Analysis of the dependence structure in econometric time series

Aurélien Hazan, Vincent Vigneron

Analysis of the dependence structure in econometric time series

Aurélien Hazan[1], **Vincent Vigneron**[1,2]
[1]IBISC CNRS-FRE 3190
91020 Evry, France
aurelien.hazan@ibisc.univ-evry.fr
[2]Équipe MATISSE-SAMOS CES CNRS-UMR 8173
75634 Paris cedex 13, France
vigneron@univ-paris1.fr

Abstract - *The various scales of a signal maintain relations of dependence the ones with the others. Those can vary in time and reveal speed changes in the studied phenomenon. In the goal to establish these changes, one shall compute first the wavelet transform of a signal, on various scales. Then one shall study the statistical dependences between these transforms thanks to an estimator of mutual information. One shall then propose to summarize the resulting network of dependences by a graph of dependences by thresholding the values of the mutual information or by quantifying its values.The method can be applied to several types of signals, such as fluctuations of market indexes for instance the S&P 500, or high frequency foreign exchange (FX) rates.*

Key words - Information theory, time series, dependence structure, correlations, mutual information

1 Introduction

A growing interest is evident in investigating the dependence relationships between complex data such as curves, spectra, time series or more generally signals. In these cases, each observation consists of values of dependent variables which are usually function of time.
The paper presents an information analysis of statistical dependencies between wavelet coefficients extracted from segmented time series on a model-free basis. These intrascale and interscale dependencies are measured using mutual information. Mutual informations depend strongly on the choice on the wavelet filters. Such dependencies have been studied intensively in image compression literature[13]. Wavelet analysis, by means of selection of criteria, take contact with self-similar fractals and iterative analysis, through other techniques of functional approximation, as radial basis functions, etc. although we don't want to enter in the arsenal of modern tools. The analysis of wavelets also allows a connection with the p-adic analysis and ultrametric criteria in general: these are important whenever the possibility of hierarchic structure with layers or levels of information arises. In the study of time series it is crucial to understand what is dependent and what independent of the temporal and space scales. The wavelet transform (WT) nearly decorrelates many time series and can be viewed as a Karhunen-Loève transform. Nevertheless, significant dependencies still exist between the wavelet coefficients.
Most algorithms focus on a certain type of dependencies, which it attempts to capture using a relatively simple and tractable model, such as the Karhunen-Loève transform (KLT), the discrete Fourier transform (DFT), and the discrete wavelet transform (DWT). Among them KLT is the most effective algorithm with minimal reconstruction error. The times series dataset is transformed into an orthogonal feature space in which each variable is orthogonal to the others. The time series dataset can be approximated by a low-rank approximation matrix by discarding the variables with lower energy [1].

DWT and DFT are powerful signal processing techniques and both of them have fast computational algorithms. DFT maps the time series data fromthe time domain to the frequency domain, and the fast Fourier transform algorithm (FFT) can compute the DFT coefficients in $\mathcal{O}(mn \log n)$ time. Unlike DFT which takes the original time series from the time domain and transforms it into the frequency domain, DWT transforms the time series from time domain into time-frequency $t - f$ domain.
The fact that the wavelet transform (WT) has the property of time-frequency localisation of the time series means that most of the times series energy can be represented by only a few wavelet coefficients. Chan and Fu used the Haar wavelet for time series representation and showed (classification) performance improvement over DFT [5]. Popivanov and Miller proposed an algorithm using the Daubechies wavelet for time series classification [17]. Lin *et al.* proposed an iterative clustering algorithm exploring the multi-scale property of wavelets [12]. Numerous other techniques for time series data reduction have been proposed such as regression tree [2], piecewise linear segmentation [11], etc. These algorithms work well for time series with few dimensions because the high correlation among time series data makes it possible to remove huge amount of redundant information. But for clustering algorithms with unlabeled data, determining the dimensionality of the feature dimensionality becomes more difficult. To our personnal knowledge, the feature dimension needs to be decided by the user.
The aim of this paper is to propose a time-series feature extraction algorithm using orthogonal wavelet capable to test for the presence of a dependence structure. The problem of determining the feature dimensionality is circumvented by choosing the appropriate scale of the WT. An ideal feature extraction technique has the ability to efficiently reduce the data while preserving the properties of the original data. However, infomation is lost in dimension reduction. The proposed feature extraction algorithm uses a information-theoretic approach for measuring dependence in time series.
The rest of the paper is organized as follows. Section 2 is a reminder on multiresolution analysis. Section 3 formulates the problem of dynamic forecasting in terms of mutual information. Section 4 contains a comprehensive experimental evaluation of the proposed algorithm on S&P and high-frequency foreign exchange time series[1]. We conclude the paper by summarizing the main contributions and perspectives in section 5.

2 A refresher on wavelet representations

WT is a domain transform technique for hierarchical decomposing techniques. It allow a sequence to be described in terms of an approximation of the original sequence, plus a set of details that range from coarse to fine. The property of wavelets is that the broad trend of the input sequence is preserved in the approximation part, while the localized changes are kept in the detail parts. More details about WT can be found in [7]. For short, a wavelet is a smooth and quickly vanishing oscillating function with good localisation properties in both frequency and time, this is more suitable for approximating time series data that contain regional structures [14, 8]. The WT uses a basis comprising n waveforms – n being the length of the data set under analysis. The basis waveforms $\psi_{j,k}$ form a set of orthogonal functions derived from scaling and translations of a mother wavelet

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k), \quad j, k \in \mathbb{Z}. \tag{1}$$

Any function $f(t) \in \mathcal{L}^2(\mathbb{R})$ can be represented in terms of this orthogonal basis as

$$f(t) = \sum_{j,k} c_{j,k}\psi_{j,k}(t), \tag{2}$$

and the $c_{j,k} =< \psi_{j,k}(t), f(t) >$ are called the *wavelet coefficients* of $f(t)$. To efficiently calculate the WT for signal processing, Mallat introduced the multiresolution analysis (MRA) and designed a family of fast algorithms based on [14]. With MRA, a signal can be viewed as being composed

[1]Chart of S&P 500 price (1950-present) can be found at `http://pages.stern.nyu.edu/čhurvich/Forecasting/Data/SNP500.CRS`

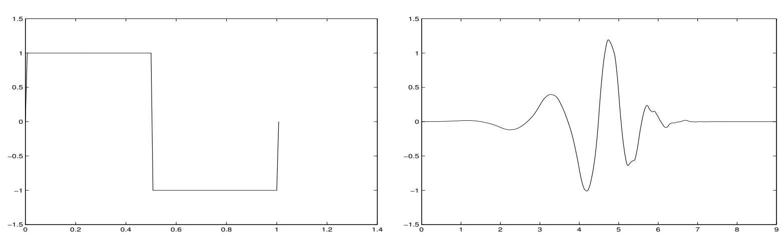
of a smooth background and fluctuations (also called *details*) on top of it. The distinction between the smooth part and the details is determined by the resolution, that is the scale below which the details of a signal cannot be discerned. At a given resolution, a signal is approximated by ignoring all fluctuations below that scale. We can progressively increase the resolution: finer details are then added to the coarser description, providing a better approximation of the signal

$$X = A_J + \sum_{j=1}^{J} D_j, \tag{3}$$

where A_j and D_j are respectively the approximation and detail at level j of the signal X. In other words, any time series can be written as the sum of orthogonal signals. Each signals lies in a common space denoted by V_0 and are of lenght n [15]. But A_j and D_j belong to spaces V_j and W_j respectively. This sequence of nested approximation spaces (V_j) involved in the multiresolution framework ($V_J \subset V_{J-1} \subset \ldots \subset V_0$). The space W_j is an orthogonal complement of V_{j-1} in V_j, *i.e.* $V_j = V_{j-1} \oplus W_{j-1}$. Then by defining

$$\underbrace{V_0 \oplus W_0}_{V_1} \oplus W_1 \oplus \ldots \oplus W_{j-1} = V_j \tag{4}$$

any signal belonging to V_j (resp. W_j) can be viewed as approximation (resp. detail) signals like A_j (resp. D_j). From a signal processing point of view, the approximation coefficients within lower scales correspond to the lower freequency part of the signal. Hence the first few coefficients A_j can be viewed as a noise-reduced signal. Thus keeping these coeffciients will not loose much information from the original time series X. Hence, normally, the first coefficients are chosen as the features: they retain the entire information of X at a particular level of granularity. The task of choosing the first few wavelet coefficients is circumvented by choosing a particular scale. The candidate selection of feature dimensions is reduced from $\{1, 2, \ldots, n\}$ to $\{2^0, 2^1, \ldots, 2^{J-1}\}$.
We use the Haar wavelet in out experiments which has the fastest transform algorithm and is the most popularly used orthogonal wavelet proposed by Haar. Figure 1 plots two wavelets: the Haar on the left and the Daubechies (db2) series on the right. Conceptually, these mother wavelet functions are analogous to the impulse response of a band-pass filter.



(a) (b)

Figure 1: The general shape of 2 wavelets commonly used in wavelet analysis. The sharp corners enable the transform to match up with local details that cannot be observed when using Fourier transform that matches only sinusoidal shapes. Fig 1(a) The Haar wavelet (b) the Daubechies wavelet.

The Haar wavelet has the function

$$\psi_H(t) = \begin{cases} 1 & \text{if} \quad 0 < t < \frac{1}{2} \\ -1 & \text{if} \quad \frac{1}{2} < t < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

So far, for many time series, the construction is not very good. In this case, no effort was made to do a good job of the decomposition, but merely to perform one MRA and to make it comprehensive.

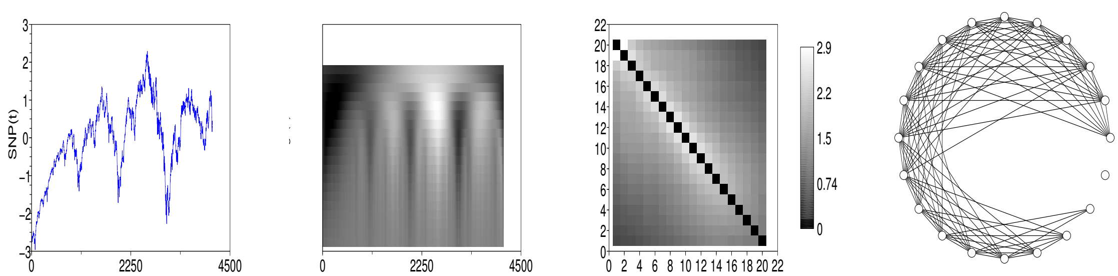
3 Methods

3.1 Related work

One frequent assumption in the wavelet domain is the absence of correlations between coefficients. As put forward by [9] when studying high-frequency time series such as exchange rates, this assumption should be questionned. In [9, 6] the underlying joint distribution is estimated thanks to a hidden Markov model, where to each coefficient of the DWT corresponds a high or low volatility level.
Besides the fact that we will consider only binary dependence, in this article we impose no predefined structure such as a Hidden Markov Tree. Rather, we propose to examine the dependence between every possible wavelet coefficients couples, in a combinatorial way.
In the litterature devoted to the statistical physics approach to financial time series, the dependence between scales is given a precise meaning, and models of the random processes are debated. For example, Arnéodo *et al.* [3] focus on an explicit model of downward causality between successive scales of the random process. Although we will not give such an inclination to this article we remark that, interestingly, mutual information plays a role in the latter work. Indeed, it allows to measure the propagation of the causal influence of a scale s_1 on scale $s_2 < s_1$.
In the following we mix the two approaches and infer the structure of dependence between the wavelet coefficients thanks to tools from information theory such as mutual information, that can be estimated from small samples.

3.2 Inference of dependence structure

One first naive approach would be to first perform the DWT or the continuous wavelet transform (CWT) of a one-dimensional signal, then to consider all the coefficients of a single scale s_1 as the realizations of a single random variables. Then, one may infer the dependence between every couples of scales (s_i, s_j) thanks to a measure of dependence between the sample $\chi_i = \{c_{i,1}, \ldots, c_{i,k}\}$ and $\chi_j = \{c_{j,1}, \ldots, c_{j,k}\}$.
This was done $S\&P500$ data, as shown by Fig. 2. One may however legitimately question the grouping of several coefficients estimates $c_{j,1}, \ldots, c_{j,k}$ belonging to the same scale, since there is no reason why they should obey the same law.



(a) (b) (c) (d)

Figure 2: Naive inference of dependence (a) Signal $S\&P(t)$ (b) Continuous wavelet transform coefficients (c) Mutual information matrix $(m_{ij}) = MI(\chi_i, \chi_j)$ where $\chi_i = \{c_{i,1}, \ldots, c_{i,k}\}$ and $\chi_j = \{c_{j,1}, \ldots, c_{j,k}\}$ (d) Dependence graph, obtained after thresholding (m_{ij})

Another approach considers every single $c_{j,k}$ as a random variable, and aims at infering the dependence between all possible couples c_{j_1,k_1} and c_{j_2,k_2}. To do so one needs several realizations $c_{j,k}$ for fixed values of j and k. Here we can take advantage of high-frequency sampling of financial time series and remark that they may display some degree of periodicity from one day to another, as shown on the example of the realized volatility of USD-DEM volatility in [9]. As illustrated by Fig.3, if we consider N high-frequency time series that correspond to N opening days, to each day i we can associate a set of wavelet coefficents $c^i_{j,k}$ and thus estimate either the pdf of $c_{j,k}$ or directly the mutual information between different coefficients c_{j_1,k_1} and c_{j_2,k_2}.

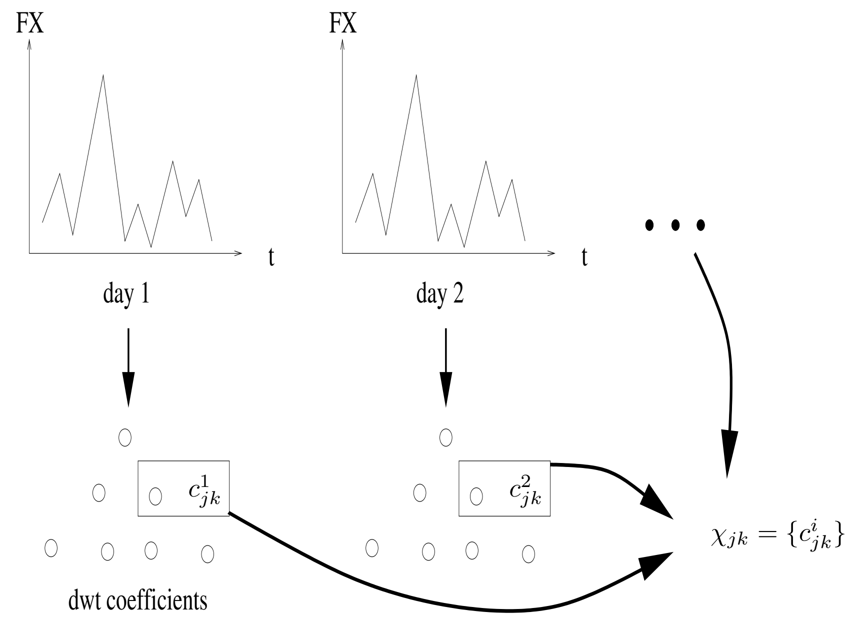


Figure 3: Sampling strategy. Each day yields a higi-frequency time series, that can be transformed in the DWT coefficients. N opening days thus allow to collect N realizations of each coefficient $c_{j,k}$

Next we describe more precisely the data under study, before laying stress on the statistical intricacies relative to the inference of the dependence structure.

3.3 EUR-USD volatility

We study the realized volatility at the sampling rate of 1-min, from January 1st, 2008, to the end of May, spanning 100 opening days[2]. The 5-min foreign exchange (FX) return can be defined as:

$$r_{i,5} = \log P_i - \log P_{i-5} \tag{6}$$

P_i being the foreign exchange price at discrete time i. The volatility will be defined below as $r^2_{i,5}$. Only the first 1024 sample out of approximately 1400 available each day are conserved, so as to fit the dyadic-length constraint.

3.4 Empirical test of dependence, graph inference

Mutual information is approximated from finite sample thanks to the estimator $I^{(1)}(x, y)$ by Kraskov *et al.* [10], based on the statistics of the k-th nearest neighbours, for a fixed maximum number of $k = 6$ neighbors. As explained in section 3.3, each sample $\chi_{j,k} = \{c^i_{j,k}\}_{i \leq i_{max}}$ for the wavelet coefficient $c_{j,k}$ will be composed of $i_{max} = 100$ realizations, which may not seem enough at first sight. The estimation error, though, was shown to remain tolerable even for such small samples in the case of a Gaussian random variable.

[2] Data made publicly available online courtesy of Forexite, see `http://www.forexite.com/free_forex_quotes/forex_history_arhiv.html`

Even if the error properties of this estimator have been well studied, its distribution as a function of the probability laws of the input random variables remains out of reach, should the latter be known. We need however to design a statistical test based on this estimation, and propose to get around this difficulty thanks to a *surrogate data* empirical test [18, p.79]. To do so, $N = 3000$ values of the mutual information $I^{(1)}(\chi_{j_1,k_1}, \chi_{j_2,k_2})$ between a sample $\chi_{j_1,k_1} = \{c^i_{j_1,k_1}\}$ and another one $\chi_{j_2,k_2} = \{c^i_{j_2,k_2}\}$ are computed, where (j_1, k_1, j_2, k_2) are randomly chosen, and the values in the samples are randomly permuted. The obtained set of mutual information estimates $\{I^{(1)}_1, \ldots I^{(1)}_N\}$ approximately follow a normal law, whose mean and standard deviation (μ, σ) can be estimated. Now, every realization of the mutual information can be tested again the null hypothesis H_0 that it follows a normal law $\mathcal{N}(\mu, \sigma)$, under the $\alpha = 0.05$ significance threshold.
We can further proceed to test the significance of the dependence between c_{j_1,k_1} and c_{j_2,k_2}, for every couples (j_1, k_1) and (j_2, k_2). This problem must be recognized as a an occurrence of a *multiple comparison problem*, where the $\alpha = 0.05$ significance can't be guaranteed if each comparison is made separately. Thus a p-value correction is made necessary, and fulfilled thanks to an Fdr correction [4], under the $\alpha_{fdr} = 0.05$ significance threshold.

4 Results

We now summarize the results, obtained on a subset of the full 1024×1024 mutual information matrix, for clarity reasons. Fig.4(a) represents the p-values matrix (p_{ij}) for the statistical test of independence between two coefficients. Low values in dark stand for values such that it is unlikely that the coefficients indexed are independent. After thresholding by the $a = 5\%$ significance level, the couples that failed to pass the test are shown in light color by Fig.4(b).
Remark that the main diagonal is rejected, which is consistent with the fact that a coefficient is hardly independent with itself. Now, other unexpected structures appear along the $y = ax$ line, for different slopes though. Their presence can be explained thanks to the wavelet coefficient dependence tree, displayed in a decimated way by Fig.4(d). A connection between two nodes means that two coefficients failed to pass the independence test, and we can note that nodes pertaining to a given scale s are connected with their closest neighbour. Such nodes correspond precisely to the $y = ax$ lines with $a \neq 1$ in the rejection matrix.
Comparaison of the full wavelet coefficient dependence tree Fig.4(c) with models in the litterature, such as thee Hidden Markov Tree (HMT) used by [9] shows that the dependence structure imposed on the model should be compared with the model-free structure obtained by direct inference. Indeed, in HMT models, associations are allowed between scales but not within scales, which would not fit the data under study. Furthermore, associations between non adjacent scales should be taken into account, as suggested by [6, p.892], in consideration of the complex interplay between scales and times in 4(c).

5 Conclusions and perspectives

We have presented a descriptive framework whose aim is to infer the dependence structure of a set of times series thanks to their Wavelet Transform, and dependence measures from Information theory. We show that this structure can be inferred without model, and apply the method to a high-frequency financial time series. It appears that intrascale dependence play an important role, between adjacent coefficients in the wavelet coefficients tree; and that dependencies span scales and time locations.
In future research we plan to discuss the following topics:

- can this method succeed in inferring the structure of processes studied in the litterature, where the relations between scales are available, such as the "causal cascade" depicted in [3]. For such processes where the pdfs may be available, can we derive explicit expressions of the mutual information between coefficients ?

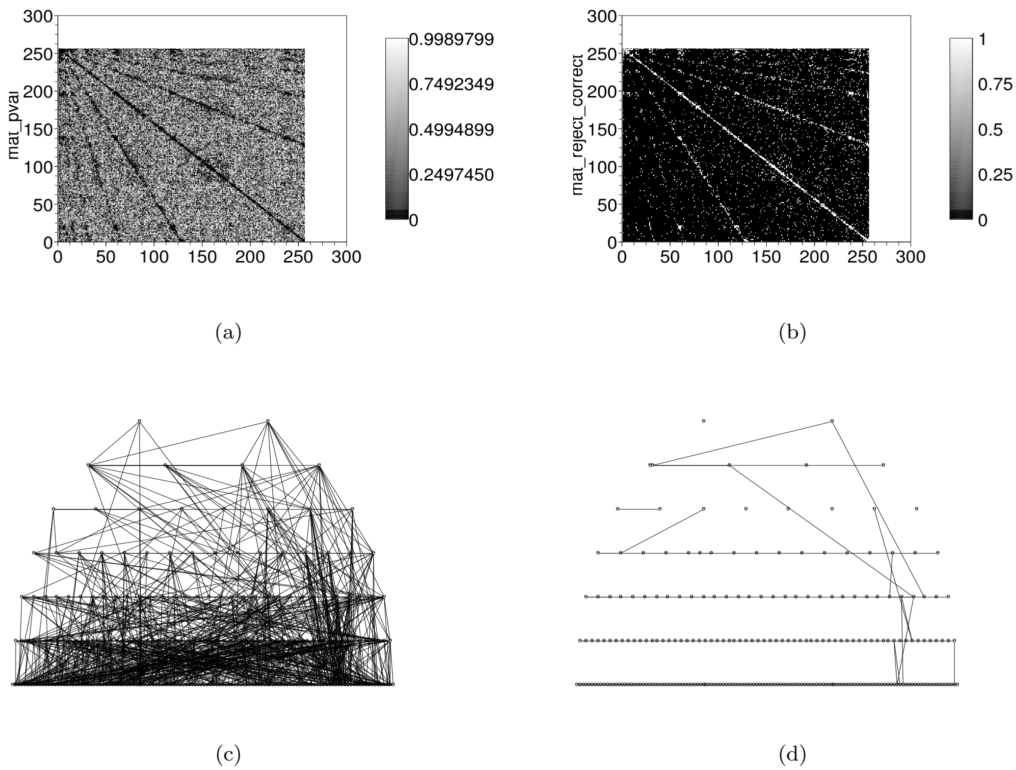


Figure 4: Inferred structure of dependence, for the first 256 coefficients, out of 1024 (a) p-value matrix (p_{ij}), for the test against the null hypothesis of independence between two DWT coefficients indexed by i, j (b) rejection matrix, $r_{ij} = 1$ if the two coefficients can't be considered as independent (c) Wavelet coefficients dependence graph (d) Decimated dependence graph.

- can we minimize mutual information estimation error, e.g. thanks to resampling techniques, for a reasonable computational cost ?
- can this framework be extended to oriented or causal measures of dependence ?
- do the structure of the obtained graphs have an explanatory power, for example through the study of communities [16].
- what applications can we envision, from machine learning (clustering, rupture detection) to multiscale physical systems analysis (e.g. robotics, . . .) ?

References

[1] A. Antoulas. Frequency domain representation and singular value decomposition, 2002.

[2] J. S. Armstrong and J. G. Andress. Exploratory analysis of marketing data: Trees vs. regression. *Journal of Marketing Research*, 7:487–492, 1970.

[3] A. Arnéodo, J.-F. Muzy, and D. Sornette. "Direct" causal cascade in the stock market. *European Physical Journal B*, 2:277–282, May 1998.

[4] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.

[5] K.P. Chan and A.W. Fu. Efficient time series matching by wavelets. In *Proceedings of the 15th International Conference on Data Engineering*, pages 126–133, 1999.

[6] M. Crouse, R. Nowak, and R. Baraniuk. Wavelet-based statistical signal processing using hidden markov models. *IEEE Trans. Signal Processing*, 46:886–902, 1998.

[7] I. Daubechies. Orthonormal bases of compactly supported wavelets. ii. variations on a theme. *SIAM J. Math. Anal.*, 24(2):499–519, 1993.

[8] I. Daubechies, S. Mallat, and A.S. Willsky. Introduction to the spatial issue on wavelet transforms and multiresolution signal analysis. *IEEE Trans. Inform. Theory*, pages 529–532, 1992.

[9] R. Gencay and F. Selcuk. Asymmetry of information flow between volatilities across time scales. Econometric Society 2004 North American Winter Meetings 90, Econometric Society, August 2004.

[10] A. Kraskov, H. Stogbauer, and P. Grassberger. Estimating mutual information. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 69(6):066138, 2004.

[11] Daniel Lemire. A better alternative to piecewise linear time series segmentation. In *SIAM Data Mining 2007*, 2007.

[12] J. Lin, M. Vlachos, E. Keogh, and D. Gunopulos. Iterative incremental clustering of time series, 2004.

[13] J. Liu and P. Moulin. Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients. *IEEE Transactions on Image Processing*, 10(11):1647–1658, November 2001.

[14] S. Mallat. A thiry for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. and Mach. Intelligence*, 11(7):674–693, 1989.

[15] Michel Misiti, Yves Misiti, Georges Oppenheim, and Jean-Michel Poggi. Clustering signals using wavelets. In *IWANN*, pages 514–521, 2007.

[16] P. Pons and M. Latapy. Computing communities in large networks using random walk. *Computer and Information Sciences - ISCIS 2005*, 3733:284–293, 2005.

[17] Ivan Popivanov and Renée J. Miller. Similarity search over time-series data using wavelets. In *ICDE*, pages 212–, 2002.

[18] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. Doyne Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D Nonlinear Phenomena*, 58:77–94, September 1992.