Appeared in Proc. of First Int. Conf. on High Frequency Data in Finance, Zürich, March, 1995

Self-similarity of high-frequency USD-DEM exchange rates

Carl J.G. Evertsz

Center for Complex Systems and Visualization, University of Bremen
FB III, Bibliothekstrasse 1, D-28359 Bremen, Germany

Abstract – High frequency DEM-USD exchange rate data (resolution $>$ 2 seconds) are analyzed for their scaling behavior as a function of the time lag. Motivated by the finding that the distribution of 1-quote returns is rather insensitive to the physical time duration between successive quotes, lags are measured in units of quotes. The mean absolute returns over lags of different sizes, shows three different regimes. The smallest time scales show no scaling, followed by two scaling regimes characterized by Hurst exponents $H = 0.45$ and $H = 0.56$, with a crossover occuring at lags of ≈ 500 quotes. The up-down correlation coefficient, defined here, shows strong anti-correlations on scales smaller than 500. The lack of convergence to a large deviation rate function, convex tails in the logarithm of the probability distributions, strong up-down correlations and $H < 0.5$, show that the dynamics on small scales is more complicated than random walk models with i.i.d. increments. Nevertheless, for both scaling regimes there is a very high degree of distributional self-similarity. For the $H = 0.56$ this self-similarity satisfies the same scaling rules as those for stable distributions with characteristic exponent $\alpha = 1/H$. The large deviation analysis shows that the probabilities for large returns (negative or positive) decays less fast than exponentially as a function of the lag. This sets the DEM-USD rates in a higher risk-class then suggested by the Gaussian. *April 15 1995*

Introduction and summary

The central concept in fractal geometry is that of self-similarity[1]. For a geometric object, such as a coast line, self-similarity means that the geometries of subsets of different sizes resemble each other when rescaled to the same size. For stock price records the self-similarity is a temporal one, and qualitatively manifests itself in the virtual impossibility to determine the time scale and price scale of a record when these are absent. In other words, quoting Benoit Mandelbrot[1], "no time lag is really more special than any other."

The quantitative analysis of self-similarity involves the measurement of scaling relations between properties (N) measured at different scales ϵ. These are powerlaw relations of the form $N(\epsilon) \sim \epsilon^D$ where D is called a fractal dimension[1]. For stock and Foreign exchange (FX) prices the Hurst exponent H[1, 2, 3, 4, 5] or (equivalently) the power law decay of the Fourier power spectrum[2] replaces the fractal dimension. In more complicated situations involving self-similar distributions in space, like for instance the charge distribution on a fractal surface[6], the multifractal[7, 8, 9, 10] properties are described by a function $f(\alpha)$, instead of the quantity D. An equivalent sort of analysis for time series and graphs[4], using large deviation analysis, is also applied in this paper to FX prices.

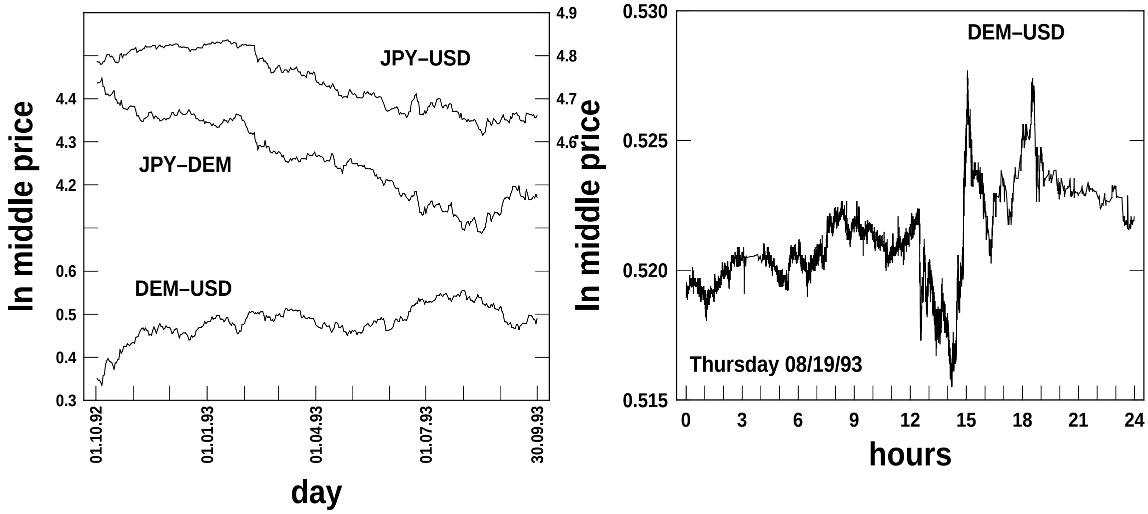

Figure 1: *left)* The daily logarithmic middle price records for the YEN-DEM, YEN-USD and DEM-USD. The study presented in this paper concentrates on scales within one day. This is shown in the *right* plot, which is the intra-day logarithmic middle price record for DEM-USD on Thursday August 19, 1993. Note the anti-correlations on the small time scales in the plot on the right.

This paper investigates quantitatively the self-similarity of the DEM-USD exchange rate on physical time scales from 2 seconds to to approximately 2 days. The data used is the foreign exchange data HFDF93 from *Olsen & Associates*[12] in Zürich, which contains of the order of 5000 bid-ask quotes per business day for the Deutschmark-US Dollar exchange over the period October 1 1992 till September 30 1993. The number of quotes equals 1 472 241 in this time period, with a highest resolution of 2 seconds. This amounts to an average of 1 quote every 21.42 seconds. We have used the filter flag provided with the data to take out extreme outliers, such as bid and ask prices 100 times the normal price. In Figure 1 we plotted the daily record of the logarithmic middle price (Equation 1) for the JPY-USD, JPY-DEM and DEM-USD foreign exchange data, for the period October 1 1992 till September 30 1993. The logarithmic middle price plotted is the one closest to 3 PM Greenwich Mean Time (GMT).

As was pointed out in Reference [3] the prices in the HFDF93 data set are quoted prices, not actual trading prices. Especially on the short time-scales, they are contaminated by transmission delays and transmission breakdowns. Nevertheless it may be reasonable to assume that this is the record that most active traders see, and that it is the record available for computer assisted real time trading. Therefore its properties are worthwhile unraveling and understanding.

It is important to note that the results discussed in this paper are not based on solely looking at the shape of the distribution on one particular scale. All analysis involve the scaling relations between the distributions on various scales (lags k). This is different from many of the analysis that have been done in the past, where conclusions are drawn from fitting known densities to the empirical density of say daily price jumps.

All the analysis presented in this paper is done in units of number of prices quoted, i.e., a lag of size $k = 512$ means that there are 512 prices quoted between the two prices considered. This notion of time, which we call *quote time*, differs from e.g. the ordinary physical time or business time[13]. One reason for using this notion of time unit is that we do not find significant empirical evidence for a strong dependence of the price jumps across consecutive quotes, and the actual physical time between the quotes. This suggests that there is no intra-quote dynamics of any relevance.

A preliminary study of the distribution of waiting times between quotes shows that these distributions have powerlaw tails which indicate that the concept of a mean or average waiting time between quotes should be handled with utmost caution. The mean value of 1 quote every 24.42 seconds mentioned above has nothing whatsoever to do with what a speculator will experience in real time during say two hours of trading. Therefore, the relation between quote-time and physical time is a statistical one with strong fluctuations.

A method is proposed, based on transforming the data into sequence of symbols 0 (price goes down) and 1 (price goes up), to analyze explicitly the up-down correlations on all time scales. The up-down correlation coefficient C at coarse-graining level k defined here, is a number between -1 and 1. For a simple random walk this number is 0, independent on the level of coarse-graining and independent of the word size used in the analysis. In general, the up-down correlation coefficient is negative for processes where the *signs* of the jumps are negatively correlated, and positive when they are positively correlated. The DEM-USD price record has a negative up-down correlation coefficient for $k < 512$. For $k > 512$ it seems to be positively correlated.

We find that the DEM-USD logarithmic middle price record has a high degree of statistical self-similarity. However this self-similarity can not be described by a single Hurst exponent H because of the presence of *crossovers*. Below a lag $k = 16$ (quotes) no exponent H is defined and there are strong anti-correlation between the jumps in the log-middle price. For $k = 32$ to 512, we find $H \approx 0.45$, and also for these quote-time scales we explicitly show anti-correlations using the above mentioned method. For $k = 512$ to 4096 we find $H \approx 0.56$ and positive correlation between the jumps. The anti-correlations on the small time scales explain why typically the intra-day logarithmic middle price records look much rougher on these scales than on larger ones (see Fig. 1).

The existence of a crossover at $k = 512$ (which roughly corresponds with a physical time of 1 hour during peak periods) introduces a time scale into this problem, which makes it possible to measure whether one is below or above that scale. However, as we show by looking at the full distributions of the logarithmic middle price jumps for different time scales k, this crossover is not from self-similar to non-self-similar: on both sides of the crossover at $k = 512$, we find a very high degree of distributional self-similarity. This concept of self-similarity is that defined and used by Mandelbrot and van Ness in the definition of fractional Brownian motions with Hurst exponents $0 < H < 1$. These are strongly correlated random processes. The renormalization rules for the distribution of jumps over lag k for these processes is the same as that for stable distributions with characteristic exponents $1 < \alpha = 1/H < 2$.

Our results show that the logarithmic middle price record is in a completely different class than Gaussian random walk models. Furthermore the observed self-similarity can certainly not be explained with a simple stable random walk model for $k < 512$, because of the strong anti-correlations and the fact that $H = 0.45 < 0.5$; the latter being impossible for a random process with independent identically distributed increments. Neither is the self-similarity encountered, close to that of fractional Brownian motions, since the basic distribution onto which the distributions for different k values collapse, has tails fatter than that of a double exponential. We assessed the full extend of the anti-correlations by time scrambling the DEM-USD record, after which it became a Gaussian process with an estimated Hurst exponent 0.5, and a well-behaved empirical convergence to a large deviation rate function falling right onto the exact solution for a Gaussian process. The non-convergence of the suitably plotted distributions of returns for lag k, to a large deviation rate function, implies that there are either long tails or strong correlations in the DEM-USD record. Indeed both are the case for $k < 512$. The large deviation analysis shows that the probabilities for large returns (negative or positive) decays less fast than exponentially as a function of the lag. This sets the DEM-USD rates in a higher risk-class then suggested by the Gaussian model.

Even though the non-convergence to a large deviation rate function, and the $H = 0.56$ distributional self-similarity are necessary conditions for the stability, there is not enough evidence, to decide whether this scaling behavior found for the larger time scales $k > 512$ is best modeled by a stable process or that correlations play an important role.

Pareto distribution of physical times between quotes

Let us denote the bid and ask prices of US dollar prices in Deutsch Mark respectively by $\{B(t_i)\}_{i=1}^{T}$ and $\{A(t_i)\}_{i=1}^{T}$, where t_i is the physical time and i is the quote time. The Olsen DEM-USD data contains $T = 1\ 472\ 241$ quotes registered at physical times

$$\{t_i\}_{i=1}^{T}$$

which are in units of seconds.

Let $\Delta_i = t_{i+1} - t_i$ be the *waiting time* between the two consecutive quotes at physical times t_i and t_{i+1}. Figure 2 depicts the estimates of the probability densities $P(\Delta)d\Delta$ of the waiting times

$$\Delta \in \{\Delta_i\}_{i=1}^{T-1}.$$

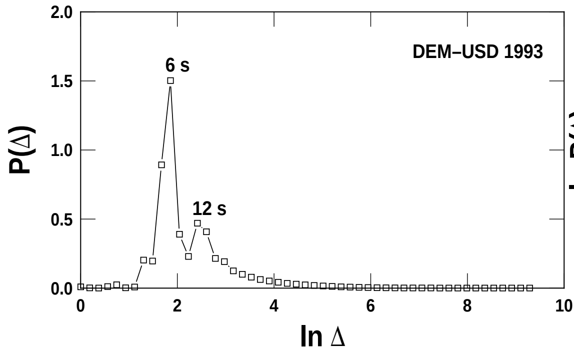

Figure 2: Semi-log plots of the probability densities of the waiting times Δ between successive quotes for DEM-USD.

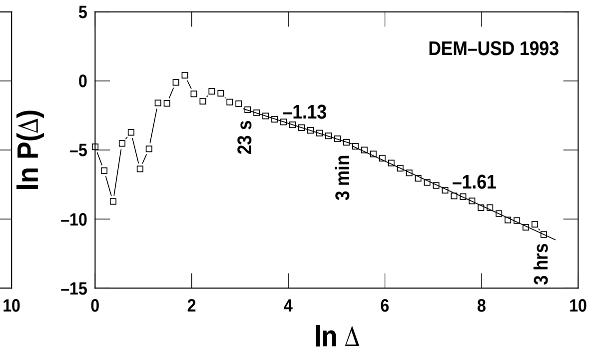

Figure 3: Double logarithmic plots of the same data shown in the left plot. The numbers -1.13 and -1.61 are least square estimates of the slopes of the tail in the regions indicated.

One notices a large peak at 6 seconds and a subsequent one at 12 seconds, indicating an effective granularity of 6 seconds. Figure 3 is a double logarithmic plot of the same density. The approximate straightness of the tails closely resembles that of Pareto distributions (power law distributions)

$$P(\Delta) \sim \Delta^{-\lambda-1}.$$

From 23 seconds to 3 minutes the estimated exponent is $\lambda \approx 0.13$. There is a crossover at 3 minutes, and for 3 minutes up to 3 hours $\lambda \approx 0.61$. Pareto distributions appear in many economic quantities like for instance the size distribution of companies and the distribution of insurance claim sizes[1, 14]. The perhaps surprising result is that the values of the exponent $\lambda \leq 1$ encountered here belong to distributions without mean or variance. The shape of the empirical distribution found, shows that the usage of the mean of the waiting times, and therefore certainly also their variance, should be avoided. Therefore, the average physical time interval between two consecutive quotes $< \Delta > \approx 21.42$ seconds, mentioned above, is not indicative of what one would experience in real market conditions (here $<>$ denotes sample averaging). Plots of the waiting times as a function of the physical time of the day during normal trading hours indeed show that this is a strongly fluctuating quantity.

What happens between successive quotes

For daily stock prices (X) and foreign exchange prices there is much empirical evidence[15, 2, 5, 3, 4, 16] showing that the average size of the absolute returns over k day lags grows like $< |\ln X(t+k) - \ln X(t)| > \sim k^H$ where H is an exponent somewhere around 0.5 (usually larger than 0.5.) This behavior makes random processes (Gaussian, stable, or others) good candidates for models for these price records on time-scales above one day[17, 18, 15, 19]. But what happens on the smallest available time-scales?

Based on the random walk model, where the number of increments is proportional to the physical time, it is sometimes argued that the average absolute price change will be larger the more one waits (the more time there is for important news to arrive.) This is the explanation often used for large price changes on the markets over weekends. We now address the question whether there is evidence for the existence of a dynamical process (economical, psychological or else) on the very smallest time scales, namely, the period between two successive quotes in the DEM-USD data. In particular, also whether such a process could naturally be modeled by an independent identically distributed (i.i.d.) random process.

To find out whether there is such an effect on the microscopic scales at hand, we analyze whether the distributions of the 1-quote price jumps depend on the waiting time between the successive quotes. In particular, we investigate whether there is a powerlaw relation between the jumps across quote intervals of the order of t seconds and those of the order of t' seconds, with $t \neq t'$?

The logarithmic middle price at physical time t_i is well-known to be defined as

$$M(t_i) = \frac{\ln A(t_i) + \ln B(t_i)}{2}. \tag{1}$$

Up to a sign change, $M(t_i)$ is equal for the USD-DEM and DEM-USD exchange rates. Let us denote the (1-quote) jumps in the logarithmic middle price by

$$l(i) = M(t_{i+1}) - M(t_i).$$

A logarithmic subdivision of the quote intervals $\Delta_i = t_{i+1} - t_i$ in bins

$$I_\kappa = 6 \times [3^{\kappa-1}, 3^\kappa]$$

is made, with $I_0 = [1, 6]$. That is,

interval	min	max
I_0	1 s	6 s
I_1	6	18 s
I_2	18	54 s
I_3	54 s	2.7 min
I_4	2.7 min	8 min
I_5	8 min	24 min
I_6	24 min	73 min
I_7	73 min	3.6 hrs
I_8	3.6 hrs	11 hrs

For the quote intervals Δ_i which lie in bin I_κ, the probability density $P_\kappa(l)$ were estimated. $P_\kappa(l)dl$ is the probability that the jump across a quote interval lying in

I_κ, has a magnitude between l and $l + dl$. If there would be some process going on that could be modeled as a i.i.d. random additive process, then one would expect that the distributions $P_k(l)$ would broaden as a function of the quote interval sizes k, as k^H, with H somewhere around $\frac{1}{2}$.

The top-right insert in figure 4 is a plot of the probability densities of the log-

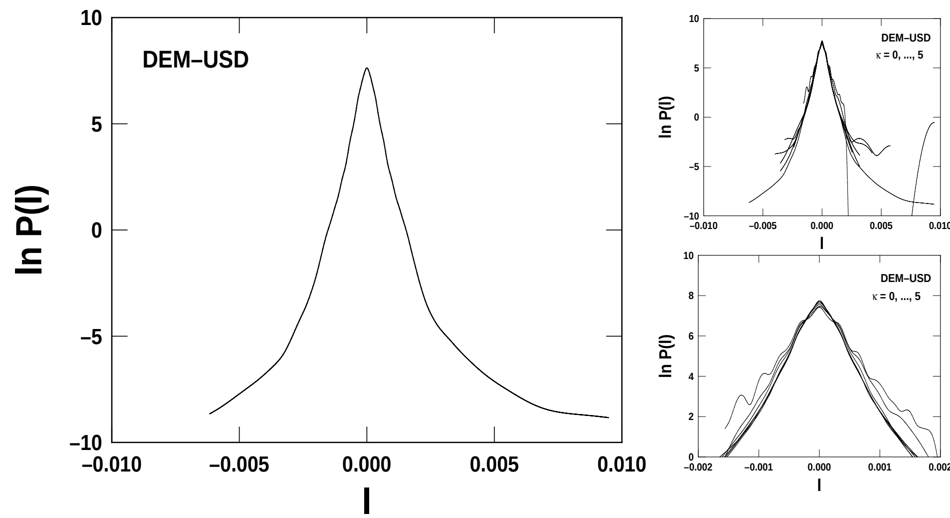

Figure 4: *Left)* Probability densities for 1-quote jumps in the DEM-USD logarithmic middle price. *Top right)* Probability densities for 1-quote jumps, but now as a function of the physical time interval, indexed by κ, between the quotes. Independent from the physical time interval, all 1-quote jumps have approximately the same distribution. *Bottom right)* Blow up of the center portion of the top figures.

arithmic price jumps for each of the I_κ. The bottom-right plot is a blow up of the central part. The left plot shows the probability density of all logarithmic middle price jumps, irrespective of the size of the waiting time. All densities have been estimated with an adaptive kernel method[20] with a Gaussian kernel.

One notices immediately that there is very little dependence on the quote interval bin I_κ. A more careful indication of possible dependencies is found in Figure 5, which is a plot of the empirical dependence of the mean absolute middle price jumps as a function of the logarithm of the size I_κ of the waiting time. The width of the middle price jumps across 18-54 second waiting times is 0.00018% while that across 3.6 hrs-11 hrs waiting times is 0.00055%. If these two results had to be linked by a powerlaw dependence, Δ^H, then the exponent $H \approx \ln(0.00055/0.00018)/\ln(11 \times 3600/54) \approx 0.02$. This is 25 times smaller than the random walk exponent. Furthermore, plotting $\ln \mathrm{E}|l|$ versus k does not yield a straight line. This shows that there is no powerlaw behavior of the logarithmic middle price jumps as a function of the waiting time between two quotes. Instead the typical jump size is rather independent of the time interval between two successive quotes. Only successive quotes which are above 4 hours apart show a weak dependency.

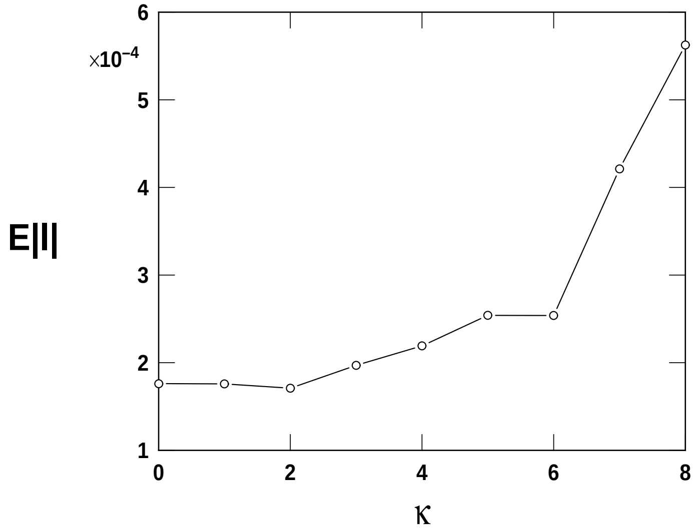

Figure 5: The empirical dependence of the width of the distribution of 1-quote jumps as a function of the waiting time Δ. The dependence on the logarithm of the physical time κ, is very weak.

Based on the lack of dependence of P_k on k, we conclude that there is no clearly measurable accumulative random process going on between quotes, and that the physical time duration between the quotes is rather irrelevant. Therefore we will use the number of quotes as a measure of time. In this new time scale, called quote-time, the interval between 2 successive quotes is 1.

The Hurst exponent H for the logarithmic middle price

From the Gaussian central limit theorem it is well-known that the expectation of the absolute value of the sum of k i.i.d. random variables with finite variance scales like k^H with $H = \frac{1}{2}$. If such a random walk model would hold for stocks, one would expect that the mean absolute return over a k-day investment, $< |L_k| >$, would behave like

$$\frac{< |L_{k'}| >}{< |L_k| >} = \left(\frac{k'}{k}\right)^H$$

as a function of the number of days. This is a scaling relation between observations at different scales, and provides quantitative description of the self-similarity of the process. The exponent H has many names in the literature[1, 21, 5, 3]: (surface) roughness exponent, drift exponent, Hurst exponent, and has many applications in the description of rough surfaces in the material sciences[22]. Empirical values of H for stock and FX records, are usually found to be larger[15, 5, 4, 16] than the random walk value > 0.5.

We now present the results of such an analysis for the DEM-USD HFDF data. This analysis differs in two ways from that presented in Ref. [3]. First it concentrates on time scales from 2 seconds to approximately 12 hours while the analysis in Ref. [3] involves larger time scales. Second, it is based on quote time, and involves no interpolation of prices, while the analysis in Ref. [3] is based on physical time with an imposed granularity of 10 minutes, and requires interpolation of data.

Denoting the return between two quotes at distance k by

$$L_i(k) = M(t_{i+k}) - M(t_i), i = 0, \ldots, T - k$$

one computes the mean absolute jump as the average $< |L(k)| >$ over all $L(k)$

$$L(k) \in \{ \ M(t_{i+k}) - M(t_i) \ \}_{i=0,\gamma k, 2\gamma k, \ldots, <(T-k)} \tag{2}$$

where the parameter $\frac{1}{k} \geq \gamma \leq 1$ determines the overlap. For $\gamma \geq 1$ there is no overlap, while for $\gamma = 1/k$ there is maximal overlap between the successive elements $t = 0, \gamma k, \ldots$ in the set. Here we used the "one element in common" case $\gamma = (k-1)/k \approx 1$ which we refer to as *no overlap*. Figure 6 shows a double logarithmic plot of $< |L(k)| >$ versus k. We distinguish three regimes. A non-scaling one for

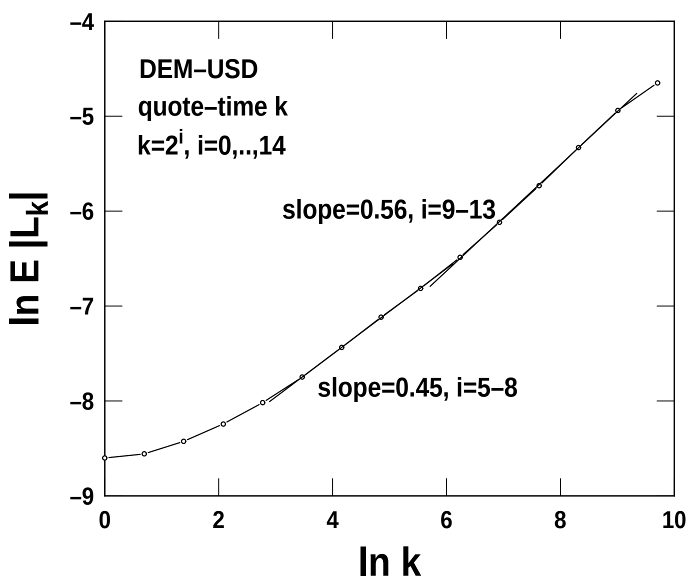

Figure 6: The behavior of the mean absolute logarithmic middle price jumps as a function of the number k of intermediate quotes. There is a clear crossover in the scaling behavior at lags of 512 quotes. What happens if the price record is scrambled is shown in Figure 10.

small k. This is followed by a scaling regime with $H < \frac{1}{2}$, which at a crossover point $k = 512$ gives way to a regime with $H > \frac{1}{2}$. In Ref. [16], a similar plot is obtained using R/S analysis. This shows that the crossover behavior discussed here is independent of the method.

As can be inferred from the shape of the tails in the left semi-logarithmic plot in Figure 1, the distribution of the 1-quote returns has a tail fatter than that of a double exponential distribution, whose tails would have formed straight lines. This and the value $H = 0.45 < 0.5$, seems to indicate anti-persistence (negative correlations) on the quote-time-scales of $k = 2^5 = 32$ quotes up to $k = 2^9 = 512$ quotes. This anti-persistence is especially strong for quote-times below $k = 32$, and fits very well with the rough (up-down) movement visible on the short time scale in Figures 1 (right).

On the other hand, there seem to be persistence (positively correlated) and/or long tails for quote-time scales of $k = 512$ up to $k = 8192$. The latter limit $k = 8192$ is of the order of two days. This exponent $H = 0.56$ is compatible with the value $H = 0.59$ found in Ref. [3] for 2 hours up to 3 months.

The existence of 3 regimes, and the values of the Hurst exponent H found, exclude simple random walk models for time scales lower than 512 quotes. The implications of the existence of these crossovers for the self-similarity of the DEM-USD record are discussed in the last section. However, even when the exponent H is well-defined over many decades, one needs additional analysis in order to assess the possible mechanism responsible for the observed scaling behavior. For example, exponents $0.5 < H < 1$ can result from both from stable processes and from correlated Gaussian processes like fractional Brownian motions[4, 23, 24]. Before giving an interpretation to the above finding for H, we now explicitly analyze the data for correlations.

The up-down correlation coefficient

We now discuss a method to systematically estimate the existence of correlations or anti-correlations on all time scales. The first step is a coarse-graining procedure, in which the k-quote logarithmic price jump time series is replaced by a time series $a_j(k)$

$$\{\ L_i(k) = M(t_{i+k}) - M(t_i)\ \}_{i=0,k,2k,\ldots,jk,\ldots,<(T-k)} \rightarrow \{\ a_j(k)\ \}_{j=0,1,2,\ldots,\approx(T-k)/k}\ , \quad (3)$$

with $a_j(k) = 0$ where the sign of the k-quote logarithmic price jumps $L_{jk}(k)$ is negative or zero, and a 1 where its sign is positive. Note that $L_i(k) \times 100\%$ is the return on buying at physical time t_i and selling k quotes later. The time series then becomes a *word* $w(k)$, consisting of the approximately $(T-k)/k$ consecutive symbols in the coarse-grained time series $a(k)$, i.e.,

$$w(k) = a_0(k)\ a_1(k)\ a_2(k)\ldots.$$

A typical word w of size $n = 5$ may look like 00101. A symbol 0 (or 1) in this word means that there has been a loss (or gain) at the end of a period containing k quotes.

Clearly words like 00000 and 11111 indicate persistence while words like 01010 indicate anti-persistence on quote-time scale k. A quantity $Q(w)$ capturing this is the number of symbol changes in a word w. For example, $Q(01011) = 3$, because there are three symbol changes (or *dislocations*) occurring at the positions denoted here $0|1|0|11$ by vertical bars. Similarly $Q(110) = 1, Q(1111111) = 0$. For words of size n, this number varies between 0 and $n-1$, but will now be normalized to vary between -1 and 1 independent of n.

For a completely random distribution of 0 and 1, like would be the case if the (coarse-grained) DEM-USD price record were a pure random walk of size n, the expected number $\mathrm{E}Q_n$ of symbol changes can be computed exactly. For a word of size n there are $n-1$ possible positions where a dislocation can occur. Each of these positions has equal probability to be an actual dislocation or not. Therefore, the expected number of dislocations $\mathrm{E}Q_n$ in a random chain of n symbols 0 or 1, is

$$\mathrm{E}Q_n = \frac{n-1}{2} \tag{4}$$

We now define the *up-down correlation* coefficient for a word of size n with q dislocations as

$$C = 1 - \frac{q}{\mathrm{E}Q_n} = 1 - \frac{2q}{n-1}. \tag{5}$$

This coefficient, always lies between -1 and 1. For example the word 010101 has coefficient -1 and 00000 has coefficient 1. The up-down correlation coefficient for a pure Gaussian random walk of size n is (by definition) 0, independent of the coarse-graining k.

For the DEM-USD, the word size is approximately $n = (T-k)/k$ for coarse graining k. Figure 7 contains the up-down correlations coefficients for both the sequence $a(k)$, with $k = 2^i$, derived from the DEM-USD exchange rates and the time-scrambled DEM-USD. This time-scrambling procedure, which is discussed in detail in the next section, scrambles the DEM-USD inter-quote jumps, and then integrates to a new price record. Therefore, this scrambled record is a realization of an i.i.d. random process. The straight line corresponds to the exact solution ($C = 0$) for the perfectly random case (say a Gaussian random walk.) Except for $k = 1, (i = 0)$, the up-down correlation coefficient for the scrambled DEM-USD data coincides with that the random case. Also, a plot (not shown here) of the up-down correlation coefficient for a simulated random process of the same size as the DEM-USD time records, indicates that after $i = 9$ there is not sufficient statistics. The deviations from zero for $k > 512$ are not significant.

However, for $k < 512, (i = 9)$ there is enough statistics and the above plot very clearly shows that the DEM-USD price records is strongly anti-correlated on scales smaller than 512 quotes.

Note: That $C(0)$ is no equal to zero for the scrambled case is due to the fact that the probability for a negative or zero interquote jump is $\mathrm{Prob}(0) = 0.563$, and that for a positive is $1-p$. The former are coded by a 0, and therefore the probability of for the interquote letters 0 and 1 are not equal. One can easily derive that the probability for a

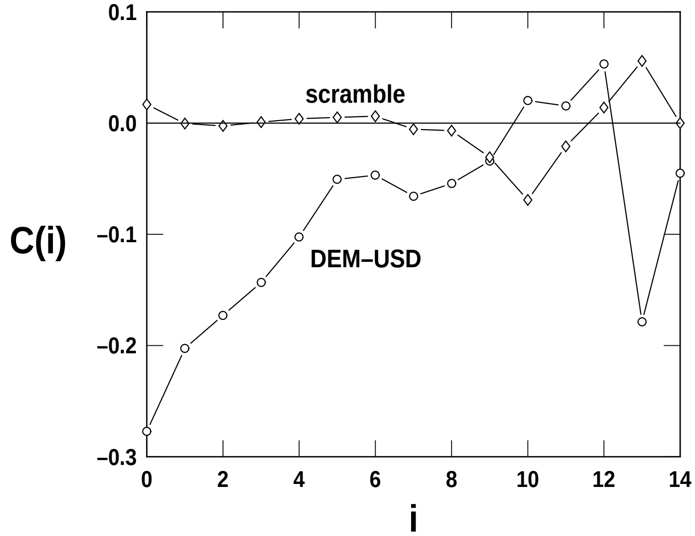

Figure 7: The up-down correlation coefficient for the DEM-USD price record and its scrambled version, as a function of various scales $k = 2^i$. The coefficient is always between -1 and 1. It is -1 for the completely anti-correlated case, 1 for the fully correlated case, and 0 for a pure random walk. The DEM-USD is anti-correlated on scales smaller than 512. For higher scales the situation is not so clear, because as can be seen from the scrambled data, the statistics does not seem enough.

dislocation in a Bernoulli trial with $\text{Prob}(0) = p$ and $\text{Prob}(1) = 1-p$ is $q = 2p(1-p)$. The correlation coefficient is therefore $C \approx 0.016$, which is in good agreement with the empirical finding for $C(0)$ for the scrambled data. The probability for zero price changes drops rapidly when aggregating quotes, and this spurious effect due to the coding convention, disappear rapidly for $k > 1$.

Large deviation analysis of the middle price

The Hurst exponent is highly degenerate, in the sense that very different processes can give rise to the same exponent. For example, the Gaussian central limit theorem implies that $H = \frac{1}{2}$ for sums of i.i.d. random variables, independent of the shape of the underlying distribution, as long as the second moment is finite. Large deviation theory[25, 26] provides a much more detailed analysis of such processes. It is used in the quantitative description of multifractal measures[9, 10] and can also be applied to the price records[4].

We only give the ideas, referring to references [4, 10] and [25, 26] and books on probability theory for further details. Consider random processes

$$l(t) = l(0) + \sum_{i=1}^{t} \mathbf{X}_i$$

where $\mathbf{X}_i$ are random variables, and for simplicity we take $\text{E}\mathbf{X}_i = 0$ for all i. The large deviation principle asserts that for a wide range of conditions on the $\mathbf{X}_i$, the probability

$$\text{Prob}\left\{\frac{l(t+k)-l(t)}{k} > \rho > 0\right\} \sim e^{kC(\rho)}$$

decays exponentially, where the large deviation rate function $C(\rho)$ is concave and negative and has a maximum value 0 at $\rho = 0$. The same holds for $\text{Prob}\{\frac{1}{k}[l(t+k)-l(t)] < \rho < 0\}$. Therefore, if price records would satisfy such a large deviation principle with rate function $C(\rho)$, this would imply that the probability for a gain larger than $\rho \times 100\%$ per unit time, would decay exponentially as $e^{kC(\rho)}$. Large deviation rate functions do not exist for all sorts of random processes. For example, it does not exist for fractional Brownian motion ($H \neq \frac{1}{2}$), nor for stable processes, nor for the 30 stocks comprising the German DAX index[4].

To do a large deviation analysis[4, 10] for the logarithmic middle price (Equation 1), we first estimate the probability densities $\text{P}_k(\rho)$ of the rate of returns ρ

$$\rho \in \left\{ \frac{M(t_{i+k})-M(t_i)}{k} \right\}_{i=1,\gamma k,2\gamma k,\ldots,<(T-k)} \tag{6}$$

for several values of k. The parameter $\frac{1}{k} \geq \gamma \leq 1$ determines the overlap.

If the large deviation principle holds, then these densities can be collapsed by plotting

$$\frac{\ln \text{P}_k(\rho)}{k} - \varepsilon(k) \ \text{ versus } \ \rho \tag{7}$$

The term $\varepsilon(k)$ is a correction term for small values of k which in the case where $\mathbf{X}_i$ are i.i.d. Gaussian random variables can be computed exactly[4]. These corrections turn out to be just vertical shifts of the graphs of $\frac{1}{k}\ln P_k(\rho)$, positioning there maximum value at 0. Because of lack of theory for general cases, also in the following we subtract the value of the maximum, i.e., we take

$$\varepsilon(k) = \frac{\ln \text{P}_k(0)}{k}$$

so that the top of the curves are all at 0.

In Figures 8 we show the result for the DEM-USD FX. Clearly there is no sign of convergence to a large deviation rate function. This non-convergence shows that the DEM-USD FX process is totally different from a simple Gaussian process, where the convergence would occur very rapidly[4]. The fact that the tails of the distributions rise as a function of k, implies that the probability for per-unit-time-returns larger than $\rho \times 100\% > 0$, decays less fast then exponential (i.e. less fast than $e^{kC(\rho)}$ for any C) as a function of the number of quotes k during which the position is held. Such an exponential decay is what should come out for a Gaussian market[4]. Therefore, the risk structure of the DEM-USD exchange record is very different from that of a Gaussian.

There are various possible reasons for the lack of convergence to a large deviation rate function. One is that the successive random variables are not independent.

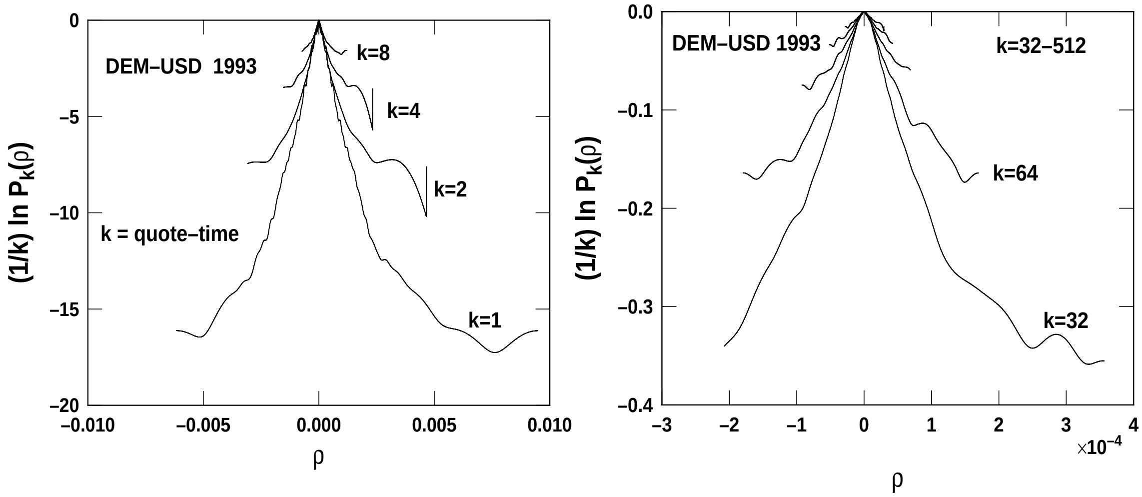

Figure 8: Large deviation analysis for the intra-day DEM-USD quotes over the year 1993. The left plot is for quote time intervals $k = 1, 2, 4, 8$ and the right plot for $k = 32, 64, 128, 256, 512$. The lack of a collapse of these different distribution onto a single curve and the clear broadening of the rescaled distributions, shows that the probabilities for large deviations decays less fast then exponentially. This is very different from what would be expected from an i.i.d. finite variance random walk.

The second is that the distribution underlying the 1-quote jumps has long tails. The third, which we do not consider here, is that the process may not be stationary. Both lack of independence and long tails are the case for stocks on the DAX index[4]. Also in the present FX case, there is direct evidence shown in Figure 9, which is a double logarithmic plot of the distribution of the absolute values $|l|$ of the logarithmic jumps across quotes. From this plot one finds that the tail behavior is an approximate power law with an exponent -6. Note that as long as the exponent is smaller than -3, such a random variable is still in the domain of attraction of the Gaussian distribution[27]. Much more accurate methods for determining this tail exponent have been developed in Ref. [28], and show that the real exponent lies around -4, i.e., it is fatter than suggested in the above Figure 9.

To assess the effect of the up-down anti-correlations on the shape of the distributions in Figures 8, we turn off these correlations by *time scrambling* the price record. To time scramble the logarithmic middle price record $\{M(t_i)\}_{i=1}^{T}$ (Equation 1,) one first determines the 1-quote jump series $\{l(i) = M(t_{i+1}) - M(t_i)\}_{i=1}^{T-1}$. Let $\{\pi(t)\}_{t=1}^{T}$ be a randomly chosen realization among the $n!$ possible random permutations of the quote-times $\{0, \ldots, T-1\}$. Than

$$\{l(\pi(1)), l(\pi(2)), \ldots, l(\pi(T-1))\}$$

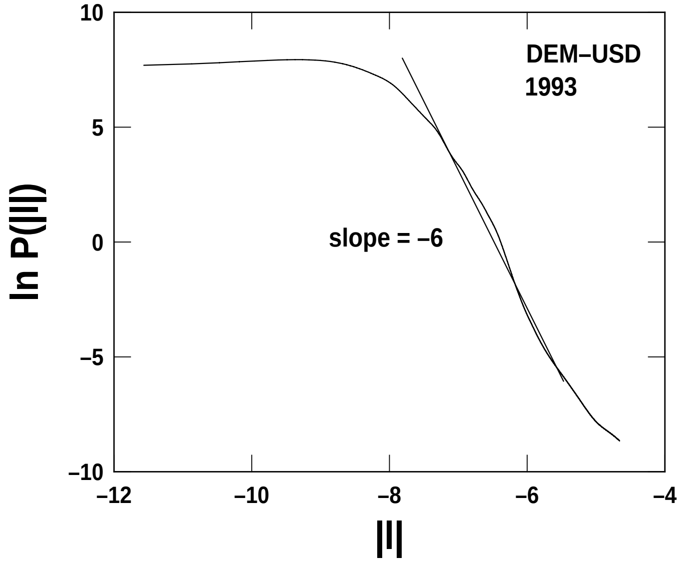

Figure 9: The distribution of the absolute values of middle price jumps across single quotes.

is a time scrambled version of $\{l(i)\}_{i=1}^{T-1}$. Each of the possible price records

$$M_i = M_0 + \sum_{j=1}^{i-1} l(\pi(j)).$$

is then a time-scrambled realization of the original DEM-USD record. Note that the scrambled record is now a sum of i.i.d. random variables.

The result of the analysis of the Hurst exponent for this scrambled DEM-USD price record, plotted in Figures 10, shows that, now that the anti-correlation have been taken away from the original data, the distributions become broader. The Hurst exponent is now $H = 0.5$, which is the simple Gaussian random walk result. The larger slope in the first 3 points is due to the long tails in the distribution of the random variables added (shown in Figure 9.)

The left plot in Figure 11 shows the results of a large deviation analysis on the scrambled data. In comparing this plot with that in Figure 8 one clearly sees the effect of the negative up-down correlations on the distributions. First of all, the scrambled record does satisfy a large deviation principle, and leads to a Gaussian market. Namely, the smooth concave curve is the exact large deviation rate function for a Gaussian process $\mathbf{X}_i$ with standard deviation $\sigma = 0.000246$. All the empirical distributions $P_k^{\mathrm{scrm}}(\rho)$ for the scrambled DEM-USD record collapse onto this large deviation rate function, using the collapse rule Equation 7. The dotted curve is the $k = 64$ result for the original DEM-USD record taken from Figure 8. From the change in the shape of the distribution for $k = 64$ in going from the original to the

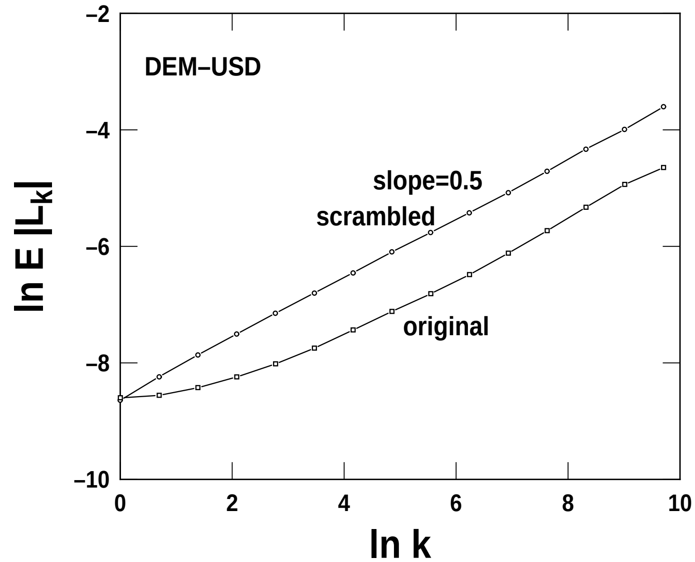

Figure 10: The behavior of the mean absolute logarithmic middle price jumps as a function of the number k of intermediate quotes for the scrambled time serie. All crossovers disappear, and the process becomes a simple Gaussian random walk. The original data have been reproduced from Figure 6

scrambled data, one infers that the up-down anti-correlations have a pronounced narrowing effect on the distribution of returns over 64 quotes. Nevertheless, the shape of this (narrowed down) distribution is far from being Gaussian, and has tails fatter than a double exponential distribution. This is not totally unexpected, since alternatingly adding and subtracting positive long-tailed random variables is not going to take away the long tail. However, this will narrow the distribution.

Similarly, the right plot in Figure 11 shows a 100 fold blow up of the top part of the left plot, now containing only the exact Gaussian large deviation rate function and both the scrambled and non-scrambled $k = 2048$ distributions. Also here the narrowing effect is still present. But now the (concave) tails of the distribution decay faster than double exponential. However, one should note that for $k = 2048$ the number of independent sample values of ρ is of the order of 700. A longer data set is needed to determine the behavior more accurately at these larger time scales.

The self-similarity of the returns

We found two different Hurst exponents. One, $H = 0.45$, relates the mean absolute returns over different lags $32 < k < 512$ with each other. The other, $H = 0.56$, relates the mean absolute returns over lags on higher scales $512 < k < 8192$. The large deviation analysis showed that this dual scaling behavior is distinctly non-Gaussian, with tail probabilities decaying less fast then exponential as a function of

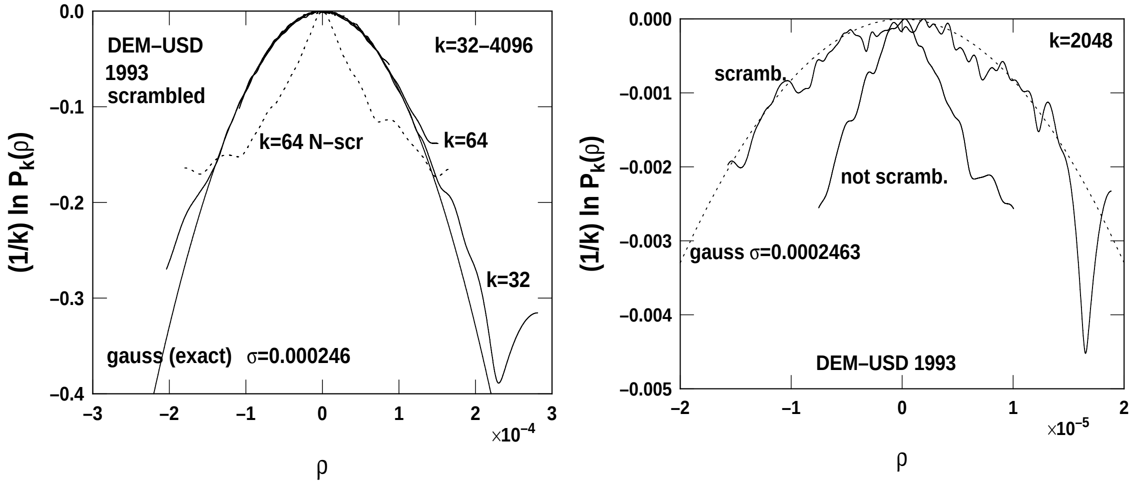

Figure 11: *left)* Large deviation analysis for the **scrambled** intra-day DEM-USD quotes over the year 1993. The smooth concave curve is the large deviation rate function for a simple Gaussian process with standard deviation $\sigma = 0.000246$. The dotted curve is the $k = 64$ result for the original DEM-USD record and one clearly sees the narrowing effect of the up-down anti-correlations. *right)* is the same as left, but now for $k = 2048$. Also here there is a substantial broadening with respect to the original. But here the correlations plot is the same as the dotted curve is the exact Gaussian result

the holding period k.

Given that the mean absolute values seem to scale, with a crossover at $k = 512$, the question remains whether the *distributions* of logarithmic middle price jumps for different quote-time intervals k are self-similar. Self-similar[23, 4] in the sense that a randomly picked k-quote return $M(t_{i+k}) - M(t_i)$ is identically distributed to a randomly picked k'-quote return $M(t_{j+k'}) - M(t_j)$ when the latter is rescaled by a factor $(k/k')^H$. Formally, if the return over a lag k is written as a sample value of a random variable $\mathbf{Y}_k$, and that of k' as a sample value of $\mathbf{Y}_{k'}$, then the process is self-similarity[23] if

$$\mathbf{Y}_k \;\text{ i.d. }\; \left(\frac{k'}{k}\right)^{-H} \mathbf{Y}_{k'}. \tag{8}$$

To accommodate for the returns per unit time (ρ) used in the large deviation analysis, we rewrite as like

$$\frac{\mathbf{Y}_k}{k} \;\text{ i.d. }\; \left(\frac{k'}{k}\right)^{1-H} \frac{\mathbf{Y}_{k'}}{k'} \tag{9}$$

which implies – by a simple change of variable in the densities – that, the process is self-similar if plots of densities $\mathrm{P}_k(\rho)$ (Equation 6) collapse, when plotting

$$(H-1)\ln k + \ln \mathrm{P}_k(\rho) \;\text{ versus }\; k^{1-H}\rho. \tag{10}$$

With the values of H known for the two different regimes, one could now check

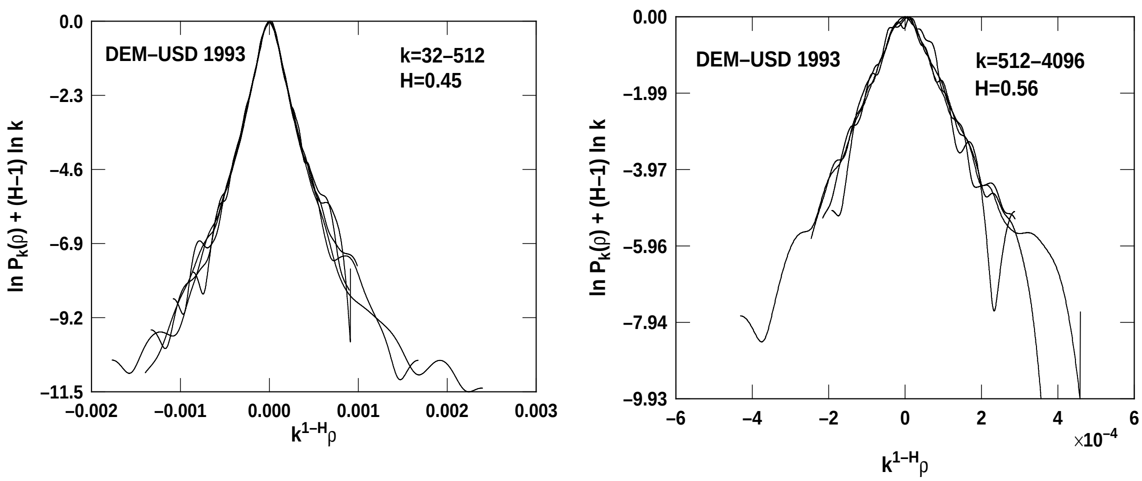

Figure 12: Self-similarity analysis for the anti-persistent and persistent regime for the DEM-USD exchange rate. The left plot is for the anti-persistent regime and has been rescaled with $H = 0.45$. The right plot has been rescaled with $H = 0.56$. In both plots we added one distribution from the other regime, to show that indeed it does not collapse onto the rest. Note that the top of the distributions have been shifted vertically so that the absolute maximum values lie at 0.

whether the distributions are self-similar, by rescaling the densities estimated from Equation 6 with the renormalization rule Equation 10. Note that the term $(H - 1)\ln k$ is a k dependent vertical shift of the distributions. Because the number of independent samples used to estimate $P_k(\rho)$ decreases with increasing k, the tails of these estimated $P_k(\rho)$ will be shorter. So, even though for a self-similar process, the shape of the rescaled distributions will be identical for those overlapping values of $k^{1-H}\rho$ that have been suitable sampled for both, the rescaled distribution for the larger k will be located above the other, because, notwithstanding its relatively shorter tails, it should normalize to 1. In order not to be disturbed by these finite sample size corrections, also here we vertically translate all the maxima of $(H - 1)\ln k + \ln \mathrm{P}_k(\rho)$ to 0.

The results are shown in Figure 12. The left plot is for the anti-persistent regime $k = 32-512$, and according to the result in Figure 6 has been rescaled with $H = 0.45$. The right plot is for the persistent regime $k = 512 - 4096$, and has been rescaled with $H = 0.56$. The non-concavity of the tails of the basic distribution[4] onto which everything collapses is again an indication of long or fatty tails. The symmetry of the basic distribution and the fact that both the left and the right tail have the same scaling behavior, fits well with the finding in Reference [3], that the mean absolute positive and negative return fluctuation have the same scaling behavior.

What these two plots show is that in both the short time-scale regime ($k < 512$, $H = 0.45$) and the long time-scales regime ($k > 512$, $H = 0.56$,) the DEM-USD returns are self-similar to a very high degree. In the short time scale regime, it is very clear that the self-similarity we measured is arising from a complicated dynamics, characterized by strong anti-correlations and by fat tails. Because $H < 0.5$ this certainly is not a stable process.

More analysis has to be done on the correlations for scales above $k > 512$ in order to assess their influence. From the smallness of the up-down correlation coefficients on scales $k > 512$, we know that that sort of correlations do not seem to play a major role. Reference [3] finds no correlations between hourly price changes, but does find significant correlations for their absolute values (see also [2]). This complicates the understanding of the origin of $H > 0.56$ even further.

By definition, the rescaling rule in Equation 10 collapses the lag-k distribution of jumps in fractional Brownian motion with $0 < H < 1$ onto a concave function of the form $-x^2$[23, 4]. This rescaling rule is also identical to that for symmetric stable random variables[27] with zero expectation and with characteristic exponent $1 < \alpha = 1/H < 2$, i.e., $0.5 < H < 1$. This implies that also for such processes one gets a collapse of the distributions onto a basic distribution, using Equation 10. The difference is that in the latter case the basic distribution has convex tails. The exponent $H = 0.56$ and the collapse in the right plot in Figure 12 are therefore necessary, but not sufficient conditions to conclude that the self-similarity observed is that of a stable process.

Acknowledgment – I would very much like to thank Heinz-Otto Peitgen for his support of this research. I am grateful to Michael Reincke for writing most of the computer codes used in the analysis, and for his help with the plots. The data has been plotted using a graphics software package (VP) developed by Richard F. Voss.

References

[1] B.B. Mandelbrot: *The fractal geometry of nature* W.H.Freeman, San Francisco (1982)

[2] R.F. Voss: 1/f noise and fractals in Economic time series In: *Fractal geometry and computer graphics* J.L. Encarnação, H.-O. Peitgen, G. Sakas, G. Englert (Eds.) Springer-Verlag, 45-52 (1992)

[3] U.A. Müller, M.M. Dacorogna, R.B. Olsen, O.V. Pictet, M. Schwarz, C. Morgenegg: Statistical study of foreign exchange rates, empirical evidence of price change scaling law, and intra-day analysis, *Journal of Banking and Finance* **14**, 1189-1208 (1990)

[4] C.J.G. Evertsz and K. Berkner: Large deviation and self-similarity analysis of curves: DAX stock prices *Chaos, Solitons & Fractals* **6** 121-130 (1995)

[5] E.E. Peters: *Chaos and order in the capital markets* John Wiley & Sons, (1991)

[6] C.J.G. Evertsz and B.B. Mandelbrot: Harmonic measure around a linearly self-similar tree *Journal of Physics* **A 25**, 1781-1797 (1992)

[7] U. Frisch, G. Parisi: Fully developed turbulence and intermittency, in *Turbulence and Predictability of Geophysical Flows and Climate Dynamics*, Proc. of the International School of Physics "Enrico Fermi,", Course LXXXVIII, Varenna 9083, edited by M. Ghil, R. Benzi and G. Parisi, North-Holland, New York, p. 84 (1985)

[8] T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia and B.I. Shraiman, Fractal measures and their singularities: The characterization of strange sets *Phys. Rev. A* **33** 1141 (1986)

[9] B.B. Mandelbrot: An introduction to multifractal distribution functions, in *Random fluctuations and pattern growth* H.E. Stanley and N. Ostrowsky, Kluwer Academic Publishers, Dordrecht, 279-291 (1988)

[10] C.J.G. Evertsz and B.B. Mandelbrot: Multifractal measures in Ref [11], 921-953 (1992)

[11] H.-O. Peitgen, H. Jürgens, D. Saupe: *Chaos and Fractals* Springer-Verlag, New York, 1992

[12] The High Frequency Data in Finance 1993 data set have been obtained from Olsen & Associates "Research institute for applied Economics", e-mail:hfdf@olsen.ch

[13] M.M. Dacorogna, U.A. Müller, R.J. Nagler, R.B. Olsen, O.V. Pictet: A geographical model for the daily and weekly seasonal volatility in the FX market *Journal of International Money and Finance* **12**, 413-438 (1993)

[14] D. Zajdenweber, Extreme values in business interruption insurance, to appear in *Journal of Risk and Insurance* (1995)

[15] B.B. Mandelbrot: The variation of certain speculative prices. *The Journal of Business of the University of Chicago* **36**, 394-419 (1963)

[16] J. Moody and L. Wu, Statistical Analysis and Forcasting of High Frequency Foreign Exchange Rates, in Proc. High Frequency Data in Finance, Zürich, 1995, to appear in *Journal of Empirical Finance*

[17] L. Bachelier: *Theory of Speculation* thesis (1900) reprinted in Ref [19] 17-78

[18] M.F.M. Osborne, Brownian motion in the stock market, reprinted in Ref [19] 100-128 (1958)

[19] P.H. Cootner: *The random character of stock market prices* The M.I.T. press, Cambridge, Massachusetts, (1964)

[20] B.W. Silverman: *Density estimation for statistics and data analysis* Chapman & Hall, London, New York (1992)

[21] J. Feder: *Fractals* Plenum Press (1988)

[22] F. Family and T. Vicsek(Eds.): *Dynamics of fractal surfaces* World Scientific, (1991)

[23] B.B. Mandelbrot and J.W. van Ness: Fractional Brownian motions, fractional noises and applications, *SIAM Review* **10**, 4, 422-437 (1968)

[24] B.B. Mandelbrot and J.R. Wallis: Computer experiments with fractional Gaussian noises I, II, III, *Water resources research* **5**, 1,228-267 (1969)

[25] R.S. Ellis: *Entropy, Large Deviations, and Statistical Mechanics* Springer-Verlag (1985)

[26] J.A. Buckley: *Large deviation techniques in decision, simulation and estimation* Wiley (1990)

[27] B.V. Gnedenko and A.N. Kolmogorov, *Limit distributions for sums of independent random variables* Addison-Wesley (1968)

[28] J. Danielsson and C.G. de Vries: Robust tail index and quantile estimation, Proc. High Frequency Data in Finance, Zürich, 1995, to appear in *Journal of Empirical Finance*