MULTISCALE REPRESENTATION OF AGENTS HETEROGENEOUS BELIEFS IN ANALYSIS OF CAC40 PRICES WITH FREQUENCY DECOMPOSITION

Abstract

This paper focuses on the time series' decomposition and economic representation of its constituent parts. Wavelet transforms are used for adaptive analysis of local behaviour of heterogeneous agents.

Unlike fully revealing equilibrium of homogeneous beliefs, in the environment with heterogeneous beliefs prices are driven by prevailing expectations of market participants. Thus, forecasting future prices, one must form expectations of others forecasts. Evolution of agents' expectations largely governs the adaptive nature of market prices. Overlapping beliefs of heterogeneous agents prevent the effective examination of expectation formation and price forecasting by traditional methods. In the approach proposed in this paper, a time series is decomposed into a combination of underlying series, representing beliefs of major clusters of agents. The analysis of individual parts improves statistical inference, isolating effectively nonstationary and nonlinearly features. Emergent local behaviour is also more receptive to prediction. The overall forecast (weighted combination of individual forecasts) is found to be determined and evolved depending on specific market conditions.

On the statistical level, the data generating mechanism is considered as complex multi-structured system, with individual layers corresponding to particular frequencies. Reflecting the time preferences of agents, trading strategies being homogeneous intra-type are heterogeneous inter-type for agents with distinct time preferences. Overall market activity at each moment, providing the dynamic feedback across agents' types, generates market prices. The frequency decomposition of a time series identifies the local and global structures and separates short and long time dynamics.

The Genetic Algorithm is applied to determine the optimal decomposition of the signal and representation of heterogeneous traders. The Artificial Neural Network is trained to learn information at the scale level that is hidden in the aggregate. The resulting models seek to enhance the understanding of the underlying data generating mechanisms of financial time series and to develop new approaches for financial forecasting.

Keywords: Artificial Neural Networks; Genetic Algorithm; Wavelet Transform

Introduction

Considering that different financial decisions occur at different scales, analysis of separate scales of a complex signal provides a valuable source of information. Wavelet transform decomposition of a complex time series into separate scales and their economic representation is a focus of this study. An evolutionary / artificial neural network (E/ANN) is used to learn the information at separate scales and combine it into meaningfully weighted structures. Potential applications of the proposed approach are in financial forecasting and trading strategies development based on individual preferences and trading styles.

Data Generating Mechanism as Complex Multi-Structured System

A recent move in financial research from homogenous to heterogeneous information added realism to the analysis of market efficiency. At the same time information heterogeneity does not necessarily imply and lead to beliefs heterogeneity. It is in fact the latter that seems to determine the empirical facts observed in financial markets. Unlike fully revealing equilibrium of homogeneous beliefs, in the environment with heterogeneous beliefs prices are driven by prevailing expectations of market participants. Thus forecasting future prices one must form expectations of others forecasts. Evolution of agents' expectations to a large extent governs the adaptive nature of market prices. Overlapping beliefs of heterogeneous agents largely prevent the effective examination of expectation formation and price forecasting by traditional (time-series) methods.

The objective is thus to decompose a time series into a combination of underlying series, representing beliefs of major clusters of agents. Isolating effectively nonstationary and nonlinearly features, analysis of individual parts is expected to improve statistical inference. Emergent local type of behaviour is anticipated to be more receptive to forecasting. Weighted combination of individual forecasts is determined and evolves in accordance with specific market conditions, providing an aggregate forecast. Such multiscale analysis is based on multiresolution framework, focusing on the development of efficient numerical algorithms for signals decomposition and reconstruction.

Agents' heterogeneity is considered through trading strategies realisation, focusing on traders' time preferences, rather than through the market organization, in relation to the market functions (market-makers, intraday traders, and the like). Financial markets are populated by players with different time preferences. Such differences are reflected in diversity of opinions about future events. The resulting trading strategies, being homogeneous intra-type, are heterogeneous inter-type of traders with distinct time preferences. Overall market activity at each moment, providing the dynamic feedback across agents' types, generates market prices.

Traders with similar strategies take comparable actions, forming a particular layer of multi-structured DGM. Thus the search is for an approach capable learning the underlying structures with different time-scales. Frequency decomposition of a time series is used for such structures identification. An inverse operation, synthesizing appropriate structures generates better characterised DGM with further applications for modelling and forecasting. In addition, frequency decomposition permits to separate short from long time dynamics, as well as to discover their consequence on the local and global structures.

Traders' Heterogeneity

Economic agents' behaviour is determined by their intrinsic opportunity sets. One of the main factors affecting traders' opportunity sets is the time dimension, realised in differentiating market participants, according to their time horizons, into short and long term traders. Agents with distinct time scale dimensions respond in a different way, with different reaction time to the news. Low frequency shocks are likely to affect all market participants, though with some time delay. On the other hand high frequency shocks may be ignored, at least for a while, by long term traders. Such information processing 'inefficiency' is heterogeneity-driven asymmetry by the market efficiency hypothesis. Identified casual relationships are potentially exploitable in trading model development.

Consider a scaling law that relates traders time horizons and frequency of price fluctuations. In these settings rapid price movements that corresponds to high frequency fluctuations are reflected in the frequent trading positions' revaluations by agents with short term horizons, such as (inter-day) speculators. By trading upon the information of high frequency signals short term traders frequently execute transactions, supplying high frequency information to the market.

On the other hand slow price movements, corresponding normally to larger price shifts (Müller, Dacorogna et al. 1990), are more apparent in low frequency signals, with less noise interference. Thus traders with long term horizons (e.g. certain institutional investors) tend to trade upon the information of low frequency signals. Long term traders, reducing their risk exposure with derivatives, stop-loss limits and the like act, to a certain extent, according to the market fundamentals (Dacorogna, Gencay et al. 2001). Long term agents, trading upon low frequency signals, provide low frequency information to the market that is used adaptively by traders with similar time scale.

The interactions between heterogeneous agents in relationship to each other actions rather than to the market news produce endogenous dynamics in the market[1]. Such dynamics provide reasonable explanation to some common empirical facts in Finance, as trend persistence or volatility clustering. Differentiating economic agents' expectations according to their time dimension has valuable consequences for forecasting. Since the time scale of traders is the key characteristic of the market, the adaptive dynamics of prices reflect beliefs and behaviour of the dominant agents on the market. Therefore, establishing a (trading) frequency signal of the dominant agents can be used for tuning into it and exploiting it in profitable predictions.

Frequency Decomposition

Wavelet Transforms

Representing a signal in the *time-scale*[2] domain provides a fine time resolution, being narrow in time, to zoom in high frequencies with less ambiguity regarding their exact position in time. On the other hand a good frequency resolution is achieved by being short in frequency (long in time) with less ambiguity about the exact value of the frequency (to deal with low frequencies). To accomplish this, a signal is multiplied with the *wavelet* function[3] and the transform is computed for different segments of TD signal. Well localized in time, without constant resolution and infinite duration, wavelets adapt the time and frequency support, reaching the boundary of Heisenberg Uncertainty Principal[4]

---

[1] Kurz (1994) calls it 'endogenous effects'.

[2] Consider scale is an inverse of frequency.

[3] A wavelet refers to a complex function (with a zero integral over the real line) of a small wave, i.e. compactly supported (finite length) oscillatory function.

[4] Heisenberg Principal imposes limits on precising an event appearance with arbitrary accuracy in both time and period terms.

By applying modified versions of a prototype, the *mother wavelet* (MW) to the time series, wavelet transform convolves the data with a series of local waveforms to discover correlated features or patterns. MW is a source function, from which translated and scaled wavelet functions (with different regions of support) are constricted. Nonorthogonal wavelet functions are used with both Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT), whereas orthogonal[5] wavelet basses imply exclusively DWT.

To apply a wavelet function to a signal, it is *scaled*[6] by varying the scale, *s* and t*ranslated*[7] along the localized time index, *n*. Consider scaling of the wavelet *V* with the expansion of its range by a multiplicative factor, *V(1/s)* and translation of *V* with shifting its range to the right, *V(n'-n)*:

$$V(n,s) = s^{-1/2} V\left(\frac{n'-n}{s}\right), \tag{1}$$

where $s^{-1/2}$ is energy conservation weighting value, a normalization to keep the total energy of the scaled wavelet constant.

By projecting a time series of interest, *x(t)* onto the wavelet function *V(n,s)*, decomposition occurs when correlated features (similar frequency content) are identified:

$$CWT(n,s) = \int_{-\infty}^{\infty} x(t) V(n,s) * (t) dt. \tag{2}$$

Providing that a wavelet function satisfy the admissibility condition, given by time-frequency space localization and zero mean, an inverse operation can be performed. Reconstruction (synthesizing) of the original signal from its wavelet transform is done with integrating over all scales and locations:

$$x(t) = \int_{0}^{\infty} \int_{-\infty}^{\infty} CWT(n,s) V(n,s)(t) dn \frac{ds}{s^2}, \tag{3}$$

---

5 Orthogonal wavelet functions characterised by no overlapping or projection.

6 Scaling is a mathematical operation that either dilates or compresses a signal. In wavelet analysis dilated (stretched out) signals are achieved with s > 1 and related to larger scales; whereas compressed signals are accomplished with s < 1 and corresponded to small scales.

7 Translation indicates the location of MW.

CWT of a discrete time series x_n is defined as the convolution of the sequence x_n with scaled and translated versions of a wavelet function $V_{n,s}$ and in direct form (in TD) is given by

$$CWT_{n,s}^{dir} = \sum_{n'=0}^{N-1} x_{n'} V_{n,s} * \left[\frac{n' - n\delta t}{s} \right], \tag{4}$$

where n is the localized time index, N is the data series' length, s is the wavelet scale, (*) is the complex conjugate and δt is the sampling interval. Thus a simultaneous presentation of the amplitude versus the scale and fluctuation of the amplitude over time is constructed by varying the wavelet scale and translating along the localized time index.

A function of two continuous parameters (scale and location), CWT is characterised by overlapping (nonorthogonality). High amount of extra information requires extensive computation and memory, though producing particularly accurate time-frequency spectrum. For computer processing, a discretised CWT (semi-discrete or wavelet series) considers a limited amount of scales with a varying quantity of wavelet coefficients at each scale.

DWT is defined by a square matrix of filter coefficients with the information being sparse in the transform space through compression. A wavelet characterised by a scaling function that spans the space Λ_2^j is represented by a linear combination of functions from the next subspace Λ_2^{j+1}. Since the subspaces are nested and self-similar a relationship between scaling functions of any two neighbouring subspaces defines the filter coefficients. Thus orthogonal wavelets are characterised by a (finite) set of filter coefficients, relating scaling functions of different subspaces to each other and DWT uses a pair of filters to isolate high and low frequencies.

Common sampling of DWT coefficients from CWT on a dyadic[8] grid (s_o = 2 and n_o = 1) yields $s = 2^j$ and $n = k2^j$. Following Mallat (1989), dilating by $s = 2^j$, $V_{2^j}(t) = 2^j V(2^j t)\big|_{\forall j \in \mathbb{Z}}$ and translating by $n = k2^j$ define an orthonormal[9] wavelet basis:

$$\forall j,k \in \mathbb{Z}, \quad V_{j,k} = 2^{-j/2} V_{2^j}(n - k2^{-j}). \tag{5}$$

DWT analyse a signal at different frequency bands with different resolutions by successive decomposing it into a (coarse) approximation and detail information. Consider a^j as a coarser

[8] Dyadic is a two-fold relationship between resolution scales.

[9] A complete set with its individual components being orthogonal and with a unit length.

approximation at the resolution J of x_n and d^j as the detail information that gets lost if one go from the approximation a^j to the coarser approximation a^{j+1}. Scaling and wavelet functions, associated with low and high pass filters, are used for filtering of the time domain signal. *Filtering* changes the resolution of the signal through convolution of the signal with the impulse response of the filter. After passing a (finite length) signal through half bands high/low pass filters the amount of detail information in the signal is halved. Following such filtering according to the Nyquist rule, a half of the samples can be discarded since the signal has a highest frequency of $\pi/2$ radians (rather than π). *Subsampling* the signal by two, halves the number of points and double the scale (reducing the sample rate). If $y^H{}_k$ and $y^L{}_k$ denote the output of high, h_n and low pass, l_n filters the decomposition of the signal, x_n is thus

$$\begin{aligned} y_k^H &= \sum_n x_n h(2k-n); \\ y_k^L &= \sum_n x_n l(2k-n). \end{aligned} \tag{6}$$

The length of the signal determines the maximum number of decomposition levels. The full transform demands the data series length to be of a power 2 and the partial transform requires it to be an integer of 2^j. Since financial data is not of dyadic length or even an integer multiple of it, one would need to truncate the series to the closest integer multiple of 2^j or 'pad' it with zeros.

Particularly for large series the DWT provides an efficient data reduction scheme, where the unimportant frequency bands, characterised by low amplitude are eliminated without significant loss of information. For an orthogonal wavelet basis, the number of convolution at each scale is proportional to its width at that scale. As a result a wavelet spectrum is confined by discrete blocks of wavelet power. Since this number of blocks at each scale is a function of wavelet width, the compact representation of the signal is achieved. Both orthonormal and biorthogonal[10] wavelet basis result in non-redundant transforms.

The reconstruction of the original signal from its DWT is done through reversing the order of the decomposition procedure, though the perfect reconstruction is only attainable for ideal halfband filters. Otherwise, Daubechies filters Daubechies (1992) might provide perfect reconstruction under certain conditions.

---

[10] Separate bases orthogonal to each other but without individually forming orthogonal sets.

Multiresolution Analysis

Multiresolution analysis considers a signal at different frequencies with different resolutions. It examines data at various levels of microscopic magnification (resolutions) by projecting a function on a set of closed subspaces $\{... \subset \Lambda_{-1} \subset \Lambda_0 \subset \Lambda_{-1} \subset \Lambda_1 \subset \Lambda_2 ...\}$. $\Lambda_2{}^j$ is the set of all possible approximations of functions in $L^2(\mathbb{R})$ at resolution 2^j. If O_2 is an approximation (projection) operator on the vector space $\Lambda_2{}^j \subset L^2(\mathbb{R})$, then the best approximation to the original signal at resolution 2^j is given by

$$\forall y(t) \in \Lambda_{2^j}, \qquad \|y(t) - x(t)\| \geq \|O_{2^j} x(t) - x(t)\|. \tag{7}$$

An approximation of the signal at resolution 2^{j+1} contains all information necessarily to make an approximation at lower resolution 2^j:

$$\forall j \in \mathbb{Z}, \qquad \Lambda_{2^j} \subset \Lambda_{2^{j+1}}. \tag{8}$$

Scaling the approximated functions by the ratio of the respective levels of resolution, one derives the spaces of approximated functions:

$$\forall j \in \mathbb{Z}, \qquad x(t) \in \Lambda_{2^j} \Leftrightarrow x(2t) \in V_{2^{j+1}}. \tag{9}$$

At the lower limit the resolution approaches to zero with the information of the approximated signal converging to zero:

$$\lim_{x \to -\infty} \Lambda_{2^j} = \bigcap_{j=-\infty}^{+\infty} \Lambda_{2^j} = \{0\}. \tag{10}$$

At the upper limit the resolution increases to $+\infty$ with the approximation converging to the original signal:

$$\lim_{x \to +\infty} \Lambda_{2^j} = \bigcup_{j=-\infty}^{+\infty} \Lambda_{2^j} - \text{dense in } L^2(\mathbb{R}), \tag{11}$$

where one set is dense in another set if its closure contains the second set; with the closure being the smallest closed set.

Considering the approximation $O_2^j x(t)$ of the signal $x(t)$ by 2^j sampled observations per unit length, a translation of the approximation for the resolution j = 0 is given by

$$\forall m \in \mathbb{Z}, \quad O_{2^0} x_m(t) = O_{2^0} x(t-m), \text{ where } x_m(t) = x(t-m). \tag{12}$$

For the discretely sampled observations the translation becomes

$$I(O_{2^0} x(t)) = (a_i)_{i\in\mathbb{Z}} \Leftrightarrow I(O_{2^0} x_m(t)) = (a_{i-m})_{i\in\mathbb{Z}}, \tag{13}$$

where I is an isomorphism from $\Lambda_2{}^0$ onto $\ell^2(\mathbb{Z})$ and the vector space of square-summable sequences, $\ell^2(\mathbb{Z}) = \left\{ (a_i)_{i\in\mathbb{Z}} : \sum_{i=-\infty}^{+\infty} |a_i|^2 \prec \infty \right\}$.

The difference in information between two levels of resolution is 'detailed' information that would be lost when moving progressively to low resolution. Representing the 'detailed' information as the orthogonal complement of $\Lambda_2{}^j$ in $\Lambda_2{}^{j+1}$, $W_2{}^j$, any higher-level of approximation is achieved by adding the appropriate details back into the approximation. Since the subspaces are nested, $\Lambda_2{}^{j+1}$ can be represented as the direct sum of the coarsely approximated subspace $\Lambda_2{}^j$ and its orthogonal complement $W_2{}^j$:

$$\Lambda_{2^{j+1}} = \Lambda_{2^j} \bigoplus_{j=-\infty}^{J} W_{2^j}. \tag{14}$$

The multiscale analysis utilises filters to split up a function into the components of subspaces and their orthogonal complements, representing different scales. Considering filtering as user-defined extraction of partial information from the input signal, a filter is an algorithm that divides a time series into individual components. In discrete setting a pair of sequences, $\{h(k); g(k)\}_{k\in\mathbb{Z}}$ represents the lowpass and highpass filters. $h(k)$ smoothes the data, keeping low frequencies (longer-terms structures), while $g(k)$ preserves the detailed information, high frequencies (transitory patterns). The above filters are related to each other in the following way

$$g(n) = (-1)^n h(1-n). \tag{15}$$

Multiresolution analysis provides an intuitive understanding of the different levels of a time series, representing separate components of the underlying DGM. Decomposed series is a combination of an approximation (mean) and detail information at specified levels of resolution. Thus economic meanings can be attributed to different levels of approximation and details.

Experimental Design

Filters

It is useful to represent a wavelet transform, characterised by a real and an imaginary part, in polar coordinates. The norm (the magnitude of the transform related to the local energy) is of primary interest, while the polar angle (phase) completes the representation. Let $\Re(V_{n,s})$ stand for in-phase or real part and $\Im(V_{n,s})$ represent in-quadrature or imaginary part of the wavelet transform. Then amplitude is $|V_{n,s}|$ and phase is $tan^{-1}[\Im(V_{n,s}) / \Re(V_{n,s})]$. A complex wavelet function, with information about both amplitude and phase, is well suited for analysing oscillatory behaviour, whereas a real function is sufficient for isolating discontinuities or peaks. The time series under consideration determines the choice of the wavelet function. For a smoothly fluctuating signal, a smooth function (e.g. a damped cosine) is appropriate. On the other hand, for a signal with sudden jolts a boxcar function (e.g. the Haar) is better suited. Comprehensive overview of the criteria determining the choice of the wavelet function is presented in Farge (1992).

Consider the period of a wavelet, p as the number of complete cycles or the rate of change of a signal. Rapid appearance/disappearance entail rapid rate of change of a signal, implying a short period, whereas a slow rate of change imply a long period. Varying the scale of the wavelet by changing its width, the wavelet is tuned to identify certain periods and their positions in time. Further consider the width of the filter, μ as the number of sample points on each side of its centre over which the filter gathers more than negligibly small information. Outside this limit the filter is modified to force its response to zero. A wide wavelet covers more time series and is capable of better identifying a particular period but is worse in detecting its exact position in time. On the other hand a narrow wavelet has broad response to features that are moderately close to the specified period with a good localisation in time.

The maximum length of the series considered by a filter is limited to twice the width plus one. For the current application the periodic sensitivity of a filter is given by the following relationship, $p \leq \mu \leq 2p$. $p \leq \mu$ is the minimal periodic sensitivity that is practically useful. μ

> $2p$ characterises a narrow sensitivity filter, typically seeking a particular period. $\mu = p$ appears to be useful for pattern detection in financial time series.

When a filter is applied to a part of a time series in which only the information about its past is available, the lag need to be determined explicitly. The specified lag of the filter determines the peak of its response. From theoretical perspective to have samples on both sides of the centre, the filter should be lagged at its width length to avoid (unknown) future interferences. Although for practical applications such extreme might be not necessary.

To differentiate between the presence of an event and its state a combination of the real and imaginary filters, given by $A = \sqrt{\Im(V_{n,s})^2 + \Re(V_{n,s})^2}$ is used. Its values are at zero when the event is not present, are significantly positive when the feature sought is strong and goes to zero as the event passes into the history. To answer the question of how much (amplitude) does the data resemble a wave of a chosen width, a new series of the projection amplitude versus time is constructed by sliding the wavelet along the time series.

Significance Testing

Peak-based critical limit significance is used for signal analysis/processing. Two backgrounds are considered: a white noise, H_0: AR(1) = 0 and a red noise (the signal power decreases with increased frequency), H_0: AR(1) > 0. A 95% (99.9%) peak-based critical limit implies that in 1 out of 20 (1000) random noise signals would the largest peak reach this height by a random chance. Monte Carlo simulation is developed to generate the peak-based critical limits[11]. The simulated data is then fitted to bivariate, univariate or trivariate polynomials, depending on the number of factors affecting the significance.

Experimental Results

Data

The data considered includes prices for CAC 40® share index measuring the evolution of a sample of 40 equities listed on Euronext® regulated markets in Paris. Created in June 1988, the CAC 40 is designed to represent the French stock exchange listed companies and serve as a support for derivatives markets. During the trading session 9.00 – 17.30, index levels are

[11] The Monte Carlo method is used to determine significance levels for a process with unknown statistical distribution.

calculated in real time using the last trade quoted[12] and are disseminated every 30 seconds. Thus the change in the index is equal to the sum of the change in each of component equity times its weight in the index, where the market price of equity is adjusted for corporate actions taking effect, but not for dividend paid.

The period under investigation runs from 01.03.90 through 07.03.05 with the business time scale, which excludes holidays, Saturdays and Sundays from the physical time scale used in the experiment. The length of the data series is driven by the objective to explain the present behaviour of the index, where the data prior to 1990 is considered to refer to a different from the current phenomena. The original data, consisting of the series with tick[13] frequency of 30 seconds was obtained from Euronext.

Data Analysis

Considering that financial markets are populated by heterogeneous agents, the dominant traders' behaviour determines the adaptive nature of the prices on that market. Under such approach, one needs to identify the most appropriate sampling intervals for the analysis and predictions of the dynamics in the market under consideration. A selected sampling interval should reflect the information about the (dominant) traders' reaction time to the news and their dealing frequency. At the same time for this information to be apparent for the analysis the amount of (random) noise in the series ought to be minimise.

Through determining the most prominent (informative) frequency bands in the original signal the appropriate sampling interval is identified. Examining the DWT coefficients using Daubechies wavelet basses with the maximum levels of decomposition, the main energy is found to be around the scale 2^{10} in dyadic scaling, which corresponds to the frequency between 8 and 9 hours. This result justifies the data subsampling with 8.5 hours frequency that produce a relatively low noise signal without significant loss of information.

Thus from the original data after subsampling, the series containing 3779 eight and a half hours prices was extracted. The analysis of such frequency well relates to the objective of examining the behaviour of heterogeneous agents that strategically fulfil certain goals and

---

[12] On the exceptional basis in the event that at least 35% of the market capitalization of the index has not been traded, a price indicator - the forerunner is substituted for the index.

[13] A 'tick' is one logical unit of information.

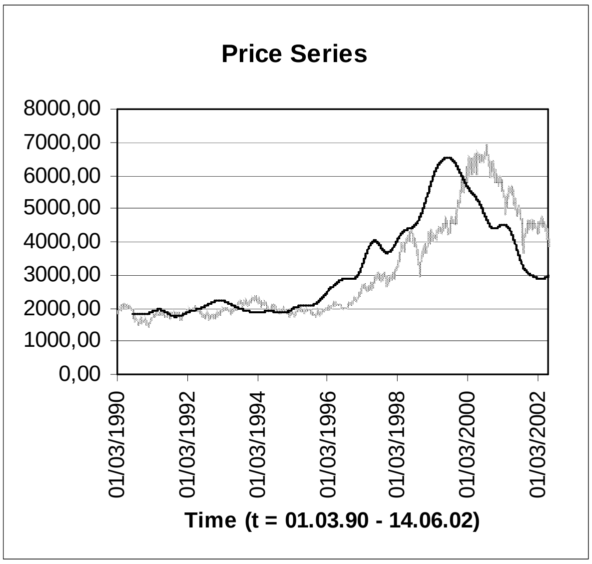
cluster according to some (nontrivial) time horizons, optimal for their economic type. Kaastra and Boyd (1996) assert a similar sampling frequency as the most appropriate in designing an ANN by a typical off-floor trader. Furthermore, for practical purposes of construing a model, generating reliable predictions, the issues of the realistic time needed to execute a strategy makes the choice of such a sampling interval justified.

Low and High Frequencies

Assuming that lower level decomposition capture the long range dependencies whereas the higher levels identify short term relationships, the signal is decomposed into low and high frequencies. Experiments with different decomposition techniques, namely CWT, DWT and WPT demonstrate that different transforms are appropriate for the data with different sampling frequencies. Specifically it was found that for a high frequency signal the best results were achieved by DWT with Daubechies wavelet basses, whereas for the signal with longer sampling interval the most promising results were with CWT using Gabor wavelet function. The results for high frequency signals confirm that moving averages and moving differences (given by DWT coefficients) are useful tools for identifying short term dependencies in the noisy environment. In addition it appeared to be an effective data reduction scheme working with large data sets, due to DWT effective subsampling operations. The efficiency of the pyramid algorithm of Mallat (1989) in the Daubechies-Mallat paradigm is ideal for rapid processing in real time. Smooth and continuous longer range relationships are more susceptive to CWT that provides a greater accuracy. Disjunct nature of Daubechies wavelets contrast with smooth motions that underline the data. Furthermore, being computationally intensive CWT appears to be appropriate for small and medium size data sets. Inferior results for Wavelet Packet Transform were found with high and low frequency signal considered. A further research into the behaviour of Wavelet Packet Transform coefficients is required before any conclusion can be drawn on this matter.

To capture and exploit the features of interest with minimum delay to be of value for financial applications, CWT with Gabor wavelet function generated the best performance due to its ability of efficient filtering with minimal lag. A number of techniques to reduce lag are considered. The *Slow taper* approach proposes to increase the frequency domain period and reduces the frequency domain width. Under a worse case scenario the lag is equal to the width (reduced by the slow taper). Another approach considered is to *truncate the filter* before its effective response has dropped to zero. *Mirroring* as a lag reduction technique appends historical points in reverse order. Without knowing the TD impulse response, it simulates a half-impulse filter with all weights doubled, except the highest (the most recent) ones. Being straightforward to implement mirroring comes at cost for asymmetric signals[14].

Considering a lowpass filter data smoothing, the best performance was obtain by the Gabor wavelet (μ = 350; p = 350; l = 350) for both price and return series. Balancing spurious frequency responses against lag reduction benefits (real) mirroring technique was found to be the most effective for the current application The Figure 1 presents the original series with the filtered outputs overlaid. Although high quality of the smoothing is achieved in both cases, significant lag produced by a fully lagged filter (1a) was successfully reduced with mirroring technique (1b). Note that the issue of balancing costs and benefits of the mirroring is application-dependent and should be considered for the problem in hand.



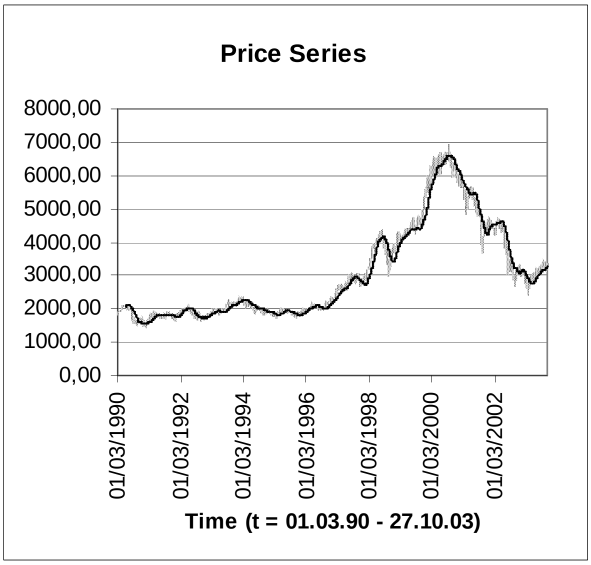


Figure 1. Lowpass Filters. (a) Fully Lagged Real Gabor; (b) Gabor with Real Mirroring. The original series - dotted curve; filtered output – solid curve.

In high frequency filtering a highpass real Gabor (μ = 3; p = 5; l = 2) generates the most accurate result for both price and return series. Applying a very narrow in TD filter, Morlet shape modification was used to avoid contamination from slow trends. Such modification centres the filter and eliminate/reduce the problem. Figure 2 presents the original return series overlaid with the filter output, extracting extremely local information from the original signal.

---

[14] To overcome problems related to the extension of the time-domain wavelet functions over the entire temporal domain in discrete wavelet analysis settings Pollock (2004, 2005) propose to use wrapping the wavelet or filter coefficients around a circle of circumference and adding the overlying coefficients.

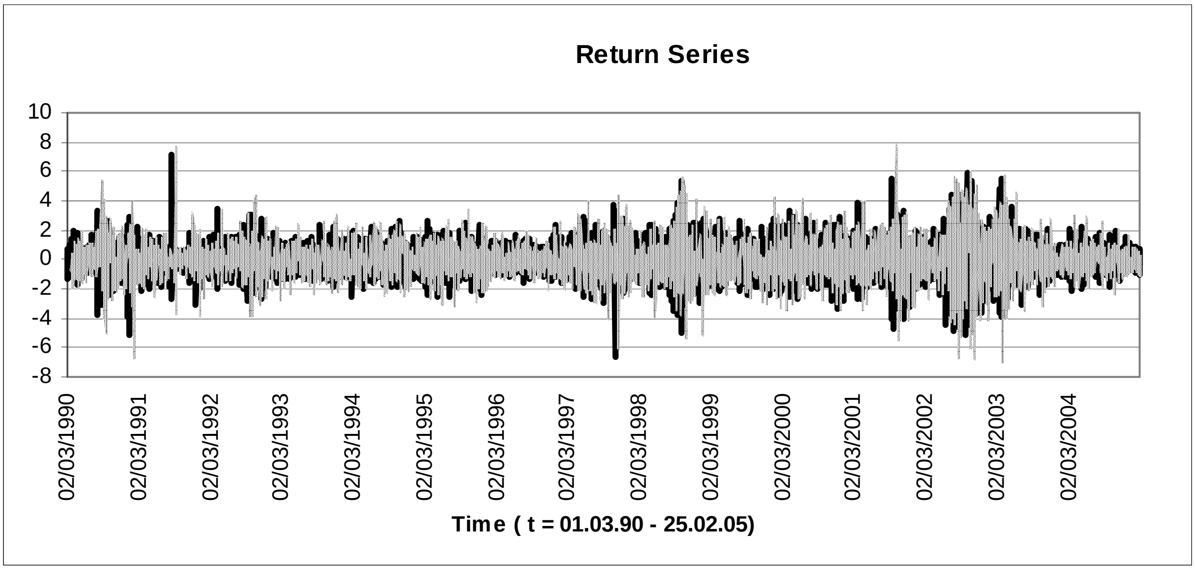


Figure 2. Highpass Filter. Gabor with Morlet Shape Modification. The original series - dotted curve; filtered output – solid curve.

The memory analysis of two series obtained indicates that low frequency signal now curries most of the memory (the Hurst exponent, $H =$ 0.098). Low persistency for high frequency signal is confirmed by the value of H close to 0.5. These results indicate that the original data series is successfully decomposed into signals currying separately main characteristics and allowing their individual analysis and modelling without undesirable interferences from each other.

Considering the individual outputs of the lowpass and highpass filters, the experiment detect that low and high frequencies move into and out of phase with each other. Figure 2 presents low and high frequencies on the same plot[15]. Around 03-09.00 [2501-2648] and 03-04.03 [3262-3302] one can see apparent phase shifts, rather then structural breaks, as it was claimed previously Hayward (2005).

[15] Note that the graphs are shifted so as to cause the features to align in a visually obvious manner.

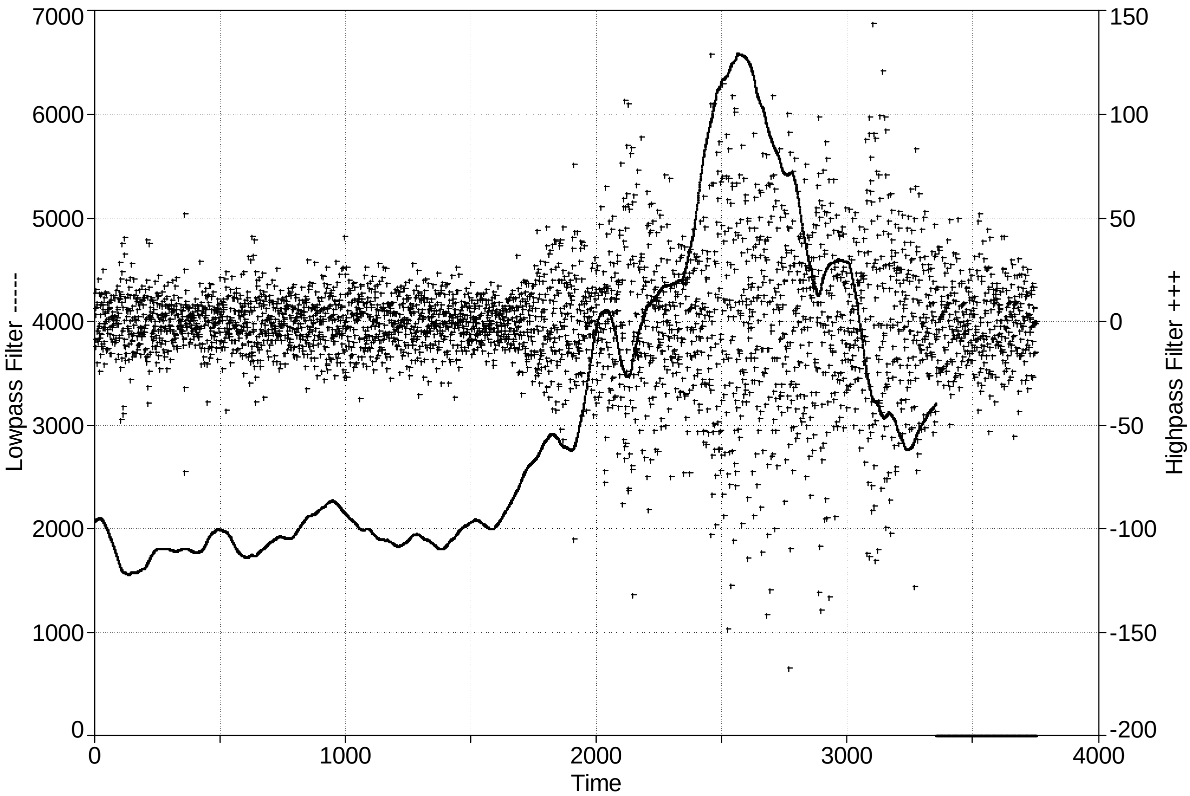


Figure 3. Low and High Frequencies.
Low frequency - solid curve; - high frequency - dotted curve.

Multiscale Prediction

The prediction scheme used in this study, steaming from the autoregressive (AR) methodology, is based on extracting important occurrences (such as turning points) from the data to determine the time of their possible reappearance in the future. In AR model fitted to individual scales of multiresolution transform, wavelet coefficients are used as a series of independent variables to forecast future values of the variable of interest. If the DGP is of autoregressive nature, the prediction converges to the optimal forecast and is asymptotically equivalent to the best forecast. ANN then learns the behaviour of the time series by changing the weights of the interconnected neurons comprising the network.

A heterogeneous beliefs model with expectations differentiated according to the time dimension is developed through decomposing a time series into a combination of the underlying series, representing beliefs of major clusters of market participants. In adaptive analysis of local behaviour of heterogeneous agents the high and low frequencies signals are distinguished to represent the short and long term traders. By separately investigating different frequency frames, the aim is to identify the dominant cluster of traders on the market considered and the adaptive nature of such market prices. Using observed values for each frequency, equations are evolved, resulting in separate trading signals for low (*S-L*), high (*S-H*) and combined (*S-LH*) frequencies:

$$\begin{aligned} &S\text{-}L = \{P_{t\text{-}i}^{LF}, fc(P_t^{LF}), R_t^{LF}\}, \\ &S\text{-}H = \{P_{t\text{-}i}^{HF}, fc(P_t^{HF}), R_t^{HF}\}, \\ &S\text{-}LH = \{P_{t\text{-}i}^{LF}, P_{t\text{-}i}^{HF}, fc(P_t^{LF}), fc(P_t^{HF}), R_t^{LF}, R_t^{HF}\}, \end{aligned} \tag{16}$$

where $P_{t\text{-}i}^{LF}, P_{t\text{-}i}^{HF}$ are price, $fc(P_t^{LF}), fc(P_t^{HF})$ are forecast and R_t^{LF}, R_t^{HF} are return series for low and high frequencies respectively. Weighted combination of individual frequencies provides the aggregate signal, *S-LH*. Weighting of individual frequencies is determined by GA and is specific for the market under investigation. As benchmarks two strategies are considered: buy and hold strategy (B/H) and a strategy based on undecomposed price series (S-P) with independent variables identical to those presented in (16).

For this experiment the structure developed in Hayward (2005) is adopted. A posterior optimal rule signal[16] is modelled with an EANN. A dual ANN consists of forecasting network feeding into the acting network. The information set of the acting network includes the output (one period ahead) of the forecasting (AR) network, the inputs used for forecasting (to relate the forecast to the data upon which it was based), as well as the log return, (given by $r(t) = r(\Delta t; t) = x(t) - x(t - \Delta t)$, where *x(t)* is a price series). In evolutionary settings a population of networks, representing agents facing identical problems but generating different solutions, is assessed on the basis of their fitness. A single hidden layer Time-Lag Recurrent Network with the number of the hidden layer's neurons determined genetically and GA with tournament selection, size 4; probability of uniform mutation [0, 0.05] and probability of uniform crossover [0.7, 0.95] are used in the experiment

To simulate 'true' forecasting and actual trading, the performance of the models evaluated on previously unseen data. To compare the developed models among themselves, the identical out-of-sample period is used and annualised results are presented. To achieve compatibility with similar studies, the internal error, determined by (widely adopted) MSE loss function is produced, comparing ANN output to the desired response, given by the next period price in

[16] The rule, using future information to determine the best current trading action, returns a buy/sell signal if prices tomorrow have increased/decreased.

forecasting and the current strategy in signal modelling. Directional accuracy (DA) is used for performance surface optimisation with GA. Optimising the learning the sign of the desired output is adopted for the reason of its established links with profitability (Leitch and Tanner 2001; Hayward 2004; Hayward 2005).

The experiment considered 75 forecasting and 120 trading strategies settings with multiple trials run for each settings. GA was capable to identify 'optimal' settings on the average in 80% of 10 individual runs. By simulating the traders' price forecast and their trading strategy evolution, eleven strategies were able to outperform the B/H benchmark in economic terms with an investment of €10,000 and TC of 0.3 of trade value. Those eleven strategies are represented by five strategies based on combined low and high frequencies; three on high frequency signal; two on low frequency and one strategy developed with undecomposed price series. Average return improvement over the B/H benchmark is 550% for *S-LH*; 370% for *S-H*; 74% for *S-L* and 20% for *S-P*.

Statistical characteristics, presented in Table 1 identify the best performance in forecasting and signal modelling by the unfiltered price series and low frequency signal. At the same time the statistical measures for the two most profitable strategies are far from satisfactory. These results confirm that economic profitability has only a weak relationship with common statistical measures. Therefore selecting a trading model on the basis of its statistical characteristics would result in overlooking potentially profitable models. Simulation with EANN provides a reliable platform for off-line models testing.

Table 1. Statistical Characteristics

Stats./Signals	S-P	S-L	S-FH	S-ALH
Accuracy (%)	58.3	50.7	43	34
Correlation	0.51	0.58	0.02	0.21
MSE Error	0.85	0.71	1.01	1.03

Accuracy: percentage of correct predictions ; Correlation: between desired and ANN output.

Table 2 presents out-of-sample economic performance of the main trading signals. The primary and secondary (in profitability terms) signals outperform *S-P* benchmark by 542% and 392% respectively. The risk exposure, measured by the Sharpe ratio, demonstrates that the primary and secondary strategies are respectively 15.75 and 14.5 times less risky than *S-P* benchmark. The largest percentage loss occurred during open trades, given by the Maximum Drawdown indicates that two primary strategies characterised by considerably lower downside risk. Superior performance of S-L and S-P signals in the percentage of winning trades should be viewed together with their infrequent trading and high downside risk exposures. The primary and particular secondary signals are characterised by high number of trades, indicating that those signals are exploited by short term speculators.

Even accounting for transaction costs of 0.3% , the extra returns achieved with the primary and secondary signals make those strategies the most profitable despite their high trading frequencies. The dealing frequency of the most active traders (making positive profit) confirms the right choice of the sampling frequency adopted in the experiment and so no re-sampling is necessary. For practical applications the information about dealing frequencies of profitable simulated traders could be used to readjust sampling interval (previously selected by some rule or even ad hoc).

Consider the length of trades, measured by the average number of bars the given type of trade is active. *S-L* signal is characterised by the longest holding periods, confirming that this signal is used by long term investors. On the other hand *S-H* signal display very short length of trades that might be particular suitable for short term speculators. Notice that a cumulative signal *S-LH* is also characterised by short holding periods suggesting that the market is likely to be dominated by short term traders. Such information about the behaviour of the dominant traders on the market under investigation is of a great value for predicting adaptive nature of the market prices. Furthermore, providing good guidance on how actively a signal needs to be traded such experiment can also be considered for the user making a decision to tune into the cycle best suited to an individual trading style.

Finally observing the length of each signal out of market, one can see that all signals based on the decomposed series spend very little time in out of market position in comparison to the signal based on undecomposed price series. This result is attributed to the improved abilities of EANN to learn information from signals decomposed on different frequencies. Comparing the signal constructed by a weighted combination of high and low frequencies with the signal developed using the price series, the former is more informative spending very time out of market (being unable to predict market behaviour).

Table 2. Economic Performance

	S-LH (Primary)	S-H (Second.)	S-L	S-P
Return Annual.	35.10	25.38	9.40	6.48
Sharpe Ratio	0.63	0.58	0.2	0.04
Profit/Loss	10,87	6,81	3.3	2.03
Max. Drawdown	-0.2	-1.94	-23.27	-14.19
Annual Trades	81	122	5	11
% Win. Trades	75.7	61.7	88.6	76.8
%Winning up Periods	49.77	50.28	50.19	47.53
Length of Trades Win/Loss/Out of Market	3/2/1	1/2/1	147/44/3	29/21/109

Annualised Return: $r = 1/N \sum_{t=1}^{N} R_t * 252$; Sharper Ratio: $SR = r / \sigma$; Max. Drawdown: $\xi = \underset{i=1,\ldots,t=1,\ldots,N}{Min} \left(\sum_{j=1}^{t} X_j \right)$; % Winning Trades: $WT = \text{Number of } R_t > 0 / (\text{Total Number of Trades}) * 100$

EAAN display good generalisation abilities, as indicated by percentage of winning up periods close to a half for most models considered. EAAN has successfully learned the in-sample information (constructed to be representative of upward and downward movements). It also displays an adequate performance in predicting a fall and a rise on out-of-sample data. To examine the stability of the proposed model different in/out-of-sample periods were considered. In particular the memory of the lengths of visually acknowledged complete cycles were adopted in the experiment, signifying that previous cycle behaviour might not be necessary informative for modelling of the current state. Only slight improvement in profitability was observed, given by average increase in annualised return by 1-3% and in Sharpe ratio by 2-3%. Thus it appears that decomposing the time series on the frequency signals (representing the time dimension of the major clusters of the market participants) is more effective for profitable models development than the search for the 'optimal' in and out of sample length (representing traders past and forward time horizons). In particular wavelet transforms, lending themselves to shifting along time in a consistent way, reveal features that have appeared in the past (without the necessity to determine with reliable precision the appropriate length of this past).

GA optimization results in average improvement in annualised return by 70%; Sharpe ratio riskness by 35%; and P/L by 270%. Optimized signals trade consistently less frequently with higher percentage of winning trades. At the same time the downside risk, measure by ξ is on average twice higher for signals optimised with GA. The last observation illustrates certain limitations with applications of computational intelligence in finance. Results obtained sometimes lacking clear reasoning. This shortcoming is often exacerbated by restrictions on using common statistical inference techniques (e.g. significance testing) with resultant dependence on heuristics.

Examining the weighting of the aggregate strategy *S-LH* the analysis identifies that high frequency signal dominate significantly the low frequency. Experiments with weights modification results in annualised return decrease, indicating that EAAN was capable to find the optimal frequencies combination. Therefore the market considered appears to be dominated by agents, trading primarily on the high frequency signals. The contribution of the agents using the low frequency although small, is not negligible. The emergent from signals decomposition local type of behaviour is proven to be more receptive to forecasting and modelling.

Conclusion

Assuming that lower level decomposition capture the long range dependencies whereas the higher levels identify short term relationships, the signal is decomposed into low and high frequencies. Considering the individual outputs of the lowpass and highpass filters, the experiment detect that low and high frequencies move into and out of phase with each other. Two apparent phase shifts (rather then structural breaks, as it was claimed previously) were identified.

A heterogeneous beliefs model with expectations differentiated according to the time dimension is developed through decomposing a time series into a combination of the underlying series, representing beliefs of major clusters of market participants. In adaptive analysis of local behaviour of heterogeneous agents the high and low frequencies signals are distinguished to represent the short and long term traders. The results of the experiment demonstrate improved abilities of EANN to learn information from signals decomposed on different frequencies. Comparing the signal constructed by a weighted combination of high and low frequencies with the signal developed using the price series, the former is more informative spending very time out of market (being unable to predict market behaviour). Examining the primary strategy, the analysis identifies that high frequency signals dominate significantly the low frequencies. Therefore the market considered appears to be dominated by agents, trading primarily on the high frequency signals. Though the contribution of the agents using the low frequency being small is not negligible. The emergent from signals decomposition local type of behaviour is proven to be more receptive to forecasting and modelling.

To examine stability of the proposed model different in/out-of-sample periods were considered. In particular the memory of the lengths of visually acknowledged complete cycles were adopted in the experiment, signifying that previous cycle behaviour might not be necessary informative for modelling of the current state. Only slight improvement in profitability was observed. Thus it appears that decomposing the time series on the frequency signals (representing the time dimension of the major clusters of the market participants) is more effective for profitable models development than the search for the 'optimal' in and out of sample length (representing traders past and forward time horizons). In particular wavelet transforms reveal features that have appeared in the past (without the necessity to determine with reliable precision the appropriate length of this past).

In search for the wavelet analysis appropriate to nonstationary data, generated by evolving structures that fall within non dyadic frequency bands, research related to Shannon wavelets (Pollock and Lo Cascio 2005) inspire their use in further experiments. Shannon wavelets are characterised by amendable dilations by arbitrary factors. Unlike popular Daubechies wavelets, Shannon wavelets are defined by a simple analytic function and are available in discretely sampled versions. Furthermore, their ordinates constitute the coefficients of symmetric digital filters that have no phase shift. Lastly the connection between the continuous and discrete time analyses is uniquely straightforward for Shannon wavelets. The filter coefficients of the discrete time analysis are also the coefficients of the dilation relationships given by the sampled ordinates and scaling functions. These features of Shannon wavelets motivate their use for financial applications in further research.

References

Dacorogna, M., R. Gencay, et al. (2001). An Introduction to High-Frequency Finance, Academic Press.

Daubechies, I. (1992). Ten Lectures on Wavelets, Society for Industrial and Applied Mathematics.

Farge, M. (1992). "Wavelet Transformations and their Applications to Turbulence." Ann. Rev. Fluid Mech.(24): 397-457.

Hayward, S. (2004). Evolutionary Artificial Neural Network Optimisation in Financial Engineering. Hybrid Intelligent Systems. M. Ishikawa, et al, IEEE Computer Society: 210-215.

Hayward, S. (2005). Quantitative Forecasting and Modeling Stock Price Fluctuations. Practical Fruits of Econophysics: Business Models in the 21st Century - Risk Management and Expectations for Econophysics. H. Takayasu, Springer.

Hayward, S. (2005). "The Role of Heterogeneous Agents' Past and Forward Time Horizons in Formulating Computational Models." Computational Economics: in press.

Kaastra, I. and M. Boyd (1996). "Designing a Neural Network for Forecasting Financial and Economic Time Series." Neurocomputing 10: 215-236.

Kurz, M. (1994). Endogenous Economic Fluctuations, Studies in the Theory of Rational Beliefs. Berlin, Springer.

Leitch, G. and E. Tanner (2001). "Economic Forecast Evaluation: Profits Versus the Conventional Error Measure." American Economic Review 81: 580-590.

Mallat, S. G. (1989). "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation." IEEE Transactions on Pattern Analysis and Machine Intelligence 11(7): 674-693.

Müller, U., M. Dacorogna, et al. (1990). "Statistical Study of Foreign Exchange Rates, Empirical Evidence of a Price Change Scaling Low, and Intraday Analysis." Journal of Banking and Finance 14: 1189-1208.

Pollock, D. S. G. and I. Lo Cascio (2004). "Adapting Discrete Wavelet Analysis to the Circumstances of Economics."

Pollock, D. S. G. and I. Lo Cascio (2005). Orthogonality Conditions for Non-Dyadic Wavelet Analysis. Queen Mary, University of London, Department of Economics, Working Papers.