Statistical analysis of financial time series under the assumption of local stationarity

Clémençon Stéphan
MODAL'X - Université Paris X Nanterre
Laboratoire de Probabilités et Modèles Aléatoires
UMR CNRS 7599 - Universités Paris VI et VII

Slim Skander
Théorie Economique Modélisation et Applications
UMR CNRS 7536 - Université Paris X Nanterre

Abstract

The aim of this paper is to apply a nonparametric methodology developped by Donoho, Mallat, von Sachs & Samuelides (2003) for estimating an autocovariance sequence to the statistical analysis of the return of securities and discuss the advantages offered by this approach over other existing methods like fixed-window-length segmentation procedures. Theoretical properties of adaptivity of this estimation method have been proved for a specific class of time series, namely the class of locally stationary processes, with an autocovariance structure which varies slowly over time in most cases but might exhibit abrupt changes of regime. This method is based on an algorithm that selects empirically from the data the tiling of the time-frequency plane which exposes best in the least squares sense the underlying second-order time-varying structure of the time series, and so may properly describe the time-inhomogeneous variations of speculative prices. The applications we consider here mainly concern the analysis of structural changes occuring in stock market returns, VaR estimation and the comparison between the variation structure of stock indexes returns in developed markets and in developing markets

1 Introduction

The modeling of the temporal variations of stock market prices have been the subject of intense research for a long time now, starting with the famous *Random Walk Hypothesis* introduced in Bachelier (1900), which claims that the successive variations $(X_{t+1} - X_t)_{t \geqslant 0}$ of a stock price are i.i.d. Gaussian random variables. As numerous statistical studies showed, even if X_t is replaced by $\log(X_t)$, this classic model does not allow to explain some prominent features of return series, such as the number of large price changes observed, that is much larger than predicted by the Gaussian (see Lo & Mackinlay (1988) for instance). As emphasized by many statistical works, that are far too numerous to mention (refer to Campbell, Lo & Mackinlay (1997) for a comprehensive overview), the following features of stock price series mainly came into sight (see Fig. 1).

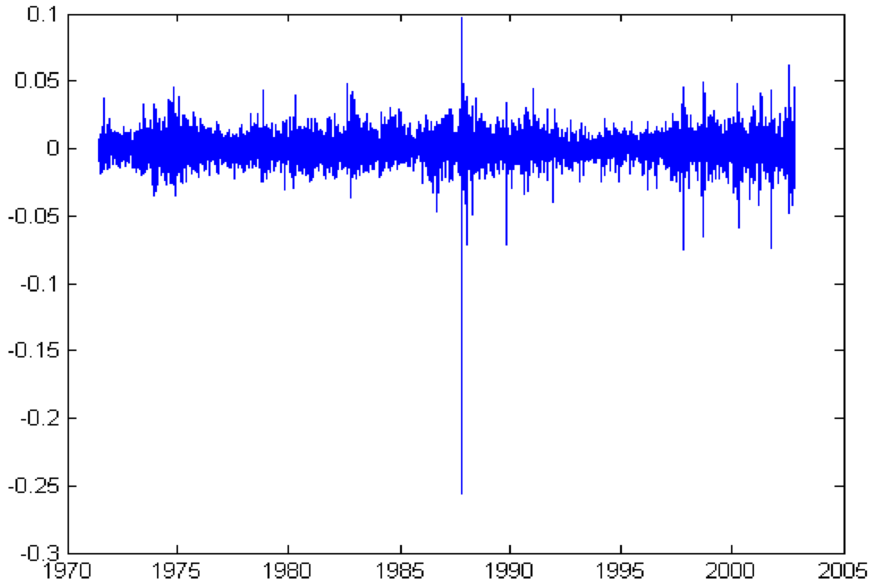

Figure 1: *Daily return series of the DJIA index, 1971-2002*

1. Spells of small amplitude for the price variations alternates with spells of large amplitude. This phenomenon is traditionally called *volatility persistence* or *volatility clustering*.
2. The "efficient markets assumption", which claims, roughly speaking, that financial returns are unforecastable, seems to be contradicted by the existence of very localized periods when return sequences exhibit strong positive autocorrelation.
3. The magnitude of the variations evolves in the long run so as to reach an "equilibrium" level, one calls this feature *mean-reversion* in a stylized manner.

Although the classic *Random Walk model* provides explicit formula for asset pricing and the economic doctrine is able to interpret it, the limitations mentioned above motivated the emergence of an abundant econometric literature, with the object to model structure in financial data. Portfolio selection/optimization, Value at Risk estimation, hedging strategies are the main stakes of this research activity, still developing. One popular approach consists in specifying an explicit model for the dynamic of return series, reflecting some patterns such as heteroskedasticity. Even if the seminal contribution of Engle (1982), which introduced the *ARCH model*, has been followed by a large number of variants (*GARCH models* for instance), and although so called *regime switching models* (see Hamilton (1989), Hamilton & Susmel (1994) for instance) and *stochastic volatility models*, assuming in some cases *long memory* in the process ruling the amplitude of the temporal variations of return series, were taken into consideration, the whole complexity underlying these data has not been captured yet by any parsimonious model and let the field of statistical analysis of financial time series open to further investigation. Therefore, selecting a statistical procedure, which allows to deal properly with the time-inhomogeneous character of return series, is not an easy task, as Mandelbrot emphasized (1963): "Price records *do not* "look" stationary, and statistical expressions such as the sample variance take very different values at different times; this nonstationarity seems to put a precise statistical model of price change out of the question". According to the estimation method chosen, one can either enhance specific patterns in the data or else make them disappear (see the discussion in Cont (2001)). This strongly advocates for the application of recent adaptive nonparametric procedures to the statistical analysis of financial return series, which, by selecting adaptively from the data a "best" representation among a large (nonparametric) class of models and including the type of structure that contributes significantly to the fit of the model only, allow to achieve more flexibility (refer to Ramsey (1996)). This alternative approach has been followed by some authors for several years now. Among these attempts to deal with non stationarities in financial data, one may mention the following works. Greenblatt (1996) considered the use of fast algorithms such as the *Matching Pursuit*, the *Method of Frames* and the *Basis Pursuit* (refer respectively to Mallat & Zhang (1993), Daubechies (1992) and Donoho & Chen (2001)) to select adaptively, from a dictionary, the superposition of "atoms", that is to say elementary functions localized both in time and frequency (wavelets for instance), that best exhibit structure in a financial time series, viewed as a noisy signal. The author applied this methodology to obtain sparse/parsimonious representations of exchange rate data in order to analyze the evolution of the frequency content of the underlying data generating process. In a closely related work, Ramsey & Zhang (1997) also applied the Matching Pursuit algorithm over a larger dictionary, namely a waveform dictionary, to decompose more efficiently exchange rates using tick-by-tick data, and noted particularly the presence of significant low frequency components. In Capobianco (2002), the presence of pronounced GARCH effects in high frequency financial time series is investigated after a preliminary denoising of the data using the *wavelet shrinkage* procedure (see Donoho & Johnstone (1998)). Several statistical procedures have been based on an explicit "functional" modeling of the nonstationarities occuring in financial time-series. Härdle, Herwartz & Spokoiny (2001) introduced and studied a method consisting in a sequence of nonparametric tests to identify periods of second-order homogeneity for each moment in time, as in Granger & Starica (2001), who considered the application of a test based on the integrated periodogram (see Picard (1985)). In Fryzlewicz, van Bellegem & von Sachs (2002), further developments of the general formalism defining the class of *locally stationary wavelet processes* (described at greater lenght in Kroisandt, Nason & von Sachs (2000)) and applications to the prediction and the time-varying second order structure estimation of the DJIA index are considered. The present paper aims to promote the use of an adaptive nonparametric methodology developped by Donoho, Mallat von Sachs & Samuelides (2003) for estimating the covariance of specific second-order nonstationary processes in the field of financial time-series statistical analysis. This method mainly amounts to analyze the data to find which out of a specific massive library of bases, namely *local cosine packets bases*, comes closest to diagonalizing the empirical covariance and make use of the latter to perform estimation. The bases of this library have localization properties both in the time domain and in the frequency domain, and the selected basis may conveniently exhibit the time-varying character of the second-order structure of the time-series. The paper is organized as follows. In section 2, the class of time-series for which the inference method mentioned above has been shown to surpass traditional procedures is described both in a qualitative and quantitative manner. Its relevance for modeling economic and financial phenomena is also discussed. The principle of the methodology is explained in section 3, and insights into the theoretical arguments explaining its performance are also given. In section 4, several empirical studies based on this analytical tool are carried out. From the resulting estimates, we investigate the temporal inhomogeneities in the fluctuation of financial returns. The estimation procedure is also applied via a simple *plug-in* approach to Value at Risk forecasting, and is shown to have advantages over less flexible methods based on moving averages. Finally, some concluding remarks are collected and several lines of research are sketched in section 5.

2 Second order local stationarity

2.1 Heuristics

As emphasized in several papers (see Ramsey (1996) for instance), a prominent characteristic of economic and financial data is the presence of temporal inhomogeneities. As a matter of fact, a significant part of the information carried by economic and financial time series consists in non-stationarities: beginning or end of certain phenomena, ruptures due to shocks or structural changes, drifts reflecting economical trends, business cycles ... Stationarity is a concept introduced to mean the independence of statistical properties from time. Hence, nonstationarity is a "non propriety", simply expressing the need for reintroducing time as a necessary description parameter, so as to be able to speak about the evolution through time of some properties of the time series and compute meaningful statistics. A constructive fashion to deal with nonstationary time-series consists in restricting oneself to a class of time series, for which one is able to specify precisely how they diverge from stationarity, while keeping a certain level of generality. Thus, many approaches may be considered. On grounds of parsimony, statistical analysis of stock prices variations mainly focused on the second order properties (that is by no means restrictive in the gaussian case), which amount to the covariance structure, since the assumption that financial returns are zero mean is beared out by both empirical evidence and theoretical economic arguments, and is carried unanimously. Consequently, it may be relevant to start with making assumptions on the autocovariance function. This approach has been followed by many authors, who considered the nonstationary framework. Let us consider a zero mean second order time series $X = (X_n)_{n \geqslant 0}$ with autocovariance function Γ_X $(\Gamma_X(n,m) = E(X_n X_m)$, for all $(n,m) \in \aleph^2)$. Note that the covariance between two observations at times n and m may be viewed as a function $C_X((n+m)/2, m-n)$ of the length between these time points and their midpoint (notice that C_X is independent from its first argument in the stationary case). It seems natural to call X a *locally stationary* time series, when it is "approximately stationary" (i.e. in a sense that should be precised) on time intervals of varying size and the variables are uncorrelated outside these intervals of quasi-stationarity. As this class of time series is supposed to describe random phenomena, which mechanism may evolve through time, it is legitimate to assume that the size $l(n)$ of the interval of quasi-stationarity may depend on the time n on which it is centered. Hence, a qualitative characterization of *locally stationary processes* could be as follows: on each time interval $[n - l(n)/2, n + l(n)/2]$, the covariance between observations X_m and $X_{m'}$ at times m and m' may be well approximated by a function depending only on $m' - m$ as soon these time points are close enough

$$\Gamma_X(m,m') \simeq C_X(n, m'-m) \text{ if } |m'-m| \leqslant l(n)/2$$

and is approximately zero when the length between the time points considered is larger than a certain threshold $d(n)$ measuring somehow the "decorrelation rate" of the time series

$$\Gamma_X(m,m') \simeq 0 \text{ if } |m-m'| > d(n)/2.$$

Under these assumptions it it can be shown that for any time points $m \in [n - l(n)/2,\ n + l(n)/2]$ and $m' \geqslant 0$

$$C_X((m+m')/2, m'-m) \simeq C_X(n, m'-m).$$

Set out in such general terms, the concept of local stationarity seems to be relevant for modeling financial data and account for the features 1-3 recalled in Section 1. As a matter of fact, the returns of a security (or a market index) are known to decorrelate rapidly when the market behaves in an "efficient way", on equilibrium, but when the latter is "evolving", when a change of business cycle occur for instance, the autocorrelation structure may evolve too and then one may attend changes of regime.

2.2 Assumptions

As recalled in the subsection above, there are many concepts of local stationarity. Even if their goal is almost the same, namely to allow to extend the statistical tools and concepts (mainly stemmed from Fourier spectral analysis) available in the stationary framework to the class of time series to which they are restricted (see Dalhaus (1997), Priestley (1965), Kozek (1996), Kroisandt, Nason & von Sachs (2000) or Mallat, Papanicolaou & Zhang (1998) for instance), not all the approaches yield a tractable statistical procedure for which a precise study of its performance regarding error risks may be carried out. In this respect, the one chosen in this paper combines several advantages. It has been worked out by Donoho, Mallat, von Sachs & Samuelides (2003) (refining the statistical method introduced in Donoho, Mallat & von Sachs (1998), which extended the approach developped by Mallat, Papanicolaou & Zhang (1998)), who both developped a full machinery to process the data (see Section 3 below) and provided theoretical arguments (rates of convergence) to support it (see also Mallat & Samuelides (2001) and Samuelides (2001)). Precisely, this methodology applies to Gaussian triangular arrays of second-order processes $X^{(T)} = (X_{t,T})_{0\leqslant t\leqslant T}$ (so as to formulate the properties of the time series with respect to the length of the observation) for $T = 2^{\tau}$, $\tau = \tau_1, \tau_1 + 1, ...$ obeying the assumption of *uniform decay of the autocorrelation*

$$\sum_{n=-t}^{T-t} \left(1 + 2\,|n|^{\delta_1}\right)^2 \Gamma^2_{X^{(T)}}(t, t+n) \leqslant c_1, \tag{1}$$

and the assumption of *quasi-stationarity of the covariance*

$$\frac{1}{T}\sum_{t=0}^{T} \|\Gamma_{X^{(T)}}(t, t+.) - \Gamma_{X^{(T)}}(t+h, t+h+.)\|_{l^2} \leqslant c_2 \left(\frac{|h|}{T}\right)^{\delta_2}, \tag{2}$$

for any h, where $\delta_1 > 1/2$, $0 < \delta_2 \leqslant 1$, c_1 and c_2 are constants. Beyond the scaling character of these assumptions, their main attraction is due to the averaging component in (2): a stochastic process $X^{(T)}$ obeying this constraint has a covariance matrix, which nearby rows $\Gamma_{X^{(T)}}(t, t+.)$ are, *on average*, very similar, but might occasionnally be very different, thus, allowing for sudden changes of regime. Note that the parameter δ_1 controls the *decorrelation rate*, while δ_2 affects the magnitude of the number of changes of regime.

3 The statistical procedure

3.1 Insights

Estimating the covariance matrix $\Gamma = \left(\gamma_{t,s}\right)_{0\leqslant t,s\leqslant T-1}$ of a zero mean Gaussian sequence $X = (X_t)_{0\leqslant t\leqslant T-1}$ from the observation of n independent realizations $X^{(i)}$ is a classical problem in traditional statistical analysis, when

T is fixed and n tends to infinity. In such a setting, the empirical covariance $C = (c_{t,s})_{0 \leqslant t,s \leqslant T-1} = n^{-1} \sum_{i=1}^{n} X^{(i)} X^{(i)'}$ is known to perform well: the expected error $E\left(l\left(C, \Gamma\right)\right)$ is in order of $O(T/n)$ when measured by the per-coordinate loss function

$$l\left(C, \Gamma\right) = T^{-1} \left\| C - \Gamma \right\|_{HS}^{2} = T^{-1} \sum_{t,s} \left(c_{t,s} - \gamma_{t,s}\right)^{2},$$

where $\left\| A \right\|_{HS} = \left(tr(AA')\right)^{1/2}$ denotes the Hilbert-Schmidt norm. This method is no longer successful when one gets out of this asymptotic framework, that is of course the case for the statistical analysis of financial returns, where $n = 1$ and T tends to infinity. In a *post-classical* setting, that is to say for T tending to infinity with n remaining fixed, it is neithertheless possible in some cases to develop a consistent estimation methodology. In the well-known case of a stationary time series for instance, an inference procedure may be built classically on the Toeplitz structure of the covariance matrix $\gamma_{s,t} = \gamma_{s+u,t+u}$ to recover: estimating first the covariance operator in the Fourier basis by its empirical counterpart, getting the periodogram, and then going back into the time domain (after a possible smoothing of the periodogram in the case when the underlying spectral density is known to be smooth), yields an estimate with an expected error in order of $O(1)$ at least (refer to Chapter 8 in Anderson (1971) for more details). A useful (with respect to further generalization to other settings) way of considering the reason of the success of such a methodology in the stationary situation has been carried in Mallat, Papanicolaou & Zhang (1998) (see also chapter 10 in Mallat (1998)): the authors observed that the crucial and paradygmatic point of the methodology lies in the fact that estimation is performed in a basis, namely the Fourier basis, that diagonalizes the covariance matrix, and thus by using such a *sparse* representation (in the sense that T coefficients only are required to characterize the covariance in the Fourier basis, instead of $T(T+1)/2$ in the original basis) one gets a statistical procedure with drastically reduced bias (avoiding this way a substantial component of estimation error). Mallat *et al.* (1998) noted also that for a time series satisfying the "qualitative" assumptions of local stationarity (see subsection 2.1), it is reasonable to expect the existence of a basis in the library $\mathcal{L}$ of cosine packets bases introduced by Coifman & Wickerhaüser (1991), that "almost diagonalizes" its covariance matrix. Thus, in the case when one disposes of a method to select such a basis empirically from the data, it becomes possible to make use of the sparse representation so provided to compute an estimate of the covariance matrix with low bias. Continuing this approach, Donoho *et al.* (2003) showed that a specific *tree pruning* algorithm, as introduced in Coifman & Wickerhaüser (1991), may be used to find empirically a basis in the library $\mathcal{L}$ that almost diagonalizes the best (in the least squares sense over $\mathcal{L}$) the covariance matrix of a time series, and yields a statistical procedure with provable performance properties (refer to Mallat & Samuelides (2001) and Samuelides (2001) for mathematical proofs of consistency) for the class of locally stationary processes they introduced (see the quantitative assumptions (1) and (2) in subsection 2.2).

3.2 The methodology

We now give precise details about the tools used in the statistical procedure and the underlying principles so that the resulting estimate may be conveniently interpreted.

3.2.1 Definitions

The library of cosine packets bases The crucial point in the statistical analysis of a stationary time series $(X_t)_{t\in\aleph}$ consists in viewing it as a linear superposition of uncorrelated periodic elementary time series $A_r e_r(t)$, where the e_r's denote the functions of the Fourier basis and the weights A_r are square integrable r.v.'s. The estimation of the variances of the A_r's from the record of the past observations of the time series yields both a low bias estimate of the covariance function and a spectral tool to analyze the structure of the time series (that is not evolving since stationarity is assumed): a current "physical" interpretation consists in measuring the relative importance of each periodic component $A_r e_r$ in the mechanism ruling the fluctuations of the time series by the variance of A_r. Since the 60's, spectral analysis has been in current use in econometrics for investigating the structure of economic series and computing predictions (see Granger (1964) for instance). The idea underlying the use of the Coifman & Wickerhaüser system to describe locally stationary processes is to keep the notion of an expansion of the time series in a basis made of mutually orthogonal cosine functions, while introducing the point of view of *temporal localization*. Hence, the construction of this system amounts to concatenate adequately the sequences

$$\xi_{M,m}(t) = \sqrt{\frac{2}{M}} \cos\left(\omega_m (t + 1/2)\right), \ 0 \leqslant t < M,$$

where $M = 2^j$ is a dyadic integer, $0 \leqslant m < M$ and $\omega_m = \pi(m + 1/2)/M$ (note that $\left\{\xi_{M,m}\right\}_{0\leqslant m<M}$ is an orthonormal basis of $\Re^M$). For reasons of a computational nature, the concatenations are induced by *recursive dyadic partitions* (RDP) of the time interval $\{0, 1, ..., T-1\}$, supposed to be of dyadic length $T = 2^\tau$. We recall that a RDP of $I_{0,0} = \{0, 1, ..., T-1\}$ is any partition reachable from the trivial partition $\mathcal{P}_0 = \{I_{0,0}\}$ by successive application of the following rule: choose a dyadic subinterval $I_{j,k} = \left\{kT/2^j, ..., (k+1)T/2^j - 1\right\}$ in the current partition and split it into two (dyadic) subintervals $I_{j+1,2k}$ and $I_{j+1,2k+1}$ of same size, creating a new (finer) partition of the time interval in this way (see Fig. 2).

Note that recursive dyadic partitioning may generate a very inhomogeneous segmentation of the time interval, with both very short subsegments and much longer ones for instance, so as to possibly properly describe the successive regimes of a nonstationary time series. Given a RDP $\mathcal{P}$ of the time axis $\{0, 1, ..., T-1\}$, one defines a *local cosine packets* basis $\mathcal{B}_\mathcal{P}$ of $\Re^T$ by setting

$$\varphi_{I_{j,k},m}(t) = \begin{cases} \xi_{2^{\tau-j},m}(t - kT/2^j) & \text{if } t \in I_{j,k} \\ 0 & \text{if } t \notin I_{j,k} \end{cases},$$

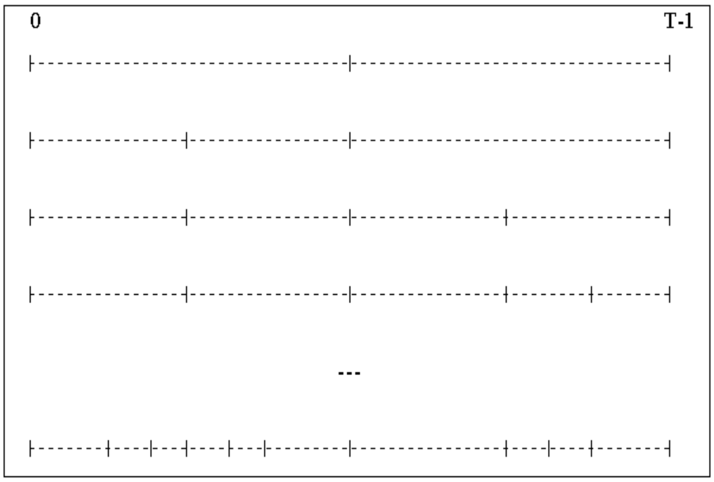

Figure 2: *Dyadic recursive partitioning scheme for a time interval* $\{0, ..., T-1\}$, $T = 2^{\tau}$.

for all $I_{j,k}$ in $\mathcal{P}$, and $0 \leqslant m < 2^{\tau-j}$. Beyond their orthonormal character, the vectors of such a basis have the crucial property of being localized both in time and in frequency (see Fig. 3): $\varphi_{I_{j,k},m}$ is supported on the subinterval $I_{j,k}$, on which it oscillates at the frequency ω_m.

Time-frequency representation Hence, every random sequence $X^{(T)} = (X_0, ..., X_{T-1})$ may be expanded in the local cosine packets basis $\mathcal{B}_{\mathcal{P}}$. Let $I_1, ..., I_{n_{\mathcal{P}}}$ be the subintervals forming $\mathcal{P}$. Then, one may write for $0 \leqslant t < T$

$$X_t^{(T)} = \sum_{u=1}^{n_{\mathcal{P}}} \sum_{m=0}^{2^{j_u}-1} \left\langle X^{(T)}, \varphi_{I_u,m} \right\rangle \varphi_{I_u,m}(t),$$

where $\langle ., . \rangle$ denotes the usual scalar product in $\Re^T$ and 2^{j_u} the length of the subinterval I_u. Thus, on each subsegment I_u of the time interval, one has a "Fourier type" decomposition of the time series into periodic time series,

$$\begin{aligned} \text{if } t &\in I_1,\ X_t^{(T)} = \sum_{m=0}^{2^{j_1}-1} \left\langle X^{(T)}, \varphi_{I_1,m} \right\rangle \varphi_{I_1,m}(t), \\ &\ldots, \\ \text{if } t &\in I_{n_{\mathcal{P}}},\ X_t^{(T)} = \sum_{m=0}^{2^{j_{n_{\mathcal{P}}}}-1} \left\langle X^{(T)}, \varphi_{I_{n_{\mathcal{P}}},m} \right\rangle \varphi_{I_{n_{\mathcal{P}}},m}(t). \end{aligned}$$

If, for each subinterval I_u, the components $\left\langle X^{(T)}, \varphi_{I_u,m} \right\rangle$ were almost un-

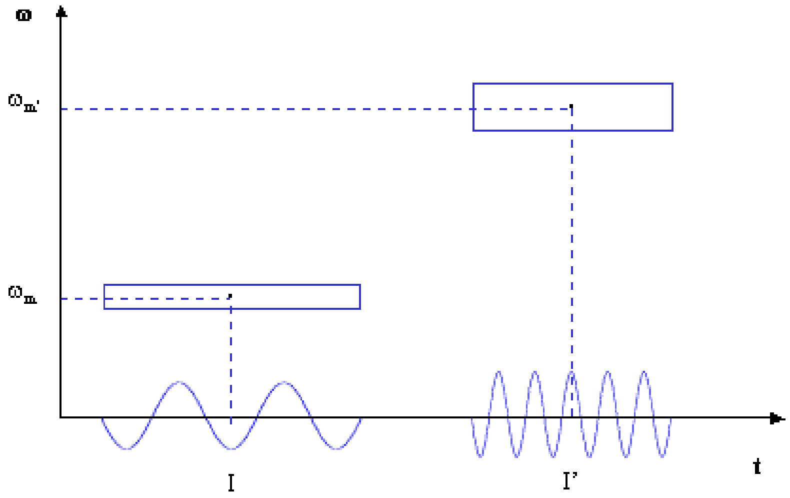

Figure 3: *Heisenberg tiles representing the time-frequency localization of local cosine functions.*

correlated, that is to say

$$E\left(\left\langle X^{(T)}, \varphi_{I_u,m}\right\rangle \left\langle X^{(T)}, \varphi_{I_u,m'}\right\rangle\right) \simeq 0 \text{ for } m \neq m',$$

or if $\mathcal{B}_{\mathcal{P}}$ almost "diagonalizes" $\Gamma_{X^{(T)}}$ in an equivalent way (since $\mathcal{B}_{\mathcal{P}}\Gamma_{X^{(T)}}\mathcal{B}_{\mathcal{P}}$ is the covariance matrix of the $\langle X^{(T)}, \varphi_{I_u,m}\rangle$'s), then one could interpret the segments $I_1, ..., I_{n_{\mathcal{P}}}$ as successive regimes of quasi-stationarity for the time series $X^{(T)}$. As recalled in subsection 3.1, Donoho *et al.* (1998) proved that for a *locally stationary* time series, there always exists such an "almost" diagonalizing basis and a *data-driven* method by complexity penalization has been introduced in Donoho *et al.* (2001) to select such a basis and yield a consistent covariance estimation procedure (see § 3.2.2 below).

3.2.2 The estimation algorithm

Let us now recall the inference procedure proposed by Donoho *et al.* (2003). We are interested in estimating the covariance matrix $\Gamma = \left(\gamma_{s,t}\right)$ of a zero mean *locally stationary* gaussian sequence on the basis of the observation of a single realization $(X_0, ..., X_{T-1})$ of this sequence. The method crucially relies on the construction of the library $\mathcal{L}$ of *local cosine packets* bases on $\Re^T$ and the use of the fast dynamic programming Coifman-Wickerhaüser tree pruning algorithm (*CW* algorithm in abbreviated form) for entropy-based best basis selection (refer to Coifman & Wickerhaüser (1992)). It requires no iteration and is performed in four steps as follows. In the sequel we shall denote by $\#\mathcal{A}$ the cardinal of any finite set $\mathcal{A}$.

1. *(Computation of the empirical variances)*

   Given the data $X = (X_t)_{0\leqslant t\leqslant T-1}$, *compute the empirical variances*

   $$s^2_{I_{j,k},m} = \left\langle X, \varphi_{I_{j,k},m} \right\rangle^2 ,$$

   for $0 \leqslant j \leqslant \tau$, $0 \leqslant k \leqslant 2^j - 1$, $0 \leqslant m \leqslant 2^{\tau-j} - 1$.

2. *(Best local cosine packets basis selection by complexity penalization)*

   Apply the CW algorithm to select a basis $\mathcal{B}_{\mathcal{P}}$ *(or equivalently, a RDP* $\mathcal{P}= I_1 \cup ... \cup I_{n_{\mathcal{P}}}$*) and a subfamily* $\mathcal{M}(\mathcal{B}_{\mathcal{P}})$ *of* $\mathcal{B}_{\mathcal{P}}$ *minimizing over all bases in* $\mathcal{L}$ *the additive cost*

   $$\mathcal{C}\left(\mathcal{B}_{\mathcal{P}}\right) = \sum_{u=1}^{n_{\mathcal{P}}} \mathcal{C}\left(I_u, \mathcal{M}^*\left(I_u\right)\right),$$

   where $\mathcal{M}^*(I_{j,k})$ is the subset of the set of indices $m \in \{0, ..., 2^{\tau-j} - 1\}$ of the local cosine vectors $\varphi_{I_{j,k},m}$ supported on $I_{j,k}$ which minimizes the cost

   $$\mathcal{C}\left(I_{j,k}, \mathcal{M}\left(I_{j,k}\right)\right) = -\sum_{m\in\mathcal{M}(I_{j,k})} \left(s^2_{I_{j,k},m}\right)^2 + \lambda.\#\mathcal{M}\left(I_{j,k}\right),$$

   over the class of subsets $\mathcal{M}\left(I_{j,k}\right)$ of $\{0, ..., 2^{\tau-j} - 1\}$.

3. *(Extraction of the diagonal)*

   Build a diagonal matrix Δ *with entries,*

   $$\tilde{s}^2_{I_u,m} = \begin{cases} s^2_{I_u,m} & \text{if } m \in \mathcal{M}^*\left(I_u\right) \\ 0 & \text{if } m \notin \mathcal{M}^*\left(I_u\right) \end{cases},$$

   $0 \leqslant m \leqslant 2^{j_u} - 1$, $1 \leqslant u \leqslant n_{\mathcal{P}}$, getting an estimate of the covariance matrix of X in the basis $\mathcal{B}_{\mathcal{P}}$ using local cosine coefficients defined by the collection $\mathcal{M}(\mathcal{B}_{\mathcal{P}})$ of vectors $\varphi_{I_u,m}$, $m \in \mathcal{M}^*\left(I_u\right)$, $1 \leqslant u \leqslant n_{\mathcal{P}}$, only.

4. *(Estimation)*

   Rotate back in the canonical basis, getting the empirical best basis covariance estimate

   $$C = \mathcal{B}_{\mathcal{P}} \Delta \mathcal{B}_{\mathcal{P}}'.$$

Remarks:

- Let us give an insight into the reason why the *tree pruning* algorithm used this way leads to a (rapid and easy) nearly optimal solution to the best-basis selection problem for covariance estimation. Recall that, ideally, we would like to find a basis $\mathcal{B}_{\mathcal{P}}$ in $\mathcal{L}$ that comes closest to diagonalize the covariance Γ in the least squares sense (see § 3.2.1 above), so as to produce an estimate, say $P_{\mathcal{M}(\mathcal{B}_{\mathcal{P}})}(C)$, of Γ in this basis by estimating a few significant diagonal coefficients only (corresponding to a set $\mathcal{M}(\mathcal{B}_{\mathcal{P}}) = \{\varphi_{I_u,m},\, m \in \mathcal{M}^*(I_u),\, 1 \leqslant u \leqslant n_{\mathcal{P}}\}$ of local cosine vectors of $\mathcal{B}_{\mathcal{P}}$) , with small mean squared error. Therefore, the squared error of such an estimate may be decomposed the following way

$$\begin{aligned}\left\|\mathcal{B}_{\mathcal{P}}\Gamma\mathcal{B}'_{\mathcal{P}} - P_{\mathcal{M}(\mathcal{B}_{\mathcal{P}})}(C)\right\|_{HS}^2 &= \left\|\mathcal{B}_{\mathcal{P}}\Gamma\mathcal{B}'_{\mathcal{P}} - P_{\mathcal{M}(\mathcal{B}_{\mathcal{P}})}(\Gamma)\right\|_{HS}^2 \\ &\quad + \left\|P_{\mathcal{M}(\mathcal{B}_{\mathcal{P}})}(\Gamma) - P_{\mathcal{M}(\mathcal{B}_{\mathcal{P}})}(C)\right\|_{HS}^2,\end{aligned}$$

denoting by $P_{\mathcal{M}(\mathcal{B}_{\mathcal{P}})}(\Gamma)$ the diagonal matrix obtained from the diagonal of $\mathcal{B}_{\mathcal{P}}\Gamma\mathcal{B}'_{\mathcal{P}}$ by keeping its diagonal coefficients $var(\langle X, \varphi_{I_u,m}\rangle)$ such that $\varphi_{I_u,m} \in \mathcal{M}(\mathcal{B}_{\mathcal{P}})$, and setting the others to zero. And, as we have

$$\left\|\mathcal{B}_{\mathcal{P}}\Gamma\mathcal{B}'_{\mathcal{P}} - P_{\mathcal{M}(\mathcal{B}_{\mathcal{P}})}(\Gamma)\right\|_{HS}^2 = \|\Gamma\|_{HS}^2 - e(\mathcal{B}_{\mathcal{P}}, \mathcal{M}(\mathcal{B}_{\mathcal{P}}))$$

(by using the invariance of the Hilbert-Schmidt norm by ortho-basis change), where

$$\begin{aligned} e(\mathcal{B}_{\mathcal{P}},\ \mathcal{M}(\mathcal{B}_{\mathcal{P}})) &= \sum_{u=1}^{n_{\mathcal{P}}} e(I_u,\ \mathcal{M}^*(I_u)), \\ \text{with } e(I_{j,k},\ \mathcal{M}(I_{j,k})) &= \sum_{m \in \mathcal{M}(I_{j,k})} \left(var(\left\langle X, \varphi_{I_{j,k},m}\right\rangle)\right)^2, \end{aligned}$$

an ideal strategy for constructing a covariance estimate with small mean squared error consists thus in selecting a basis $\mathcal{B}_{\mathcal{P}}$ and a subfamily $\mathcal{M}(\mathcal{B}_{\mathcal{P}})$ that minimizes the quantity

$$-e(\mathcal{B}_{\mathcal{P}}, \mathcal{M}(\mathcal{B}_{\mathcal{P}})) + E(\left\|P_{\mathcal{M}(\mathcal{B}_{\mathcal{P}})}(\Gamma) - P_{\mathcal{M}(\mathcal{B}_{\mathcal{P}})}(C)\right\|_{HS}^2),$$

in which the first term plays the role of the bias of the statistical method chosen and the second term the role of the variance in the language of statistical estimation theory. In Mallat & Samuelides (2001) it is proved that the variance term may be sharply bounded by $\#\mathcal{M}(\mathcal{B}_{\mathcal{P}})$ (which somehow measures the complexity of the estimator by the number of coefficients required to construct it) up to a multiplicative constant λ. This allows to reduce the best-basis selection problem to the search for a basis $\mathcal{B}_{\mathcal{P}}$ and a subfamily $\mathcal{M}(\mathcal{B}_{\mathcal{P}})$ that minimizes the theoretical cost

$$\mathcal{C}_{\mathcal{T}}(\mathcal{B}_{\mathcal{P}}, \mathcal{M}(\mathcal{B}_{\mathcal{P}})) = -e(\mathcal{B}_{\mathcal{P}}, \mathcal{M}(\mathcal{B}_{\mathcal{P}})) + \lambda.\#\mathcal{M}(\mathcal{B}_{\mathcal{P}}).$$

Given its additivity property, in the case when this theoretical cost could be calculated (that requires the knowledge of Γ) for any basis

$\mathcal{B}_{\mathcal{P}}$, the *CW* algorithm based on the penalized entropy $\mathcal{C}_{\mathcal{T}}$ would allow to solve this optimization problem and thus to find a quasi-optimal basis for covariance estimation. But, as Γ is precisely the object we try to estimate, the entropies $e(\mathcal{B}_{\mathcal{P}}, \mathcal{M}(\mathcal{B}_{\mathcal{P}}))$ are unknown and the selection has to be based on their empirical counterparts $\hat{e}(\mathcal{B}_{\mathcal{P}}, \mathcal{M}(\mathcal{B}_{\mathcal{P}}))$, so as to minimize the empirical cost $\mathcal{C}(\mathcal{B}_{\mathcal{P}}, \mathcal{M}(\mathcal{B}_{\mathcal{P}})) = -\hat{e}(\mathcal{B}_{\mathcal{P}}, \mathcal{M}(\mathcal{B}_{\mathcal{P}})) + \lambda.\#\mathcal{M}(\mathcal{B}_{\mathcal{P}})$. In Mallat & Samuelides (2001), the deviation between $\mathcal{C}_{\mathcal{T}}(\mathcal{B}_{\mathcal{P}}, \mathcal{M}(\mathcal{B}_{\mathcal{P}}))$ and $\mathcal{C}(\mathcal{B}_{\mathcal{P}}, \mathcal{M}(\mathcal{B}_{\mathcal{P}}))$ has been investigated, so as to show that implementation of the CW algorithm based on the empirical cost $\mathcal{C}$ produces an estimate with a mean squared error nearly as small as the error of an ideal estimate resulting from the minimization of the theoretical cost $\mathcal{C}_{\mathcal{T}}$.

- We observe that the algorithm does not fully rely on the data alone. The user has to choose the tuning parameter λ in the complexity penalization term. Building on the computation of the mean squared error, Mallat & Samuelides (2001) proposed a choice based on the top eigenvalue Λ_0 of the covariance matrix Γ. More precisely, they showed that, by taking $\lambda = c.\Lambda_0^2 \log(T)^2$ with a constant c large enough, the resulting covariance estimate has theoretical properties (regarding to the mean squared error) that surpass other methods for locally stationary processes.

- It is noteworthy that the assumptions (1) and (2) may be straightforwardly generalized so as to define bivariate (and even trivariate) *locally stationary* sequences $(X, Y) = ((X_0, Y_0), ..., (X_{T-1}, Y_{T-1}))$, and that the algorithm above may be adapted so as to analyze the latter. As a matter of fact, the only difference consists in replacing the entropy $\hat{e}(I_{j,k}, \mathcal{M}(I_{j,k}))$ by its two-dimensional analogue

$$\hat{e}_2(I_{j,k}, \mathcal{M}(I_{j,k})) = \sum_{m \in \mathcal{M}(I_{j,k})} \left(s(X)^2_{I_{j,k},m} + s(Y)^2_{I_{j,k},m} \right)^2,$$

  $s(X)^2_{I_{j,k},m}$ (respectively $s(Y)^2_{I_{j,k},m}$) denoting the empirical estimate of $var(\left\langle X, \varphi_{I_{j,k},m} \right\rangle)$ (resp., of $var(\left\langle Y, \varphi_{I_{j,k},m} \right\rangle)$). Difficulties arising from the extension of this methodology to higher dimensions are discussed in Section 5.

- Whereas the variations of a stationary time series may be analyzed in the frequency domain via the periodogram, as we recalled in subsection 3.2.1, a very useful representation of the variations of a locally stationary time series in the time-frequency plane may be provided by a "best basis" $\mathcal{B}_{\mathcal{P}}$ in the following way. Note first that, by assigning to each local cosine function $\xi_{I_{j,k},m}$, $0 \leqslant j \leqslant \tau$, $0 \leqslant k < 2^j$, $0 \leqslant m < 2^{\tau-j}$, the rectangle $H_{j,k,m} = I_{j,k} \times I_{\tau-j,m}$ in the time-frequency plane $I_{0,0} \times I_{0,0}$, each basis $B_{\mathcal{P}}$ may be described by a specific tiling $\mathcal{T}_{\mathcal{P}}$ of the time-frequency plane (see Fig. 4). Then, given a "best basis" $B_{\mathcal{P}}$, by assigning to each rectangle $H_{j,k,m}$ in $\mathcal{T}_{\mathcal{P}}$ the variance of $\left\langle X, \xi_{I_{j,k},m} \right\rangle$, one gets a "portrait" of the covariance, in that it describes how much variance of the time-series is associated to the frequency $(m + 1/2)/2^{\tau - j}$ on a quasi-stationary interval $I_{j,k}$ in $\mathcal{P}$.

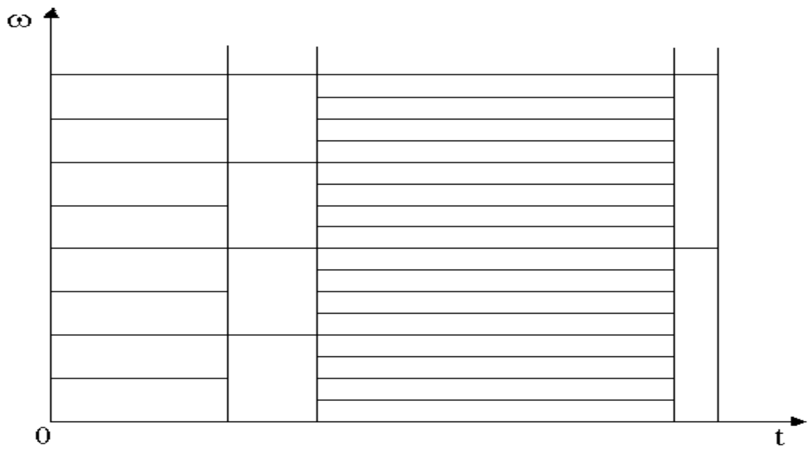

Figure 4: *Time-frequency tiling related to a local cosine packets basis. The collection of all the Heisenberg tiles forms a partition of the time-frequency plane.*

4 Applications - Empirical studies

We now turn to the application of the "best basis" method for covariance estimation to the statistical analysis of financial time series. Throughout this section, various data sources are used and several questions arising in market risk management are considered. Value at risk (VaR) techniques for risk management aim to quantify the amount of uncertainty about the return of a portfolio or a single asset over a given period of time. In a gaussian framework, this uncertainty is exhaustively described by the autocovariance structure. Thus, estimating accurately the autocovariance of financial returns is of great importance for assessing market risk. In an evolutionary context, the matter is to find a proper time window to compute meaningful statistics, reflecting the economic reality at a given time. Roughly speaking, depending on the way statistical averages are computed, one may either enhance specific relevant historical features in the data, or else obscure them, and even make them disappear. As shown in the following applications, the ability of the "best-basis" estimation method to identify, empirically from the historical data, periods of approximate stationarity, allows to exhibit patterns in the stock market volatility, that are attenuated or rubbed out when applying other methods based on the stationarity assumption or placing more emphasis on more recent historical data in a rigid fashion (as in the RiskMetrics^{TM} approach or by using fixed-window-length segmentation). As a matter of fact, by construction the latter cannot reflect nonlinearities such as jumps in stock returns for instance. The algorithm for best-basis covariance estimation (see 3.2.2) has been implemented by using routines of *Wavelab .701*, a *Matlab*^{TM} toolbox for wavelet and cosine packets analysis.

4.1 Covariance estimates

4.1.1 Statistical analysis of the DJIA index

We now present the results of the analysis of the DJIA index daily returns from 1971 to 2002 through the best-basis method for covariance estimation. In Figure 5, the tiling of the time-frequency space characterizing the empirical "best basis" selected via the algorithm detailed in 3.2.2 is represented, the variance of each component of the time-series in this basis is indicated by a proportional level of gray, and the variations over time of the (unconditional) volatility (*i.e.* the diagonal of the covariance matrix) are plotted in Figure 6.

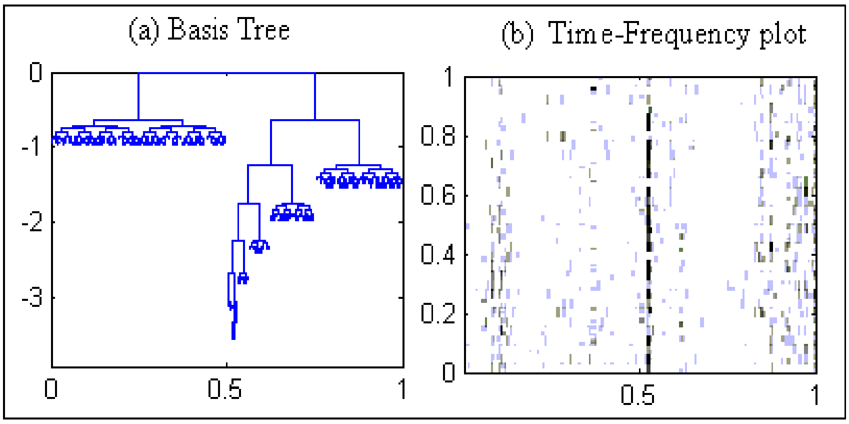

Figure 5: *Results of the best-basis method for the estimation of the autocovariance of the DJIA over the period 1971-2002 , $\lambda = 3.5 \cdot 10^{-8}$. (a): recursive dyadic partition of the time interval obtained by the CW algorithm, (b) time-frequency representation of the autocovariance*

As a comparison, we plotted the results of the variance estimation by using a moving average (see Taylor (1986)) with a fixed window length. Clearly, the best-basis method for covariance estimation reveals much more contrast in the volatility fluctuation and vouch for the global inhomogeneity of the variations of the DJIA index on a large scale. However, as Figure 5 shows, most of the basis functions of the "best-basis" for the DJIA returns are supported on time intervals of short length and the variation of the DJIA returns is mainly concentrated

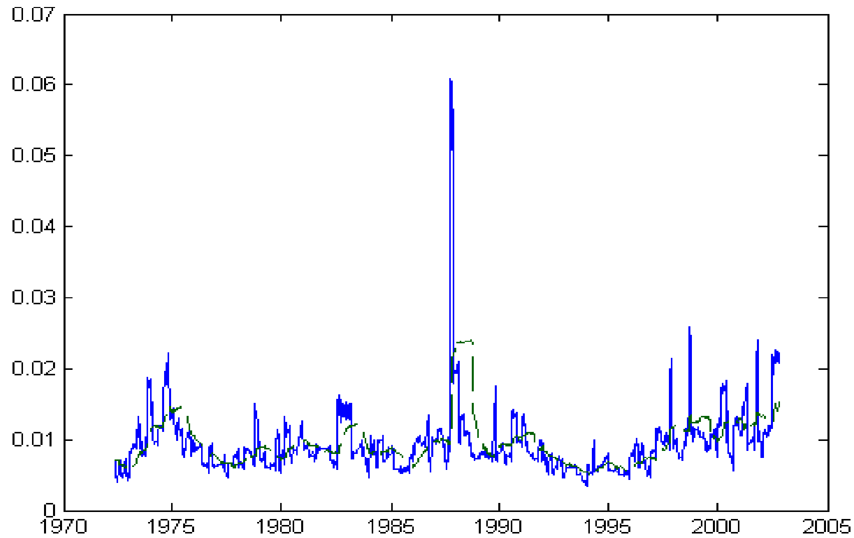

Figure 6: *Estimate of the daily unconditional volatility of the DJIA index using the best-basis method (solid line) and the moving average method with fixed window length n = 250 trading days (dashed line).*

at high frequency levels. This somehow accounts for the *efficiency* of the U.S. stock market in these times when the return series behaves nearly as a random walk: the return time-series decorrelates very rapidly and is thus unpredictable on a short horizon over these periods of time. Besides, some peculiar periods and the specific structure of the autocovariance of the DJIA return series over these periods clearly come into sight by examining the estimation results. It becomes visible that over some of these periods the basis functions describing the variations of the DJIA returns are supported on much longer time intervals, and that low frequency components significantly contribute to the covariance of the time-series. Among such singular periods detected by the best-basis selection, one counts several approximatively stationary time-intervals of 128 trading days, which exhibit a long persistence for the autocorrelation (see Fig. 7 for instance): from May 1971 to November 1971, from August 1974 to December 1974, from October 1976 to April 1977, from September 1982 to May 1983 and from July 1990 to January 1991. Political and economic considerations may provide helpful interpretation for these results. For instance, the structure in the DJIA returns between August 1974 and December 1974 may be linked of course to the quadrupling of the oil price by OPEC in 1973 and to the increase of debt that followed. During this period, a significant credit crisis in the U.S.has also been recorded, as well as the bankrupt of the Franklin National Bank. For the period running from September 1982 to May 1983, it may be noteworthy that the inflation rate has considerably lowered and that in December 1982, the unemployment rate has reached its highest level since 1940. Besides, the events of the Gulf War may partially explain the persistence of autocorrelation observed between July 1990 and January 1991 and the substantial variance of some low frequency components of the return series from November 1997 to February 1998 possibly reflects the impact of the Asian crisis on the U.S. stock market. Generally speaking, the covariance estimate rends fairly well the sudden changes of regime in the DJIA volatility, when the market switches from a calm regime to a volatile one. Beyond the krach in October 1987, the decomposition of the covariance estimate in the time-frequency plane allows to discern clearly the shock caused by the Russian financial crisis in August 1998 and the burst of the speculative information and technology bubble in 2001 for instance.

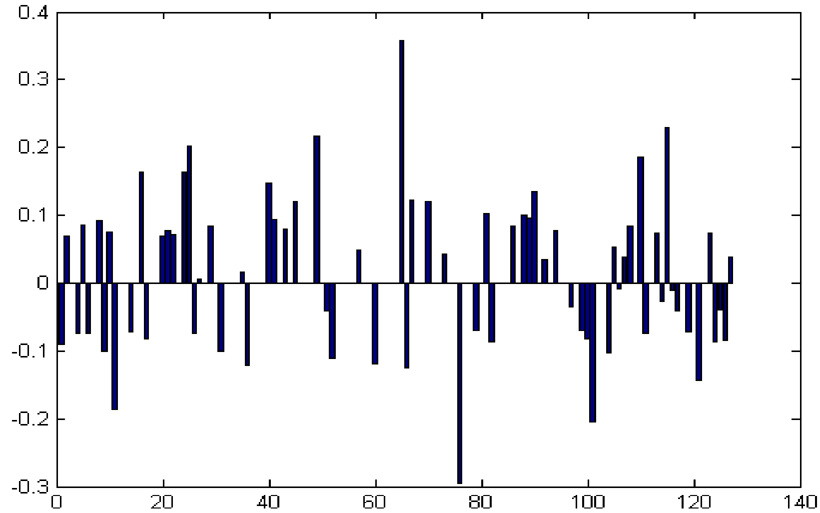

Figure 7: *Estimate of the autocorrelation of the DJIA return series on a quasi-stationary interval of length* $n = 128$ trading days.

4.1.2 Correlation between the DJIA index and the CAC 40

As we noticed in 3.2.2, the best-basis methodology may be extended to the case of bivariate time series. The latter is here applied to the estimation of the joint autocorrelation of the DJIA daily returns and the CAC 40 daily returns over the period 1992 - 2001. Figure 8 shows the plot of the estimate based on the best-basis selection of the time-varying correlation between these two series. Although this estimate tells us that, on average, the CAC 40 and the DJIA return series are strongly positively correlated, as a simple moving average with fixed window length would do, it allows to exhibit changes in their joint behaviour very localized in time, such as the one that occured in the second part of the year 1995. According to the latter, these two return series were strongly negatively correlated at this time. This phenomenon may be partly explained by the wave of social protest, which occured then in France.

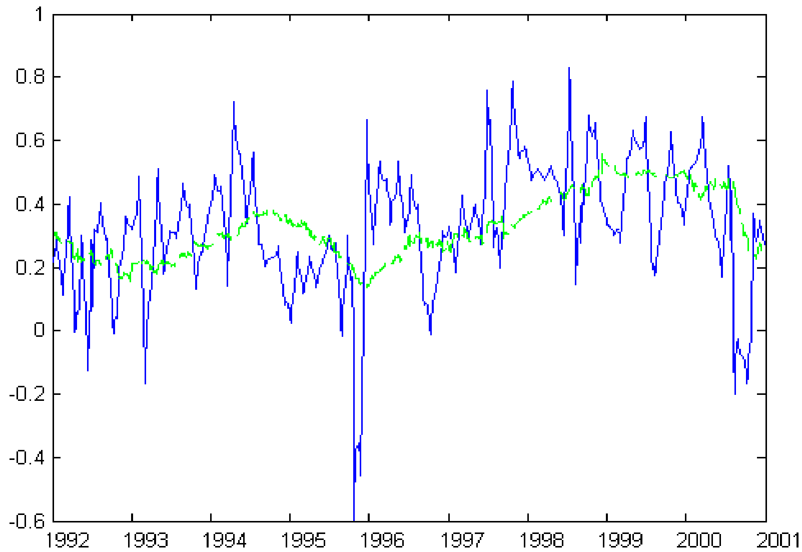

Figure 8: *Estimate of the daily correlation between the DJIA and the CAC40 using the best-basis method (blue solid line),* $\lambda = 10^{-7}$ *and the moving average method with fixed window length* $n = 250$ *trading days (green dashed line).*

4.1.3 Comparing developing and developed markets

The best-basis method for covariance estimation provides also empirical evidence to support that the temporal behaviour of stock market returns usually diverges more from efficiency in emerging markets than in developed markets. To this concern, we plotted the time-frequency representation of the return series of the IGPA index (Chile) for the period 1986 - 2002 (see Fig. 9) as estimated by the best-basis methodology, in comparison with the covariance estimate of the DJIA return series (refer to Fig. 5). As may be confirmed by application to many other series to our own experience, quasi-stationary time intervals, on which the return series exhibits a strong autocorrelation, are larger and much more frequent in developing markets (that suggests more forecastability from past observations for these time series, see also subsection 4.2). Beyond the difference related to the form of the time-frequency tiling representing the covariance, these estimates indicate also that large and abrupt volatility movements occur much more frequently for the IGPA return series (see also Fig. 10). Such empirical findings may motivate a careful study of the difference between the price-generating mechanisms and give rise to different economic modeling. This is beyond the scope of this paper, but will be the focus of further investigation.

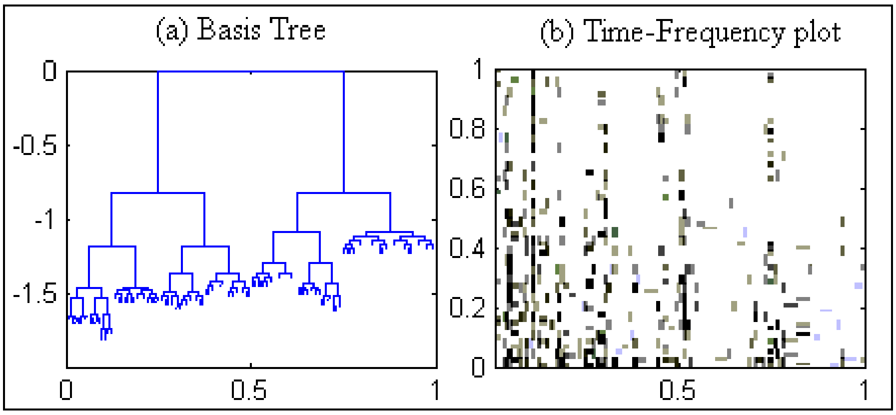

Figure 9: *Results of the best-basis method for the estimation of the autocovariance of the IGPA index over the period 1986-2002, $\lambda = 6 \cdot 10^{-8}$. (a): recursive dyadic partition of the time interval obtained by the CW algorithm, (b) time-frequency representation of the autocovariance*

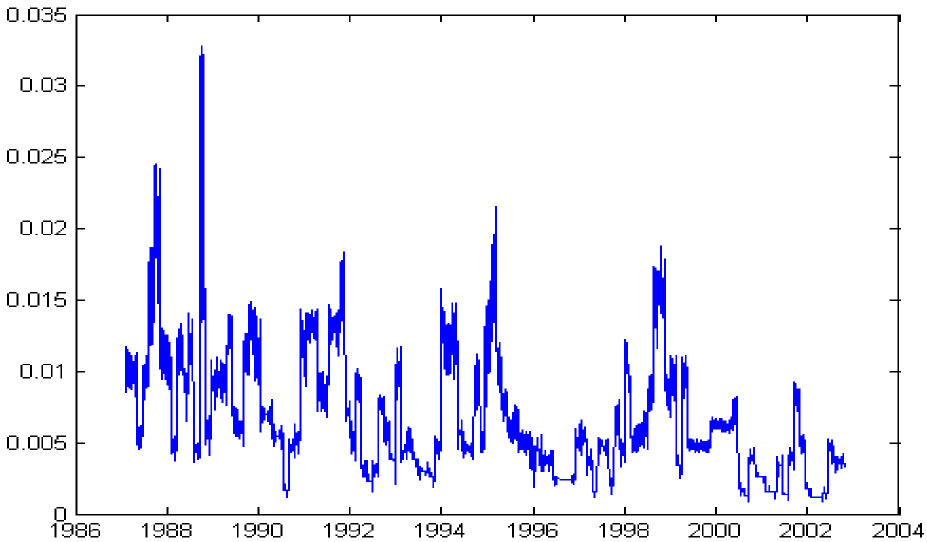

Figure 10: *Estimate of the daily unconditional volatility of the IGPA index using the best-basis method*

4.2 Value at Risk

4.2.1 VaR estimates based on the local stationarity assumption

As recalled above, VaR techniques intend to quantify the risk for an asset (respectively, a stock index, a portfolio), by measuring, in most cases, the level of loss $VaR_{t,h}(\alpha_0)$ that the asset price I_t could loose over a given time horizon h with a given degree of confidence $1-\alpha_0$ at time t conditionally on the information available $\mathcal{I}_t$:

$$P\left((I_{t+h}-I_t)/I_t \geqslant VaR_{t,h}(\alpha_0) \mid \mathcal{I}_t\right) = 1-\alpha_0.$$

In practice, all amounts to compute a forecast distribution $\mathcal{L}_{t,h}$ conditioned on updated segments of historical data, typically calculated through a *plug-in* approach. As, operationally, risk is often assessed at a 1-day horizon, we focus here on the case $h=1$. Since, we restricted ourselves to Gaussian sequences in the specific definition of locally stationary processes we considered here (see subsection 2.2), the problem reduces from a practical point of view to the estimation of the mean and variance of the return $X_{t+1}=(I_{t+1}-I_t)/I_t$ conditioned on the history $\mathcal{I}_t$ of the return series X at time t. The method we propose consists then in calculating estimates m_{t+1} and $\sigma^2_{t+1}$of the conditional mean and variance on the basis of the historical data $(X_{t-2^j+1},..., X_t)$ of the latest quasi-stationary interval of the return series (notice that the length 2^j of this interval varies with time t), adaptively selected according to the best-basis methodology (see 3.2.2) from a forward rolling data history of dyadic length $\mathcal{H}_t=(X_{t-2^\tau+1},..., X_t)$, and the Toeplitz covariance matrix estimate C_t of $\Gamma_t=(\Gamma_{r,s})_{t-2^j+1\leqslant r,s\leqslant t}=var(X_{t-2^\tau+1},..., X_t)$ thus obtained, as if the segment $(X_{t-2^\tau+1},..., X_t, X_{t+1})$ were a realization of an exactly stationary sequence. This yield the VaR estimate

$$\hat{VaR}_{t,1}=m_{t+1}+q_{\alpha_0}\sigma_{t+1},$$

where q_{α_0} denotes the α_0-quantile of the standard normal distribution.

Evaluation by backtesting Many other models have been suggested to forecast the return distribution and so to compute VaR estimates (refer to Dempster (2002) for an overview). One may compare the accuracy of different VaR statistical forecast systems by using the *Proportion of Failure test* (PF test) and the *Time Until First Failure test* (TFF test) considered in Kupiec (1995), as regulators do for the analysis of internal models. The PF test is based on the probability under the binomial distribution of observing that the number of times the observed value x_t for the return is lower than the forecast $VaR_{t,1}$ in a sample of size T is equal to n,

$$B\left(n;\alpha,T\right)=\binom{T}{n}\alpha^n\left(1-\alpha\right)^{T-n}.$$

Let H_0 be the hypothesis stipulating that the unconditional coverage α of a given VaR estimate equals the theoretical coverage level α_0. This hypothesis is tested by using the log-likelihood ratio statistic, asymptotically distributed according to the chi-squared distribution with one degree of freedom under H_0,

$$LR_{PF} = 2\log(\hat{\alpha}^{n}(1-\hat{\alpha})^{T-n}) - 2\log(\alpha_0^n(1-\alpha_0)^{T-n}),$$

where $\hat{\alpha}$ equals to $\mathbf{n}/T$, denoting by $\mathbf{n}$ the number of times the observed value for the return is lower than the forecasted Value at Risk $VaR_{t,1}$ over the sample. The TFF test is based on the duration before the first time the observed return is lower than the forecast $VaR_{t,1}$

$$\mathbf{t} = \inf\{0 \leqslant t < T/\ X_t < VaR_{t,1}\}.$$

The null hypothesis H_0 is tested by using the log-likelihood ratio statistic, asymptotically distributed according to $\chi^2(1)$ under H_0,

$$LR_{TFF} = -2\log(\hat{\alpha}(1-\hat{\alpha})^{\mathbf{t}-1}) + 2\log((1/\mathbf{t})(1-1/\mathbf{t})^{\mathbf{t}-1}).$$

Despite their limitation regarding their power to distinguish among alternative hypotheses (see Kupiec (1995)), we used here these two methods to compare the results of the VaR estimate based on the locally stationary assumption we proposed above with the performance of two classical approaches: the VaR estimate based on the simple moving average (MA) and the VaR estimate based on the Riskmetrics variance-covariance model (built by an exponential weighted moving average (EMA) with a decay factor $\lambda = 0.94$, see Riskmetrics (1996)), using for both a moving window with fixed length of 250 observation days. The data used in this statistical analysis are daily returns (based on closing prices) for 1994-2002 of 11 market indexes, among which 5 are related to developed stock markets and 6 to developing stock markets. The results of the PF and TFF tests are displayed in Tables 1 and 2. For high confidence levels, the null hypothesis H_0 is rejected in almost all cases, when VaR is forecasted by using either the moving average or the exponential weighted moving average. The VaR forecast based on the best-basis method clearly performs better. It generally presents higher P-values for both the PF test and the TFF test and leads to accept H_0 more frequently. Moreover, its advantage over the MA and EMA methods seems more obvious when dealing with emerging markets, which are characterized by frequent sudden changes of regime with large stock movements (see § 4.1.3) and for which the covariance estimate resulting from the empirical best-basis selection may provide a fairly better fit to the data by capturing these changes. Besides, it is noteworthy that even if, for lower degrees of confidence, the best-basis method does not always perform better than the MA and EMA methods regarding to the PF and TFF tests anymore, it does not tend to overestimate the risk (see Fig.11).

Table 1
P-values of the proportion of failure test (PF) for each method.

PF test	Degree of Confidence (%)	Best-Basis *P-value* (%)	EMA *P-value* (%)	MA *P-value* (%)
	2.5	3.64	3.64	3.64
France	1.0	23.06	0.10	0.16
	0.5	64.22	0.00	0.00
	2.5	83.84	17.39	3.64
Germany	1.0	36.21	4.31	0.04
	0.5	21.62	0.03	0.79
	2.5	84.05	0.51	55.09
Japan	1.0	1.00	0.04	2.23
	0.5	66.39	0.79	0.00
	2.5	8.37	42.93	0.00
U.K.	1.0	23.06	0.51	0.04
	0.5	10.71	0.28	0.01
	2.5	24.11	0.51	0.30
U.S	1.0	74.65	0.02	0.04
	0.5	66.39	0.03	0.00
	2.5	24.11	0.01	0.05
Argentina	1.0	53.77	0.00	0.00
	0.5	39.79	0.00	0.00
	2.5	40.50	1.43	0.51
Brazil	1.0	75.44	0.10	0.41
	0.5	66.39	0.00	0.00
	2.5	55.09	42.93	83.84
Chile	1.0	31.36	36.21	2.23
	0.5	12.58	21.62	2.80
	2.5	1.00	8.37	42.93
Hong-Kong	1.0	74.65	0.01	0.06
	0.5	21.62	0.00	0.00
	2.5	24.11	12.22	83.84
Mexico	1.0	13.90	0.51	13.90
	0.5	39.79	0.079	10.71
	2.5	42.93	1.43	3.64
Singapore	1.0	23.06	0.10	0.10
	0.5	10.71	0.00	0.00

Table 2
P-values of the time untill the first failure test (TFF) for each method.

TFF test	Degree of Confidence (%)	Best-Basis *P-value* (%)	EMA *P-value* (%)	MA *P-value* (%)
	2.5	20.47	8.45	2.65
France	1.0	7.05	4.82	1.40
	0.5	17.03	2.89	3.75
	2.5	9.91	0.87	68.05
Germany	1.0	12.86	0.40	58.26
	0.5	22.09	0.38	87.55
	2.5	67.02	33.19	3.19
Japan	1.0	60.50	77.44	18.73
	0.5	36.01	71.91	21.72
	2.5	53.87	12.84	27.54
U.K.	1.0	50.35	8.42	25.26
	0.5	87.14	5.17	31.52
	2.5	17.31	13.95	19.84
U.S	1.0	4.28	12.97	11.16
	0.5	52.98	7.62	7.62
	2.5	71.41	46.96	1.57
Argentina	1.0	52.25	79.99	6.63
	0.5	29.80	91.65	15.81
	2.5	62.80	46.82	7.64
Brazil	1.0	36.22	22.42	1.28
	0.5	96.94	1.44	2.88
	2.5	34.50	50.44	0.31
Chile	1.0	12.00	42.13	2.19
	0.5	98.55	28.97	8.58
	2.5	16.37	0.93	18.66
Hong-Kong	1.0	62.98	0.63	26.36
	0.5	74.82	4.84	8.22
	2.5	46.19	8.42	53.02
Mexico	1.0	67.00	0.66	2.52
	0.5	68.36	0.29	5.88
	2.5	46.53	38.11	55.33
Singapore	1.0	93.74	42.86	67.63
	0.5	62.98	22.64	83.84

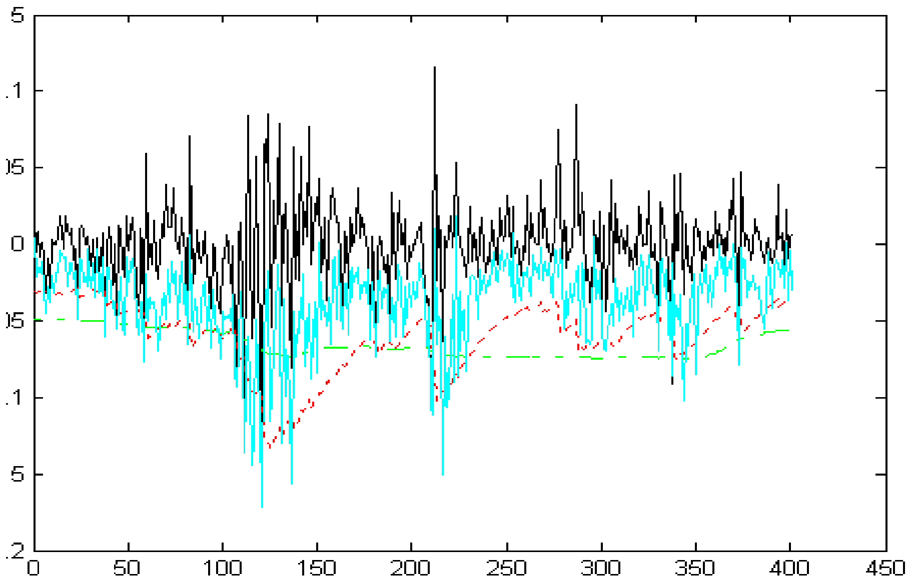

Figure 11: VaR forecast *of the Argentina market index using the best-basis method (cyan dotted line), the EMA method (red dash-dot line) and the MA method (green dashed line).*

5 Conclusion - Work for further research

Although in this paper we are far from having covered the whole range of applications of the best-basis method for covariance estimation to financial data (refer for instance to Fouque, Papanicolaou & Sircar (2000) for the application of a closely related algorithm to build option pricing tools) and having determined the limitations of such an approach, we end up by pointing out some issues and consider several questions open for further investigation. As emphasized in 3.2.2, the estimate resulting from the best-basis methodology, taking account of the present degree of advancement of the latter, depends on a tuning parameter λ, of which the value seems difficult to pick in practice. As a matter of fact, knowing that an optimal choice consists in choosing it from the top eigenvalue of the underlying covariance is useless from a practical point of view, since the latter is unknown and cannot be estimated in the nonstationary framework we consider. This suggests to study the properties of best-basis estimates thoroughly and to modify eventually the selection algorithm, so that a "universal" threshold may be found. Beyond this practical issue, the need for modeling return series by non Gaussian processes, justified by empirical evidence, so as to deal precisely with a possible evolution of skewness and kurtosis, could lead to introduce a tractable concept of locally stationary time series at orders higher than two and determine efficient "sparse" representations for their distributions and inference procedures based on the latter. This defines a second ambitious direction for further research. A third, and more ambitious problem would consist in building convenient nonparametric statistical tools to deal with multivariate financial time series of high dimension. This problem is of crucial importance in Portfolio selection/optimization for instance, which relies on the statistical estimation of the covariance matrix of the time-series of returns of D securities from historical data, when based on the *mean-variance approach* introduced by Markowitz (1952). Therefore, the asymptotic properties of the best *local cosine packets* basis method described in this paper depend heavily on the dimension D of the observations (one may refer to Donoho (2000) for a discussion of the "curse of dimensionality" phenomenon), and considerably lower as D increases. As a matter of fact, the uncertainty about the covariance matrix and its spectral properties is dramatically large when its dimensions are not small compared to the number of observations N on which statistical inference is based (and this is the case in practice, since we have for instance $D \approx 400$ securities and $N \approx 250$ observation days for the covariance matrix publicly posted daily by *RiskMetrics*^{TM}). This may lead to investigate which libraries contain bases of $\Re^D$ that may describe properly (and "sparsely") the variations of such multivariate time series on "almost" stationary periods. If such libraries were known, as well as an algorithm to select a proper basis through them, then one could try to elaborate a kind of "double best-bases" method for covariance estimation combining both the selection of approximatively time-homogeneous periods for the time series with the selection of bases representing efficiently the variations of the time series on these periods of time.

References

[1] Anderson, T. (1971). *The Statistical Analysis of Time Series.* Wiley & Sons.

[2] Bachelier, L. (1900). *Théorie de la spéculation.* Annales Scientifiques de l'Ecole Normale Supérieure, 3ème série, **17**, 21-88. Translation: *The Random Character of Stock Market Prices.* Ed. Paul Cootner, Cambridge, MA: MIT Press.

[3] Campell, J., Lo, A., Mackinlay, A.C. (1997). *The Econometrics of Financial Markets.* Princeton University Press, New Jersey.

[4] Capobianco, E. (2002). Multiresolution approximation for volatility processes. *Quantitative Finance*, **2**, No2, pp 91-110.

[5] Coifman, R., Wickerhaüser, (1992). Entropy based algorithms for best-basis selection. *IEEE Trans. Inf. Theor.* , **38**, 713-718.

[6] Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, **1**, No 2, pp 223-236.

[7] Dahlaus, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics,* **25**, 1-37.

[8] Daubechies, I. (1992). *Ten lectures in wavelets.* Philadelphia: SIAM.

[9] Dempster, M.A.H. (2002). *Risk management: beyond value at risk.* Cambridge University Press.

[10] Donoho, D. (2000). The curses and blessing of dimensionality. *Lecture delivred at the AMS concerns (mathematical challenges of the 21^{st} century).*

[11] Donoho, D.L., Chen, S. (2001). Atomic decomposition by basis pursuit. *SIAM Review,* **43**, No 1, 129-159.

[12] Donoho, D.L., Johnstone, I. (1998). Minimax estimation via wavelet shrinkage. *Annals of Statistics, 26, No 3, 879-921.*

[13] Donoho, D.L., Mallat, S., von Sachs, R. (1998). Estimating Covariances of Locally Stationary Processes: Rates of Convergence of Best Basis Methods. *Technical report 517, Statistics Department. Stanford University.*

[14] Donoho, D., Mallat, S., von Sachs, R. & Samuelides, Y. (2003). Locally Stationary Covariance and Signal Estimation with Macrotiles. *To appear in IEEE Transactions on Signal Processing.*

[15] Fouque, J.P., Papanicolaou, G., Sircar, K.R. (2000). Mean-Reverting Stochastic Volatility. *International Journal of Theoretical and Applied Finance*, **3**, No. 1, 101-142.

[16] Fryzlewicz, S., van Bellegem, S., von Sachs, R. (2002). Forecasting nonstationary time series by wavelet process modeling. *Technical Report*, Department of Mathematics, University of Bristol.

[17] Granger, C.W.J. (1964). *Spectral Analysis of Economic Time Series.* Princeton University Press.

[18] Greenblatt, S.A., (1996). Atomic Decomposition of Financial Data. *Computational Economics,* ***12,*** No 3, 275-293. Kluwer Academic Publisher.

[19] Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, **57**, 357-384.

[20] Hamilton, J.D., Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, **64**, 307-333.

[21] Härdle, W., Herwartz, H., Spokoiny, V. (2001). Time Inhomogeneous Multiple Volatility Modelling. Discussion Paper 7, Sonderforschungsbereich 373, Humboldt University, Berlin.

[22] Kozek, W. (1996). *Matched Weyl-Heisenberg Expansions of Nonstationary Environments.* PhD thesis, Dept. Communication and RF engineering, TU Vienna.

[23] Kroisandt, G., Nason, von Sachs, R. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectra. *J. R. Stat. Soc. Ser.* B, **62**, 271-292

[24] Kupiec, P.H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 73-84.

[25] Lo, A.W., Mackinlay, A.C. (1988). Stock prices do not follow random walk: evidence from a simple specification test. *Review of Financial Studies,* **1**, 41-66.

[26] Mallat, S. (1999). *A Wavelet Tour of Signal Processing.* 2nd Edition, Academic Press.

[27] Mallat, S., Papanicolaou, G., Zhang, Z. (1998). Adaptive covariance estimation of locally stationary processes. *The Annals of Statistics,* **26**, No 1, 1-47.

[28] Mallat, S., Samuelides, Y. (2001). Non-stationary covariance estimation with macrotiles. *Technical Report, Ecole Polytechnique.*

[29] Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business,* **36**, 394-419.

[30] Markowitz, H. (1952). Portfolio Selection. *Journal of Finance*, 77-91.

[31] Picard, D. (1985). Testing and estimating change-points in time-series. *Adv. Appl. Prob.,* **17**, 841-867.

[32] Priestley, M.B. (1965). Evolutionary spectra and non-stationary processes. *J. R. Statist. Soc. B,* **27**, 204-237.

[33] Ramsey, J.B. (1996). The Contribution of Wavelets to the Analysis of Economic and Financial Data. Unpublished manuscript.

[34] Ramsey, J.B., Usikov, D., Zaslavsky, G.M. (1995). An analysis of U.S. stock price behavior using wavelets. *Fractals,* **3**, No 2, 377-389.

[35] Ramsey, J.B., Zhang, Z (1997). The analysis of foreign exchange data using waveform dictionaries. *Journal of Empirical Finance,* **4**, 341-372

[36] RiskMetrics, (1996). *Technical Document*, Fourth Edition, JP Morgan. New York, USA.

[37] Samuelides, Y. (2001). *Macrotile Estimation and Market Model with Jumps.* PhD thesis, Ecole Polytechnique.

[38] Starica, C., Granger, C. (2002). Non-stationarities in stock returns. *Unpublished manuscript.*

[39] Taylor, S. J. (1986). *Modelling Financial Time Series.* Wiley & Sons.