is parameterized by γ and the term structure estimate is thus γ-dependent.

For every fixed vector γ, problem (5) is a minimization of a convex function over a convex set. This means that for every fixed γ, the Kuhn-Tucker conditions for this problem are necessary and sufficient and can be written as[13]

$$2\,(P - Ad)' \, \Gamma \, A \, D \;\; = \;\; \mathbf{0} \tag{6a}$$

$$2\,(P - Ad)' \, \Gamma \, A \;\; \leq \;\; \mathbf{0} \tag{6b}$$

$$d \;\; \geq \;\; \mathbf{0}, \tag{6c}$$

[9] It is well known from mathematics that every positive definite matrix induces a norm. Note though that not all norms can be induced by a positive definite matrix.

[10] Here $\mathbb{I}$ denotes an identity matrix.

[11] If required, this assumption can be relaxed to allow a more general form of the matrix.

[12] Since Γ is assumed to be diagonal and positive definite, the vector of its diagonal elements, γ, must be positive, i.e., $\gamma > \mathbf{0}$.

[13] For the derivation of these results see Appendix 1.

where D is defined as

$$D = \begin{bmatrix} d_1 & 0 & 0 & \dots & 0 \\ 0 & d_2 & 0 & \dots & 0 \\ 0 & 0 & d_3 & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & 0 & \dots & d_J \end{bmatrix}.$$

Sufficiency of the Kuhn-Tucker conditions means that for a given vector γ^*, every $d(\gamma^*)$ satisfying (6) will be an optimal solution to (5). In order to "let the data speak" we will not prespecify the shape of the norm, i.e., we will not take γ as given. Instead, we will allow the data tell us what γ is. We will achieve this by estimating γ together with the sought-after discount function $d(\gamma)$. By using the Kuhn-Tucker conditions and allowing γ to vary, all possible pairs of the implied norm and corresponding term structure estimate are revealed. This is because for every pair $(\gamma_{imp}, d(\gamma_{imp}))$ satisfying (6), $d(\gamma_{imp})$ will be an optimal solution to

$$\Psi(\gamma_{imp}) = \inf_{\epsilon,d} \{\epsilon' \, \Gamma_{imp} \, \epsilon \mid Ad + \epsilon = P, \quad d \geq \mathbf{0}\}.$$

By allowing γ to vary in (6) we have obtained a data-induced set of all possible pairs of norms and term structure estimates. However, given the structure of the problem, there are many pairs $(\gamma_{imp}, d(\gamma_{imp}))$ satisfying (6) and the question becomes one of selection. This is where we utilize the information contained in bonds with embedded options. The criterion we use to pick a pair $(\gamma_{NA}, d(\gamma_{NA}))$ out of those satisfying the Kuhn-Tucker condition is

$$\begin{aligned} Bd &\leq p_e \qquad \text{for extendible} \\ Bd &\geq p_c \qquad \text{for callable.} \end{aligned}$$

In other words, of all possible estimates of the term structure, we choose one that does not induce the "puzzle". We require these constraints to hold for every bond. However, since bonds with embedded options may also have erroneous prices, it may be necessary to accommodate them. In principle, these constraints can be relaxed so they hold only in a statistical sense, i.e., so that the constraints hold for the "majority" of the bonds rather than for all of them. This can be done by solving a few iterations of our methodology. Each iteration represents the estimation of an implied term structure with the selection criterion being a "no-puzzle" constraint that is allowed to break down only for a statistically insignificant number of bonds. The number of bonds for which the constraint is allowed to break down depends on the level of statistical significance desired and is the same across iterations. The number of iterations as well as the bonds for which the break down is allowed can be determined by applying combinatorics. This less-stringent route imposes a level of numerical complexity that is not justified to achieve the aim. Nonetheless, it can be accomplished in principle.

We can now write the full term structure estimation problem. Its solution will provide an implied norm together with an arbitrage-free implied valuation operator.[14]

$$
\begin{aligned}
2\,(P - Ad)'\,\Gamma\,A\,D &= \mathbf{0} && (7a)\\
2\,(P - Ad)'\,\Gamma\,A &\leq \mathbf{0} && (7b)\\
d &\geq \mathbf{0} && (7c)\\
\gamma &> \mathbf{0} && (7d)\\
Bd &\geq p_c \quad \text{for callable} && (7e)\\
Bd &\leq p_e \quad \text{for extendible} && (7f)
\end{aligned}
$$

It is critical to recognize that this approach is not the same as adding constraints (7e) or (7f) to the estimation of the term structure in problem (5) and solving directly. Not only are these two approaches mathematically different problems, they differ in their philosophies. In our approach we obtain a set of all possible pairs of implied norms and implied term structure estimates, and then use either (7e) or (7f) as a criterion to pick a reasonable pair, $(\gamma, d(\gamma))$. Using the other approach requires prespecifying γ which is not available *a priori*. Moreover, if an estimate of the term structure is obtained from problem (5) with γ prespecified in the latter approach, it will not necessarily price bonds with embedded options in a reasonable way. If, however, an estimate of the term structure is obtained from problem (5) with the γ induced by the data in our approach, bonds with embedded options will be priced in a sensible way. Another disadvantage of adding constraints (7e) or (7f) to the problem (5) and prespecifying γ is that it will not be possible to reveal the whole spectrum of the term structures. It would also be very rigid with respect to errors, treating all observations in a prespecified way.

If one were to solve directly problem (5) together with an added "no-puzzle" constraint one necessarily would obtain a term structure estimate that does not produce the puzzling results. However, such an estimate of the term structure would not necessarily be an optimal solution to (5) and the resultant implicit option values might be negative. Rather, we are looking for an implied norm such that, if used in (5) for the term structure estimation, would produce an estimate that is arbitrage free without imposing the no-puzzle constraint.

There is an important (and convenient) property that problem (5) possesses. The objective function of this problem is homogeneous of degree one in γ, and γ does not enter the constraints of the problem. This means that multiplying γ by a positive constant will not change the optimal solution to the problem. By the same token, multiplying γ by a positive constant will not change the solution to (7). In particular, it would not change the set of otherwise obtained feasible pairs of implied norm and corresponding implied term structure. This quality of the problem allows us,

[14] In general, there is no unique solution to this system of equations and one needs to use a criterion to pick a pair $(\gamma,\, d(\gamma))$. We will discuss a choice of criteria in Section 4.2.

for example, to add a requirement that $\mathbf{1}'\gamma = c$ to the problem, where c is some positive constant, without changing the solution to the problem.

By setting $c = 1$, the reader may already sense that there is a connection between our methodology and the problem of weighted regression. Weighted regression is a classical technique. We hope that understanding of the connection between it and our methodology will shed light on our approach to the problem. Consider the problem

$$\inf_{\epsilon,d} \{ \epsilon'\Gamma^* \epsilon \mid Ad + \epsilon = P, \quad d \geq \mathbf{0} \}, \tag{8}$$

where $(\gamma^*, d(\gamma^*))$ is a solution to (6). The optimal solution to this problem is identical to the optimal solution to the following problem:

$$\frac{1}{\mathbf{1}'\gamma^*} \inf_{\epsilon,d} \{ \epsilon' \, \Gamma^* \, \epsilon \mid Ad + \epsilon = P, \quad d \geq \mathbf{0} \}, \tag{9}$$

and is $d(\gamma^*)$. Defining $\omega^* = \frac{\gamma^*}{\mathbf{1}'\gamma^*}$ we can see that $\mathbf{1}'\omega = 1$. The vector ω can be thought of as a vector of weights. In other words, given a solution to problem (6), a pair $(\gamma^*, \, d(\gamma^*))$, problem (9) will now give the same results as

$$\inf_{\epsilon,d} \{ \epsilon'\Omega^* \epsilon \mid Ad + \epsilon = P, \quad d \geq \mathbf{0} \},$$

where Ω^* is a diagonal matrix and ω^* is the vector of its diagonal elements. This means that given a solution to (6) we have a solution to a weighted regression problem wherein the weights are given by the vector ω^*. In our approach the weights are not chosen *a priori*. Rather, they are imputed in the estimation process.

It is not necessary to have callable or extendible bonds available in the market in order to utilize our methodology. One only needs to know market prices of some cash flows on which theoretical arbitrage bounds can be placed. The obtained hierarchy can be utilized to pick an implied term structure estimation that does not violate sensible intuition concerning the pricing hierarchy.

4 Numerical Implementation

4.1 Continuous Approximation

It is desirable to obtain a continuous approximation of a discount function: we need to know the discount factors not only for the times at which coupon payments occur but also for time periods between coupon payments. We could estimate discount factors associated with each and every bond payment date and then extrapolate between two consecutive dates. This technique is not very accurate. A more accurate estimate can be obtained by assuming a continuous approximation of a discount function as is done below.

It can be assumed that a discount function belongs to a $(n+1)$-dimensional functional space, the basis for which is $\{\varphi_i\}_{i=0,...,n}$. The discount function then can be represented as

$$d(t) = \sum_{i=0}^{n} \alpha_i \varphi_i(t). \tag{10}$$

There are a few approximations that can be chosen to implement this approach. One is to assume that the functional space is the space of polynomials and that its basis is $\{1, t, t^2, t^3, ..., t^n\}$. This is the approach of Chambers, Carleton and Waldman (1994). The discount function then would be represented as

$$d(t) = \sum_{i=0}^{n} \alpha_i t^i. \tag{11}$$

There are a few problems with this choice of approximation. In general, it requires special treatment.[15]

We will implement the representation of the discount function in (10) based on the Phillips and Taylor (1970) approximation of a convex function.[16] It is assumed that the discount function is of the form (10), where $\alpha_i \geq 0$ for $i = 1, ..., n$, $\alpha_0 = 1$, and the φ_is are chosen to be

$$\varphi_i(t) = \sum_{l=0}^{n-i} (-1)^{l+1} \begin{pmatrix} n-i \\ l \end{pmatrix} \frac{t^{i+l}}{i+l} \tag{12a}$$

for $i = 1, ..., n$ and

$$\varphi_0(t) = 1. \tag{12b}$$

These φs are first derivatives of the basis of the approximating space used by Phillips and Taylor. In their paper, the approximation of a convex (concave) function is derived on $[0, 1]$. Without loss of generality, it is assumed here as well that the time is measured on the interval $[0, 1]$, where 1 corresponds to the longest maturity.[17]

No matter what type of approximation is used, the estimate must possess all the economic characteristics that are inherent in the term structure of interest rates. In particular, an estimate of a discount function, $d(\cdot)$, must be monotonically decreasing in order to preclude the occurrence of negative forward rates, $d(0) = 1$ and $d(t) \geq 0$ for all $t \in [0, 1]$. It can be shown that the approximation approach based on Phillips and Taylor possesses all of these qualities. The monotonicity is demonstrated by noting two things. First, the first derivative of a concave function is monotonically decreasing, since the second derivative is always non-positive. Second, a non-negative linear combination of monotonically decreasing functions is monotonically decreasing. By substituting $t = 0$ into

[15] The constraint dealing with the monotonicity of the discount function needs to be imposed explicitly. This may lead to economically unappealing assumptions.

[16] See Schaefer (1981).

[17] It is true that the discount function must be positive for all future times. However, empirically we have found that the estimate of the term structure is not sensitive to the choice of the longest maturity or some later time as "infinity".

the approximation one will see that $d(0) = 1$. Finally, positivity is guaranteed by requiring that

$$\sum_{i=0}^{n} \alpha_i \varphi_i(1) \geq 0.$$

Another reason for choosing this approximation is that any monotonically decreasing function can be approximated by a non-negative combination of φ_is as in (12a) with whatever desired level of accuracy.

Under the assumption that discount function is of the form presented in (12a) through (12b) the matrix A is transformed into matrix $\tilde{A} = A\Phi$, where Φ is a $J \times n+1$ matrix. Every row j of Φ corresponds to the basis vector of the functional space evaluated at some time j at which a bond payment occurs, i.e., $\Phi_j = \{\varphi_0(j), \varphi_1(j), ..., \varphi_n(j)\}$. The choice of n is mostly arbitrary. In this paper we choose n to be 11.[18]

4.2 Numerical Specification of the Problem

Taking into account the continuous approximation just introduced, the problem to be solved can now be written as

$$(2\ (P - \tilde{A}\alpha)'\ \Gamma\ \tilde{A} + \xi\ \varphi(1)\ )\ I_\alpha = \mathbf{0} \tag{13a}$$

$$2\ (P - \tilde{A}\alpha)'\ \Gamma\ \tilde{A} + \xi\ \varphi(1) \leq \mathbf{0} \tag{13b}$$

$$\xi\alpha'\varphi(1) = 0 \tag{13c}$$

$$-\alpha'\varphi(1) \leq 0 \tag{13d}$$

$$\tilde{B}\alpha \geq p_c \quad \text{for callable} \tag{13e}$$

$$\tilde{B}\alpha \leq p_e \quad \text{for extendible} \tag{13f}$$

$$\alpha \geq \mathbf{0}, \qquad \gamma > \mathbf{0}, \qquad \xi \geq 0, \tag{13g}$$

where ξ is a Lagrange multiplier for a non-negativity constraint of the discount function,

$$I_\alpha = \begin{bmatrix} \alpha_0 & 0 & 0 & \dots & 0 \\ 0 & \alpha_2 & 0 & \dots & 0 \\ 0 & 0 & \alpha_3 & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & 0 & \dots & \alpha_n \end{bmatrix}$$

(i.e., I_α is a diagonal matrix whose non-zero elements are the coefficients of the approximating functions that need to be estimated), and $\tilde{B} = B\Phi$ is the transformed matrix of cash flows from bonds with imbedded options. For the details on the derivation of these conditions see Appendix 2.

[18] This choice of n corresponds to that used in Schaefer (1981) and in Ioffe (1998).

MATLAB is used to solve the problem numerically. The system of equations (13) is solved. We obtain a feasible set of implied norms, γ, together with corresponding estimates of the term structure, $\alpha(\gamma)$. There are, however, many pairs of $(\gamma, \alpha(\gamma))$ that satisfy this system of equations and hence we require a criterion with which to make a selection. This situation is similar to the one found to exist in the derivatives literature. In that literature, many implied valuation operators exist and a criterion is used to choose among them. For example, Jackwerth and Rubinstein (1997) consider a few possible criteria, e.g. a "goodness-of-fit" function and a "sum of absolute deviation function". There are other selection criteria and corresponding reasons for employing one and not another in a given situation. In general, the selection criterion is chosen based on its intuitive appeal given the problem at hand.

In this paper we use two selection criteria: the sum of squared deviations and the sum of absolute deviations. In other words, we chose a pair $(\gamma, \alpha(\gamma))$ that satisfies (13) and that minimizes either $\sum_i \epsilon_i^2$ or $\sum_i |\epsilon_i|$. The intuition behind the use of the sum of squared deviations criterion is to maximize the explanatory power of the model in the sense of the classical regression model. In other words, out of all possible pairs $(\gamma, \alpha(\gamma))$ we choose one that maximizes R^2. An absolute deviation criterion is chosen since it is also well known and widely used. It is argued sometimes to be more appropriate than the sum of squared errors since, for example, it is not as affected by outliers as the sum of squared errors. It is of interest to see how sensitive the implied norm and the implied term structure are to the choice of the selection criteria. The sensitivity of the implied norm and implied term structure to the selection criterion employed can be analyzed by considering the "distance" between implied term structures chosen by different criteria. We calculate an L^2 norm between the implied term structures d^*_{abs} and d^*_{sq}, chosen by the sum of absolute and of squared deviations criteria, respectively. In particular, we consider

$$\left[\int_0^1 \left\{d^*_{abs}(t) - d^*_{sq}(t)\right\}^2 dt\right]^{1/2}. \tag{14}$$

We do not report these results because the calculated value was zero at all times.

In addition to solving (13), we also solve the usual regression problem

$$\min_{\epsilon,\alpha}\{\ \epsilon'\epsilon \mid \tilde{A}\alpha + \epsilon = p,\ \alpha'\varphi(1) \geq 0,\ \alpha \geq \mathbf{0}\ \} \tag{15}$$

and the regression problem with the addition of a "no-puzzle" constraint

$$\min_{\epsilon,\alpha}\{\ \epsilon'\epsilon \mid \tilde{A}\alpha + \epsilon = p,\ \alpha'\varphi(1) \geq 0,\ \alpha \geq \mathbf{0},\ \tilde{B}\alpha \leq p_e\}. \tag{16}$$

As we have explained before, both of these problems are meaningless within the framework of the implied valuation philosophy. Moreover, their optimal solutions do not satisfy the conditions of problem (13) (unless by chance there were no puzzling results produced in the regression problems). However, an optimal solution from (16) is "less infeasible" than one from problem (15) since it satisfies at least some constraints, in particular $\tilde{B}\alpha \leq p_e$. When the former is used as an initial solution for numerical implementation of (13), it allows for a faster convergence of the numerical procedure used for solving (13).

Imposing strict inequality constraints in a numerical problem is impossible. Here, such a constraint must be imposed on the vector of coefficients of the implied norm: $\gamma > \mathbf{0}$. For the purposes of this paper, we have assumed that γ is greater than or equal to some fixed $\delta > \mathbf{0}$. In some types of problems this may not be reasonable to do since a solution may depend on the choice of δ. However, in our problem this does not present any difficulties since multiplying γ by a positive constant in (5) will not alter the optimal solution. This is because problem (5) is homogenous in γ, as discussed in Section 3. Equivalently, multiplying γ by a positive constant will not alter the problem that we solve numerically, (13).[19]

This latter quality allows us to rescale obtained γ^* to satisfy the following requirements:

$$\mathbf{1}'\gamma^{**} = I,$$

where I is the number of bonds in the sample and

$$\gamma^{**} = I \frac{\gamma^*}{\mathbf{1}'\gamma^*}.$$

This is done in order to compare the implied norm induced by Γ^{**} with the Euclidean norm used in regression that is induced by $\mathbb{I}$, the identity matrix. In order to see how far apart the implied norm and regression estimates are, we consider the "distance" in the L^2 sense between the implied discount function, d^*_{sq}, and the one obtained from (15), d^*_{Rgr}:

$$\left[\int_0^1 \left\{ d^*_{sq}(t) - d^*_{Rgr}(t) \right\}^2 dt \right]^{1/2}.$$

In some sense we are interested in knowing how far the economically reasonable estimate and the estimate that induces arbitrage are from each other. Another criterion under which we can compare the two estimation procedures is R^2. It is clear that the R^2 of the regression will not be smaller than the R^2 of the implied estimation approach. After all, the regression approach is devised to maximize R^2. However, since the regression approach produces an arbitrage-inducing estimate of the term structure, even its highest R^2 is meaningless to economists. It would be wrong to select an estimation model in this case based on the highest R^2 alone.

4.3 The Data

The data used for the empirical application is the Dalhousie Bond Tape of the Government of Canada bonds. The tape contains monthly observations of all outstanding bonds denominated in Canadian

[19] We say it loosely here, since the problem is actually homogeneous in (γ, ξ).

dollars with no imbedded options (straight bonds) as well as callable[20] and extendible bonds. Callable and extendible bonds that were available during the period from January 1968 to December 1985 were separated from the straight bonds in order to implement the methodology. For every month the data set includes the observation date, prices of the bonds, coupon rates, maturity dates, bond numbers and yields. This data set provides only the mid-market prices and not the bid and ask prices.

5 Empirical Results

This estimation technique that allows the particular bond data for a given time period to "speak for itself" proves very effective. Results show that our estimation of the term structure is "good" using the usual statistical measures and that it produces sensible prices of options implicit in bonds.

Results were obtained for April and October of each year of the sample. We report only results for April 1970, October 1974, June 1979 and October 1982. We consider these to be representative and further insight is not gained from examination of additional months. For each month, we estimated a term structure of interest rates via constrained regression and via our implied valuation technique. For each estimation method we used both minimized squared errors and minimized sum of absolute deviations as selection criteria. The estimation proves to be insensitive to the selection criterion employed, at least numerically.

Table 1 reports the R^2 obtained from our estimation method, "Implied", and from the regression method. As expected, that from the regression estimation method is higher than the one from our approach for each of the four months of results reported. Caution should be exercised in using the R^2 alone to judge the superiority of one estimation technique over another in this case. One should question the usefulness of an estimation that produces results implying negative option prices implicit in bonds, even if its R^2 measure is high. The estimates from regression induce arbitrage and are therefore not admissable estimates, in our view.

Table 2 compares the estimated option prices via each of the two estimation methods. Option prices are estimated using the methodology of this paper, "Implied OP", and employing the constrained regression method, "Regression OP". In each case, the term structure is used to calculate the present value of the cash flows for each of the extendible Government of Canada bonds outstanding in that month. This estimated price of the bond as a straight bond with no imbedded option feature is compared to the observed price of the extendible bond. The difference represents the estimated implied option value via that term structure estimation method. The negative option pricing puzzle is clearly evident in the implied option prices estimated using constrained regression methods. In only one of the four months, June 1979, do all of the implicit options have positive

[20] The Government of Canada issued only one callable bond denominated in Canadian dollars. It matured on March 15, 1998.

value using this estimation technique. By definition, using the implied term structure estimation method developed and advocated in this paper, all of the options implicit in the extendible bonds have estimated positive value.

The coefficients, α_i, of the discount function are shown in Table 3 and the discount functions are plotted in Figures 1 through 4. Table 4 reports the L^2 norm between the constrained regression and implied discount functions. In those figures, as in Table 4, it seems that the "distance" between the two methods is not drastic. The difference between the discount functions from the two estimation approaches is apparent from Figures 5 through 8 that depict the corresponding term structures. The term structures are expressed in terms of annual semiannually compounded interest rates obtained from the discount function via

$$r(t) = 2\left\{d(t)^{-1/2t} - 1\right\}.$$

It is important to notice that the estimated term structures differ not only in value, but in shape. This is crucial when trying to estimate the parameters of the stochastic processes assumed to govern an evolution of the term structure of interest rates. If an estimate resulting in negative option prices is used, the fitting becomes meaningless, since the models become internally inconsistent. This is most obvious if one considers the Heath-Jarrow-Morton (HJM) model. HJM is a no-arbitrage based model and it takes a whole term structure of interest rates as an initial condition. If this term structure induces arbitrage, clearly it should not be used as an initial condition for a model that assumes no arbitrage.

To give a more general sense of how our estimation results compare to the regression methodology, plots of ϵ (error terms) and γ (the implied norm) obtained from both estimation techniques are presented. Consider first the pattern of the difference between the error terms in the two estimations as shown in Figures 9 through 12. The errors for bonds of very short and very long maturity do not differ much across the two estimation methods. However, for bonds in the medium range maturity, the difference in errors from the fitted estimation between the two methodologies differ markedly.

Consider finally the pattern of the implied norm, the component terms of the vector γ as shown in Figures 13 through 16. The figures clearly show that the observations are not all given equal weight in the estimation. (It may also appear that some observations receive no weight at all, i.e., that the γ are zero. Rather they are merely very small.[21])

The implied term structure estimation method outlined and advocated in this paper has been shown to produce a term structure that results in sensible prices for options implicit in bonds. This outcome was achieved by construction. As was mentioned previously, the existence of bonds with embedded options is not necessary for implementation of our methodology. The method is more widely applicable, as discussed in the concluding section to the paper.

[21] Rescaling of the numerical values in the figure to make clear that no γ is zero is not possible without eliminating the larger values of γ.

6 Conclusions

This paper marks the first application of the recent trend in pricing that originated in the derivatives market to the estimation of the term structure. The implied valuation philosophy is applied to the estimation of the term structure. In so doing, we obtain an estimate of the term structure that does not induce arbitrage.

Despite the importance of the term structure in the fields of economics and finance, an arbitrage-inducing term structure has been used in the literature as a component of no-arbitrage-based and equilibrium-based models for many years. In this paper we have proposed a methodology that eliminates the inconsistency. We allow the data to "speak " by harnessing the information contained in the callable and extendible bonds to produce an arbitrage free estimate of the term structure.

Our results demonstrate clearly that a judicious estimation of the term structure results in prices for options implicit in bonds that are positive. In other words, the puzzle of negative option prices implicit in bonds no longer exists. We argue that using R^2 alone as a criterion for choosing a model to fit the term structure in the observed data is economically senseless and should not be used. Is a high R^2 for a term structure estimate really useful if it does not result in a sensible pricing hierarchy for straight bonds and those with imbedded options? Such an estimate misprices financial securities and causes bias in the estimation of the parameters of the process that governs the term structure of interest rates. However, utilizing the methodology outlined in this paper, a researcher can produce no-arbitrage estimates of the term structure and can select one that maximizes R^2. This way an economically sensible estimate is obtained and the explanatory power of the model is maximized.

While in this paper we have utilized the information contained in extendible bonds, the existence of a negative option pricing puzzle is not necessary for our methodology to be meaningful and sensible. For example, corporate bonds rather than government bonds with implicit options may be used in our methodology as an information source for the determination of the selection criterion. A no-arbitrage hierarchical structure among the prices of corporate bonds and of default-free securities with otherwise identical cash flows might be used instead of the bonds with embedded options. Setting up the structure of the implied norm methodology will then proceed by imposing a constraint requiring that the prices of corporate bonds obtained from the default-free estimated term structure should be larger than their actual observed market prices. Other securities for which a similar pricing hierarchy can be established could also (or instead) be used as an information point.

While the empirical results in this paper were obtained by requiring the "no-puzzle" constraints to hold for all bonds, this requirement can be relaxed as was mentioned before. The methodology can be iterated, as explained in the paper, in a search for an estimate that induces the absence of the "puzzle" for a statistically significant number of bonds. We have shown that the prices of implicit options obtained via the regression methodology and via our implied valuation approach differ drastically. Moreover, it was demonstrated that term structure estimates obtained by the two approaches differ. This latter result should be taken into consideration when trying to estimate the stochastic process assumed to be followed by interest rates. The parameters of the stochastic processes assumed to govern the term structure of interest rates are sensitive to the choice of term structure used as an initial condition. It would be interesting to see how sensitive the parameters actually are and what implications that sensitivity has for pricing interest-rate contingent securities.

Table 1: R^2 of Implied and Constrained Regression Estimation

April 1970		October 1974		June 1979		October 1982	
Implied	Regression	Implied	Regression	Implied	Regression	Implied	Regression
0.839904	0.915205	0.748746	0.862462	0.785373	0.786441	0.784196	0.820926

This table shows the R^2 for each of the four months of reported results for two estimation methods: the one proposed in this paper, "Implied", and the constrained regression method, "Regression". For each month, the R^2 is higher for the regression method. As explained in the text of the paper, this does not imply that the regression estimation method should be preferred. Rather, one should be wary of judging the usefulness of an estimation by its measured R^2.

Table 2: Estimated Option Prices

This table reports estimated option prices. The term structure is estimated using standard constrained regression methods and using the implied approach advocated in this paper. Each term structure is used to calculate the price estimate for the option implicit in the bond. The first column of each panel of the table reports the observed prices for the extendible Government of Canada bond. The final two columns report the maturity date and coupon rate for that bond. The two middle columns report the prices, each based on a different estimation technique. In each case, the estimate of the term structure is used to calculate the price of the cash flows from the extendible bond as if it were a straight bond. The option price is then the difference between the observed price for the extendible bond and its estimated price as a straight bond.

April 29, 1970

Price	Implied OP	Regression OP	Maturity	Coupon (%)
99.325	2.497266	0.527996	01/04/71	6.00
96.875	4.414840	-0.085988	01/12/73	6.25
99.938	2.073376	-2.733511	01/04/74	7.25
102.688	0.000011	-5.095273	01/10/74	8.00

On the observation day in April 1970, there were four extendible Government of Canada bonds outstanding. Of those, for three, the implicit option had negative value when its price was estimated using regression techniques. All four options had positive value under the methodology outlined in this paper.

October 30, 1974

Price	Implied OP	Regression OP	Maturity	Coupon
99.650	2.772280	-0.416941	15/12/75	7.25
97.700	4.567043	0.145574	01/08/76	6.25
105.700	2.663632	-2.928867	01/02/77	9.25
98.450	3.874862	-1.875529	01/07/77	7.00
105.750	0.000489	-6.782067	01/04/78	9.25
99.750	1.259442	-5.255497	01/10/78	7.75
97.375	3.130598	-3.066244	01/04/79	7.00
98.438	0.535899	-4.641652	01/12/80	7.50

On the observation day in October 1974, there were eight extendible Government of Canada bonds outstanding. Of those, for seven, the implicit option had negative value when its price was estimated using regression techniques. All eight options had positive value under the methodology outlined in this paper.

Table 2 (continued)

June 27, 1979

Price	Implied OP	Regression OP	Maturity	Coupon
99.075	0.147272	0.144206	01/10/79	7.5
99.650	0.783631	0.746559	01/10/80	9.0
96.525	1.400280	1.355781	01/12/80	7.5

On the observation day in June 1979, there were three extendible Government of Canada bonds outstanding. For this observation date, all three options had positive value under the methodology outlined in this paper.

October 24, 1982

Price	Implied OP	Regression OP	Maturity	Coupon
104.450	4.296384	2.039846	01/10/84	12.50
106.375	3.634266	0.890062	01/02/85	13.25
108.375	3.815600	0.892125	15/03/85	13.75
106.125	4.085419	1.041816	01/05/85	13.00
108.125	4.665093	2.597912	01/08/84	13.75
104.375	3.594117	-0.205556	01/02/86	12.50
112.125	2.370706	-1.865110	01/05/86	14.50
116.500	3.121760	-1.276190	01/06/86	15.25
114.125	2.725184	-1.676541	01/07/86	14.75
128.250	0.570364	-4.462150	01/10/86	18.00
116.875	0.196529	-4.700897	01/02/87	15.50
115.375	0.267166	-4.724403	01/07/87	15.00
111.750	0.322967	-4.577033	01/09/87	14.25

On the observation day in October 1982, there were thirteen extendible Government of Canada bonds outstanding. Of those, for eight, the implicit option had negative value when its price was estimated using regression techniques. All thirteen options had positive value under the methodology outlined in this paper.

Table 3: Estimated Coefficients of the Discount Function

	April 1970		October 1974	
	Implied	Regression	Implied	Regression
α_1	6.478783	5.436676	6.345445	5.260758
α_2	18.386258	30.159767	0.000010	0.000010
α_3	0.000041	0.960837	111.916638	258.265788
α_4	275.372905	303.119312	260.097165	0.000010
α_5	81.269459	0.000010	0.000010	0.000010
α_6	0.000113	0.000010	0.000010	0.000010
α_7	0.000354	0.000010	0.000010	0.000010
α_8	0.000022	0.000010	0.000010	0.000010
α_9	0.000405	0.000010	0.000100	0.000010
α_{10}	0.008071	0.000010	0.000379	0.000010
R^2	0.839904	0.915205	0.748746	0.862462
	June 1979		October 1982	
	Implied	Regression	Implied	Regression
α_1	7.040361	7.038074	7.854809	7.656564
α_2	5.140986	4.570964	15.539234	2.720156
α_3	115.165835	116.120348	0.000010	138.213868
α_4	106.353845	110.925055	190.950527	0.000010
α_5	0.000011	0.000010	0.000010	0.000010
α_6	0.000011	0.000010	0.000011	0.000010
α_7	0.000162	0.000010	0.000119	0.000010
α_8	0.000011	0.000010	0.000012	0.000010
α_9	0.000484	0.000010	0.000108	0.000010
α_{10}	0.000060	0.000010	0.000101	0.000000
R^2	0.785373	0.786441	0.784196	0.820926

This table presents the estimated coefficients of the discount function and the associated R^2 measure for each of two estimation methods: constrained regression (Regression) and the implied valuation approach developed and advocated in this paper. There are ten estimated coefficients (since n was chosen to be 11 and the first coefficient is set equal to 1). These are the estimated coefficients of (10) based on the choice made in (4.3) for the approximating functions. The discount functions for both estimation methods for these four observation dates are plotted in Figures 10 through 13.

Table 4: The L^2 Norm between Regression and Implied Discount Functions

$$\left[\int_0^1 \left\{d^*_{sq}(t) - d^*_{Rgr}(t)\right\}^2 dt\right]^{1/2},$$

where d^*_{sq} is the implied discount function and d^*_{Rgr} is obtained from regression.

	April 1970	October 1974	June 1979	October 1982	Average all Runs†
L^2	0.084251	0.102627	0.026640	0.064277	0.086600

† The estimation technique was run for April and October of every year from 1968 through 1985.

This table shows the L^2 distance between the regression-estimated discount function and the discount function estimated via our implied technique. Figures 9 through 12 show even more clearly the distance between the two estimation methods.

Appendix 1: Kuhn-Tucker Conditions for the Term Structure Estimation Problem

The Kuhn-Tucker (KT) conditions for an optimization problem

$$\begin{aligned} \max \quad & f(x) \\ s.t. \quad & g_i(x) \geq 0 \quad i = 1, ..., m_1 \\ & g_i(x) = 0 \quad i = m_1 + 1, ..., m \end{aligned}$$

can be stated as follows:

$$x^* \text{ is feasible,}$$

there exist multipliers $\lambda_i \geq 0$, $i = 1, ..., m_1$, and unconstrained multipliers λ_i $i = m_1 + 1, ..., m$, such that

$$\lambda_i g_i(x^*) = 0 \quad i = 1, ..., m_1$$

and

$$\bigtriangledown f(x^*) + \sum_{i=1}^{m} \lambda_i \bigtriangledown g_i(x^*) = 0.$$

Our problem is

$$\begin{aligned} & \min_{\epsilon} \quad \epsilon' \Gamma \epsilon \\ & Ad + \epsilon = P \\ & d \geq \mathbf{0}, \end{aligned}$$

or equivalently

$$\begin{aligned} & -\max_{\epsilon} \quad -\epsilon' \Gamma \epsilon \\ & Ad + \epsilon = P \\ & d \geq \mathbf{0} \end{aligned}$$

Then KT conditions for this problem can be written as follows:

$$\begin{aligned} Ad + \epsilon &= P \\ d &\geq \mathbf{0} \\ \beta &\geq \mathbf{0} \\ \beta_j d_j &= 0 \quad \forall j \\ -2\epsilon_i \gamma_i + \lambda_i &= 0 \quad \forall i \\ \sum_i \lambda_i a_{ij} + \beta_j &= 0 \quad \forall j, \end{aligned}$$

where λ and β are the Lagrange multipliers corresponding to equality and inequality constraints respectively. Equivalently, the Kuhn-Tucker conditions can be written as

$$\begin{aligned}
\sum_i 2(P-Ad)_i\gamma_i a_{ij} + \beta_j &= 0 \quad \forall j \\
\beta_j d_j &= 0 \quad \forall j \\
d &\geq \mathbf{0} \\
\beta &\geq \mathbf{0},
\end{aligned}$$

since $\epsilon = P - Ad$ and $-2\epsilon_i\gamma_i + \lambda_i = 0 \quad \forall i$ and both ϵ and λ are unconstrained. Notice also that since $\beta_j = -\sum_i 2(P-Ad)_i\gamma_i a_{ij}$ for all j the conditions can be simplified further to be:

$$\begin{aligned}
\left\{-\sum_i 2(P-Ad)_i\gamma_i a_{ij}\right\} d_j &= 0 \quad \forall j \\
\sum_i 2(P-Ad)_i\gamma_i a_{ij} &\leq 0 \quad \forall j \\
d &\geq \mathbf{0}
\end{aligned}$$

In a matrix form we can write it as

$$\begin{aligned}
-2\ (P-Ad)'\ \Gamma\ A\ diag(\ d\ ) &= \mathbf{0} \\
2\ (P-Ad)'\ \Gamma\ A &\leq \mathbf{0} \\
d &\geq \mathbf{0},
\end{aligned}$$

where

$$diag(\ d\ ) = \begin{bmatrix} d_1 & 0 & 0 & \dots & 0 \\ 0 & d_2 & 0 & \dots & 0 \\ 0 & 0 & d_3 & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & 0 & \dots & d_J \end{bmatrix}.$$

Appendix 2: Kuhn-Tucker Condition with Continuos Approximation

The problem of the estimation of the term structure with continuous approximation is

$$\begin{aligned}
&\min_{\epsilon} \quad \epsilon'\Gamma\epsilon \\
&\tilde{A}\alpha + \epsilon = P \\
&\alpha'\theta(1) \geq 0 \\
&\alpha \geq \mathbf{0}
\end{aligned}$$

or equivalently

$$
\begin{aligned}
&-\max_{\epsilon} \quad -\epsilon'\Gamma\epsilon \\
&\alpha \geq \mathbf{0} \\
&\alpha'\theta(1) \geq 0 \\
&\tilde{A}\alpha + \epsilon = P
\end{aligned}
$$

The KT conditions to this problem can be written as

$$
\begin{aligned}
\tilde{A}\alpha + \epsilon &= P \\
\alpha &\geq \mathbf{0} \\
\alpha'\varphi(1) &\geq 0 \\
\beta &\geq \mathbf{0} \\
\beta_k \alpha_k &= 0 \quad \forall k \\
\xi &\geq 0 \\
\xi \alpha'\varphi(1) &= 0 \\
-2\epsilon_i \gamma_i + \lambda_i &= 0 \quad \forall i \\
\sum_i \lambda_i \tilde{a}_{ik} + \xi \varphi_k(1) + \beta_k &= 0 \quad \forall k,
\end{aligned}
$$

where λ and β are the Lagrange multipliers corresponding to equality and inequality constraints respectively and ξ is a Lagrange multiplier for the non-negativity constraint of the discount function. By an argument similar to the one in Appendix 1, these KT conditions can be equivalently written as

$$
\begin{aligned}
\alpha_k \times (2\sum_i (P - \tilde{A}\alpha)_i \gamma_i \tilde{a}_{ik} + \xi\varphi_k(1)) &= 0 \quad \forall k \\
2\sum_i (P - \tilde{A}\alpha)_i \gamma_i \tilde{a}_{ik} + \xi\varphi_k(1) &\leq 0 \quad \forall k
\end{aligned}
$$

$$
\begin{aligned}
\xi\alpha'\varphi(1) &= 0, \quad \alpha'\varphi(1) \geq 0 \\
\alpha &\geq \mathbf{0}, \quad \xi \geq 0.
\end{aligned}
$$

In a matrix notation these conditions can be written as

$$
\begin{aligned}
(2\ (P - \tilde{A}\alpha)'\ \Gamma\ \tilde{A} + \xi\ \varphi(1)\ )\ I_\alpha &= \mathbf{0} \\
2\ (P - \tilde{A}\alpha)'\ \Gamma\ \tilde{A} + \xi\ \varphi(1) &\leq \mathbf{0} \\
\xi\alpha'\varphi(1) &= 0
\end{aligned}
$$

$$
\begin{aligned}
-\alpha'\varphi(1) &\leq 0 \\
\alpha &\geq \mathbf{0} \\
\gamma &> \mathbf{0} \\
\xi &\geq 0
\end{aligned}
$$

where

$$
I_\alpha = \begin{bmatrix} \alpha_0 & 0 & 0 & \dots & 0 \\ 0 & \alpha_1 & 0 & \dots & 0 \\ 0 & 0 & \alpha_2 & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & 0 & \dots & \alpha_n \end{bmatrix}.
$$

References

[1] Aït-Sahalia, Yacine and Andrew W. Lo, 1998, "Non-parametric Estimation of State-Price Densities Implicit in Financial Asset Prices", *Journal of Finance,* 53, 2, 499-547.

[2] Athanassakos, George, Peter Carayannopoulos and Yisong Tian, 1994, "Negative Option Values in Extendible Canadian Treasury Bonds", *Advances in Futures and Options Research*, 97 (9), 83-110.

[3] Boyle, Phelim R., 1996, "Dynamic Asset Allocation", presentation to The Fields Institute Seminar on Financial Mathematics, November 27, 1996, Toronto, Canada.

[4] Chambers, Donald R., Willard T. Carleton and Donald W. Waldman, 1984, "Estimation of the Term Structure of Interest Rates Using a Simple Polynomial", *Journal of Financial and Quantitative Analysis*, 19(3), 233–252.

[5] Chen, Zhiwu and Peter J. Knez, 1994, "A Pricing Operator-Bowed Testing Foundation for a Class of Factor Pricing Models", *Mathematical Finance*, 4, 2, 121–141.

[6] Dermody, J.C. and Eliezer Z. Prisman, 1988, "Term Structure Multiplicity and Clientele in Markets with Transaction Costs and Taxes", *Journal of Finance*, 43, No. 4, 893-911.

[7] Dermody, J. C. and T. R. Rockafellar, 1991, "Cash Stream Valuation in the Face of Transaction Costs and Taxes", *Mathematical Finance*, 1, 1, 31–54.

[8] Duffie, Darrell, 1992, *Dynamic Asset Pricing Theory*, Princeton University Press, New Jersey.

[9] Elton, E.J. and T.C. Green, 1998, "Tax and Liquidity Effects in Pricing Government Bonds", *Journal of Finance*, 53, No. 5, 1533-1562

[10] Hansen, Lars Peter and Ravi Jagannathan, 1997, "Assessing Specification Errors in Stochastic Discount Factor Models", *Journal of Finance*, 52, 2, 557–590.

[11] Heath, D., R. Jarrow and A. Morton, 1992, "Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation", *Econometrica*, 60, 1, 77-105.

[12] Ho, Thomas S.Y., Sang-Bin Lee and Kyu-Hyun Son, 1992, "Term Structure Estimation and Pricing of Callable Treasury Bonds", *Review of Quantitative Finance and Accounting*, 2, 127-143.

[13] Ioffe, Ioulia D., 1998, "Arbitrage Bounds in Markets with Noisy Prices and the Puzzle of Negative Option Prices Implicit in Bonds", working paper.

[14] Itzikowitz, Shmuel, Alexandra E. MacKay and Eliezer Z. Prisman, 1996, "Negative Option Values Implicit in the Israeli Bond Market", *Bank of Israel Banking Review*, 5, 44-58.

[15] Jackwerth, Jens Carsten and Mark Rubinstein, 1996, "Recovering Probability Distributions from Option Prices", *Journal of Finance*, 5, 51, 1611-1631.

[16] Jordan, Bradford D., S.D. Jordan and R.D.Jorgensen, 1995, "A Reexamination of Option Values Implicit in Callable Treasury Bonds", *Journal of Financial Economics,* 38, 2, 141-162.

[17] Jordan, Bradford D. and David R. Kuipers, 1997, "Negative Option Values are Possible: The Impact of Treasury Bond Futures on the Cash US Treasury Market", *Journal of Financial Economics* , 46, 1, 29-65.

[18] Jordan, James V., 1984, "Tax Effects in Term Structure Estimation", *Journal of Finance*, 39, 2, 393–406.

[19] Katz, Eliakim. and Eliezer Z. Prisman, 1991,"Clientele Effects, Arbitrage and Term Structure Estimations", *Journal of Financial and Quantitative Analysis*, 26, 435-443.

[20] Longstaff, Francis A., 1992, "Are Negative Option Prices Possible? The Callable U.S. Treasury-Bond Puzzle", *Journal of Business*, 65, 4, 571–592.

[21] MacKay, Alexandra E. and Eliezer Z. Prisman, "Estimating Valuation Operators in Incomplete Market with Noises: Can Noise Complete the Market", presented at the AFA meetings, 1996.

[22] MacKay, Alexandra E., Eliezer Z. Prisman and Yisong Tian, , 2000, "Non-Segmented Equilibria Under Differential Taxation: Evidence from the Canadian Government Bond Market", working paper.

[23] Phillips, G.M. and P.J. Taylor, 1970, "Approximation of Convex Data", *BIT*, 10, 324-332.

[24] Prisman, Eliezer Z., 1990, "A Unified Approach to Term Structure Estimation: A Methodology for Estimating the Term Structure in a Market with Frictions", *Journal of Financial and Quantitative Analysis*, 25, 1, 127–142.

[25] Prisman, Eliezer Z. and Y. Tian, 1994, "Immunization in Markets with Tax Clientele Effect: Evidence from the Canadian Market", *Journal of Financial and Quantitative Analysis*, 29, 301-321.

[26] Rockafellar, R. T., 1970, *Convex Analysis*, Cambridge, MA: Princeton University Press.

[27] Rubinstein, Mark, 1994, "Implied Binomial Tree", *Journal of Finance,* 49, No. 3, 771-818

[28] Schaefer, S.M., 1981, "Measuring a Tax-Specific Term Structure of Interest Rates in the Market for British Government Securities", *The Economic Journal*, 91, 415-438.

[29] *Summary of Government of Canada Direct Securities and Loans*, 1993, Bank of Canada.

[30] Tuckman, Bruce, 1996, *Fixed Income Securities: Tools for Today's Markets*, John Wiley & Sons, Inc., New York.

[31] Zangwill, Willard I., 1969, *Nonlinear Programming: A Unified Approach,* Prentice-Hall, Inc., Englewood Cliffs, N.J.

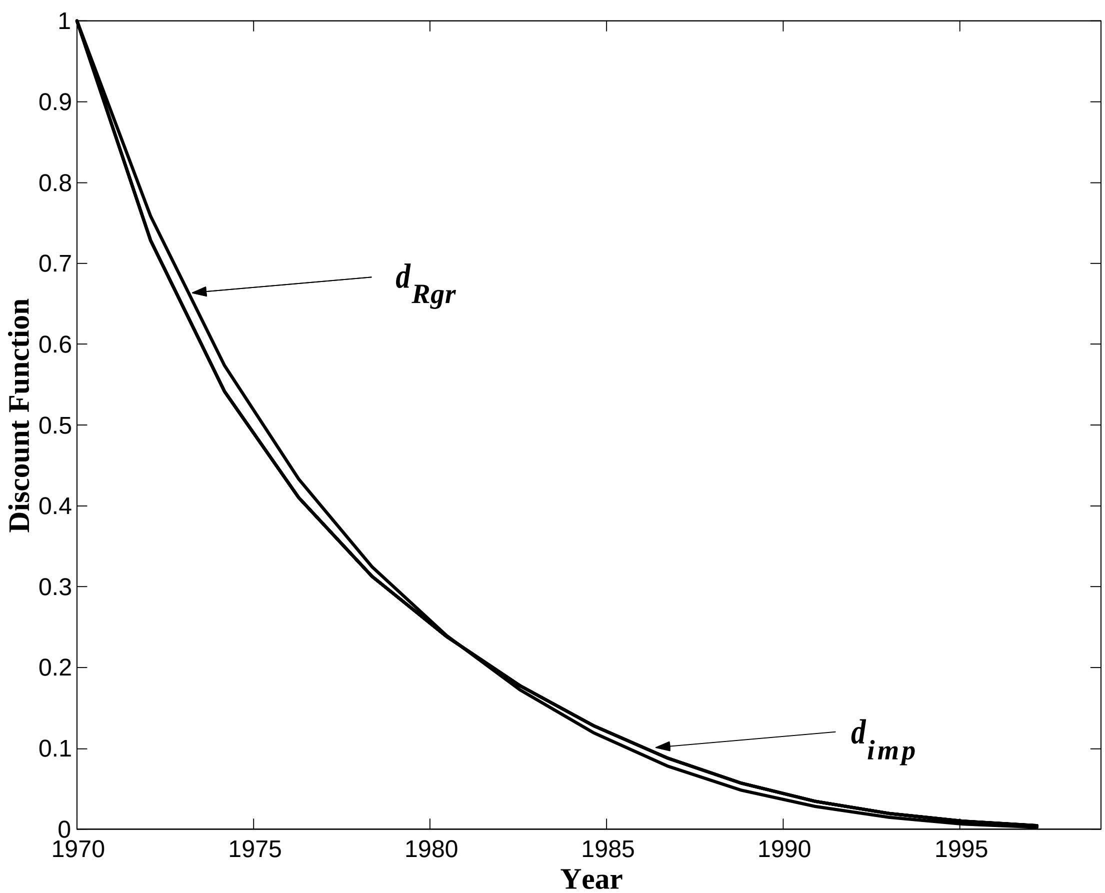

Figure 1: The Discount Functions on April 29, 1970

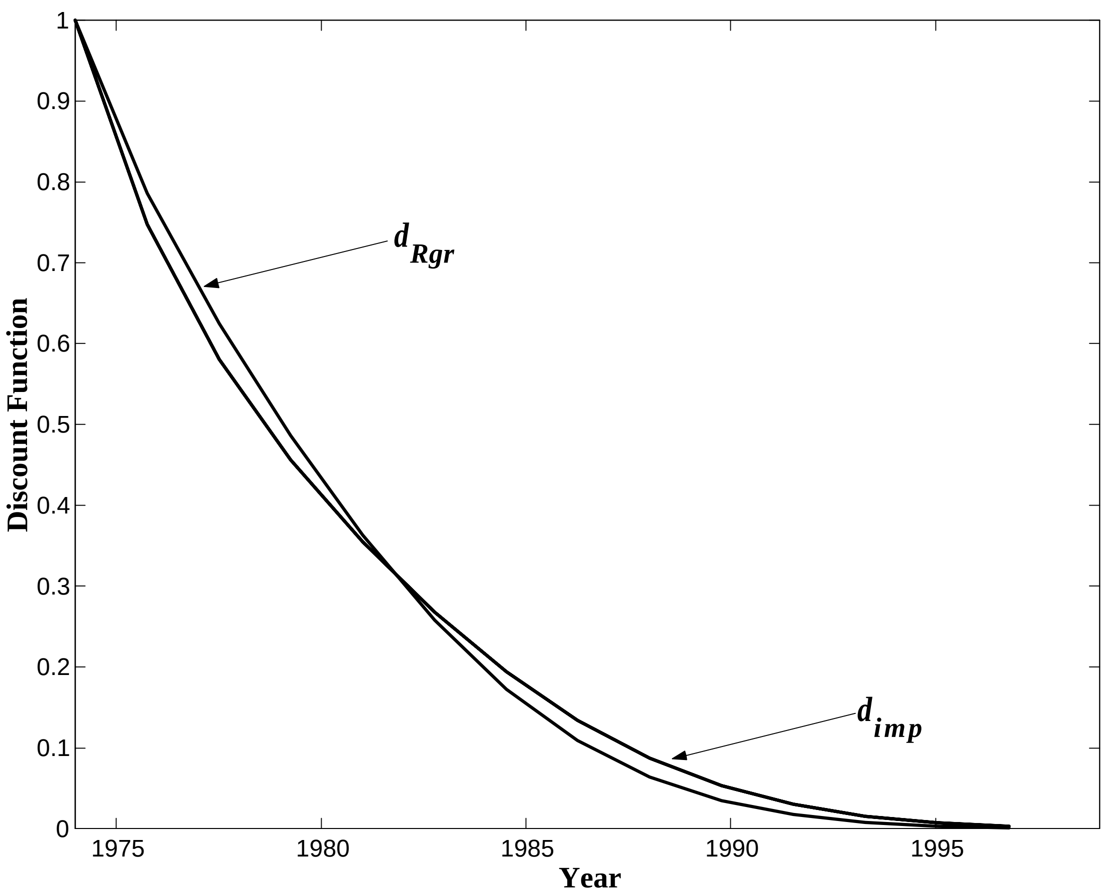

Figure 2: The Discount Functions on October 30, 1974

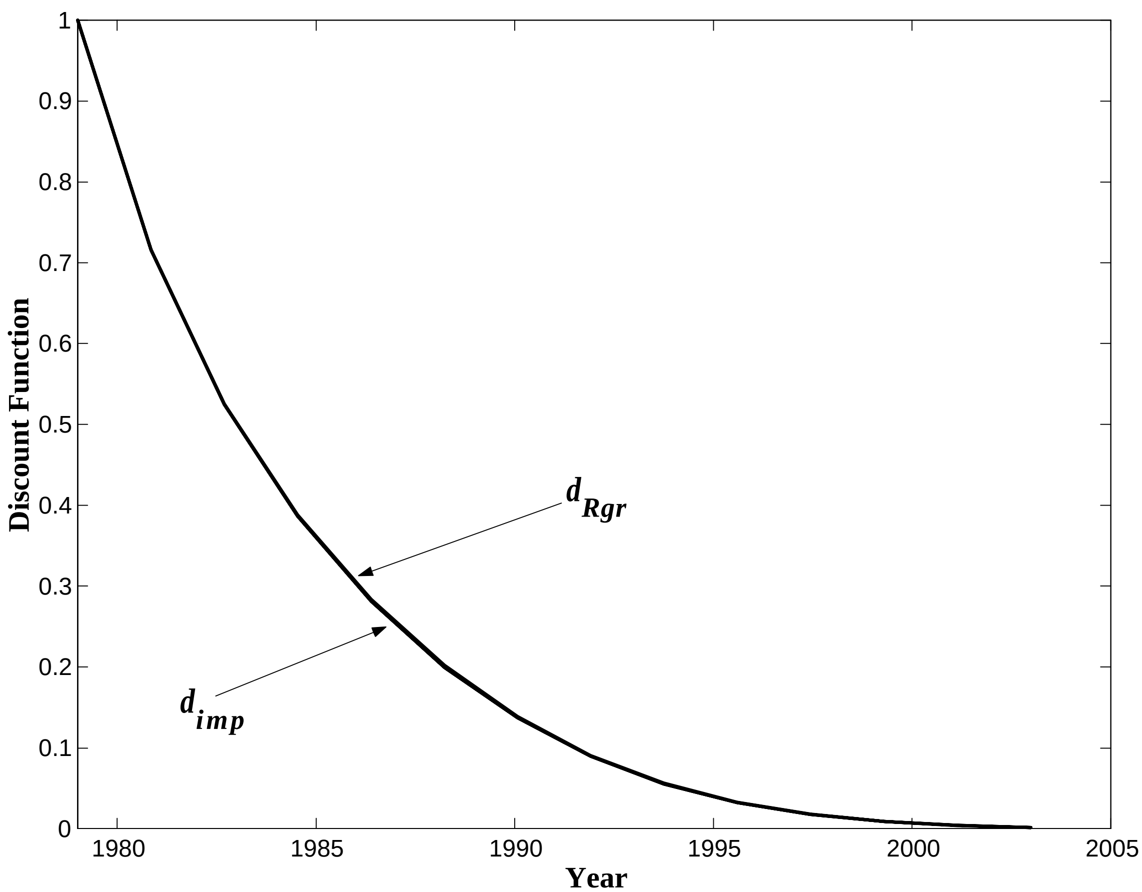

Figure 3: The Discount Functions on June 27, 1979

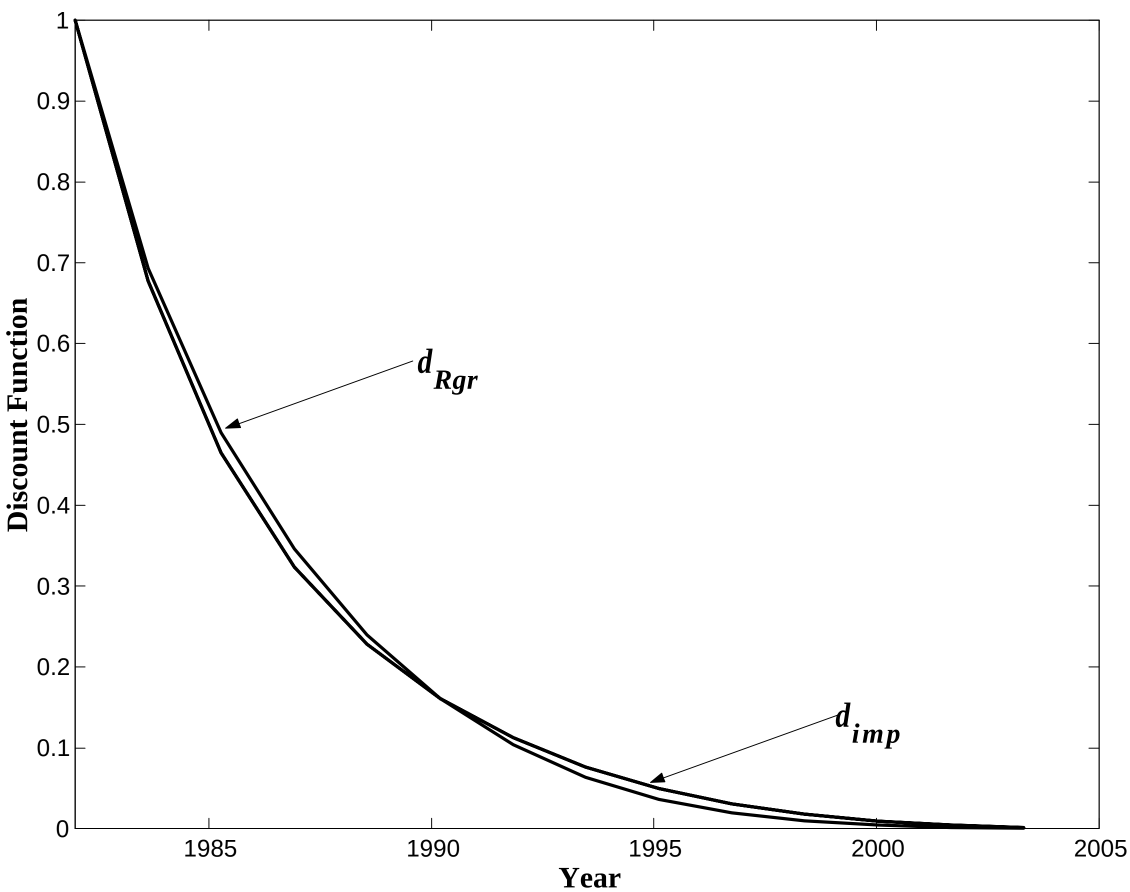

Figure 4: The Discount Functions on October 24, 1982

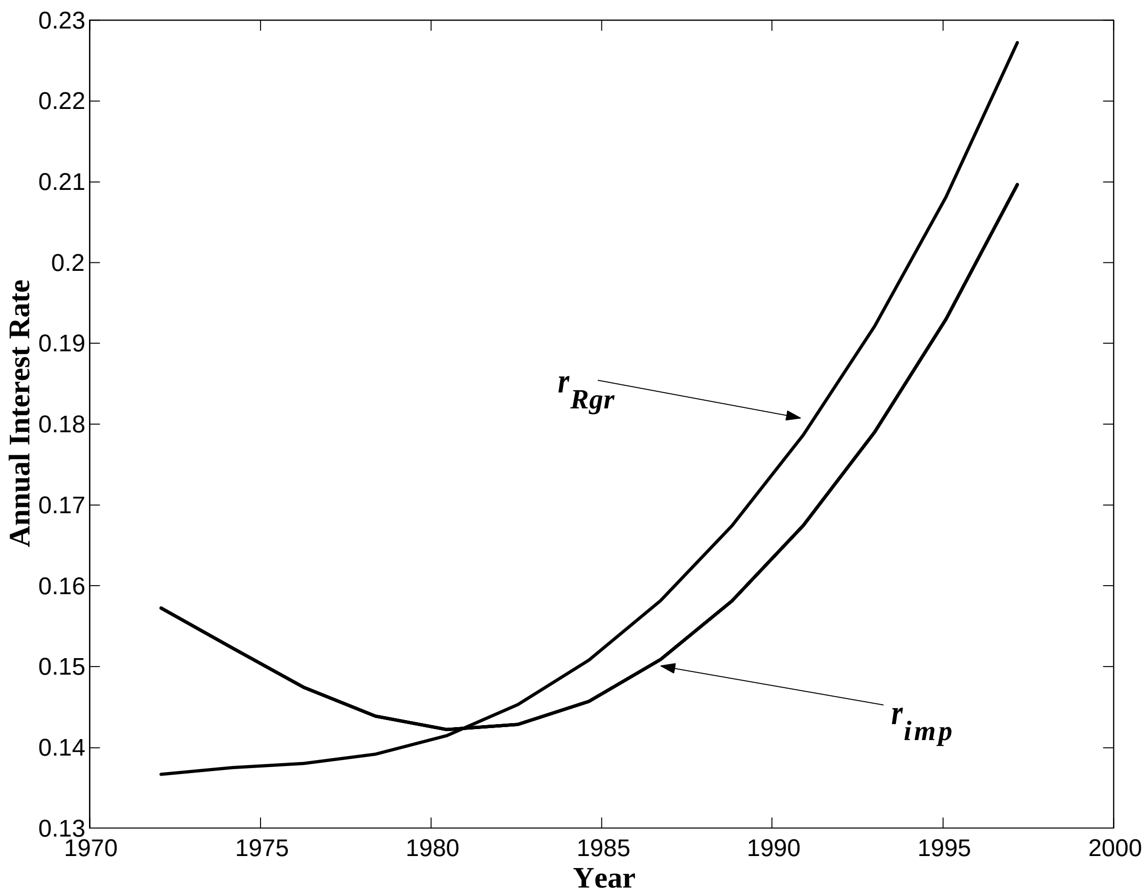

Figure 5: The Term Structure of Interest Rates on April 29, 1970

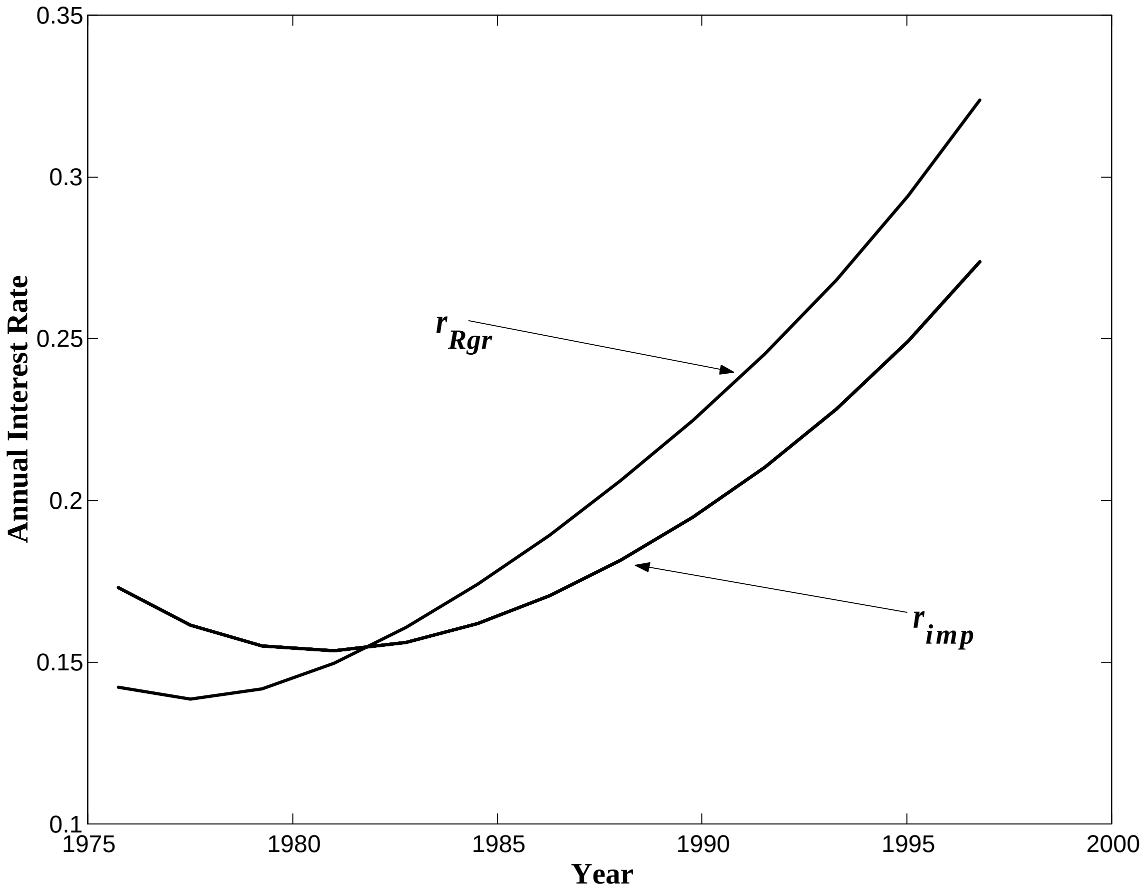

Figure 6: The Term Structure of Interest Rates on October 30, 1974

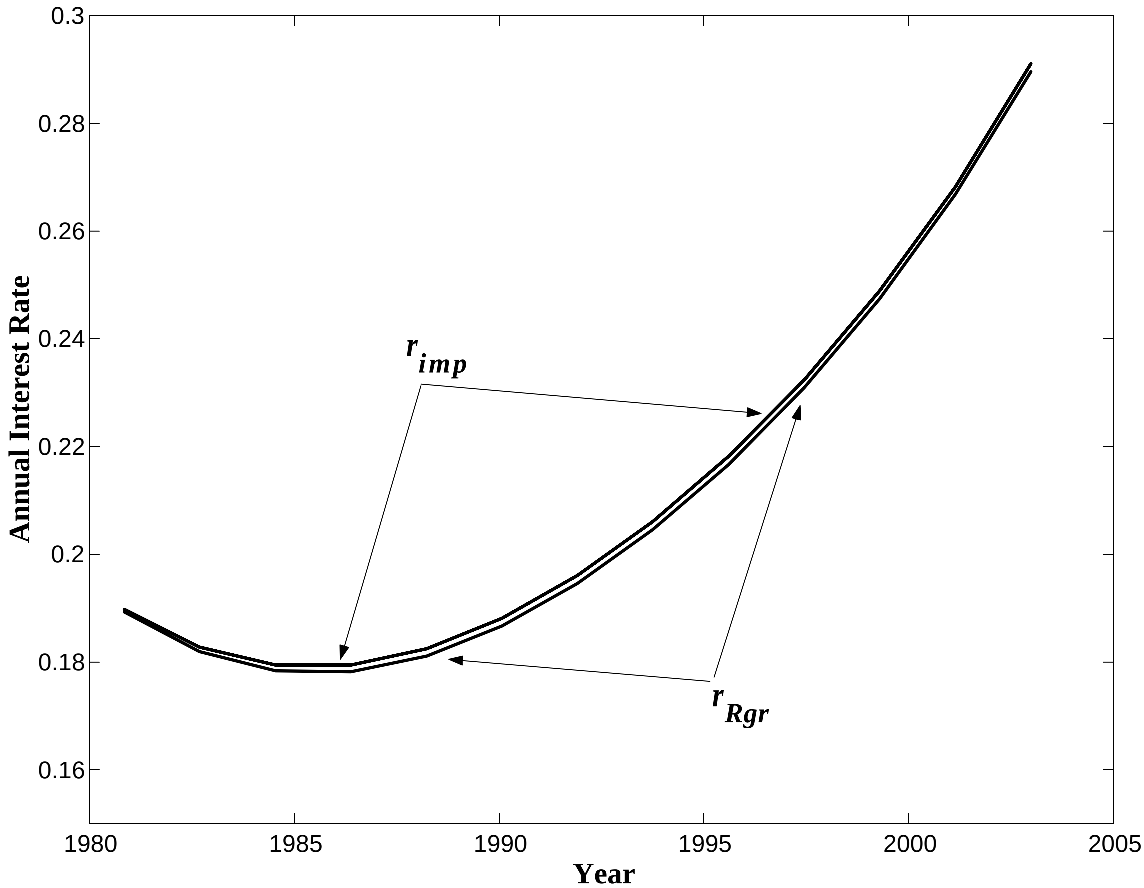

Figure 7: The Term Structure of Interest Rates on June 27, 1979

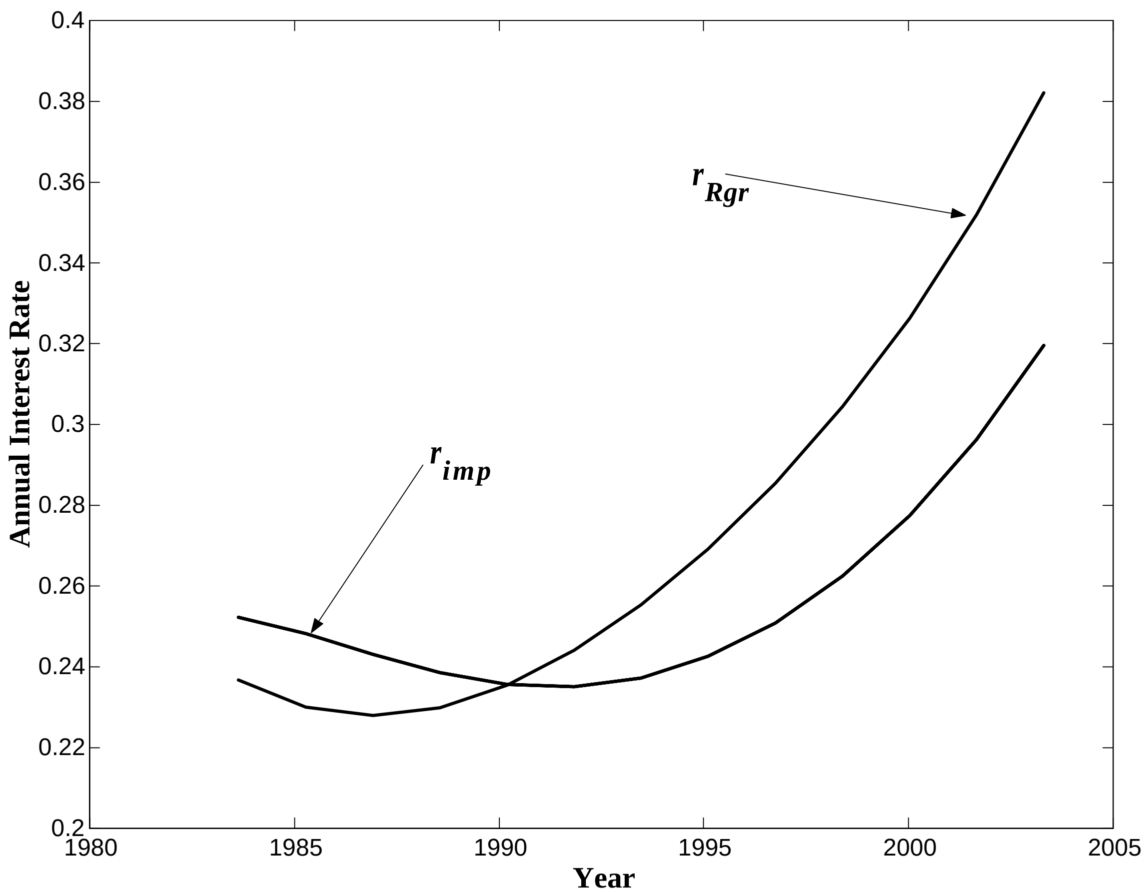

Figure 8: The Term Structure of Interest Rates on October 24, 1982

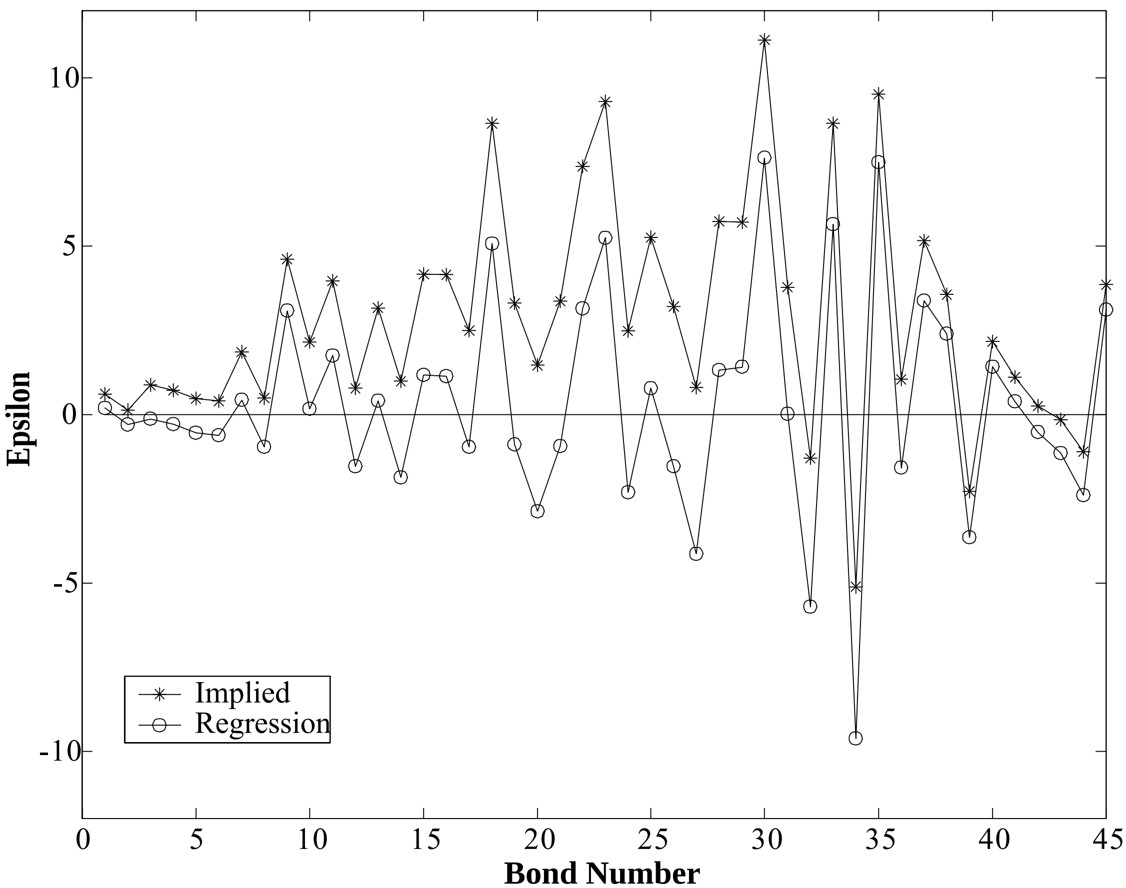

Figure 9: This figure shows the difference between the observed market price and the fitted price for each bond outstanding on April 29, 1970 for each of two estimation methods.

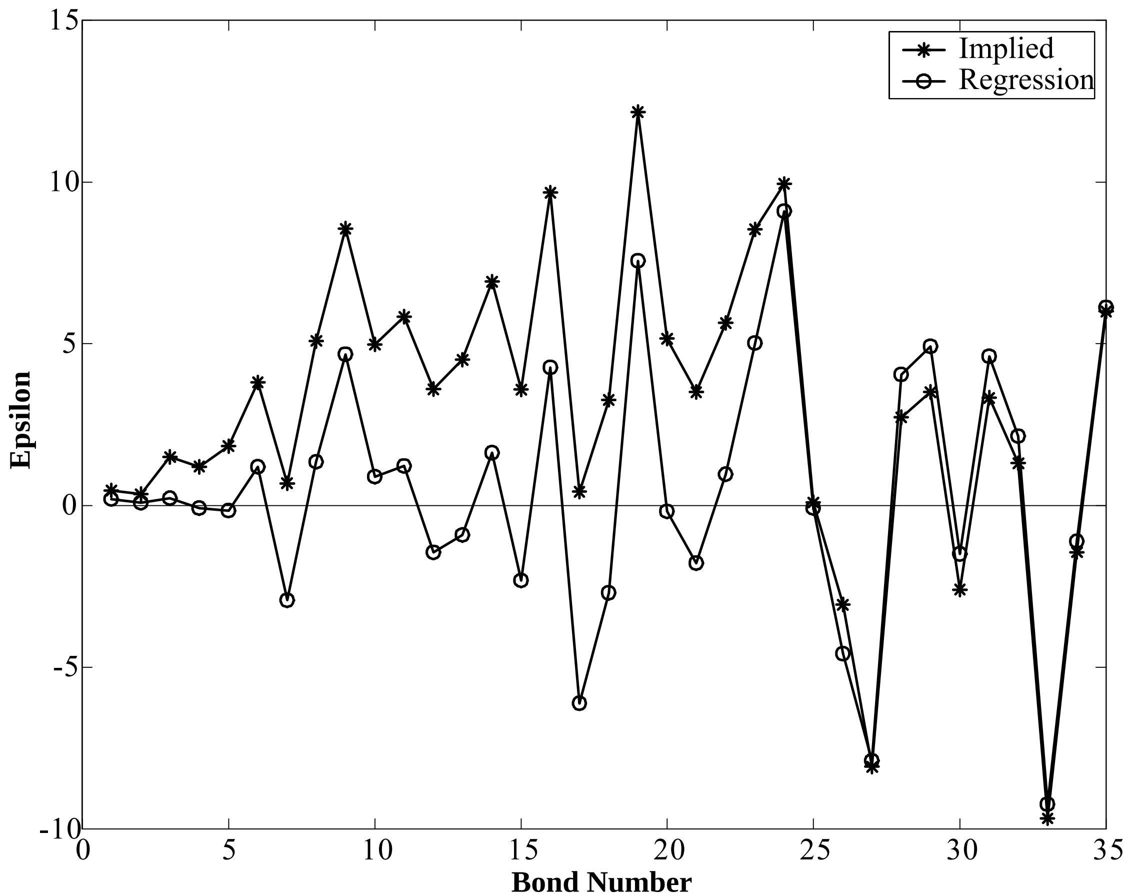

Figure 10: This figure shows the difference between the observed market price and the fitted price for each bond outstanding on October 30, 1974 for each of two estimation methods.

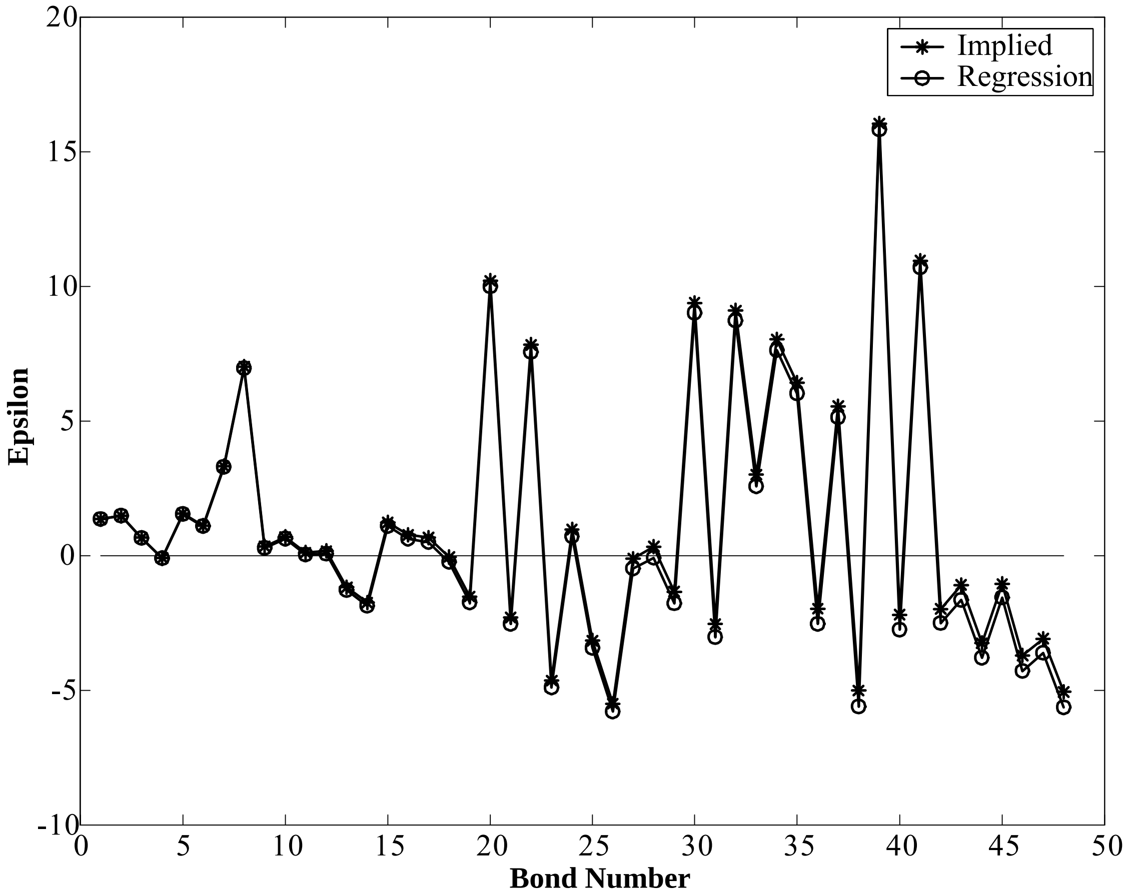

Figure 11: This figure shows the difference between the observed market price and the fitted price for each bond outstanding on June 27, 1979 for each of two estimation methods.

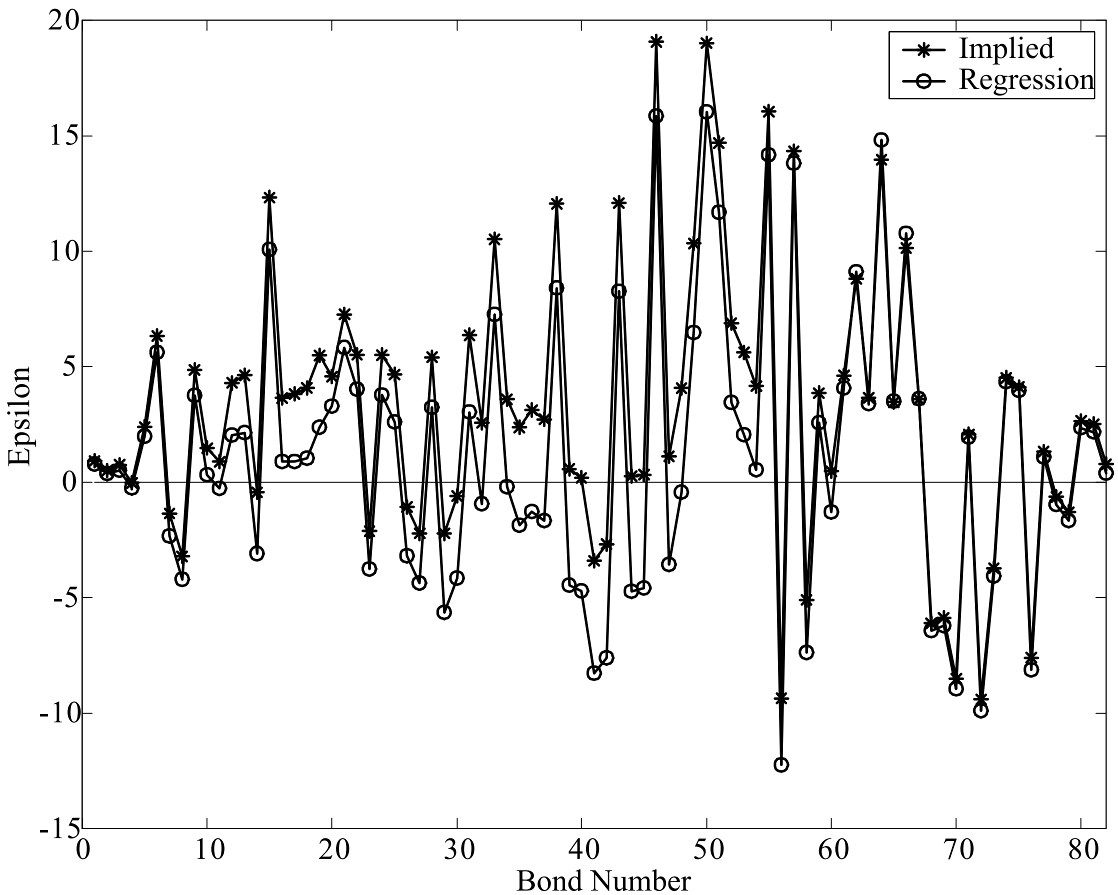

Figure 12: This figure shows the difference between the observed market price and the fitted price for each bond outstanding on October 24, 1982 for each of two estimation methods.

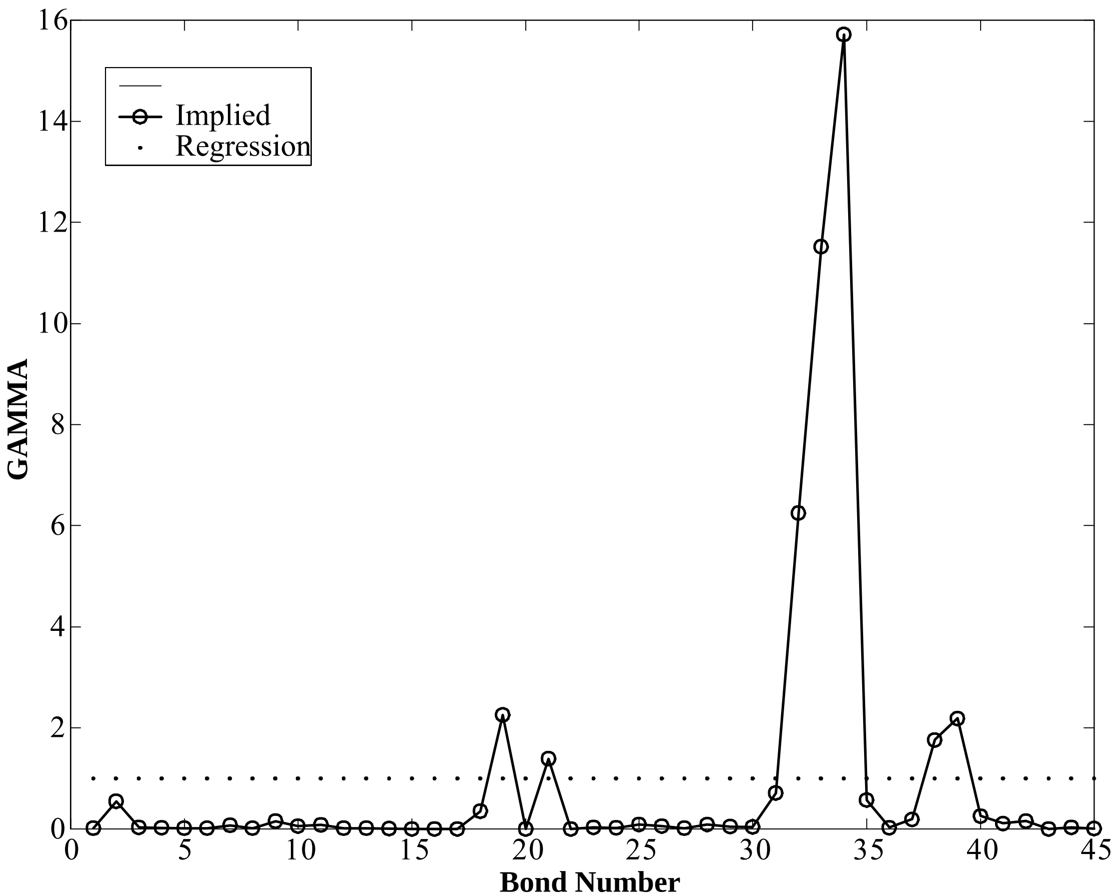

Figure 13: In this figure the diagonal elements of the matrix Γ for the estimation via our implied methodology are plotted for April 29, 1970. These are represented by the circles. The horizontal line of dots at 1 are the corresponding elements of the identity matrix. There are as many diagonal elements as there are bonds outstanding.

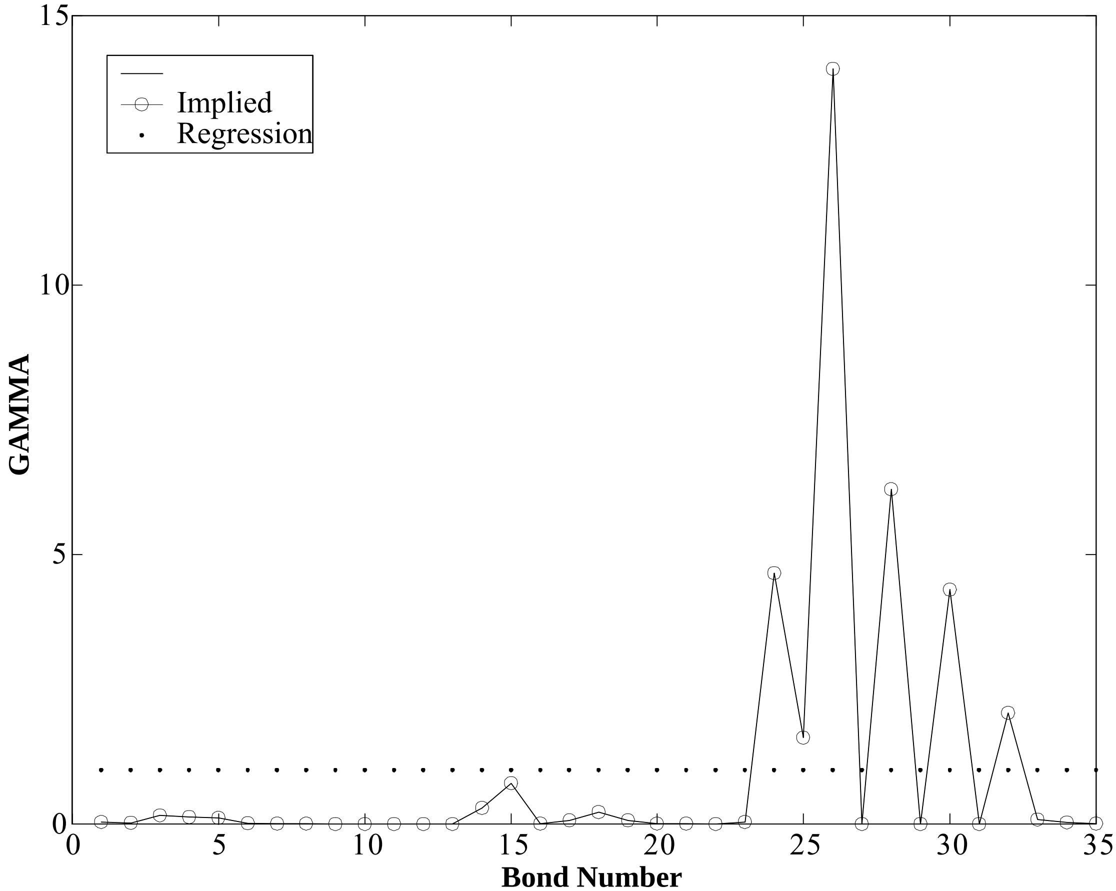

Figure 14: In this figure the diagonal elements of the matrix Γ for the estimation via our implied methodology are plotted for October 30, 1974. These are represented by the circles. The horizontal line of dots at 1 are the corresponding elements of the identity matrix. There are as many diagonal elements as there are bonds outstanding.

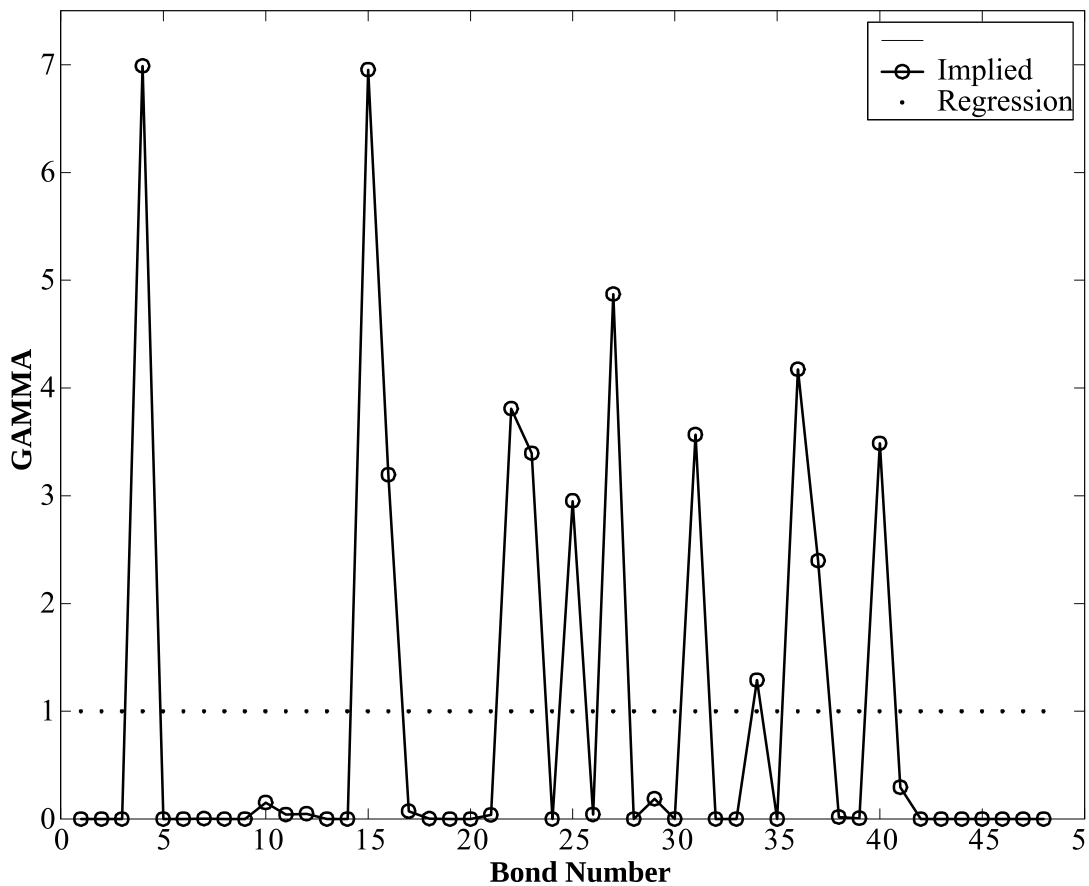

Figure 15: In this figure the diagonal elements of the matrix Γ for the estimation via our implied methodology are plotted for June 27, 1979. These are represented by the circles. The horizontal line of dots at 1 are the corresponding elements of the identity matrix. There are as many diagonal elements as there are bonds outstanding.

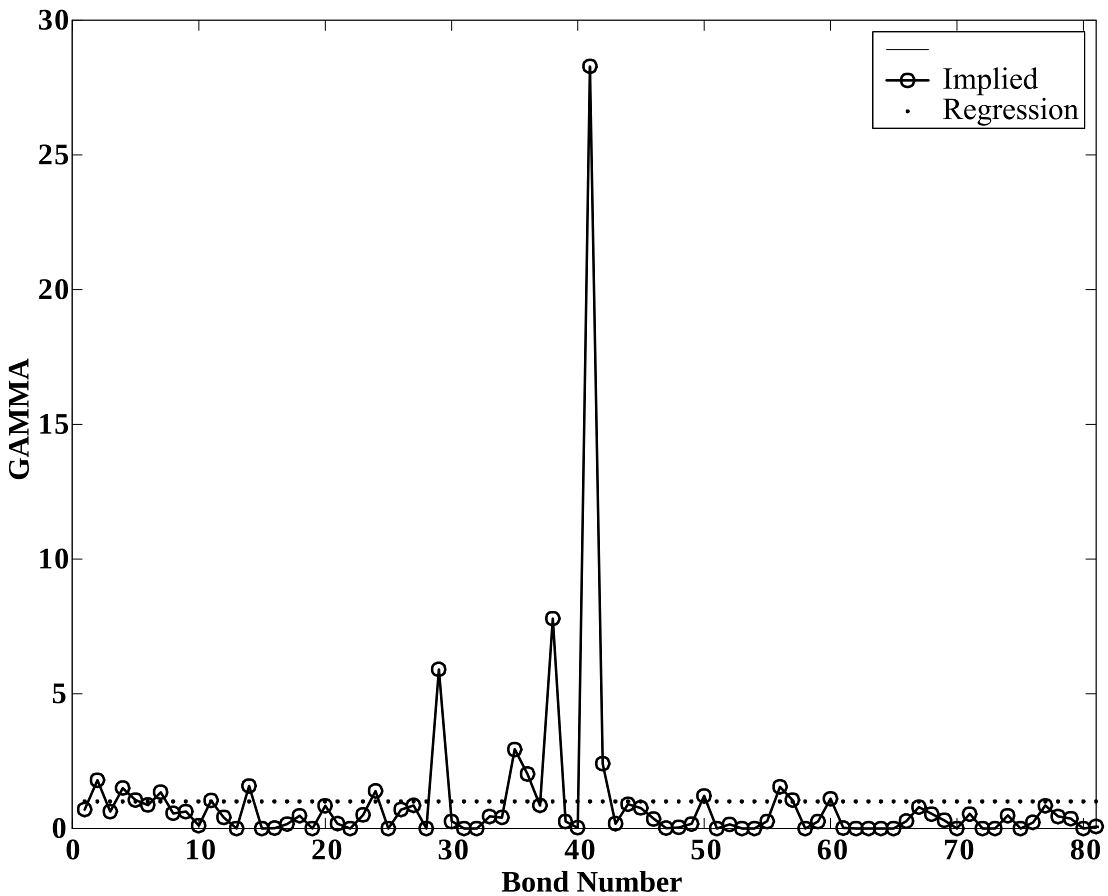

Figure 16: In this figure the diagonal elements of the matrix Γ for the estimation via our implied methodology are plotted for October 24, 1982. These are represented by the circles. The horizontal line of dots at 1 are the corresponding elements of the identity matrix. There are as many diagonal elements as there are bonds outstanding.