Physica A 312 (2002) 565–576

PHYSICA A

www.elsevier.com/locate/physa

Antipersistent Markov behavior in foreign exchange markets

Roberto Baviera[a,*], Michele Pasquini[b], Maurizio Serva[b], Davide Vergni[c], Angelo Vulpiani[c]

[a] *Derivative Products, MPS Finance Banca Mobiliare, V.le Mazzini 23, I-53100 Siena, Italy*
[b] *Dipartimento di Matematica, Università dell'Aquila, and INFM, Unità dell'Aquila, I-67100 Coppito, L'Aquila, Italy*
[c] *Dipartimento di Fisica, Università di Roma "La Sapienza" and INFM, Unità di Roma 1, I-00185 P.le A. Moro 2, Roma, Italy*

Received 16 April 2002

Abstract

A quantitative check of efficiency in US dollar/Deutsche mark exchange rates is developed using high-frequency (tick by tick) data. The antipersistent Markov behavior of log-price fluctuations of given size implies, in principle, the possibility of a statistical forecast. We introduce and measure the *available* information of the quote sequence, and we show how it can be profitable following a particular trading rule. © 2002 Elsevier Science B.V. All rights reserved.

PACS: 05.45.Tp; 02.50.Ga

Keywords: Markov process; Exchange rate; Shannon entropy; Forecasting

1. Introduction

A challenging question in financial markets is whether there exist correlations which provide useful information for a speculator. We show that an elementary Markov model can grasp some essential features of foreign exchange markets and how this is related to a profitable strategy.

* Corresponding author.

E-mail addresses: roberto.baviera@mpsfinance.it (R. Baviera), pasquini@serva.dm.univaq.it (M. Pasquini), maurizio.serva@aquila.infn.it (M. Serva), davide.vergni@roma1.infn.it (D. Vergni), angelo.vulpiani@roma1.infn.it (A. Vulpiani).

PII: S0378-4371(02)00968-8

Since the celebrated work of Fama [1] a big effort has been made to empirically test markets' efficiency. The debate is still open: on one side, academics mostly believe in market efficiency, at least in the weak form, on the other, chartists are convinced of the feasibility of financial series forecast. In particular, currency exchange markets seem to be the natural subject for an efficiency test. Their large liquidity should imply efficiency; however, practitioners widely use chartist techniques as documented by Allen and Taylor [2].

In this paper we show how, in an even simpler framework, a Markov model can be an useful tool in forecasting FX fluctuations of given size.

First, to obtain statistical significant results and estimations which are stable over time, we analyze a 1 year high-frequency dataset of the US dollar/Deutsche mark exchange, the most liquid market. Our data, made available by Olsen and Associates, contains all worldwide 1,472,241 bid–ask US dollar/Deutsche mark exchange rate quotes registered by the inter-bank Reuters network over the period October 1, 1992–September 30, 1993. In fact, FX markets have changed significantly over 15 years: the BIS [3] reports a change even in the micro-structure of the FX market. In the eighties transactions occurred only by telephone. However, since 1992 three electronic broking systems have been operating in London, the most important FX exchange, and automated dealing systems have been mostly used in the second half of the nineties. A 1-year lag should allow the stationarity of the essential features of the market, avoiding most of these problems.

Second, one cannot be sure that considering today's information for tomorrow's forecast (as in a Markov model of order 1) one has included all the relevant information. For example, it should be checked that the information on yesterday's prices does not add anything in tomorrow's forecast. A natural question is then: how far in the past one has to go to get all the relevant information?

Finally, we link the forecast to profits. In particular, we show how the speculator can exploit this information in an optimal trading strategy if correlations are present.

For these reasons, in order to quantify the degree of (eventual) inefficiency of the market, we define and measure the *available* information of the returns' time series. This *available* information is related to the Shannon entropy of a symbolic sequence associated to the time series. We also discuss the relation with the growth rate of the capital of a speculator which follows an optimal trading rule, i.e., a trading rule which makes the largest profit in the long run using the *available* information. *Available* information reminds ε entropy introduced by Kolmogorov [4] in the context of information theory.

The *available* information is not uniquely defined since there are different ways of associating a symbolic dynamics to a time series. In this paper, we propose a simple but very reasonable strategy which consists in fixing the resolution Δ for the return in spite of the time lag. Before trading again the speculator waits that prices vary of a given amount Δ and then he rebalances his position.

The paper is organized as follows. First, in Section 2, we define the *available* information. In Section 3, we compute this quantity for our prescription for any fixed resolution Δ and measure it in Section 4. In Section 5, we relate the *available* information to the profitability by means of a particular trading rule. In Section 6, we discuss the results.

2. Available information: mathematical background

In this section we show how information theory provides the natural tools in financial forecasting, selecting the correlations relevant for a speculator and giving a quantitative criteria to determine how far in the past one should go to include all the relevant information.

Information theory can be applied providing that the dataset is properly codified in a symbolic sequence. The standard procedure for constructing symbolic sequences from time series is substantially a coarse graining. The range of variability of the filtered data is partitioned, and a conventional symbol is assigned to each element of the partition.

Nevertheless, there are many ways to codify the dataset in a symbolic sequence. In particular, as we will see later, our choice brings a binary codification with a minimal loss of information.

Whatever may be the resulting codification, one has a discrete symbolic sequence:

$$c_1, c_2, \ldots, c_i, \ldots ,$$

where each c_i takes only a finite number, say m, of values.

Let us now recall some basic concepts of information theory. Consider a sequence of n symbols $C_n = \{c_1, c_2, \ldots, c_n\}$ and its probability $p(C_n)$. The block entropy H_n is defined by

$$H_n \equiv -\sum_{C_n} p(C_n) \ln p(C_n) .$$

The difference

$$h_n \equiv H_{n+1} - H_n$$

represents the average information needed to specify the symbol c_{n+1} once the sequence $\{c_1, c_2, \ldots, c_n\}$ is known.

The series of h_n is monotonically non-increasing and for an ergodic process one has

$$h = \lim_{n \to \infty} h_n ,$$

where h is the Shannon [5] entropy.

It is possible to demonstrate [6] that if the stochastic process $\{c_1, c_2, \ldots\}$ is Markovian of order v, then $h_n = h$ for $n \geqslant v$. This property provides a criteria to check if a sequence can be properly described by a Markov process of order v: v is the value where a plot of h_n versus n reaches a plateau.

Moreover, measuring h_n on a finite sequence of N elements, it can be proven [6] that the statistical analysis does not give the proper value of h_n for n larger than

$$n^* \approx \frac{1}{h} \ln(N) . \tag{1}$$

For this reason, entropy measures, on one hand assure the statistical relevance of the results obtained up to $n \approx n^*$, but on the other they are data consuming and can be applied only to hi-frequency datasets. Since the upper bound depends on the log of the sequence length, one needs in practice an intraday dataset to distinguish between statistical properties of a sequence and finite size effects.

Summarizing, or h_n goes to zero for increasing n (which means that a symbol is completely predictable knowing all the past sequence) or it tends to a positive value. The maximum possible value of h is $\ln(m)$ which occurs if the process has no memory and the m symbols occur with the same probability.

The difference between $\ln(m)$ and h is intuitively the quantity of information one may use to predict the next result of the observed phenomenon, i.e., the market behavior. It is then natural to define the *available* information as

$$\mathscr{I} \equiv \ln(m) - h \,. \tag{2}$$

In the next sections, we introduce and measure the *available* information for a sequence where the codification preserves most of the information of the series, and in Section 5 we discuss its financial meaning showing how *available* information is related to the growth rate of the capital following a trading strategy.

3. The fixed resolution binary codification

Suppose that a speculator selects a typical fluctuation size for his trading strategy, i.e., he updates his position only when the market has a fluctuation of size Δ: he is not interested in the length of the lag between two trades.

According to his philosophy, the best procedure to create a symbolic sequence from the bid (marks per dollar) quotes $\{S_t\}$ is

- fix a resolution value $\Delta > 0$ and define
$$r_{t_1, t_0} \equiv \ln \frac{S_{t_1}}{S_{t_0}} \,,$$
- where t_0 is the initial time, and t_1 is the first next time such that
$$|r_{t_1, t_0}| \geqslant \Delta \,,$$
- then repeat the procedure starting from t_1 and so on, obtaining the sequence of returns
$$\{r_{t_1, t_0}, r_{t_2, t_1}, \ldots, r_{t_k, t_{k-1}}, \ldots\} \,.$$
- consider the symbolic dynamics:
$$c_k = \begin{cases} - & \text{if } r_{t_k, t_{k-1}} < 0, \\ + & \text{if } r_{t_k, t_{k-1}} > 0 \,, \end{cases} \tag{3}$$
- c_k is the sign of the fluctuation, the quantity we would like to forecast.

In Fig. 1, we show an example of evolution of the $r_{t_k, t_{k-1}}$ in the US dollar/Deutsche mark exchange. Let us note that the variable $|r_{t_k, t_{k-1}}|$ has a narrow distribution close to the threshold value Δ and, for all practical purposes, $|r_{t_k, t_{k-1}}|$ can be well approximated with Δ. In other words, very little information is lost in this codification.

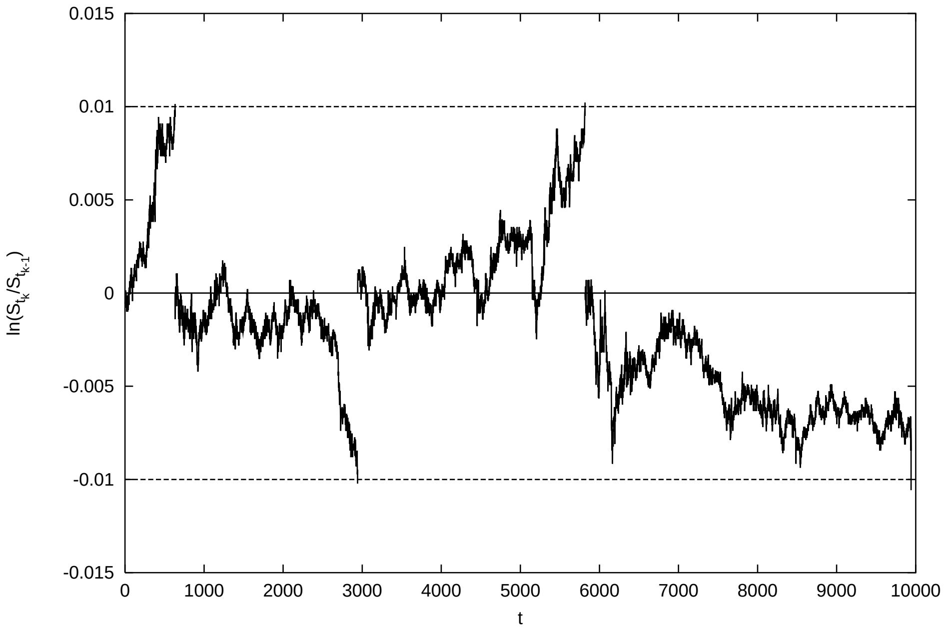

Fig. 1. Evolution of $r_{t_k,t_{k-1}}$ with $\Delta = 0.01$. $t_0 = 0$ corresponds to 00:00:14 of October 1, 1992 and $t_4 = 9939$ corresponds to 11:59:28 of October 2, 1992.

In Fig. 2, it is shown that the Shannon entropy h is clearly different from ln(2) in a wide range of Δ, i.e., the *available* information (see Eq. (2)) is non-vanishing. It is also possible to observe that h_n reaches approximately a plateau (corresponding to h) after the first step. As discussed in the previous section, this fact implies that the process is approximately Markovian of order 1. The entropic analysis has showed that the major contribution to forecast c_k (at the trade $k-1$) comes from the last fluctuation's sign. Entropy allows also to quantify the error we are committing considering this Markov approximation.

Furthermore in Fig. 2 we can observe the folding of h_n for large n, which depends on the finite number of sequence elements as discussed in the previous section (see Eq. (1)).

A Markov model of order 1 is completely described by the transition probability matrix

$$\begin{pmatrix} p(c_k = +|c_{k-1} = +) & p(c_k = -|c_{k-1} = +) \\ p(c_k = +|c_{k-1} = -) & p(c_k = -|c_{k-1} = -) \end{pmatrix} \equiv \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}, \tag{4}$$

which gives the probability of the sign k given the sign of the fluctuation $k-1$.

A consequence is that the probability to have a positive or a negative sign is

$$\begin{pmatrix} p(c_k = +) \\ p(c_k = -) \end{pmatrix} \equiv \frac{1}{\alpha+\beta} \begin{pmatrix} \beta \\ \alpha \end{pmatrix}.$$

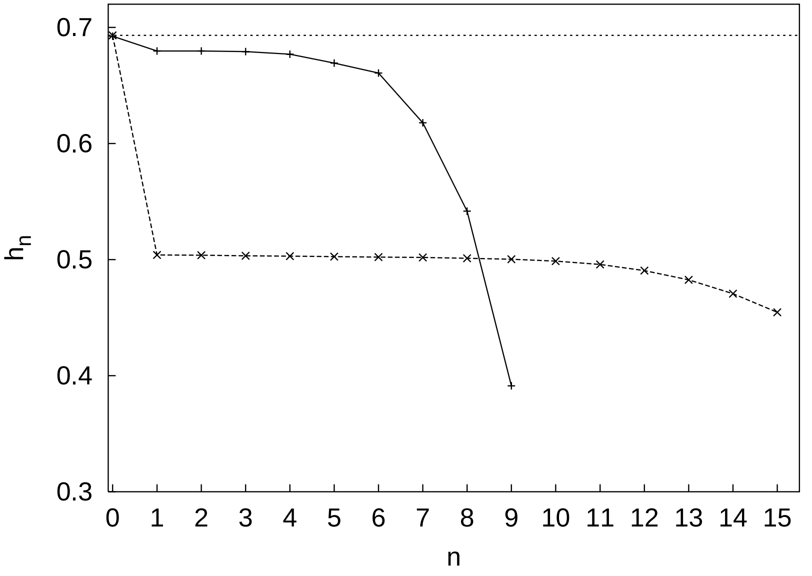

Fig. 2. h_n versus n. The different plots correspond to the values of Δ 0.0004 (+) and 0.004 ($\times$). One can observe that h_n reaches a plateau for $n = 1$, i.e., a Markov process of order 1 is a good approximation; the bend for n large is due to finite size effects (see text). The horizontal dashed line indicates ln(2), the maximum Shannon entropy.

As we show in the next section, in a foreign exchange such as the US dollar/Deutsche mark, the number of up and down movements of size Δ are equal (inside the statistical error) for Δ larger than 10^{-3}; these values of Δ are the most interesting ones in speculator's perspective, as discussed in Section 5. This symmetry ($\alpha = \beta$) is equivalent to say that no trend can be forecasted. In this simple case all information is encoded in the probability α of sign change and the *available* information is

$$\mathscr{I} = \alpha \ln \alpha + (1 - \alpha) \ln(1 - \alpha) + \ln(2) . \tag{5}$$

Therefore, since $\mathscr{I} \neq 0$, one must have $\alpha \neq \frac{1}{2}$. Indeed, we show in the next section that the symbolic dynamics is an antipersistent Markov process ($\alpha > 1/2$).

4. Empirical estimation

In this section we measure the *available* information and its standard deviation in the US dollar/Deutsche mark exchange rate quotes over the period October 1, 1992–September 30, 1993.

First, we test the asymmetry of the transition probability matrix (4) for a given Δ. The simplest way is to consider the sequence's trend

$$\mu = \frac{1}{N} \sum_i c_i$$

Table 1
Symmetry test of the Markov process and sequence's length N for different Δ

Δ	μ	μ/σ_μ	N
0.0001	0.0101	17.0103	974816
0.0002	−0.0017	−2.4481	516812
0.0003	−0.0014	−1.7828	387570
0.0004	−0.0052	−4.5133	188548
0.0006	−0.0044	−2.2713	87320
0.0008	−0.0082	−2.5118	37976
0.0010	−0.0096	−2.0699	21812
0.0015	−0.0138	−1.7039	8830
0.0020	−0.0169	−1.3992	4386
0.0025	−0.0209	−1.3249	2674
0.0030	−0.0259	−1.3154	1854
0.0035	−0.0299	−1.2474	1336
0.0040	−0.0356	−1.3238	1010

and to check whether it is different from zero. As before, N is the sequence's length.

It is simple to show that its expected value and variance are

$$E[\mu] = \frac{\beta - \alpha}{\beta + \alpha}, \quad \sigma_\mu^2 = \frac{1}{N}\frac{4\alpha\beta}{(\alpha+\beta)^3}(2 - \alpha - \beta) .$$

In Table 1, we show that we can reject the symmetry with a 5% level of confidence for $\Delta \leqslant 10^{-3}$ (except for $\Delta = 3 \cdot 10^{-4}$) and in four cases ($\Delta = 10^{-4}, 2 \times 10^{-4}, 4 \times 10^{-4}$ and 8×10^{-4}) with a 1% level of confidence. We use a Gaussian statistics due to the sufficiently large sequence's length N.

In Table 2, we provide the estimations of the *available* information $\mathscr{I}$, the transition probabilities α and β with standard deviations of their estimations (explicit formulas are reported in the appendix). In the whole range of Δ considered, the *available* information is statistically different from zero, and then it is possible to forecast the sign of the next fluctuation given the previous one.

In Fig. 3, we plot the *available* information versus Δ (left scale) and the distribution $P(\gamma)$ of the transaction costs γ (right scale). In fact, in order to select the interesting Δ range it is important to compare the results with transaction costs γ. Since commission fees are generally negligible with respect to the bid/ask spread in the foreign exchange, we have estimated γ from the bid-ask spread

$$\gamma_t = \frac{1}{2}\ln\frac{S_t^{(ask)}}{S_t^{(bid)}} \simeq \frac{S_t^{(ask)} - S_t^{(bid)}}{2S_t^{(bid)}} .$$

We observe that the maximum of the *available* information is almost in correspondence with the maximum of the distribution (typical value) of the transaction cost. For Δ larger than the typical transaction cost, the *available* information decreases with Δ, and it is almost equal to zero for large values of Δ.

Let us note that in a similar way the speculator can try to forecast the sign of price evolution after a fixed interval of time. The analysis (not reported here) shows an

Table 2
Available information and transition probabilities

Δ	$\mathscr{I}$	α	β
0.0001	0.1236	0.7357	0.7508
	(0.0005)	(0.0006)	(0.0006)
0.0002	0.1897	0.7992	0.7964
	(0.0008)	(0.0008)	(0.0008)
0.0003	0.1887	0.7982	0.7959
	(0.0009)	(0.0009)	(0.0009)
0.0004	0.1891	0.8015	0.7932
	(0.0013)	(0.0013)	(0.0013)
0.0006	0.1324	0.7548	0.7481
	(0.0016)	(0.0021)	(0.0021)
0.0008	0.0920	0.7169	0.7052
	(0.0021)	(0.0033)	(0.0033)
0.0010	0.0659	0.6860	0.6728
	(0.0024)	(0.0045)	(0.0045)
0.0015	0.0357	0.6415	0.6238
	(0.0028)	(0.0073)	(0.0072)
0.0020	0.0248	0.6211	0.6000
	(0.0033)	(0.0104)	(0.0104)
0.0025	0.0201	0.6119	0.5868
	(0.0038)	(0.0135)	(0.0133)
0.0030	0.0138	0.5969	0.5668
	(0.0038)	(0.0163)	(0.0161)
0.0035	0.0089	0.5818	0.5480
	(0.0036)	(0.0194)	(0.0190)
0.0040	0.0125	0.5975	0.5564
	(0.0049)	(0.0222)	(0.0217)

Note: Standard deviations are given in parenthesis below the estimated quantity.

available information near to zero, and then almost no forecast is possible in this case. This is the reason to prefer a fixed resolution Δ in spite of a fixed time lag approach in a forecasting tool.

For the interpretation of these results, it is necessary to show the equivalence between *available* information and capital growth rate following a particular trading rule. We show how this is possible in the next section.

5. Optimal trading strategy

Let us focus on a speculator who decides to diversify his portfolio only in a riskless asset with a given interest rate return and in the US dollar/Deutsche mark exchange market. At each trade t he invests a fraction l_t of his capital in the US dollar/Deutsche mark exchange and the remaining in the riskfree asset. The investment remains untouched until the following trade when it is updated. Our convention is that the fraction l_t is positive if he exchanges marks into dollars, negative vice versa, and we allow the speculator to borrow money.

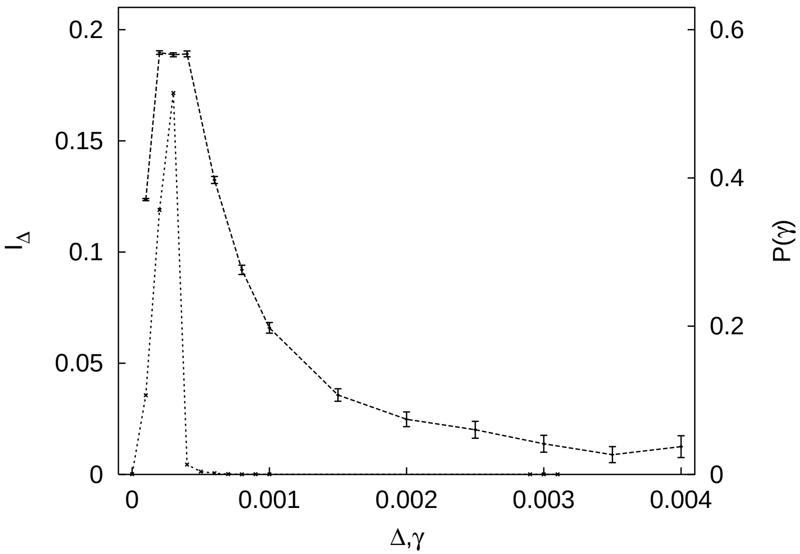

Fig. 3. *Available* information $\mathscr{I}$ versus Δ with its standard deviation (on the left), superimposed to the distribution of transaction costs, $P(\gamma)$ versus γ (on the right).

We assume a vanishing interest rate return. This hypothesis is reasonable: in the period we are dealing with, the official discount rate fixed by the Federal Reserve is 3 percent per year and fluctuates between 5.75 and 8.75 percent in the German case.

In the time lags involved, the interest rate return is about one hundred times smaller than Δ: the speculator rehedges his portfolio on average every 66 s when Δ is equal to the typical transaction cost; the largest Δ corresponds to an average time of 8.7 h of standby. The approximation of a vanishing interest rate return appears to be fair.

Assuming the investment has been performed at the trade t, immediately before the following trade $t+1$ speculator's capital becomes

$$W_{t+1} = [1 + l_t(\exp(c_{t+1}\Delta) - 1)]W_t \simeq (1 + l_t c_{t+1}\Delta)W_t \,, \tag{6}$$

where the first-order approximation in Δ is accurate enough for the Δ range considered in this paper (see Tables 1 and 2) and c_{t+1} is the sign of the fluctuation (see Eq. (3)). An intuitive consequence of anti-persistence is that the optimal strategy is to buy dollars if the mark rises, and to do the opposite otherwise.

Let us consider for simplicity the symmetric case, which well describes experimental evidences for a range of Δ values and approximates reasonably the remaining. Similar results hold in the general case. As a consequence of the symmetry of the symbolic dynamics and of its Markov nature, the optimal strategy implies that l_t may assume only two values l or $-l$ (buy dollars or marks). In other words, the same fraction of capital is invested in the forex market both if the speculator buys marks or dollars. The exact value of l is determined by maximizing the exponential rate of speculator's capital.

The strict connection between the maximal exponential rate and the *available* information was first noticed by Kelly [7], who first gave an interpretation of Shannon entropy in the context of optimal investments.

For the investment described above, the capital growth rate is

$$\lim_{N\to\infty} \frac{1}{N} \ln \frac{W_N}{W_0} = \alpha \ln(1 + l\Delta) + (1 - \alpha)\ln(1 - l\Delta)\,. \tag{7}$$

It reaches its maximum for

$$l^* = \frac{2\alpha - 1}{\Delta}\,. \tag{8}$$

From Eqs. (7) and (8) one has that the optimal growth rate is equal to the *available* information (see Eq. (5)):

$$\lambda = \alpha \ln \alpha + (1 - \alpha)\ln(1 - \alpha) + \ln(2) = \mathscr{I}\,. \tag{9}$$

This equivalence can be easily generalized to a generic (even non-symmetric) Markov process of whatever order. This means that, if we forget the costs involved in this trading rule, a speculator has the possibility, following a particular strategy, to obtain a growth rate of his capital exactly equal to the *available* information. This equivalence justifies, from a financial point of view, why we have used the previously mentioned criteria to determine the order of a "reasonable" Markov description of the financial signal, as we have done in Section 2. It is worthless to consider a larger-order Markov model if the optimal capital growth rate does not increase significantly.

We stress that we have measured the optimal growth rate of the trades' sequence. To obtain the capital growth rate of the capital in the usual time unit we have to normalize (7) with the average time between two trades for the specified Δ. For example, the greater Δ considered is characterized by an average lag of 8.7 h; the average optimal growth rate is equal to 12.6 per year in this case.

6. Discussion

We now have all the ingredients to comment on the shape of the *available* information shown in Fig. 3.

The speculator cannot have a resolution Δ lower than transaction costs, otherwise profits would be, in fact, less than costs. On the other hand, for very large Δ, everybody is able to discover the *available* information and to make it profitable with a feasible strategy: this is why there is almost no *available* information in this Δ region.

We have shown that for Δ larger than the transaction costs (but not very large) the information can be exploited by proper strategies. For example in the case of $\Delta = 0.004$, the larger value we consider, which is roughly 20 times larger than the typical transaction cost, we obtain a capital growth rate of 12.6 per year: a naive consequence could be that an efficient market hypothesis should be rejected.

Unfortunately (especially for the authors), this is not so obvious and at least two remarks hold.

First, note that for Δ larger but still near to the value corresponding to the maximum *available* information, the speculator changes his position with high frequency: it is not any more possible to neglect costs.

Second, in Eq. (8) Δ appears at the denominator, and then the values of l^* can be huge (in absolute values); the method strongly uses the leverage effect. For example for $\Delta = 0.004$ the speculator who follows the optimal growth rate strategy should borrow almost 38 times the capital he has. Lower Δ require an even larger leverage.

In conclusion, we have shown that a forecast is possible even in the very liquid FX market and how this is related to a trading rule which tries to exploit it: in this context, *available* information rises to the task of being a feasible measure of market inefficiency.

Acknowledgements

R.B. acknowledges financial support from TMR (Grant on *Financial Market's Efficiency*).

Appendix

In the more general asymmetric case, the *available* information reads

$$\mathscr{I} = \frac{\beta}{\alpha+\beta}(\alpha \ln \alpha + (1-\alpha)\ln(1-\alpha)) + \frac{\alpha}{\alpha+\beta}(\beta \ln \beta + (1-\beta)\ln(1-\beta)) + \ln(2)$$

and the leading term of the variance of its estimation reads

$$\sigma_{\mathscr{I}}^2 = \left(\frac{\partial \mathscr{I}}{\partial \alpha}\right)^2 \sigma_\alpha^2 + \left(\frac{\partial \mathscr{I}}{\partial \beta}\right)^2 \sigma_\beta^2 ,$$

where we have used α and β independence.

The variance of the transition probabilities α and β can be evaluated observing that their estimations, on a sequence of $+$ and $-$ signs of length N, can be expressed as a sum of independent variables.

In fact, the most natural estimation of α is

$$\hat{\alpha} \equiv \frac{1}{n^+}\sum_{i=1}^{n^+} x_i ,$$

where the sum is restricted to the symbols which are preceeded by a $+$ sign and

$$x_i = \begin{cases} 1 & \\ & \text{if the} + \text{is followed by a} \begin{matrix} - \\ + \end{matrix} . \\ 0 & \end{cases}$$

n^+ is the number of $+$ in the sequence.

Due to the Markov property of the sequence, the x_i are independent variables. It follows that

$$E[\hat{\alpha}] = \alpha + O\left(\frac{1}{N}\right)$$

and the leading term of the variance of its estimation is

$$\sigma_\alpha^2 = \frac{1}{N}\,\frac{\alpha+\beta}{\beta}\alpha(1-\alpha)\,.$$

Analogously one obtains

$$\sigma_\beta^2 = \frac{1}{N}\,\frac{\alpha+\beta}{\alpha}\beta(1-\beta)\,.$$

The leading term of the *available* information's variance is then

$$\sigma_{\mathscr{I}}^2 = \frac{1}{N}\,\frac{1}{\alpha+\beta}\left\{\frac{\alpha(1-\alpha)}{\beta}\left[\mathscr{I} + \beta\ln\frac{(1-\alpha)(1-\beta)}{\alpha\beta} - \ln 2(1-\beta)\right]^2 + \frac{\beta(1-\beta)}{\alpha}\left[\mathscr{I} + \alpha\ln\frac{(1-\alpha)(1-\beta)}{\alpha\beta} - \ln 2(1-\alpha)\right]^2\right\}\,.$$

References

[1] E.F. Fama, Efficient capital markets: a review of theory and empirical work J. Finance 25 (1970) 383–417.

[2] H. Allen, M.P. Taylor, The use of technical analysis in the foreign exchange market, J. Int. Money Finance 11 (1992) 304–314.

[3] Bank of International Settlements, Triennial Central Bank survey of foreign exchange and derivatives market activity, 1996, pp. 1–50 and 1999, pp. 1–43.

[4] A.N. Kolmogorov, On the Shannon theory of information transmission in the case of continuous signals, IRE Trans. Inform. Theory IT-2 1 (1956) 102–108.

[5] C.E. Shannon, The mathematical theory of communication, Bell Syst. Techn. J. 27 (1948) 379–423, 623–656.

[6] A.I. Khinchin, Mathematical Foundations of Information Theory, Dover, New York, 1957.

[7] J.L. Kelly Jr., A new interpretation of the information rate, Bell Syst. Tech. J. 35 (1956) 917–926.