ON THE INFORMATIONAL CONTENT OF ASSET PRICES

Demosthenes N. Tambakis[1]
City University Business School
Frobisher Crescent, Barbican Centre
London EC2Y 8HB, U.K.
Tel: 020-74.77.87.61, Fax: 020-74.77.86.48
d.tambakis@city.ac.uk.

June 2000

Abstract: This paper studies the predictability and weak-form informational e±ciency of eight long daily time series of major stock index, spot exchange rate and Eurodeposit rate returns. E±ciency and predictability are inversely related and respectively increasing and decreasing in a sample's self-information measure (SIM). SIM corresponds to a sample's normalized entropy. Its evolution with sample size is non-monotonic and characterized by sharp breaks corresponding to extreme events (outliers). Including such events in-sample lowers the SIM and renders the underlying data generating process less e±cient, or more predictable. It is found that Eurodeposit rate returns are relatively more predictable than stock index returns, which in turn are more predictable than exchange rate returns. The sample size at which the SIM is maximum is smallest for sterling Eurodeposit returns and greatest for dollar Eurodeposit returns. The proposed non-parametric framework o®ers a convenient measure of the contribution of new information to the predictability of asset returns processes.

Keywords: Self-information, normalized entropy, predictability, weak-form e±ciency
JEL Classi¯cation: C22, C53, G10

[1] I would like to thank participants at the Parallel Applications in Statistics and Economics conference (PASE 2000, Leuven 2-5 April 2000) for their comments. The usual disclaimer applies.

1 Introduction

Is the amount of past information to use in forecasting financial returns independent of the likelihood of extreme events? Equivalently, what is the appropriate out-of-sample separation point for optimizing the performance of a given forecasting model? A non-normal but linear data generating process only satisfies weak stationarity, which is necessary but not sufficient for strict stationarity (Granger and Newbold (1986) and Hamilton (1994)). Therefore, sample size matters. The decision to include an extreme observation outlier such as a financial crash in-sample can significant affect a forecasting model's out-of-sample performance.

The need for a simple and robust framework for quantifying the value of past financial information flows has grown following recent episodes of high volatility in various interest rate and stock markets. Indeed, against a background of increasing financial market uncertainty, the value of past information refers not only to the observed level — conditional mean — but also to the observed volatility of the underlying financial assets. In that respect, advances in GARCH and VaR methodologies for risk assessment are often sensitive to parametric assumptions about the underlying returns distributions. Unfortunately, these assumptions are likely to be violated by the non-normal features of the actual empirical distributions. Moreover, the choice of information set is independent of the following factors: (i) the likely presence of extreme events, and (ii) the risk of contagion between different financial markets documented in turbulent circumstances.

This paper proposes a finite-sample non-parametric measure of a dgp's predictability and weak-form informational efficiency. The self-information measure (*SIM*) of a finite sample is defined as its entropy normalized by its alphabet length. This measure is closely related to the normalized entropy statistic of Golan, Judge and Miller (1996). The key motivation is to develop a simple non-parametric measure of the value of past information which accounts for factor (i) above. In that respect, Tambakis (2000) proposed using a dataset's average information content (AIC) as an information-theoretic predictability measure. A discrete dataset's average information content was defined as its normalized entropy. In turn, normalized entropy was shown to be the first component of non-parametric predictability, defined as the mutual information between the random variable to be forecast and the ensemble of past observations, normalized by the alphabet length underlying the sample's empirical fequency distribution. Theoretically, ¯nite-sample predictability was shown to be increasing in normalized entropy and decreasing in the conditional entropy of the data generating process, while the general relation between forecast error probability and normalized entropy was non-monotonic. Quantifying the value of extreme events within one market is necessary in order to address cross-market risk, which is the subject of factor (ii) above.[2]

The focus is on the impact of the arrival of new information | that is of changing sample size | on the SIM of a dataset, and hence on its univariate predictability and weak-form informational e±ciency. The SIM of 8 daily time series of stock index, FX and Eurodeposit rate returns is analyzed. It is shown that SIM is non-monotonic in the sample size and is characterized by clear breaks corresponding to extreme events. In the e±cient market terminology of Fama (1970, 1991), the predictability of a ¯nancial asset's returns process is inversely related to its informational e±ciency. In information-theoretic terms, the predictability of the data generating process can be de¯ned as the mutual information between the random variable to be forecast and the ensemble of past observations, normalized by the alphabet length underlying the sample's empirical probability distribution.[3]

The main results are as follows. First, the SIM of simulated data from known pdf's is shown to be increasing in the sample size. Then, using long daily return time series for stock market averages, spot exchange rates and overnight Eurodeposit rates, it is shown that the evolution of SIM is non-monotonic in the sample size and is behavior is marked by breaks corresponding to extreme events. The relative predictability of returns is then compared across ¯nancial markets. For small to moderate sample sizes, spot FX market returns are relatively less predictable (more informationally ef¯cient) than stock market returns, which in turn are less predictable than Eurodeposit rate returns. Finally, we estimate linear autoregressive (AR) parametric models and compare the evolution of SIM against that of mean squared forecast error (MSE) for changing sample size. The results suggest that although the levels of SIM and MSE can be signi¯cantly correlated,

---

[2] This is the subject of current research which is discussed in the conclusion.

[3] See Fraser (1989) and Palus (1993).

their stationary first differences are uncorrelated. However, it is shown that the first difference of SIM Granger-causes the first difference of MSE for 6 out of the 8 time series.

The remainder of the paper is arranged as follows. In the theoretical Section 2 the information-theoretic measures of sample, conditional and relative entropy reviewed and the self-information measure obtained as one component of finite-sample predictability. The latter is shown to be increasing in the conditional entropy of the dgp and decreasing in its SIM. Section 3 discusses the financial returns data. In Section 4 the implications of the theoretical properties of SIM are explored for simulated and actual data. Section 5 concludes.

2 Sample entropy and self-information

An n–vector of observations $\{x_t, x_{t-1}, \ldots, x_{t-n+1}\}$ from a discretely-observed financial variable X is denoted $\underline{x}^n$. The sample entropy of X is $H_n^k(X) = -\sum_{i=1}^k p_i \log p_i$, where $\{p_i\}_{i=1}^k$ are the empirical probabilities of observations partitioned into equally-spaced percentiles $i = 1, \ldots, k$. The log is to base 2, so the entropy units are information bits. The percentile ensemble is defined as the alphabet of the dgp. Its length k defines the partition: a finer (coarser) partition amounts to a bigger (smaller) alphabet. For discrete random variables, the value of maximum entropy occurs for the uniform probability density function (pdf) where $p_i = 1/k$ for all i: $\max H_n^k(X) = \log k$.[4] In "normal" circumstances the alphabet length is invariant to the sample size. However, the arrival of an "extreme" observation in-sample may necessitate a marginal increase in alphabet length from k to k + 1.

The joint and conditional entropies of random variables $\{X_1, X_2, \ldots, X_n\}$ with joint pdf $p(x_1, \ldots, x_n)$ are respectively:

$$H_n^k(X_1, \ldots, X_n) = -\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} p(x_1, \ldots, x_n) \log p(x_1, \ldots, x_n) \qquad (1)$$

$$H_n^k(X_n \mid \underline{X}^{n-1}) = \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} p(x_n \mid \underline{x}^{n-1}) \log p(x_n \mid \underline{x}^{n-1}), \qquad (2)$$

[4] See Applebaum (1996) and Golan, Judge and Miller (1996).

where $p(x_n \mid \underline{x}^{n-1})$ is the conditional pdf of X_n given past observations $\underline{X}^{n-1}$ $\{x_1, \ldots, x_{n-1}\}$.

In Tambakis (2000), univariate non-parametric predictability $P_n^k(X_n; \underline{X}^{n-1})$ of random variable X_n as a function of the information set $\underline{X}^{n-1}$ was defined as its mutual information $I(X_n; \underline{X}^{n-1})$ normalized by the maximum entropy of a discrete dgp with a k-alphabet:

$$
\begin{aligned}
P_n^k(X_n; \underline{X}^{n-1}) &= \frac{I(X_n; \underline{X}^{n-1})}{\log k} \\
&= \frac{H_n^k(X_n) - H_n^k(X_n \mid \underline{X}^{n-1})}{\log k}
\end{aligned}
\qquad (3)
$$

Normalization implies that $P_n^k(X_n)$ is and bounded between 0 and 1. Fraser (1989) and Palus (1993) define the numerator of the predictability statistic to be the (non-linear) redundancy measure. Note that mutual information in the numerator of (3) is symmetric and non-negative. If X and Y are independent then $H(X \mid Y) = H(X)$, so $I(X; Y) = 0$ and Y is useless in predicting X. In contrast, if X is a deterministic function of Y then $H(X \mid Y) = 0$ and mutual information is maximized.

The entropy rate of a random sequence $\{X_i\}_{i=1}^n$ is defined as $H^k(n) = \lim_{n \to \infty} \frac{1}{n} H_n^k(X_1, \ldots, X_n)$. Thus $H^k(n)$ is the limit of the average joint entropy per observation. If the $\{X_i\}$ sequence is iid then $H_n^k(X_1, \ldots, X_n) = nH_n^k(X_1)$, implying $H^k(n) = \lim_{n \to \infty} nH_n^k(X_n)/n = H_n^k(X_n)$. Khinchin (1957) shows the existence of $H^k(n)$ for strictly stationary processes. Moreover, for strictly stationary ergodic processes conditional entropy converges to the entropy rate. Asymptotic predictability then becomes:[5]

$$
\lim_{n \to \infty} P_n^k(X_n; \underline{X}^{n-1}) = \frac{1}{\log k}\left[\lim_{n \to \infty} (H_n^k(X_n)) - H^k(n)\right] \qquad (4)
$$

The second term in (4) converges to the entropy rate. Thus asymptotic predictability is smallest (efficiency is greatest) when $\lim_{n \to \infty} H_n^k(X_n)/\log k$ is maximized, while it is zero for an iid sequence. We therefore define the

[5] For a proof see Cover and Thomas (1991).

resulting ¯nite-sample statistic as the self-information measure (SIM) of a sample of size n and alphabet length k:

$$SIM_n^k = \frac{H_n^k(X_n)}{\log k} \tag{5}$$

3 The data

The variation of SIM with sample size is examined for long daily returns series from 8 ¯nancial markets. The stock index data consist of daily returns of the Dow Jones (DJIA) and Nikkei (NIKKEI) averages over the period 1=1=1973{6=4=1998. The FX market data consist of daily returns of the deutschemark/dollar (DM/\$), yen/dollar (JPY/\$) and sterling/dollar (\$/\$) spot exchange rates over the same period. This yields a total of 6; 591 observations. The interest rate data are daily returns on overnight Eurodeposit rates denominated in US dollars (EURO{\$), German marks (EURO{DM) and sterling (EURO{\$). The Eurodeposit rate sample period is 3=1=1975{2=4=1998, a total 6; 065 observations. All returns time series are weakly stationary over the sample period.[6] The alphabet length is ¯xed at $k = 100$ percentiles, and the last in-sample observation is ¯xed at 6; 500 for stock market and FX returns, and 6; 000 for Eurodeposit rate returns. The last in-sample observation is the out-of-sample cut-o® point.

4 Returns predictability and market e±ciency

4.1 Monte Carlo simulations

The evolution of SIM_n^k with changing sample size is ¯rst illustrated using simulated data from the Gaussian (0; 1), uniform [0; 1], gamma (1; 1) and Poisson (1) distributions. For each sample size from $n = 1$ to 1000, the simulated n{vector is partitioned using a discrete alphabet of ¯xed length $k = 100$, thus yields k^n possible output signals. Given the alphabet length, the density function for a given sample size corresponds to the empirical frequency distribution. Figure 1 shows that SIM_n^k increases with sam-

[6] The results of the ADF tests have been omitted for space constraints. They are available from the author upon request.

ple size.[7] This property is robust to alternative distributions and alphabet length. Also note that, asymptotically, the maximum SIM occurs for the uniform pdf. This follows from the maximum entropy principle because the simulated data, although drawn from continuous pdf's, have been discretized. Therefore, $n^{*} = \arg\max_n SIM_n^k$ coincides with the maximum sample size, implying that predictability increases in the number of observations. Intuitively, "more is better" because the underlying (true) dgp is strictly stationary. We now turn to analyze the evolution of the self-information measure with changing sample size and its relation to predictability and weak-form efficiency for actual financial data.

4.2 Self-information and sample size

The variation of SIM with sample size yields important insights regarding the informational value of extreme observations. The empirical frequency distribution is used to compute SIM_n^k, where n denotes the variable sample size and k denotes the fixed alphabet length. For both time series, the cut-off obervation of the in-sample data is fixed at $6;500$, so the length of the out-of-sample period is fixed at 91. The sample size is increased incrementally from n_{MIN} to n_{MAX} observations by moving the first in-sample observation backward one day at a time. SIM_n^k measure is computed for each rolling sample size n. For illustration purposes, we set $n_{MIN} = 400$ and $n_{MAX} = 4;400$, or $4;000$ rolling sample sizes. The data is partitioned in $k = 100$ equally-spaced percentiles. The maximum (uniform) entropy is therefore $\log_2 100 = 6{:}64$.[8]

Figure 2 plots all eight datasets' SIM as the sample size increases from $n = 1$ (11/28/1997) to $n = 6;000$ (12/2/1974). In The sample size n is plotted against SIM_n^k and the level of each time series on the left and right vertical axes, respectively. Because the first in-sample observation is fixed at $11{=}28{=}1997$, the levels are plotted from right to left so that smaller sample sizes correspond to more recent observations. We make the following observations. First, unlike the simulated datasets in Figure 1, SIM_n^k is non-monotonic in sample size: the normalized sample entropy is very sensitive

[7] The increase is not monotonic because there is only one sample: if many random samples of length n were generated then AIC_n^k would be smoothly increasing in n.

[8] Following Golan, Judge and Miller (1996), the alphabet length k must be less than the sample size n in order for the recovery of the probability vector $\{p_i\}_{i=1}^k$ to be well-defined. This constraint is unlikely to be binding in practice.

to the arrival of extreme observations (outliers). In particular, the evolution of SIM_n^k is characterized by discontinuities corresponding to the inclusion of extreme events in-sample. Because the results of standard unit root tests for non-stationarity may be misleading, we focus on comparing the variation of SIM with sample size across markets.

For small sample sizes ($n < 50$), SIM rises from zero to a value close to its maximum over all sample sizes. The predictability of the ¯rst out-of-sample observation falls concurrently. In the limit, when $n = 1$ the empirical frequency distribution is degenerate so the sample entropy is zero. As the sample size increases, it approaches the maximum SIM over all sample sizes. This corresponds to maximum weak-form e±ciency, or minimum predictability relative to the uniform distribution benchmark. The SIM_n^k{maximizing sample size $n^{¤}$ is less than 500 obervations in all cases. A smaller $n^{¤}$ suggests that the dgp is relatively less predictable using past information, or relatively more weak-form e±cient. Table 1 shows the sample sizes $n^{¤}$ and the value of the maximum SIM for each returns series:

Table 1
Maximum SIM sample sizes

Series	$n^{¤}$	$SIM_{n^{¤}}^k$
DM/$	152	0:8683
$/$	83	0:8511
JPY/$	144	0:8422
NIKKEI	236	0:8052
DJIA	214	0:7601
EURO-DM	237	0:6543
EURO-$	66	0:5926
EURO-$	287	0:5707

The results indicate that the EURO-$ market ($n^{¤} = 66$) is relatively more e±cient, while the EURO-$ market ($n^{¤} = 287$) is relatively less e±cient. $/$ is the second most e±cient market, while the DJIA and NIKKEI are almost equally e±cient. We brie°y focus on the two stock market indices. The most dramatic break occurs in the SIM of Dow Jones returns, which drops discontinuously by 30% (from 70 to 49 percent) at sample size $n = 2;640$. This corresponds to 10=19=1987, the stock market crash of October 1987.

Intuitively, the (ex ante extremely unlikely) realization of an extreme event such as a crash occurring on any one day lowers the average information content conditional upon an extreme event. Consequently, excluding the crash from the sample increases the normalized entropy and lowers the dgp's predictability. The Nikkei returns' SIM displays two smaller breaks, the second on 10=19=1987, and the ¯rst at sample size $1;908$, corresponding to 8=19=1978. Following October 1987 and through to $n = 6;500$, the SIM of Dow Jones returns remains close to 50 percent, while that of the Nikkei declines gradually from 60 to 50 percent.

For larger sample sizes ($50 < n < 500$), the SIM of exchange rate returns exceeds that of stock market returns, which in turn exceeds that of Eurodeposit rate returns. Within FX markets, DM/$ returns are somewhat less predictable than GBP/$ returns, which in turn are relatively less predictable than JPY/$ returns. Within Eurodeposit markets, overnight EURO-DM returns are less predictable than either EURO-$ or EURO-$ returns. The SIM of EURO-$ returns is noticeably smoother than that of the EURO-$ and EURO-DM. The SIM of EURO-$ returns is characterized by two breaks: the ¯rst corresponds to the 1973 oil price shock and the second to sterling's forced exit from the European Monetary System in 1992. In each case, including the relevant observation in-sample sharply lowers the average information content and increases the ¯nite-sample predictability of the dataset. Finally, daily Nikkei returns are less predictable than Dow Jones returns. Overall, the value of past information for univariate prediction is smaller in the FX markets than in the stock markets, which in turn is smaller than the overnight Eurodeposit markets. Therefore, overnight Eurodeposit rate returns are relatively more predictable than daily stock market returns, which in turn are more predictable than daily exchange rate returns.

4.3 Self-information and parametric forecast error

Figure 3 plots the variation in levels of SIM_n^k and mean square forecasting error (MSE_n^j) with changing sample size n for the DJIA and NIKKEI time series. The plotted values are conditional upon the alphabet length, forecasting model, AR order information criterion and forecast horizon. The forecast horizon is ¯xed at $j = 10$ days-ahead and the forecasts are dynamic. The MSE_n^j statistic is computed using a linear AR speci¯cation, where the lag order is determined is determined using the modified Schwartz information criterion of Neumaier and Schneider (1997). The AR coefficients and the lag order specification are reestimated at each rolling sample size.

The sample correlation coefficients of the levels of SIM^{100}_{4000} and MSE^{10}_{4000} are $\hat{\rho}_{DJIA} = 0.80$ for Dow Jones returns and $\hat{\rho}_{NIK} = 0.005$ for Nikkei returns. The significance of these values is examined by bootstrapping the estimated correlation coefficients $\hat{\rho}(SIM^k_n; MSE^j_n)$. The right panels show the empirical distribution of each bootstrap correlation statistic for $1{,}000$ bootstrap replications of the SIM^k_n and MSE^j_n vectors. The bootstrap histograms suggest that the true correlation coefficient is significantly positive for Dow Jones returns and zero for Nikkei returns. However, as the series have strong breaks, standard ADF tests for non-stationarity have low power.[9] Indeed, the correlations between the stationary first differences of the two statistics are insignificant: $\hat{\rho}_{DJIA}(\Delta SIM^{100}_{4000}; \Delta MSE^{10}_{4000}) = 0.004$ and $\hat{\rho}_{NIK}(\Delta SIM^{100}_{4000}; \Delta MSE^{10}_{4000}) = -0.0009$. Therefore, there is insufficient evidence that selecting sample size so as to maximize SIM contributes to a lower mean squared error.

Moreover, correlation does not imply causation. Table 2 reports the results of Granger causality tests on whether the current value of the first-differenced mean squared errors (ΔMSE) can be explained by lagged values of the first-differenced self-information measure (ΔSIM), and vice versa. In each case, we report the shortest significant lag specification:

[9] See Perron (1997) and the references therein.

Table 2
¢SIM and ¢MSE: Granger causality tests[10]

DJIA	Lags/n	F-stat	Prob
H_0 : ¢MSE does not cause ¢SIM	l = 5	1:05	0:38740
H_0 : ¢SIM does not cause ¢MSE	3994	4:70	0:00028
NIKKEI	Lags/n	F-stat	Prob
H_0 : ¢MSE does not cause ¢SIM	l = 3	0:16	0:92117
H_0 : ¢SIM does not cause ¢MSE	3996	47:22	0:00000
DM/$	Lags/n	F-stat	Prob
H_0 : ¢MSE does not cause ¢SIM	l = 1	13:91	0:00019
H_0 : ¢SIM does not cause ¢MSE	3998	27:82	0:00000
$/$	Lags/n	F-stat	Prob
H_0 : ¢MSE does not cause ¢SIM	l = 1	1:45	0:22846
H_0 : ¢SIM does not cause ¢MSE	3998	5:49	0:01913
JPY/$	Lags/n	F-stat	Prob
H_0 : ¢MSE does not cause ¢SIM	l = 10	26:79	0:00000
H_0 : ¢SIM does not cause ¢MSE	3989	27:09	0:00000
EURO{$	Lags/n	F-stat	Prob
H_0 : ¢MSE does not cause ¢SIM	l = 10	9:45	0:00000
H_0 : ¢SIM does not cause ¢MSE	3989	0:46	0:91758
EURO{DM	Lags/n	F-stat	Prob
H_0 : ¢MSE does not cause ¢SIM	l = 10	0:11	0:99974
H_0 : ¢SIM does not cause ¢MSE	3989	0:29	0:98305
EURO{$	Lags/n	F-stat	Prob
H_0 : ¢MSE does not cause ¢SIM	l = 10	0:19	0:99713
H_0 : ¢SIM does not cause ¢MSE	3989	108:9	0:00000

[10] In each case the AIC and MSE statistics are di®erenced and I(0). The Granger tests computes the F{statistic for the null hypothesis that the coe±cients on all l lagged explanatory variables are zero. The estimation uses n observations.

The results indicate that for DJIA, NIKKEI, $/$ and EURO-$ returns the null hypothesis that ¢SIM does not Granger-cause ¢MSE is strongly rejected, whereas the null that ¢MSE does not Granger-cause ¢SIM is not. For DM/$ and JPY/$ returns the null of no causality is strongly rejected in both directions. For the EURO{$ returns ¢MSE Granger-causes ¢SIM but not vice versa. Finally, for EURO{DM returns the null of no causality cannot be rejected in either direction. We tentatively conclude that the forecast accuracy of linear parametric models may be explained by the variation in ¢SIM.

5 Conclusion

This paper studied the evolution of the self information measure (SIM) for ¯nancial data from the FX, stock and interest rate markets. The methodology o®ers a robust non-parametric indicator of ¯nancial return predictability and market e±ciency based on the sample size. First, it was shown that higher SIM values reduce predictability and improves market e±ciency. In particular, interest rate returns were found to be more predictable than stock markets returns, which in turn were more predictable than spot FX returns. Second, for each dataset the sample size was determined corresponding to maximum SIM, and thus to greatest weak-form e±ciency relative to the uniform distribution benchmark. Finally, the ¯nancial returns data also indicated that the levels of SIM and mean square forecast error are spuriously correlated as they are non-stationary and characterized by structural breaks. However, it was shown that on ¯rst di®erences SIM Granger-causes MSE and not vice versa.

The present theoretical framework can be extended by relating parametric and parametric predictability using Fisher information. An empirical extension involves computing bivariate predictability using the mutual information statistic between two di®erent datasets. This was the second key factor of current parametric risk measurement practice referred to in Section 1. For example, the evolution of the mutual information between stock market and interest rate data at a daily | or higher | frequency should re°ect the degree to which past stock market events a®ect current and future interest rate returns, and vice versa. The same can be said about the mutual information between stock market and spot FX data. A bivariate approach to informational content thus suggests a new and powerful tool for updating the relative value of information °ows from di®erent ¯nancial markets in real time. The °exibility of the non-parametric approach is particularly useful against the recent background of contagion between di®erent markets and across international ¯nancial centers. Such an extension is the subject of current research.

References

[1] Abarbanel, H.D.I. 1996. Analysis of Observed Chaotic Data. New York: Springer Verlag.

[2] Applebaum, D. 1996, Probability and Information Theory: An Integrated Approach, Cambridge: Cambridge University Press.

[3] Christo®ersen P., and Diebold, F.X. 1996. Further Results on Forecasting and Model Selection Under Asymmetric Loss. Journal of Applied Econometrics 11: 561-572.

[4] Cover, T. and Thomas, J. 1991, Elements of Information Theory. New York: Wiley Interscience.

[5] Fama, E.F. 1970. E±cient Capital Markets: A Review of Theory and Empirical Work. Journal of Finance 25.

[6] Fama, E.F. 1991. E±cient Capital Markets: II. Journal of Finance 46.

[7] Feder, M. and Merhav N. 1994. Relations Between Entropy and Error Probability. IEEE Transactions on Information Theory 40(1): 259-266.

[8] Fraser, A.M. 1989. Reconstructing Attractors from Scalar Time Series: A Comparison of Singular System and Redundancy Criteria. Physica D 34: 391-404.

[9] Golan, A., Judge, G. and Miller, D. 1996, Maximum Entropy Econometrics: Robust Estimation With Limited Data. New York: John Wiley & Sons.

[10] Granger, C.W.J. and Newbold D. 1986: Forecasting Economic Time Series, Cambridge: Cambridge University Press.

[11] Hamilton, J. (1994): Time Series Analysis. Princeton: Princeton University Press

[12] Khinchin, A.I. (1957): Mathematical Foundations of Information Theory. New York: Dover.

[13] Neumaier, A. and Schneider, T. 1997, Multivariate Autoregressive and Ohrstein-Uhlenbeck processes: Evidence for order, parameters, spectral information and con¯dence regions, submitted to ACM Trans. Math. Soft.

[14] Palus, M. 1993. Identifying and Quantifying Chaos by Using Information-Theoretic Functionals, in Time Series Prediction: Forecasting the Future and Understanding the Past, eds. A.S. Weigend and N.A. Gershenfeld, SFI Studies in the Sciences of Complexity, Proc. Vol. XV, Addison-Wesley.

[15] Perron, P. 1997. Further Evidence on Breaking Trend Functions in Macroeconomic Variables. Journal of Econometrics 80: 355-385.

[16] Tambakis, D.N. 2000. Information-Theoretic Sample Size Selection for Linear Prediction. Neural Network World 10(1-2): 73-79.