DYNAMICAL COMPLEXITY ANALYSIS OF MULTIVARIATE FINANCIAL DATA

Wenjun Er and Danilo P. Mandic

Electrical and Electronic Engineering Department, Imperial College London, UK
E-mail: {wenjun.er08, d.mandic}@imperial.ac.uk

ABSTRACT

Characterization of joint dynamics of multivariate financial time series calls for the analysis based on joint intrinsic temporal and information-theoretic scales. Yet a rigorous account of dynamical complexity of such time series is hampered by the univariate natures and mathematical artefacts associated with the existing methods. To that end, we employ multivariate multiscale entropy (MMSE) in order to associate multivariate complexity with long-range correlations, direct and indirect couplings, and synchronies among the data channels. Simulations on major stock indices support the approach.

Index Terms— Dynamical complexity, long term correlation, multivariate entropy, market efficiency, Hurst exponent

1. INTRODUCTION

Financial indices of the same kind tend to exhibit coupled dynamical behaviour, yet the most intricate properties and descriptors, such as the Hurst exponent, long term correlation, irregularity and entropy are typically evaluated in an index-by-index manner, over multiple univariate series within a particular portfolio [1, 2]. While linear stochastic models, such as the autoregressive moving average (ARMA), and the links with chaos and fractals, are traditionally the first step in the analysis of both univariate and multivariate financial indices, model based methods which operate on a single scale are not likely to explain complex patterns of evolution or direct and indirect causalities [3, 4].

Nonlinear time series analysis methods are well suited to finding hidden patterns in time series, together with various latencies, periodicities and fractal behaviour. However, they also typically operate on a single scale and tend to be univariate [5]. More recently, surrogate data methods in conjunction with various test statistics have been used to assess the degree of nonlinearity or uncertainty in time series. These methods also operate only on stationary data and would hence mistake nonstationarity for nonlinearity [5, 6]. One such approach is the delay vector variance (DVV) method [7,8], which is based on local predictability in phase space and has been shown to robustly quantify the degree of nonlinearity and uncertainty in time series. It has also been successfully applied to econometric time series, specifically in the context of business cycle modelling [9].

Figure 1 illustrates models spanned by the properties of stochasticity, determinism, linearity, and nonlinearity (modified from [5]). Observe that the most frequently used and mathematically tractable (i) linear stochastic models (ARMA) and (ii) nonlinear ARMA models (e.g. neural networks), are diametrally opposite from (iii) nonlinear deterministic chaotic models. Yet most real world signals occupy the areas denoted by ‘a’,’b’,’c’, and ’?’, for which models (i)-(iii) are inadequate as they assume rigid structure in data.

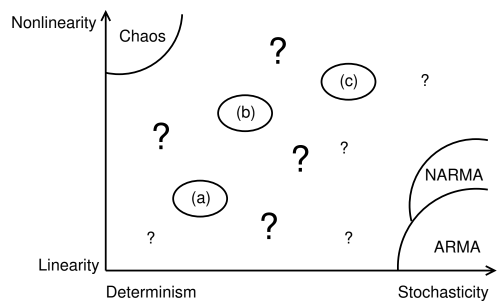


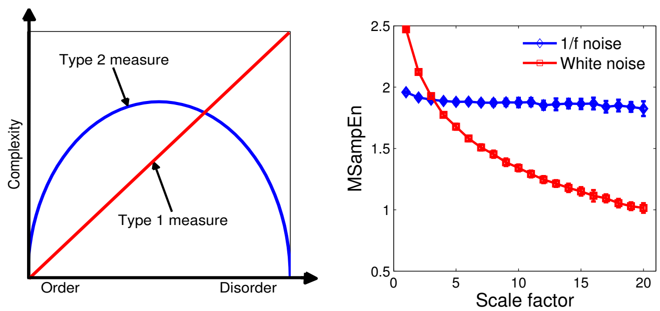
Fig. 1: Standard data analysis models and their limitations

Finding inherent structure in time series is closely related to complexity science and is typically based on delay vector (time-delay embedded) reconstruction [5, 10] which lends itself to phase space representation and allows for the estimation of dynamical characteristics including attractor dimensions, Lyapunov exponents and entropy measures [5]. While there is a lack in consistency within the field of complexity science on its exact definition, a common consensus is that a signal is considered dynamically complex when it spans the whole range between randomness and periodicity [11, 12].

This work employs some recently developed multivariate entropy-based techniques to quantify dynamical complexity of multichannel financial data. Our approach is based on the multivariate multiscale entropy (MMSE) method [13, 14], which evaluates the conditional probability that similar delay vectors remain close to one another (within a threshold expressed in a certain metric) after increasing the dimension in the state space. The multivariate sample entropy estimate is then evaluated over multiple time scales, thus reflecting the underlying multivariate dynamical complexity. The usefulness and enhanced insight offered by such an approach is illustrated on several stocks analyzed over periods of different geopolitical and socio-economics conditions.

2. PRACTICAL COMPLEXITY ESTIMATORS

Standard complexity estimators operate on a single scale (directly on sampled data) and relate complexity to the degree of randomness. These complexity estimators typically use entropy as a basis for the complexity metric, for instance, approximate entropy (ApEn) [15] and sample entropy (SampEn) [16]. In this sense, the higher the entropy the more dynamically complex the signal at hand, as illustrated in the *Type 1* measure graph in Figure 2(a).



(a) Complexity measures (b) Multiscale entropy

Fig. 2: The need for multiscale complexity measures

However, this is counterintuitive, as that would mean that the randomized (or surrogate) data would have higher complexity than the original time series, which may exhibit long range correlations and couplings - signatures of complexity. Therefore, truly complex signals are neither completely random nor deterministic, and exhibit rich dynamical properties over intrinsic spatio-temporal scales - physically meaningful complexity measures should therefore be of *Type 2* in Figure 2(a). The recent *univariate* multiscale entropy (MSE) [17] circumvents the misleading results of the standard measures by evaluating the sample entropy across multiple time scales, and offers a physically meaningful interpretation (Fig 2(b)):

- A deterministic signal has low complexity and will have low entropy across all the scales;
- A random signal has highest entropy for scale 1, but its entropy will rapidly decrease with scale, as such signals have no structure;
- A truly complex signal (e.g. 1/f noise) has lower entropy than random signal for scale 1, but will maintain its entropy values along the increasing time scales.

One limitation of the MSE measure is its inability to process multivariate time series, thus restricting the complexity analysis of multiple data channels to a channel-wise approach, which does not consider couplings, long term correlation and causality. In view of this shortcoming, the multivariate multiscale entropy (MMSE) [13] measure expands on the concepts underpinning the MSE and the multivariate sample entropy (MSampEn) [13] measures, to provide a rigorous analysis of multivariate complexity, thus accounting for within- and cross-channel correlations, and direct and indirect couplings.

3. MULTIVARIATE MULTISCALE ENTROPY

To evaluate the dynamical complexity of multivariate time series, we first need to estimate the joint probability function from the data, as a basis for multivariate entropy calculation. The first multivariate sample entropy approach was recently introduced in [13, 18], and is outlined below.

3.1. Multivariate Sample Entropy (MSampEn)

By time-delay embedded reconstruction [5], the composite delay vector is given by:

$$X_m(i) = \begin{pmatrix} X_{1,i}, X_{1,i+\tau_1}, \ldots, X_{1,i+(M_1-1)\tau_1} \\ X_{2,i}, X_{2,i+\tau_2}, \ldots, X_{1,i+(M_2-1)\tau_2} \\ \vdots \\ X_{c,i}, X_{c,i+\tau_c}, \ldots, X_{c,i+(M_c-1)\tau_c} \end{pmatrix} \quad (1)$$

where $M = [m_1, m_2, \ldots, m_c]$ denotes the multidimensional embedding vector and $\tau = [\tau_1, \tau_2, \ldots, \tau_c]$ the time delay vector of the composite delay vector containing c data channels. The parameters m_i and τ_i are the channel-wise embedding dimension and time delay, and can be found jointly [19].
The MSampEn multivariate entropy estimator [13] runs as:

1. Define $n = max(M) * max(\tau)$ and construct $(N - n)$ composite delay vectors $X_m(i) \in \mathbb{R}^c$ for $i = 1, 2, \ldots, N - n$.
2. Find the maximum norm between all delay vectors $d[X_M(i), X_M(j)] = max_{k=1,\ldots,m}\{|x_i(k) - x_j(k)|\}$.
3. For each of the composite delay vectors within the set, count the number of delay vectors that are within a tolerance r. Denote this count as a probability given by $P_i^m = \frac{1}{N-n-1} * count\{d[X_M(i), X_M(j)] \leq r\}, i \neq j$. Repeat for every composite delay vector and average the probability to obtain $P_1 = \frac{1}{N-n} * \sum_{i=1}^{N-n} P_i^m$.
4. Increase the embedding dimension from m to $(m+1)$ in as many ways as there are data channels, c. Construct the composite delay vectors for each of the c possible $(m+1)$ embedding dimensions to yield $c(N-n)$ composite delay vectors. Repeat Step 3 to arrive at probability $P_2 = \frac{1}{c(N-n)} * \sum_{i=1}^{c(N-n)} P_i^{m+1}$.
5. The MSampEn is: $MSampEn(M, \tau, r, N) = -ln\frac{P_2}{P_1}$.

This extension of sample entropy to the multivariate case requires some careful considerations [13]:

- To avoid any bias towards data channels of larger amplitudes, amplitude scaling to within $(0, 1)$ is used.
- The probability for the $(m+1)$-dimensional case is calculated in a rigorous way across all subspaces so as to account for the correlations across data channels.
- The variance of each data channel is normalized to unity, and the embedding dimensions and time delays must be large enough so that the results are not governed by noise.

4. MULTIVARIATE DYNAMICAL COMPLEXITY

The multivariate MSE measure (MMSE) [13, 18] evaluates MSampEn along multiple temporal scales defined through the so-called *coarse graining* process:

1. For a user-defined range of temporal scale factors ϵ, perform the coarse-graining process for each channel of the multivariate time series $\{x_{k,i}\}$ where $k = 1\ldots,c$ and $i = 1,\ldots,N$ for each scale factor ϵ, by averaging over the scale factor ϵ, as

$$x^{\epsilon}_{coarse\,k,j} = \frac{1}{\epsilon}\sum_{i=(j-1)*\epsilon+1}^{j*\epsilon} x_{k,i}, \qquad (2)$$

for $k = 1,\ldots,c$ and $1 \leq j \leq \frac{N}{\epsilon}$ to produce the coarse-grained time series.

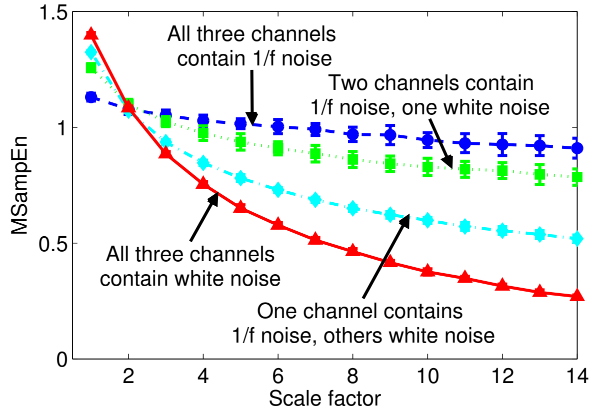
2. Compute the MSampEn as in Section 3.1 for each coarse-grained multivariate time series and plot such multiscale MSampEn against the scale factors ϵ.

4.1. Physical Interpretation of MMSE Plot

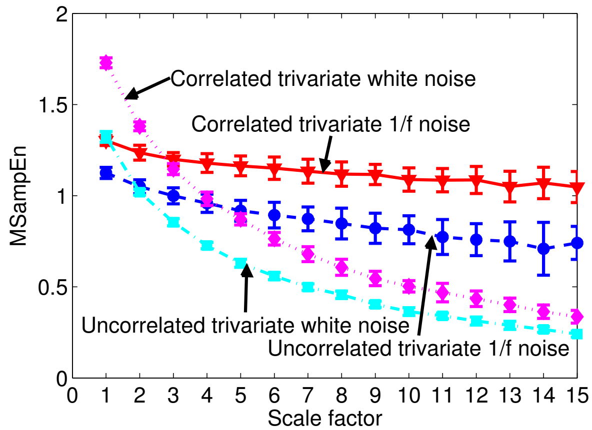
The complexity of the multivariate time series is interpreted in the following way [13]:

- A multivariate time series whose MMSE plot has higher values for larger scales than MMSE of another multivariate time series, is more dynamically complex.
- A multivariate time series whose MMSE plot exhibits a monotonically decreasing behaviour with scale factors contains useful information only at small scales - this is either purely stochastic or deterministic time series.
- A multivariate time series whose MMSE plot is either monotonically increasing or constant with an increase in the scale factor exhibits long-term correlation structures (direct or indirect), a signature of dynamically complex time series.

Figure 3 illustrates the concept of multivariate complexity on trivariate data composed of correlated and uncorrelated white (low complexity) and $1/f$ (high complexity) noises. In Fig 3(a), observe that MMSE values increased at higher scales with an increase in the number of $1/f$ noise channels. As expected, for scale 1 ($\epsilon = 1$) the entropy of the trivariate purely white signal was highest, and it decreased rapidly with the scale index. On the contrary, the complexity of the trivariate $1/f$ noise was at its lowest for scale 1, but maintained its value with an increase in scale factor, indicating a truly complex signal. Figure 3(b) shows that structural complexity manifests itself in both within-channel correlation (as shown in Figure 3(a)), and also in across-channel correlation and coupling. Observe that the the complexity of both the white and $1/f$ channels were higher when the data channels were correlated, conforming with the physics, and that (as expected) the uncorrelated trivariate white signal had lowest complexity, while the correlated $1/f$ signal had highest complexity.



(a) Complexity of trivariate white and $1/f$ channel mixtures



(b) Complexity of trivariate white and $1/f$ correlated channels

Fig. 3: MMSE plot for multivariate white and $1/f$ channels.

5. MULTIVARIATE COMPLEXITY IN FINANCE

A macro-approach to complexity analysis was employed by using major stock indices as the multivariate time series considered. The motivation was that while stock indices may have their shortcomings when used as benchmarks, major indices are widely accepted as accurate indicators of the financial health of the sectors they represent [20], and to some extent even the world economy. In addition, key events affecting the global economy are well understood and known. With this in mind, we analyzed the following four major stock indices benchmarking the economy of the United States of America: 1. *Dow Jones Industrial Average*, 2. *NASDAQ Composite*, 3. *Standard & Poor's 500*, 4. *Russell 2000*.

Based on our interpretation of key geopolitical and socio-economic events that affected the U.S. and world economy at large, financial data between 01/01/1991 to 31/12/2011 were segmented into the following four periods:

P1 from 01/01/1991 to 31/12/1999, we witnessed economic boom and exceptional levels of growth in the markets which were buoyed by optimism in technological developments, hence the term "dot-com boom".

P2 from 01/01/2000 to 31/12/2003, is characterised by uncertainty and high volatility, culminating in one of the biggest financial meltdowns since the Great Depression. The financial crisis was further compounded by the 9/11 terrorist attacks and its aftermaths.

P3 from 01/01/2004 to 31/12/2007, is a period of recovery where the renewed enthusiasm amongst investors saw huge investment opportunities in undervalued stocks, which aided market recovery.

P4 from 01/01/2008 to 31/12/2011, is characterized by bursting of the sub-prime mortgage bubble and debt crisis, which plunged the world economy into a deep recession which still has lingering effects today.

In the simulations, the closing prices of the four indices were taken as a quadrivariate time series in the MMSE analysis, with the parameters $\epsilon = 1, \ldots, 10$, $M = [3, 3, 3, 3]$ and $\tau = [1, 1, 1, 1]$, giving the MMSE plot in Figure 4. Observe

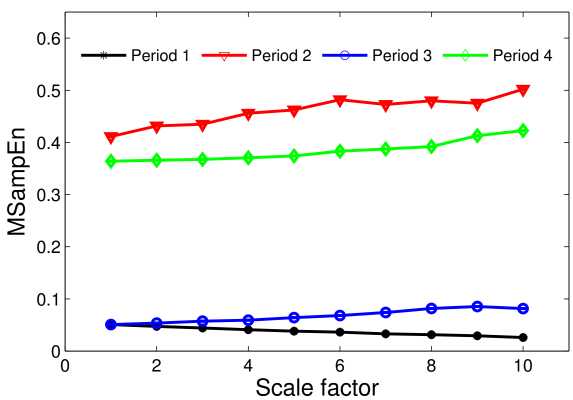


Fig. 4: Complexity of quadrivariate financial time series.

that the quadrivariate financial time series in all the four periods maintained constant trends across the temporal scale factors ϵ. This suggests that all the financial time series exhibit long-range correlation and are dynamically complex - this is intuitive as it is not expected that financial time series are purely random/deterministic in nature. The results also suggest higher complexity during periods of uncertainty and high volatility, as compared to periods of relative stability and growth. The long-range correlation in the financial indices considered are also confirmed through the Hurst exponent, H, a widely used tool for evaluating long-range correlation in a time series. Table 1 shows the results using the Rescaled Range (R/S) method [21], whereby a Hurst exponent value in the range of $0.5 < H < 1$ indicates that the time series under study possesses a long-range positive autocorrelation structure while a value in the range of $0 < H < 0.5$ indicates a long-term negative autocorrelation structure. Observe that every segment of the financial data in Table 1 exhibited long-range positive autocorrelation structure, whose values are consistent with the complexity results obtained from the MMSE analysis in Figure 4.

Table 1: Hurst Exponent Estimates by R/S Method

Period	DJIA	NASDAQ	S&P500	RUSSELL2000
1	1.0087	1.0015	0.98636	1.0018
2	0.90905	0.96882	0.97016	0.94115
3	0.99256	0.97731	0.97348	0.97529
4	0.98167	0.96966	0.99667	0.97101

Tracking complexity. The MMSE measure was next applied to quadrivariate log-prices through the highly volatile period at the turn of the millenium. The tile diagram in Fig. 5 indicates that the MMSE evaluated over 4-year sliding windows captured a trend in the degree of complexity, which increased from the low 1991-1994 level to peak at around 1999-2002, followed by a decrease to the low 2002-2005 level.

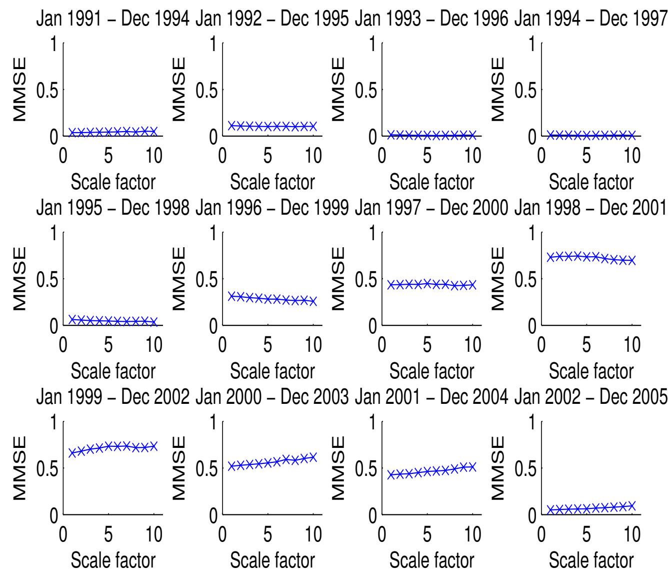


Fig. 5: Complexity of quadrivariate log-prices from 1991–2005.

5.1. Physical Interpretation

The MMSE analysis is in line with our intuition – since the prices reflect changing market conditions, higher complexity is observed during periods of uncertainty and high volatility, where events and shocks cause irregular shifts. Conversely, during periods of financial stability, index prices follow well-studied business cycles and exhibit more regularity and lower complexity. Therefore, complexity measures may act as a proxy for the volatility levels within the markets. A comparison of the results in Figure 5 with the S&P500 volatility index (VIX) shows similar trends during the period under study.

6. CONCLUSION

We have investigated the usefulness of multivariate dynamical complexity in the characterisation of coupled financial indices. The analysis conducted on selected financial time series has showed the changes in complexity for different periods of financial outlook. We have also established a link between the measure of dynamical complexity and the volatility within the markets. This is supported by both economic intuition and evidence from past economic cycles.

7. REFERENCES

[1] R. D. Edwards, J. Magee, and W. Bassetti, *Technical Analysis of Stock Trends*, Prentice-Hall, 9th edition, 2007.

[2] J. Hull, *Options, Futures and Other Derivatives*, Prentice-Hall, 7th edition, 2009.

[3] J. A. Matos, S. M. A. Gama, H. J. Ruskin, and A. A. Sharkasi, "Time and scale Hurst exponent analysis for financial markets," vol. 387, pp. 3910–3915, 2008.

[4] I. Faes and G. Nollo, "Extended causal modelling to assess partial directed coherence in multiple time series with significant instantaneous interactions," *Biological Cybernetics*, vol. 103, no. 5, pp. 387–400, 2010.

[5] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, 2000.

[6] T. Schreiber and A. Schmitz, "Surrogate time series," *Physica D*, vol. 142, pp. 346–382, 2000.

[7] D. P. Mandic, M. Chen, T. Gautama, M. M. Van Hulle, and A. Constantinides, "On the characterisation of the deterministic/stochastic and linear/nonlinear nature of time series," *Proceedings of the Royal Society A*, vol. 464, no. 2093, pp. 1141–1160, 2008.

[8] T. Gautama, D. P. Mandic, and M. Van Hulle, "A novel method for determining the nature of time series," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 5, pp. 728–736, 2004.

[9] P. M. Addo, M. Billio, and D. Guegan, "Alternative methodology for turning-point detection in business cycle: A wavelet approach," *CES Working Papers, Centre d'Economie de la Sorbonne*, pp. 1–20, submitted, 2012.

[10] J. L. Elman, "Finding structure in time," *Cognitive Science*, vol. 14, pp. 179–211, 1990.

[11] M. Costa, A. L. Goldberger, and C.K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical Review Letters*, vol. 89, no. 6, pp. 068102–1–068102–4, 2002.

[12] M. Costa, A. L. Goldberger, and C.K. Peng, "Multiscale entropy analysis of biological signals," *Physical Review E*, vol. 71, pp. 021906–1–021906–18, 2005.

[13] M. Ahmed and D. P. Mandic, "Multivariate multiscale entropy: A tool for complexity analysis of multichannel data," *Physical Review E*, vol. 84, no. 6, pp. 061918–1 – 061918–10, 2011.

[14] D. Looney, M. Ahmed, and D. P. Mandic, "Human-centred multivariate complexity analysis," *Natural Intelligence Magazine*, vol. 1, no. 3, pp. 40–43, 2012.

[15] S. M. Pincus, "Approximate entropy as a measure of system complexity," vol. 88, no. 6, pp. 2297–2301, 1991.

[16] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Am J Physiol Heart Circ Physiol*, vol. 287, no. 6, pp. 2039–2049, 2000.

[17] A. L. Goldberger M. Costa and C. K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Phys. Rev. Lett.*, vol. 89, no. 6, pp. 1–4, 2002.

[18] M. Ahmed and D. P. Mandic, "Multivariate multiscale entropy," *IEEE Signal Processing Letters*, vol. 19, no. 2, pp. 91–94, 2011.

[19] T. Gautama, D. P. Mandic, and M. M. Van Hulle, "A differential entropy based method for determining the optimal embedding parameters of a signal," *Proceedings of ICASSP*, vol. 6, pp. 29–32, 2003.

[20] F. Goltz N. Amenc and V. Le Sourd, "Assessing quality of stock market indices," *EDHEC Risk & Asset Management Research Centre Publication*, 2006.

[21] V. Teverovsky M. S. Taqqu and W. Willinger, "Estimators for long-range dependence: An empirical study," *Fractals*, vol. 3, 1995.