A Haar-Fisz Technique for Locally Stationary Volatility Estimation

Piotr Fryzlewicz,
Department of Mathematics,
Imperial College,
London SW7 2AZ,
United Kingdom.
Email: p.fryzlewicz@imperial.ac.uk

Theofanis Sapatinas,
Department of Mathematics and Statistics,
University of Cyprus,
P.O. Box 20537,
CY 1678 Nicosia,
Cyprus.
Email: t.sapatinas@ucy.ac.cy

and

Suhasini Subba Rao,
Institut für Angewandte Mathematik,
Universität Heidelberg,
Im Neuenheimer Feld 294,
D-69120 Heidelberg,
Germany.
Email: suhasini@statlab.uni-heidelberg.de

January 11, 2005

Abstract

We propose a locally stationary model for financial log-returns whereby the returns are independent and the volatility is a piecewise constant function with an unknown number and location of jumps, defined on a compact interval to enable a meaningful estimation theory. We demonstrate that our model explains well the common stylised facts of log-returns. We propose a new wavelet thresholding algorithm for volatility estimation in our model, where Haar wavelets are powerfully combined with the variance-stabilizing Fisz transform. The resulting volatility estimator is mean-square consistent with a near-parametric rate, does not require any pre-estimates, is rapidly computable and easy to implement. We also discuss important variations on the choice of estimation parameters.

In an extensive empirical study, we fit our model to a variety of currency exchange datasets, and identify versions of our algorithm which perform particularly well in producing "well-behaved" empirical residuals. We show that our algorithm significantly outperforms GARCH in long-term volatility forecasting, and performs slightly better than a model-free moving window technique. We also demonstrate that the short-term volatility forecasting ability of our algorithm matches that of the GARCH(1,1) technique. In summary, by using our methodology, we are able to obtain both good fits to the data and convincing short- and long-term volatility forecasts.

Keywords: GARCH Models; Haar Wavelets; Locally Stationary Models; Variance-Stabilizing Transforms; Wavelet Thresholding.

1 Introduction

Log-returns on speculative prices, such as stock indices, currency exchange rates, share prices, etc., often exhibit the following well known properties:

1. The sample mean of the observed series is close to zero.
2. The marginal distribution of the observed series is roughly symmetric (or slightly skewed), has a peak at zero, and is heavy-tailed.
3. The sample autocorrelations of the observed series are "small" at almost all lags; however, the sample autocorrelations of the absolute values and squares of the observed series are significant for a large number of lags.

To capture the above "stylised facts", one needs to look beyond the stationary linear framework, and in order to preserve stationarity, a large number of non-linear models have been proposed. Among them, two branches are by far the most popular: the families of "autoregressive conditionally heteroscedastic" (ARCH) models, proposed by Engle (1982), and "generalised autoregressive conditionally heteroscedastic" (GARCH) models, proposed independently by Bollerslev (1986) and Taylor (1986), as well as the family of "Stochastic Volatility" (SV) models, suggested by Taylor (1986) as an alternative to ARCH/GARCH modelling. For a review of recent advances on ARCH, GARCH and SV modelling, we refer the reader to the monograph of Fan & Yao (2003) and the review paper of Giraitis *et al.* (2003).

Although stationarity is an attractive assumption from the estimation point of view, some authors point out that the above stylised facts can be better explained by resorting to non-stationary models (see, e.g., Mikosch & Stărică, 2000, 2003, 2004). Indeed, Kokoszka & Leipus (2000) considered the problem of change-point detection in the ARCH model, whilst Dahlhaus & Subba Rao (2003) proposed a locally stationary time-varying ARCH model where the parameters were permitted to change over time in a "slow" fashion. Underlying all these approaches is the observation that given the changing pace of the world economy, it is unlikely that log-return series should stay stationary over long time intervals.

However, an interesting question which arises once one abandons the assumption of stationarity, is whether non-linearity is still needed to model log-returns accurately, or whether it is sufficient to stick to linear models, the latter being conceptually simpler and better understood. Locally stationary linear models (see, e.g., Dahlhaus, 1997; Nason *et al.*, 2000) seem to be a particularly interesting option here, as they combine linearity with a modelling approach whereby the time-varying parameters are modelled as "well-behaved" functions defined on a compact interval, which enables a meaningful asymptotic estimation theory. Indeed, some authors have applied the locally stationary linear framework to the modelling of log-returns; see, e.g., Kim (1998) (who provides various statistical analyses of log-returns in the framework of Dahlhaus, 1997), Fryzlewicz (2004) (who provides an exploratory analysis of log-returns in the framework of Nason *et al.*, 2000), Clémençon & Slim (2004) (who apply the locally stationary covariance estimation methodology of Donoho *et al.*, 2003, to log-returns), and Drees & Stărică (2003) (who use the Nadaraya-Watson estimator to estimate the time-varying log-return volatility in a locally stationary setting).

Motivated by the above discussion, we also follow the “locally stationary linear” avenue and propose a simple non-stationary model for log-returns in which the time-varying volatility (= the log-return variance) is taken to be piecewise constant: this enables the modelling of abrupt changes in the stochastic regime which are often expected to follow the arrival of good or bad news in the market; in between the shocks, the volatility is modelled as constant. Also, we assume that log-returns at different time points are independent. Our aims in this paper are as follows:

- to argue that the proposed non-stationary model can accurately account for the most commonly observed stylised facts of log-returns;
- to propose a new wavelet thresholding technique for estimating the piecewise constant volatility: the proposed method powerfully combines the Haar wavelets and the variance-stabilizing Fisz transform;
- to obtain mean-square consistency results which demonstrate the validity of our approach;
- to show that the proposed non-stationary model provides an accurate fit to a selection of currency exchange rate datasets;
- to demonstrate that the proposed estimation approach yields accurate volatility forecasts.

The paper is organised as follows: in Section 2, we introduce our non-stationary model and motivate it by arguing that it is capable of explaining the most commonly observed stylised facts of log-returns mentioned above. In Section 3, we introduce our estimation methodology for the piecewise constant volatility in the proposed non-stationary model. Our wavelet thresholding technique is based on the Haar wavelets and the variance-stabilizing Fisz transform. We prove mean-square consistency results which demonstrate the validity of our approach. Important variations on the choice of estimation parameters are discussed. In Section 4, we demonstrate that the proposed estimation technique performs convincingly on real-life data, by fitting our non-stationary model to a selection of currency exchange rate datasets. In Section 5, we compare long- and short-term volatility forecasts obtained using the proposed methodology, the benchmark stationary GARCH(1,1) model, and a simple model-free moving-window technique, for a variety of currency exchange datasets. We show that our methodology does not perform worse than the competitors, and in many cases significantly outperforms them. Concluding remarks are provided in Section 6. The proofs are deferred to three Appendices.

2 The non-stationary model and motivation

In this section, we introduce our non-stationary model and motivate it by arguing that it is capable of explaining the most commonly observed stylised facts of log-returns.

2.1 The non-stationary model

Given a financial instrument $P_{t,N}$, $t = 0, 1, \ldots, N$ (e.g., a stock index, a currency exchange rate, a share price, etc.), our object of interest is the log-return series:

$$X_{t,N} := \log(P_{t,N}) - \log(P_{t-1,N}), \quad t = 1, 2, \ldots, N.$$

We propose the following "stochastic triangular array" model for the non-stationary process $\{X_{t,N} : \ t = 1, 2, \ldots, N\}$:

$$X_{t,N} = \sigma\left(\frac{t}{N}\right) Z_t, \quad t = 1, 2, \ldots, N, \tag{1}$$

where

- $\sigma(z) : [0, 1] \mapsto \mathbb{R}_+$ is a non-parametric function,
- $\{Z_t : \ t = 1, 2, \ldots\}$ is a sequence of independent and identically distributed (*iid*) random variables such that $\mathbb{E}Z_t = 0$, $\mathbb{E}Z_t^2 = 1$.

Here, $\sigma(z)$ (or alternatively $\sigma^2(z)$) can be viewed as a time-dependent "parameter" of the proposed model (1), which needs to be estimated from a single stretch of observations $X_{1,N}, X_{2,N}, \ldots, X_{N,N}$. Note that $\sigma(z)$ is defined over the (compact) interval $[0, 1]$, which is common practice in non-parametric regression and is done in order to enable the specification of regularity assumptions for $\sigma(z)$. Indeed, without such regularity assumptions, any attempts at estimating $\sigma(z)$ in a consistent manner would not be possible.

As we are primarily interested in $\sigma^2(z)$ (the local variance, or volatility, of the proces $\{X_{t,N} : \ t = 1, 2, \ldots, N\}$) rather than $\sigma(z)$ itself, we now specify the smoothness assumption for $\sigma^2(z)$ which will be used throughout the paper.

Assumption 2.1 *$\sigma^2(z)$ is a piecewise constant function, bounded from above and away from zero, with a finite (but unknown) number of jumps.*

Additional assumptions on the innovation process $\{Z_t : t = 1, 2, \ldots\}$ will be specified later.

Let us conclude the description of our model with some explanation regarding its nature. In contrast to the ARCH/GARCH modelling approach where the volatility $\sigma^2(z)$ is modelled as a stationary stochastic process (more specifically, a function of $\sigma^2(z)$ is modelled as a linear combination of certain functions of past volatilities and past squared log-returns), the non-stationary, non-parametric, unconditional approach described by the simple model (1) allows us to avoid the restrictions imposed by the parametric structures of ARCH/GARCH models. Similarly to Drees & Stărică (2002) and Stărică & Granger (2003), by modelling volatility as a non-parametric function, we do not claim that random effects do not play any role in the volatility dynamics. In our modelling approach, we express our belief that both past and future returns are manifestations of an unspecified exogenous economic factor about which we only assume a piecewise constant nature. Since no obvious candidates for explanatory exogenous variables are at hand, we model the volatility as a non-parametric, piecewise constant function.

2.2 Explanation of the common stylised facts

Below, we demonstrate that the non-stationary model (1) is capable of explaining the most commonly observed stylised facts of log-returns, mentioned in Section 1. We introduce the following notation:

$$\bar{X}_N^p = \frac{1}{N}\sum_{t=1}^{N} X_{t,N}^p, \tag{2}$$

$$\gamma_{X^p}^N(h) = \frac{1}{N}\sum_{t=1}^{N-h} X_{t,N}^p X_{t+h,N}^p - \left(\bar{X}_N^p\right)^2. \tag{3}$$

Throughout the paper, $\stackrel{P}{\to}$ denotes convergence in probability, and $\stackrel{L_2}{\to}$ denotes convergence in the L_2 norm, i.e., $(U_N \stackrel{L_2}{\to} u$, as $N \to \infty) \equiv (\mathbb{E}\{U_N - u\}^2 \to 0$, as $N \to \infty)$. The following proposition holds.

Proposition 2.1 *Suppose that $\{X_{t,N} : t = 1, 2, \ldots, N\}$ follows model (1), and that Assumption 2.1 holds. Assume further that $\mathbb{E}(Z_t^8) < \infty$. Then, we have*

$$\bar{X}_N^1 \stackrel{L_2}{\to} 0, \text{ as } N \to \infty, \tag{4}$$

$$\frac{\bar{X}_N^4}{\left(\bar{X}_N^2\right)^2} \stackrel{P}{\to} \mathbb{E}(Z_t^4)\frac{\int_0^1 \sigma^4(z)dz}{\left(\int_0^1 \sigma^2(z)dz\right)^2}, \text{ as } N \to \infty, \tag{5}$$

$$\gamma_{X^1}^N(h) \stackrel{L_2}{\to} 0, \text{ for a fixed } h > 0, \text{ as } N \to \infty, \tag{6}$$

$$\gamma_{X^2}^N(h) \stackrel{L_2}{\to} \int_0^1 \left(\sigma^2(z) - \int_0^1 \sigma^2(\omega)d\omega\right)^2 dz, \text{ for a fixed } h > 0, \text{ as } N \to \infty. \tag{7}$$

Typically, upon assuming that our observations come from a stationary (not necessarily linear) process $\{X_{t,N} : t = 1, 2, \ldots, N\}$, which is indeed often done in log-return analysis, we would use the quantities on the left-hand sides of formulae (4) – (7) as measures of the mean, the kurtosis, the autocovariance, and the autocovariance of the squares of the process $\{X_{t,N} : t = 1, 2, \ldots, N\}$ at lag $h > 0$, respectively. As mentioned in Section 1, we would then in most cases observe (a) the sample mean of the process $\{X_{t,N} : t = 1, 2, \ldots, N\}$ being close to 0, (b) the sample kurtosis of the process $\{X_{t,N} : t = 1, 2, \ldots, N\}$ being greater than 3 (3 being the kurtosis of a Gaussian variable), (c) the sample autocovariance of the process $\{X_{t,N} : t = 1, 2, \ldots, N\}$ at lag $h > 0$ being close to 0, (d) the sample autocovariances of the squares of the process $\{X_{t,N} : t = 1, 2, \ldots, N\}$ not decaying to 0, or decaying to 0 only very slowly. Proposition 2.1 shows that these stylised facts can be well explained in our non-stationary model (1): indeed, formulae (4) – (7) provide a heuristic explanation for the stylised facts (a) – (d), respectively, even in the case of Z_t being *iid* Gaussian. In particular, note that the ratio on the right-hand side of formula (5) is always greater that 1, unless $\sigma^2(z)$ is constant when it is equal to 1. Similarly, the integral on the right-hand side of formula (7) is always positive, unless $\sigma^2(z)$ is constant when it is equal to 0.

The above discussion indicates that care must be taken when applying stationary, global tools to the analysis of log-returns, as the true underlying model might well turn out to be non-stationary, as is indeed the case here. In particular, Proposition 2.1 demonstrates that the sample autocovariance $\gamma^N_{X^2}(h)$ is not an appropriate measure of $\mathrm{cov}(X^2_{t,N}, X^2_{t+h,N})$, for $h > 0$, as the latter quantity is always 0 in our non-stationary model (1), whereas the former quantity converges to a positive number if $\sigma^2(z)$ is non-constant.

3 A Haar-Fisz estimation theory

In this section, we aim to estimate $\sigma^2(t/N)$ at time points $t = 1, 2, \ldots, N$ from a single stretch of observations $X_{1,N}, X_{2,N}, \ldots, X_{N,N}$ from the non-stationary model (1). As we assume $\sigma^2(z)$ to be piecewise constant, we base our estimator on Haar wavelets, which, being also piecewise constant, are potentially good "building blocks" for this purpose. Our estimator uses the principle of nonlinear wavelet shrinkage, thus being potentially well-suited for the estimation of $\sigma^2(z)$ even if it is spatially inhomogeneous, i.e. if its regularity varies from one region to another. For an overview of wavelet methods in statistics, we refer the reader to the monograph of Vidakovic (1999) and the review paper of Abramovich *et al.* (2000).

The starting point for these considerations is a reformulation of the proposed non-stationary model (1):

$$X^2_{t,N} = \sigma^2\left(\frac{t}{N}\right) Z^2_t, \quad t = 1, 2, \ldots, N. \tag{8}$$

Note that $X_{t,N}^2$ is an unbiased but inconsistent estimate of $\sigma^2(t/N)$ at time points $t = 1, 2, \ldots, N$, and thus needs to be smoothed to achieve consistency. Obviously, (8) can be rewritten as:

$$X_{t,N}^2 = \sigma^2\left(\frac{t}{N}\right) + \sigma^2\left(\frac{t}{N}\right)(Z_t^2 - 1), \quad t = 1, 2, \ldots, N, \tag{9}$$

so that the problem of estimating $\sigma^2(t/N)$ at time points $t = 1, 2, \ldots, N$ can be viewed as the problem of removing the "noise" $\sigma^2\left(\frac{t}{N}\right)(Z_t^2 - 1)$ from $X_{t,N}^2$ at time points $t = 1, 2, \ldots, N$.

Neumann & von Sachs (1995) used a nonlinear wavelet estimation technique in a setting similar to (9). However, their method involved finding an estimate of the local variance of the "noise" (in our case: $\sigma^2\left(\frac{t}{N}\right)(Z_t^2 - 1)$), which in our case would amount to finding a pre-estimate of $\sigma^2(t/N)$ itself at time points $t = 1, 2, \ldots, N$. This is an obvious drawback of the estimation procedure, and can hamper the practical performance of the method (see, e.g., Fryzlewicz, 2004).

In order to avoid having to find a pre-estimate of $\sigma^2(t/N)$ at time points $t = 1, 2, \ldots, N$, an obvious step would be to take the logarithmic transformation of (8):

$$\log X_{t,N}^2 = \log \sigma^2(t/N) + \log Z_t^2, \quad t = 1, 2, \ldots, N. \tag{10}$$

The logarithmic transformation transforms model (8) from mutliplicative to additive, and acts as a variance-stabilizer: note that the variance of the "noise" $\log Z_t^2$ does not depend on t. In the special case of $\{Z_t : \ t = 1, 2, \ldots\}$ being a sequence of *iid* $N(0, 1)$ random variables, this setting is similar to the representation of the log-periodogram of a second-order stationary process considered by Wahba (1980). Several authors proposed wavelet techniques for the estimation of the log-periodogram (see, e.g., Moulin, 1994; Gao, 1997; Walden *et al.*, 1998; Pensky & Vidakovic, 2004), and those techniques could easily be adapted to our framework. However, any wavelet estimator in the setting specified by (10) would possess two undesirable properties:

- Naturally enough, it would be an estimate of $\log \sigma^2(t/N)$ (and not $\sigma^2(t/N)$ itself), at time points $t = 1, 2, \ldots, N$. Exponentiating this estimate would yield *an* estimate of $\sigma^2(t/N)$ at time points $t = 1, 2, \ldots, N$; however, the statistical properties of the latter, such as mean-square consistency, would not be easy to establish.

- Any wavelet estimator in the model (10) would suffer from a bias of order $\mathbb{E} \log Z_t^2$. Since, as mentioned before, we do not assume any specific distributional form of the innovation process $\{Z_t : \ t = 1, 2, \ldots\}$, the magnitude of the bias correction factor would then be unknown.

In contrast to those unwelcome features, the Haar-Fisz estimation technique which we propose below enjoys the following properties:

- It uses a variance-stablizing step, which eliminates the need for a local variance pre-estimation;
- It yields an asymptotically unbiased, mean-square consistent estimate of $\sigma^2(t/N)$ (as opposed to $\log\sigma^2(t/N)$) at time points $t = 1, 2, \ldots, N$, which removes the need for a bias correction factor;

Moreover, as we demonstrate below, it is conceptually simple, fast, easy to code, and performs well on several exchange rate datasets.

3.1 The Haar-Fisz estimation algorithm

The input to the algorithm is the vector $(X^2_{1,N}, X^2_{2,N}, \ldots, X^2_{N,N})$: here, we assume that N is an integer power of two. To simplify the notation, we drop the subscript N and consider the sequence $\{X_t^2 : t = 1, 2, \ldots, N\} := \{X^2_{t,N} : t = 1, 2, \ldots, N\}$. We denote $J = \log_2(N)$. The estimation algorithm proceeds as follows:

1. Compute the Haar decomposition of $\{X_t^2 : t = 1, 2, \ldots, N\}$ using the following algorithm:

   (a) Let $s_{J,k} := X_k^2, \quad k = 1, 2, \ldots, 2^J$.

   (b) For each $j = J-1, J-2, \ldots, 0$, recursively form vectors $\mathbf{s}_j$, $\mathbf{d}_j$ and $\mathbf{f}_j$ with elements:

$$
\begin{aligned}
s_{j,k} &= \frac{s_{j+1,2k-1} + s_{j+1,2k}}{\sqrt{2}}, \; k = 1, 2, \ldots, 2^j, \\
d_{j,k} &= \frac{s_{j+1,2k-1} - s_{j+1,2k}}{\sqrt{2}}, \; k = 1, 2, \ldots, 2^j, \\
f_{j,k} &= \frac{d_{j,k}}{s_{j,k}}, \; k = 1, 2, \ldots, 2^j,
\end{aligned}
$$

   respectively.

2. For each $j = J-1, J-2, \ldots, 0$ and $k = 1, 2, \ldots, 2^j$, denote $\mu_{j,k} := \mathbb{E}\, d_{j,k}$. For most levels j (in a sense to be made precise later), estimate $\mu_{j,k}$ by

$$
\begin{aligned}
\hat{\mu}^{(h)}_{j,k} &= s_{j,k} f_{j,k} \mathbb{I}(|f_{j,k}| > t_j) \\
&= d_{j,k} \mathbb{I}(|f_{j,k}| > t_j) \quad \text{(hard thresholding)}, \qquad (11)
\end{aligned}
$$

   or by

$$
\hat{\mu}^{(s)}_{j,k} = s_{j,k}\, \mathrm{sgn}(f_{j,k})\, (|f_{j,k}| - t_j)_+ \quad \text{(soft thresholding)}, \qquad (12)
$$

   where $\mathbb{I}(\cdot)$ and $\mathrm{sgn}(\cdot)$ are the indicator and signum functions, respectively, and $(x)_+ = \max(0, x)$ for any $x \in \mathbb{R}$. In other words, we "kill" each $d_{j,k}$ if and only if the corresponding *Haar-Fisz coefficient* $f_{j,k}$ does not exceed (in absolute value) a certain threshold t_j (to be specified later). Note that this is different to classical wavelet thresholding in that the estimated quantity $d_{j,k}$ and the "thresholding statistic" $f_{j,k}$ are *different*.

3. Invert the Haar decomposition in the usual way to obtain an estimate of $\sigma^2(t/N)$ at time points $t = 1, 2, \ldots, N$. Call the resulting estimate $\hat{\sigma}^2_{(h)}(t/N)$ (for hard thresholding) or $\hat{\sigma}^2_{(s)}(t/N)$ (for soft thresholding).

Asymptotic normality and variance stabilization for certain random variables of the form $(X - Y)/(X + Y)$, where X, Y are nonnegative, independent random variables, were studied by Fisz (1955): hence we label $f_{j,k}$ the "Haar-Fisz coefficients". The main heuristic idea here is that the variance of $f_{j,k}$ (for most j, k) does not depend on $\sigma^2(z)$. Consider the following example: $j = J - 1$, $k = 1$. The Haar-Fisz coefficient $f_{J-1,1}$ has the form:

$$f_{J-1,1} = \frac{X_1^2 - X_2^2}{X_1^2 + X_2^2} = \frac{\sigma^2(1/N)Z_1^2 - \sigma^2(2/N)Z_2^2}{\sigma^2(1/N)Z_1^2 + \sigma^2(2/N)Z_2^2}.$$

Suppose now that $\sigma^2(1/N) = \sigma^2(2/N)$ (this is likely as $\sigma^2(z)$ is piecewise constant). We then have $f_{J-1,1} = (Z_1^2 - Z_2^2)/(Z_1^2 + Z_2^2)$; the variance of $f_{J-1,1}$ does not depend on $\sigma^2(z)$. Thus, the thresholds t_j in (11) and (12) also do not need to depend on $\sigma^2(z)$, and can therefore be selected more easily.

Note that the above algorithm is different to the *Haar-Fisz transforms* of Fryzlewicz & Nason (2003, 2004) in that it combines the transform step 1 and the estimation step 2: this greatly facilitates the theoretical analysis of the arising estimate as we shall see below.

Denote

$$v := \mathbb{E}|Z_t^2 - 1|^2, \tag{13}$$

and consider the following assumption:

Assumption 3.1 *The law of the random variable Z_t^2 has no atom at 0, and*

$$\exists\, c > 0,\ \gamma \geq 0 \quad \text{such that} \quad \mathbb{E}|Z_t^2 - 1|^n \leq c^{n-2}(n!)^{1+\gamma}v, \quad \text{for all} \quad n \geq 3. \tag{14}$$

Remark 3.1 By Lemmas B.3 and B.4 (see Appendix B), Assumption 3.1 is satisfied, in particular, if Z_t is standard normal (with $v = 2$, $\gamma = 0$) or standard Laplace (with $v = 5$, $\gamma = 2$). Assumption 3.1 can also accommodate other distributions which are leptokurtic or possess a degree of skewness. Note that, using similar arguments to those employed in the proof of Proposition 2.1, it can easily be shown that non-zero skewness of Z_t translates into asymptotically non-zero sample skewness of $\{X_{t,N}\}_{t=1}^N$.

The following two theorems demonstrate the mean-square consistency of $\hat{\sigma}^2_{(h)}(z)$ and $\hat{\sigma}^2_{(s)}(z)$, respectively.

Theorem 3.1 *(Hard thresholding.) Suppose that* $\{X_{t,N}: \ t = 1, 2, \ldots, N\}$ *follows model (1), and that Assumptions 2.1 and 3.1 hold. Construct the estimator* $\hat{\sigma}^2_{(h)}(z)$ *as follows: fix* $\delta \in (0, 1)$*. For each* $N = 2^J$*, define the set*

$$\mathcal{I}_N = \{(j, k) : j < J^*\},$$

with $2^{J^*} = O\left(N^{1-\delta}\right)$*. Define*

$$\hat{\mu}^{(h)}_{j,k} = \begin{cases} d_{j,k}\mathbb{I}(|f_{j,k}| > t_j) & (j, k) \in \mathcal{I}_N, \\ 0 & (j, k) \notin \mathcal{I}_N, \end{cases}$$

where

$$t_j = 2^{-\frac{J-j-1}{2}} \sqrt{2 \log(N)}. \tag{15}$$

Then, we have

$$\begin{aligned} \frac{1}{N}\sum_{t=1}^{N} \mathbb{E}\left\{\hat{\sigma}^2_{(h)}(t/N) - \sigma^2(t/N)\right\}^2 &= \frac{v}{N^2}\sum_{t=1}^{N}\sigma^4(t/N) + \frac{1}{N}\sum_{j=0}^{J-1}\sum_{k=1}^{2^j}\mathbb{E}\left\{\hat{\mu}^{(h)}_{j,k} - \mu_{j,k}\right\}^2 \\ &= O\left(N^{-\min\left(1-\delta, \frac{2}{v}\right)}\right), \end{aligned} \tag{16}$$

where v *is defined as in formula (13).*

Theorem 3.2 *(Soft thresholding.) Suppose that* $\{X_{t,N}: \ t = 1, 2, \ldots, N\}$ *follows model (1), and that Assumptions 2.1 and 3.1 hold. Construct the estimator* $\hat{\sigma}^2_{(s)}(z)$ *as follows: fix* $\delta \in (0, 1)$*. For each* $N = 2^J$*, define the set*

$$\mathcal{I}_N = \{(j, k) : j < J^*\},$$

with $2^{J^*} = O\left(N^{1-\delta}\right)$*. Define*

$$\hat{\mu}^{(s)}_{j,k} = \begin{cases} s_{j,k}\,\mathrm{sgn}(f_{j,k})\left(|f_{j,k}| - t_j\right)_+ & (j, k) \in \mathcal{I}_N, \\ 0 & (j, k) \notin \mathcal{I}_N, \end{cases}$$

where

$$t_j = 2^{-\frac{J-j-1}{2}} \sqrt{2 \log(N)}. \tag{17}$$

Then, we have

$$\begin{aligned} \frac{1}{N}\sum_{t=1}^{N} \mathbb{E}\left\{\hat{\sigma}^2_{(s)}(t/N) - \sigma^2(t/N)\right\}^2 &= \frac{v}{N^2}\sum_{t=1}^{N}\sigma^4(t/N) + \frac{1}{N}\sum_{j=0}^{J-1}\sum_{k=1}^{2^j}\mathbb{E}\left\{\hat{\mu}^{(s)}_{j,k} - \mu_{j,k}\right\}^2 \\ &= O\left(N^{-\min\left(1-\delta, \frac{2}{v}\right)}\right), \end{aligned} \tag{18}$$

where v *is defined as in formula (13).*

Remark 3.2 Note that, in the particular case of Z_t being Gaussian (so that $v = 2$), the mean-square error rate in (16) and (18) reduces to $O(N^{-1+\delta})$, which is arbitrarily close to the parametric rate $O(N^{-1})$.

3.2 Noise-free reconstruction

In this section, we consider the case when the errors $\{Z_t : \ t = 1, 2, \ldots\}$ in (1) are Gaussian random variables with mean zero and variance one. We construct an estimate of $\sigma^2(z)$ which possesses the following *noise-free reconstruction property*: if the true function $\sigma^2(z)$ is a constant function of z, then, with high probability, our estimate of $\sigma^2(z)$ is also constant and equal to the empirical mean of $\{X^2_{t,N}\}_{t=1}^N$.

The noise-free reconstruction property guarantees that estimates obtained using our method have a visually appealing character and do not exhibit spurious "spikes", even for a non-constant $\sigma^2(z)$. This is achieved by requiring that, asymptotically, no pure "noise" coefficients survive the thresholding procedure.

For the noise-free reconstruction property to hold, we require that the probability of any $f_{j,k}$ exceeding $\tilde{t}_j$ should tend to 0 as $N \to \infty$, more precisely:

$$P\left(\cup_{j=0}^{J-1}[\cup_{k=1}^{2^j}|f_{j,k}| > \tilde{t}_j]\right) \to 0, \quad \text{as} \quad N \to \infty \quad (J = \log_2 N), \tag{19}$$

where $f_{j,k}$ is the Haar-Fisz coefficient of $\{X_{t,N}\}$ and $\tilde{t}_j$ are appropriately chosen thresholds. We note that the thresholds used in this case are different to those given in Section 3.1.

In order to derive the appropriate thresholds $\tilde{t}_j$ we use the following lemma.

Lemma 3.1 *Let $\{X_i : \ i = 1, 2, \ldots, 2m\}$ be a sequence of iid χ^2_1 random variables, and let $X^{(1)} = \sum_{i=1}^m X_i$ and $X^{(2)} = \sum_{i=m+1}^{2m} X_i$. Then, the ratio $(X^{(1)} - X^{(2)})/(X^{(1)} + X^{(2)})$ is distributed as $2Y - 1$, where Y follows a Beta distribution with parameters $m/2$ and $m/2$ (i.e., $Y \sim \beta(m/2, m/2)$).*

We now derive the thresholds $\tilde{t}_j$. Note the use of the Bonferroni inequality (when $\sigma^2(z)$ is constant, the Haar-Fisz coefficients $f_{j,k}$ are (a) mutually uncorrelated, (b) independent and identically distributed within scales, (c) but not independent across scales). By using the independence of $\{f_{j,k}\}_k$ at each scale j we have the following bound for (19):

$$\begin{aligned} P\left(\cup_{j=0}^{J-1}[\cup_{k=1}^{2^j}|f_{j,k}| > \tilde{t}_j]\right) &\leq \sum_{j=0}^{J-1} P\left(\cup_{k=1}^{2^j}|f_{j,k}| > \tilde{t}_j\right) \\ &= \sum_{j=0}^{J-1}\left(1 - P\left(\cap_{k=1}^{2^j}|f_{j,k}| > \tilde{t}_j\right)\right. \\ &= J - \sum_{j=0}^{J-1}\prod_{k=1}^{2^j} P\left(|f_{j,k}| < \tilde{t}_j\right). \end{aligned} \tag{20}$$

As the distribution of $f_{j,k}$ does not depend on k, we can define $\alpha_j(N) = P(|f_{j,k}| < \tilde{t}_j)$, and (20) becomes:

$$P\left(\cup_{j=0}^{J-1}[\cup_{k=1}^{2^j}|f_{j,k}| > \tilde{t}_j]\right) \leq J - \sum_{j=0}^{J-1} \alpha_j^{2^j}(N). \tag{21}$$

Our objective is to choose $\{\alpha_j(N)\}_{j=0}^{J-1}$ such that $J - \sum_{j=0}^{J-1} \alpha_j^{2^j}(N) \to 0$, as $N \to \infty$. The choice of $\{\alpha_j(N)\}_{j=0}^{J-1}$, will determine how fast $J - \sum_{j=0}^{J-1} \alpha_j^{2^j}$ approaches zero and thus the rate of convergence. Probably the simplest option is to mimick standard universal thresholding in a classical *iid* Gaussian non-parametric regression setting for wavelets, where the analogue of $\alpha_j(N)$ is constant across scales (i.e. $\alpha_j(N) = \alpha(N)$) and the rate of convergence to 0 of the probability corresponding to (19) is equal to $(\pi J \log(2))^{-1/2}$. To guarantee such a rate we choose $\alpha(N)$ such that:

$$P\left(\cup_{j=0}^{J-1}[\cup_{k=1}^{2^j}|f_{j,k}| > \tilde{t}_j]\right) \leq J - \sum_{j=0}^{J-1} \alpha(N)^{2^j} = (\pi J \log(2))^{-1/2}.$$

That is $\alpha(N)$ satisfies the nonlinear equation:

$$J - (\pi J \log(2))^{-1/2} = \sum_{j=0}^{J-1} \alpha(N)^{2^j}. \tag{22}$$

Equation (22) cannot be solved analytically but note that the tangent to $p(\alpha) := \sum_{j=0}^{J-1} \alpha^{2^j}$ at $\alpha = 1$ has equation $q(\alpha) := J + (2^J - 1)(\alpha - 1)$. As $p(\alpha)$ is increasing and convex, for large enough α we have:

$$P\left(\cup_{j=0}^{J-1}[\cup_{k=1}^{2^j}|f_{j,k}| > \tilde{t}_j]\right) \leq J - \sum_{j=0}^{J-1} \alpha^{2^j} \leq J + (2^J - 1)(\alpha - 1).$$

In other words, by choosing $\alpha(N)^*$ such that

$$J - (\pi J \log(2))^{-1/2} = J + (2^J - 1)(\alpha(N)^* - 1), \tag{23}$$

we have

$$P\left(\cup_{j=0}^{J-1}[\cup_{k=1}^{2^j}|f_{j,k}| > \tilde{t}_j]\right) \leq J + (2^J - 1)(\alpha(N)^* - 1) = (\pi J \log(2))^{-1/2}.$$

The unique solution to (23) is given by

$$\alpha^*(N) = 1 - \frac{1}{(\pi J \log(2))^{1/2}(2^J - 1)}. \tag{24}$$

In the light of what was said above, $\alpha^*(N)$ guarantees the convergence of (19) to 0 at a rate of *at least* $(\pi J \log(2))^{-1/2}$.

By Lemma 3.1, $f_{j,k}$ has a $2\,\beta(2^{J-j-2}, 2^{J-j-2}) - 1$ distribution, and $\tilde{t}_j$'s are now easily found numerically by solving $\alpha^* = P(|f_{j,k}| < \tilde{t}_j)$. We note that noise free reconstruction is also possible for random variables Z_t which have distributions other than Gaussian, so long as the exact distribution of $(X^{(1)} - X^{(2)})/(X^{(1)} + X^{(2)})$ is known, where $X^{(1)} = \sum_{i=1}^{m} Z_i^2$ and $X^{(2)} = \sum_{i=m+1}^{2m} Z_i^2$.

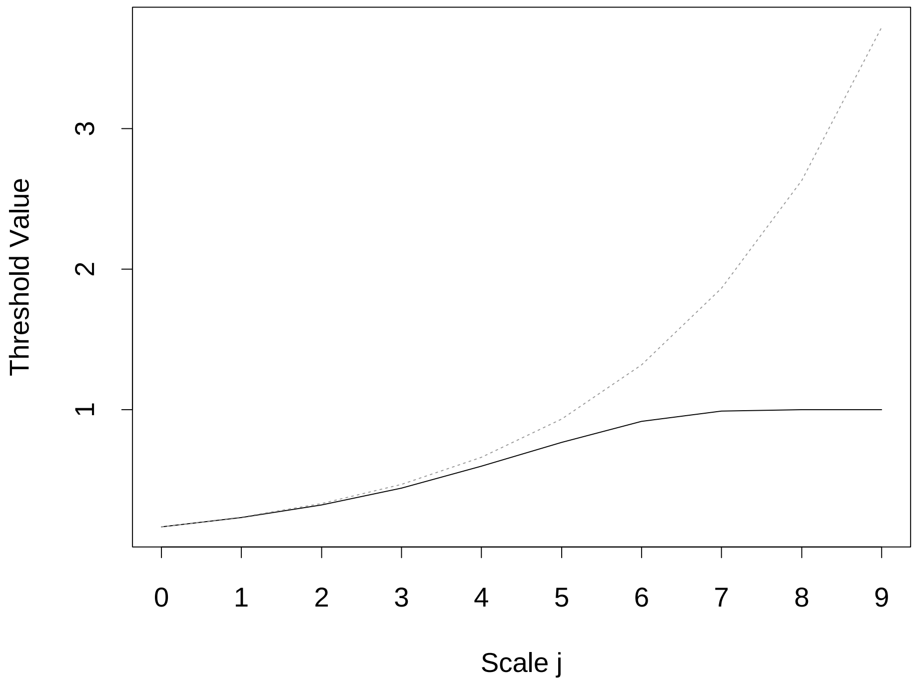

Figure 1: Thresholds $\tilde{t}_j$ (solid line) and t_j (dotted line) for $J = 10$ and $j = 0, \ldots, J-1$. See discussion in Section 3.2.

Figure 1 compares the thresholds t_j and $\tilde{t}_j$ for $J = 10$ and $j = 0, \ldots, J-1$. Note that t_j's exceed 1 at the 4 finest scales, and therefore no coefficients at these scales survive the thresholding (remember that $|f_{j,k}|$ is always bounded from above by 1).

In classical Gaussian wavelet regression, some authors argue that instead of modelling the analogue of $\alpha_j(N)$ as constant across scales, more accurate estimates are obtained by allowing it to decrease from finer to coarser scales, see e.g. Antoniadis & Fryzlewicz (2004). For simplicity, we consider a linear dependence of $\alpha_j(N)$ on j:

$$\alpha_j = \alpha_{J-1}\frac{j}{J-1} + \alpha_0\frac{J-1-j}{J-1}.$$

The equation for (α_0, α_{J-1}) is:

$$J - c_J = \sum_{j=0}^{J-1}\left(\alpha_{J-1}\frac{j}{J-1} + \alpha_0\frac{J-1-j}{J-1}\right)^{2^j}, \tag{25}$$

where $c_J \downarrow 0$ is the desired rate of convergence. By the weak convexity of $p(\alpha_0, \alpha_{J-1}) := \sum_{j=0}^{J-1}\left(\alpha_{J-1}\frac{j}{J-1} + \alpha_0\frac{J-1-j}{J-1}\right)^{2^j}$, any solution to (25) can be approximated from above by the solution to:

$$J - c_J = q(\alpha_0, \alpha_{J-1}),$$

where $q(\alpha_0, \alpha_{J-1})$ is a plane tangent to $p(\alpha_0, \alpha_{J-1})$ at $(\alpha_0, \alpha_{J-1}) = (1, 1)$, given by:

$$q(\alpha_0, \alpha_{J-1}) = J + \frac{(J-2)2^J + 2}{J-1}(\alpha_{J-1} - 1) + \frac{2^J - J - 1}{J-1}(\alpha_0 - 1).$$

One possibility is to set $\alpha_{J-1} = \alpha^*$ and then solve for α_0 to obtain the desired speed of convergence of c_J to 0.

4 Currency exchange rate examples

In this section, we exhibit the performance of various versions of our Haar-Fisz volatility estimator on two currency exchange datasets: the logged and differenced daily exchange rates between the US Dollar (USD) and the British Pound (GBP), as well as between the USD and the Japanese Yen (JPY), both running from 01/01/1990 to 31/12/1999. The data are available from the US Federal Reserve website

`http://www.federalreserve.gov/releases/h10/hist/default1999.htm`

We have also tested our estimator on other exchange rate datasets available from the above website but for lack of space we only provide graphical illustration of its performance on the USD/GBP and USD/JPY series in this section. However, the discussion below applies to *all* of the exchange rate time series available from the above website.

The length of both series is $n = 2515$, but as our estimators require the length of input to be a power of two, we only consider the last $N = 2048$ observations in both series (so that $J = \log_2(N) = 11$). Those are plotted in Figure 2.

We now single out a few specific versions of our Haar-Fisz volatility estimator:

MS-H: Our Haar-Fisz algorithm with hard thresholding and thresholds t_j (see formula (15)) which guarantee mean-square consistency. We take $J^* = J$ for simplicity. See Section 3.1 for details.

NF-100-H: Our Haar-Fisz algorithm with hard thresholding and noise-free reconstruction thresholds $\tilde{t}_j$ chosen in such a way that $\alpha_{J-1} = \alpha^*$ (see formula (24)) and $\alpha_0 = \frac{100}{100}\alpha_{J-1} = \alpha_{J-1}$. See Section 3.2 for details.

NF-98-S: Our Haar-Fisz algorithm with soft thresholding and noise-free reconstruction thresholds $\tilde{t}_j$ chosen in such a way that $\alpha_{J-1} = \alpha^*$ (see formula (24)) and $\alpha_0 = \frac{98}{100}\alpha_{J-1} = 0.98\alpha_{J-1}$. See Section 3.2 for details.

NF-98-H-TI: The translation-invariant (TI) version of our Haar-Fisz algorithm with soft thresholding and noise-free reconstruction thresholds $\tilde{t}_j$ chosen in such a way that $\alpha_{J-1} = \alpha^*$ (see formula (24)) and $\alpha_0 = \frac{98}{100}\alpha_{J-1} = 0.98\alpha_{J-1}$. See Section 3.2 for details. In TI versions of wavelet-based denoising algorithms, the final estimator is

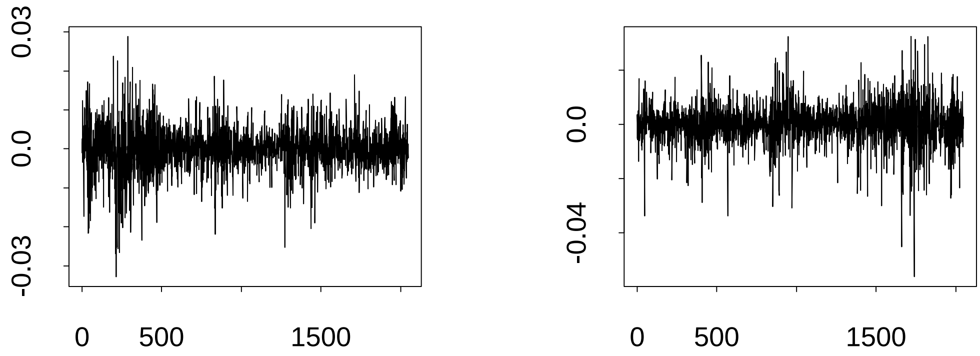

Figure 2: The last $N = 2048$ observations of the USD/GBP (left) and USD/JPY (right) exchange rate series, running from 8 November 1991 to 31 December 1999.

obtained as the average of the estimators obtained for all circular shifts of the data. This is common practice in wavelet regression (see, e.g., Coifman & Donoho, 1995). The fast $O(N \log(N))$ implementation of TI wavelet thresholding algorithms uses the Non-Decimated Wavelet Transform (NDWT) (see, e.g., Nason & Silverman, 1995).

We have tested several versions of our estimator by looking at the behaviour of empirical residuals for each of the datasets. More specifically, let $X_{t,N} = \sigma(t/N)Z_t$ denote a series of currency exchange log-returns, and let $\hat{\sigma}^2(t/N)$ be any Haar-Fisz estimator of $\sigma^2(t/N)$. We define the empirical residuals as $\hat{Z}_t = X_{t,N}/\hat{\sigma}(t/N)$. We are satisfied with the performance of $\hat{\sigma}^2(t/N)$ if the sequence $\hat{Z}_t$ (a) looks "stationary", and (b) displays only very little dependence in the squares, i.e. the sample autocorrelation function of $\hat{Z}_t^2$ is "small" for all non-zero lags.

In an extensive empirical study which compared several parameter configurations for several currency exchange datasets, we found that the NF-98-S estimator performed well for all the datasets considered. On the other hand, its hard thresholding counterpart, NF-98-H, was often extremely "spiky", and only by applying its translation-invariant version (NF-98-H-TI) did we manage to obtain satisfactory behaviour of the empirical residuals. However, the price for using a translation-invariant estimator was that, naturally enough, we lost the piecewise constant nature of the reconstructions and increased the computational effort from $O(N)$ to $O(N \log(N))$. Further, we found that the NF-100-H estimator typically gave slightly oversmoothed reconstructions, as did the MS-H version of our Haar-Fisz algorithm. As mentioned in Section 3.2, the reason for the less-than-perfect performance of the latter was that the thresholds t_j were larger than 1 at the 4 finest scales which meant that no detail coefficients $d_{j,k}$ at those scales survived the thresholding.

Figure 3 shows the results for the USD/GBP series. From top to bottom, the rows correspond to MS-H, NF-100-H, NF-98-S, NF-98-H-TI, respectively. In each row, the left plot shows the volatility estimate obtained using the given version of our estimator, and the right plot shows the autocorrelation plot of the squares of the arising empirical residuals. Figure 4 shows the analogous results for the USD/JPY series. The results confirm our observations of the previous paragraph.

In summary, NF-98-S is the recommended estimator, unless obtaining a piecewise constant estimate is not essential to the analyst despite the constraints of the model, in which case NF-98-H-TI should be considered as an alternative.

5 Forecasting currency exchange rate volatility

In this section, we describe the outcome of an empirical study designed to assess the forecasting ability of our model and compare it to that of the benchmark stationary GARCH(1,1) process with Gaussian innovations, as well as a simple "moving window" procedure (the latter in the case of long-term forecasting only). Suppose that we observe $X_{1,N}^2, X_{2,N}^2, \ldots, X_{t,N}^2$ from model (1) and want to predict $X_{t+1,N}^2, \ldots, X_{t+h,N}^2$, where $t + h \leq N$. The Mean-Square-optimal forecasts are given by:

$$\sigma_{t|t+h,N}^{2,\mathrm{HF}} := \mathbb{E}\{X_{t+h,N}^2 | X_{1,N}^2, \ldots, X_{t,N}^2\} = \sigma^2\left(\frac{t+h}{N}\right).$$

Obviously, the true value of $\sigma^2(\frac{t+h}{N})$ is unknown at time t and "the best" that we can do is to extrapolate it as:

$$\sigma_{t|t+h,N}^{2,\mathrm{HF}} = \hat{\sigma}^2(t/N),$$

where $\hat{\sigma}^2(t/N)$ is any of our Haar-Fisz estimates of $\sigma^2(t/N)$. Note that contrary to the GARCH case (see e.g. Bera & Higgins, 1993) our forecasts do not depend on the forecasting horizon. In the examples below, we use the following versions of our estimator to compute $\hat{\sigma}^2(t/N)$ for the above forecasts:

NF-98-S: The estimator described in Section 4, computed on $X_{t-1023,N}^2, \ldots, X_{t,N}^2$.

NF-100-S: Like NF-98-S, but with $\alpha_0 = \alpha_{J-1}$, where $J = \log_2(1024) = 10$.

For GARCH-based forecasts, we forecast the volatility at time $t+1, \ldots, t+h$ from $X_1^2, \ldots, X_t^2$ using the following methods:

G-SCROLL: We fit the GARCH(1,1) model with Gaussian innovations to $X_{t-1023}, \ldots, X_t$ using the S-Plus routine `garch`, and then forecast the volatility using the S-Plus routine `predict`.

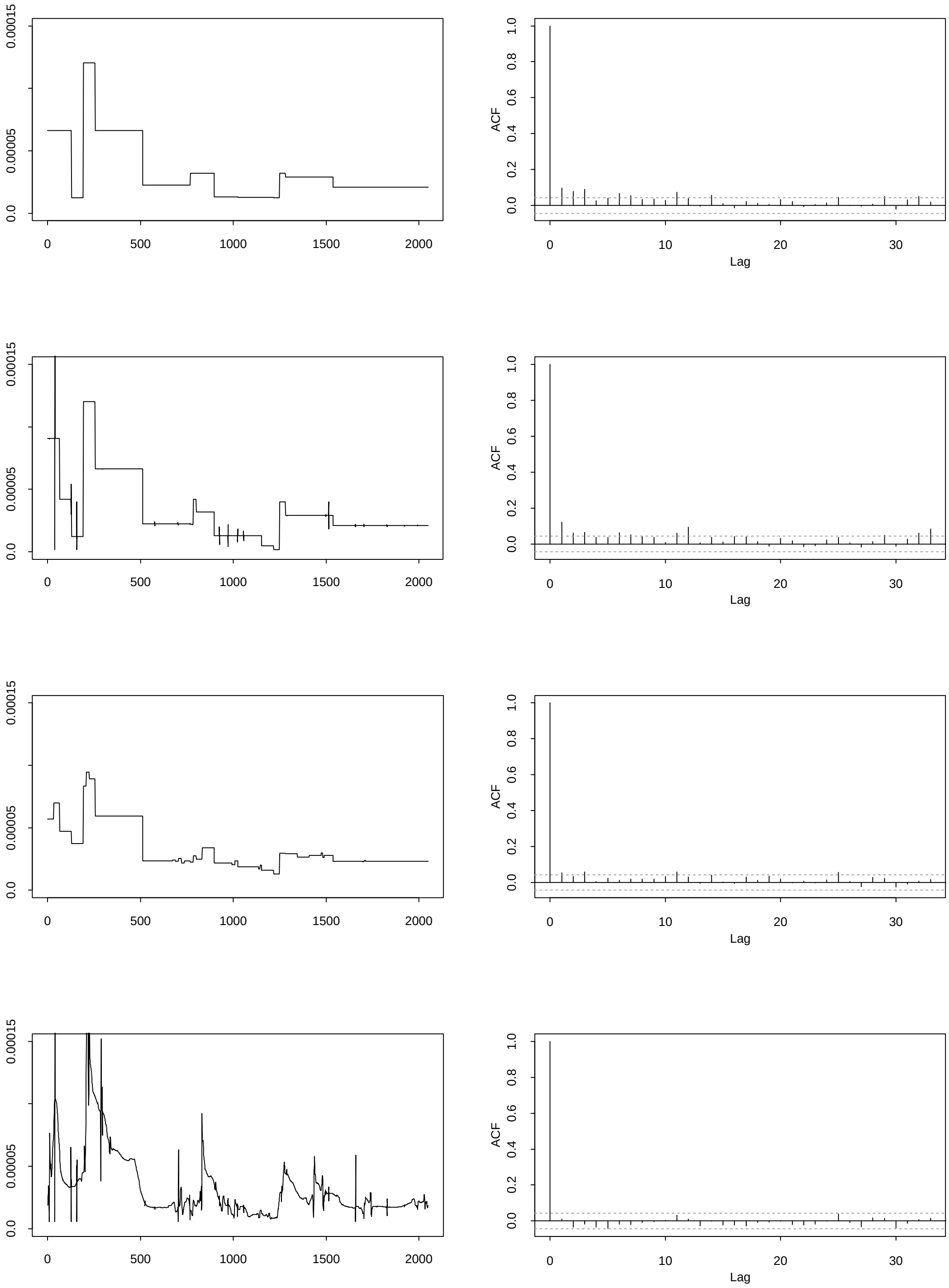

Figure 3: Empirical results for the USD/GBP series shown in Figure 2. Left column, from top to bottom: reconstructions obtained with the MS-H, NF-100-H, NF-98-S, NF-98-H-TI estimators, respectively. Right column: the acf plots of the squares of the corresponding empirical residuals. See the discussion in Section 4.

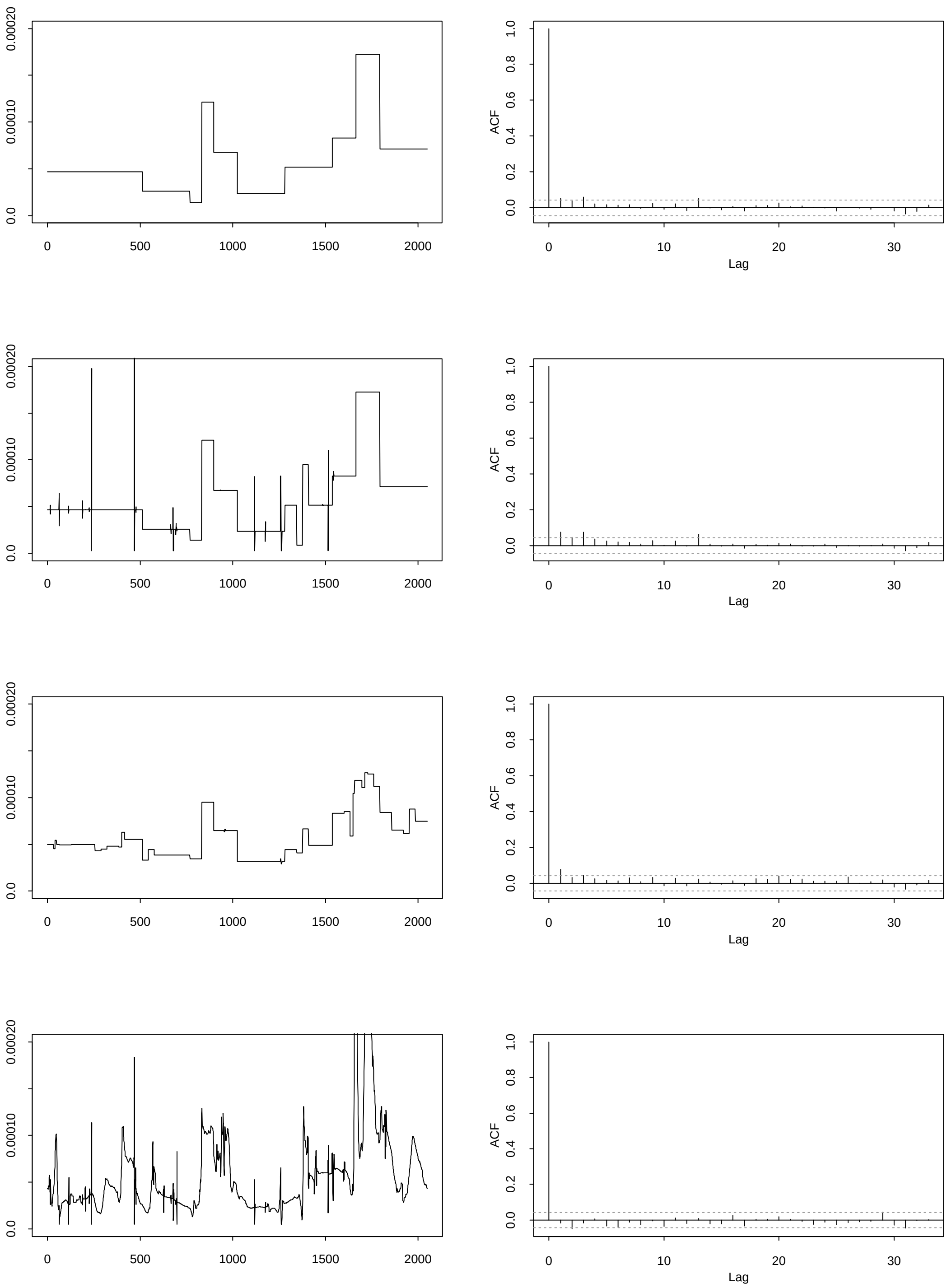

Figure 4: Empirical results for the USD/JPY series shown in Figure 2. Left column, from top to bottom: reconstructions obtained with the MS-H, NF-100-H, NF-98-S, NF-98-H-TI estimators, respectively. Right column: the acf plots of the squares of the corresponding empirical residuals. See the discussion in Section 4.

G-NSCROLL: Like G-SCROLL, but the model is fitted to $X_1, \ldots, X_t$.

For the simple moving window procedure, labelled **MW**, the predicted volatility at times $t+1, \ldots, t+h$ is the empirical mean of the vector $(X_{t-h+1}^2, \ldots, X_t^2)$. Here, we use the "rule of thumb" advocated by Hull (1997, p. 233) whereby the time period over which volatility is estimated should be set equal to the time period h over which is it to be applied.

Let $\sigma_{t|t+h}^2$ denote a generic volatility forecast, computed using any of the above procedures. For each of the currency exchange datasets tested (details are given below), we compute the following error measure: for each $t = 1024, \ldots, N - 250$ (where N is the length of each dataset; this oscillates around 2500 but varies from one dataset to another), we compute the quantity

$$\bar{\sigma}_{t|t+250}^2 = \sum_{h=1}^{250} \sigma_{t|t+h}^2,$$

and compare it to the "realised" volatility

$$\bar{X}_{t|t+250}^2 = \sum_{h=1}^{250} X_{t+h}^2,$$

using the Average Squared Error

$$\mathrm{ASE}_{250,1024,N} = \frac{1}{N - 1273} \sum_{t=1024}^{N-250} \left(\bar{\sigma}_{t|t+250}^2 - \bar{X}_{t|t+250}^2 \right)^2.$$

For a justification of this error measure, see Stărică (2003). In other words, we forecast the volatility one business year (= 250 days) ahead: this is done to compare the long-term forecasting ability of the competitors.

Our datasets are logged and differenced currency exchange rates between the USD and a variety of other currencies, available from the web address given in Section 4. The other currencies are: AUD (Australia Dollar), CAD (Canada Dollar), CHF (Switzerland Franc), DKK (Denmark Kroner), GBP (United Kingdom Pound), HKD (Hong Kong Dollar), JPY (Japan Yen), KRW (South Korea Won), NOK (Norway Kroner), NZD (New Zealand Dollar), SEK (Sweden Kronor), SGD (Singapore Dollar), THB (Thailand Baht), TWD (Taiwan New Dollar) and ZAR (South Africa Rand). Table 1 lists the (scaled by the number in the rightmost column, and rounded) ASE's attained by the competing methods for each currency. The best results, and those within 10% of the best ones, are boxed.

The stationary GARCH(1,1) model failed to fit at several points of the USD/HKD, USD/KRW, USD/THB, USD/TWD and USD/ZAR exchange rate series (this is marked by the "bullets" in the table), producing forecasts which were extremely inaccurate. The inaccuracy was due to the fact that the estimated parameters fell outside the stationarity region. Our NF-100-S method and the simple MW technique performed the best, or nearly

Currency	G-NSCROLL	G-SCROLL	MW	NF-98-S	NF-100-S	scaling
AUD	3122	3673	$\boxed{2944}$	$\boxed{2776}$	3095	10^8
CAD	$\boxed{1610}$	2280	2047	1955	$\boxed{1770}$	10^9
CHF	$\boxed{2844}$	$\boxed{2819}$	5168	4574	3541	10^8
DKK	2183	$\boxed{1658}$	2530	$\boxed{1697}$	$\boxed{1746}$	10^8
GBP	23139	15192	7796	$\boxed{6957}$	8645	10^9
HKD	•	•	$\boxed{3038}$	4740	3841	10^{13}
JPY	$\boxed{1060}$	$\boxed{1006}$	1213	1130	$\boxed{1056}$	10^8
KRW	2826	•	$\boxed{1333}$	1938	1657	10^5
NOK	3634	2812	2173	2885	$\boxed{1846}$	10^8
NZD	14502	10414	$\boxed{9293}$	$\boxed{8847}$	10849	10^8
SEK	4451	3812	$\boxed{741}$	1309	1706	10^8
SGD	$\boxed{4709}$	142174	6414	5686	$\boxed{5122}$	10^8
THB	•	•	$\boxed{3806}$	4127	$\boxed{3580}$	10^6
TWD	•	•	4419	4773	$\boxed{3953}$	10^8
ZAR	•	•	$\boxed{2017}$	3321	2600	10^7

Table 1: ASE for long-term forecasts using the methods described in Section 5.

the best, for 7 out of 15 datasets, and are clearly the two preferred options here. Either of our two methods performed the best, or nearly the best, for 10 out of 15 datasets.

In practice, our recommendation is to use our forecasting technique based on our Haar-Fisz estimation method with soft thresholding and noise-free reconstruction thresholds where $\alpha_{J-1} = \alpha^*$, and α_0 is chosen from a pre-set grid $\{\alpha_{0,l}\}_{l=1}^{L}$ by comparing the performance of the method on the observed part of the series and choosing the value of $\alpha_{0,l}$ which performs the best. We have found that $\{\alpha_{0,l}\}_{l=1}^{5} = \{\frac{95+l}{100}\alpha_{J-1}\}_{l=1}^{5}$ is a good practical choice for the grid.

In our empirical study, we have also found that forecasts based on our Haar-Fisz estimation technique with hard thresholding tend to be less accurate and therefore their use is not recommended.

Short-term volatility forecasting. It is well known that the GARCH framework (even in the simplest case of the Gaussian GARCH(1,1) model) provides excellent short-term volatility forecasts. In a brief simulation study, we have compared the performance of our NF-98-S and NF-100-S algorithms to that of the G-NSCROLL technique in forecasting one-day-ahead volatility of the above datasets (except USD/HKD, USD/KRW, USD/THB, USD/TWD and USD/ZAR, for which GARCH(1,1) does not give a good fit as explained above). We have found that the ratio of the ASE for the worst of our two algorithms and the ASE of G-NSCROLL ranged between 0.99 and 1.09 for all of the datasets, which demonstrates good performance of our technique also in the case of short-term forecasts.

6 Concluding remarks

In this paper, we have proposed a new locally stationary model for log-returns, whereby the log-returns are modelled as independent random variables (not necessarily Gaussian) whose local variance (= volatility) is modelled as a piecewise constant function with an unknown number and location of jumps. The time-dependent volatility is defined on a compact interval to enable a meaningful estimation theory. Our approach of modelling the log-return process as linear and locally stationary is in stark contrast to the classical ARCH/GARCH or Stochastic Volatility frameworks where the log-return process is modelled as non-linear but stationary.

We have demonstrated that even in the simplest case of the innovations being Gaussian, our model explains well the commonly observed stylised facts of log-returns: namely, the sample mean of the series being close to zero, the marginal distribution being leptokurtic, the returns being serially uncorrelated, and their squares displaying the well known "long memory" property. Our model can also easily accommodate skewed marginal distributions of log-returns.

We have also proposed a new nonlinear algorithm for estimating the time-dependent volatility in our model. Our algorithm powerfully combines Haar wavelet thresholding and the variance-stabilizing Fisz transform and yields an asymptotically unbiased, mean-square consistent estimate of the volatility. Its mean-square convergence rate is arbitrarily close to the parametric rate. Our algorithm does not require any pre-estimates or a bias correction factor. Moreover, it is conceptually simple and easy to code, and its computational complexity is linear in the number of data points. To allow more flexibility, we have also provided an alternative version of our estimation algorithm which enjoys the "noise-free reconstruction property" in the special case of the model innovations being Gaussian. Both versions of our algorithm allow either "soft" or "hard" thresholding rules.

In an extensive empirical study, we have fitted our model to a variety of currency exchange datasets, and identified versions of our algorithm which performed particularly well in producing "well-behaved" empirical residuals. We have shown that our algorithm significantly outperforms GARCH in long-term volatility forecasting, and performs slightly better than a model-free moving window technique. We have also demonstrated that the short-term volatility forecasting ability of our algorithm matches that of the GARCH(1,1) technique. In summary, by using our methodology, we have been able to obtain *both* good fits to the data *and* convincing short- and long-term volatility forecasts.

A Explanation of the stylised facts

Proof of Proposition 2.1. We first show (7). Denote $K_\sigma = \max_z \sigma^2(z)$. We have

$$\mathbb{E}\left\{\frac{1}{N}\sum_{t=1}^{N-h} X_{t,N}^2 X_{t+h,N}^2 - \int_0^1 \sigma^4(z)dz\right\}^2 \leq$$
$$2\,\mathbb{E}\left\{\frac{1}{N}\sum_{t=1}^{N-h}\sigma^2\left(\frac{t}{N}\right)\sigma^2\left(\frac{t+h}{N}\right)\left(Z_t^2 Z_{t+h}^2 - 1\right)\right\}^2 +$$
$$2\left\{\frac{1}{N}\sum_{t=1}^{N-h}\sigma^2\left(\frac{t}{N}\right)\sigma^2\left(\frac{t+h}{N}\right) - \int_0^1 \sigma^4(z)dz\right\} =: I + II.$$

Elementary approximation theory implies that $II = O(h^2/N^2)$. Denoting $S = \{t-h, t, t+h\} \cap \{1, \ldots, N-h\}$ and using the independence of the sequence $\{Z_t : \ t = 1, 2, \ldots\}$ and that $\mathbb{E}(Z_t^2) = 1$, we have

$$\begin{aligned} I &= \frac{2}{N^2}\sum_{t=1}^{N-h}\sum_{s\in S}\sigma^2\left(\frac{t}{N}\right)\sigma^2\left(\frac{t+h}{N}\right)\sigma^2\left(\frac{s}{N}\right)\sigma^2\left(\frac{s+h}{N}\right)\mathbb{E}\left\{(Z_t^2 Z_{t+h}^2 - 1)(Z_s^2 Z_{s+h}^2 - 1)\right\} \\ &\leq \frac{CK_\sigma^4}{N}, \end{aligned}$$

where C is a finite constant. We now turn to $\left(\bar{X}_N^2\right)^2$. It is easily seen that

$$\begin{aligned} &\mathbb{E}\left\{\left(\bar{X}_N^2\right)^2 - \left(\int_0^1 \sigma^2(z)dz\right)^2\right\}^2 \leq \\ &\quad 2\,\mathbb{E}\left\{\left(\bar{X}_N^2\right)^2 - \left(\frac{1}{N}\sum_{t=1}^N \sigma^2(t/N)\right)^2\right\}^2 + 2\left\{\left(\frac{1}{N}\sum_{t=1}^N \sigma^2(t/N)\right)^2 - \left(\int_0^1 \sigma^2(z)dz\right)^2\right\} \\ &\quad =: III + IIII. \end{aligned}$$

Again, elementary approximation theory implies that $IIII = O(N^{-2})$. Also, by Cauchy inequality, we have

$$\begin{aligned} III &= 2\mathbb{E}\left\{\left(\bar{X}_N^2 - \frac{1}{N}\sum_{t=1}^N \sigma^2(t/N)\right)^2\left(\bar{X}_N^2 + \frac{1}{N}\sum_{t=1}^N \sigma^2(t/N)\right)^2\right\} \\ &\leq 2\left\{\mathbb{E}\left(\frac{1}{N}\sum_{t=1}^N \sigma^2(t/N)\left(Z_t^2 - 1\right)\right)^4\right\}^{1/2}\left\{\mathbb{E}\left(\frac{1}{N}\sum_{t=1}^N \sigma^2(t/N)\left(Z_t^2 + 1\right)\right)^4\right\}^{1/2} \\ &=: 2\, III_1^{1/2} \cdot III_2^{1/2}. \end{aligned}$$

Note that $III_2 = O(1)$. We now compute

$$
\begin{aligned}
III_1 &= \\
&\frac{1}{N^4}\mathbb{E}\left\{\left(\sum_{t=1}^{N}\sigma^4(t/N)(Z_t^2-1)^2\right)^2 + 4\sum_{t=1}^{N}\sum_{s\neq t}\sigma^4(t/N)\sigma^4(s/N)(Z_t^2-1)^2(Z_s^2-1)^2\right\} \\
&= O(N^{-2}).
\end{aligned}
$$

Finally, (7) follows by noting that

$$\int_0^1 \sigma^4(z)dz - \left(\int_0^1 \sigma^2(z)dz\right)^2 = \int_0^1 \left(\sigma^2(z) - \int_0^1 \sigma^2(\omega)d\omega\right)^2 dz.$$

We now show (4), (5) and (6). Since the proofs of (4) and (6) use exactly the same technique, we omit the details. For (5), note that, using the same technique as above, it is easy to show that

$$\bar{X}_N^4 \xrightarrow{L_2} \mathbb{E}(Z_t^4)\int_0^1 \sigma^4(z)dz \quad \text{and} \quad \left(\bar{X}_N^2\right)^2 \xrightarrow{L_2} \left(\int_0^1 \sigma^2(z)dz\right)^2, \quad \text{as} \quad N \to \infty,$$

which also imply convergence in probability. Thus, Slutsky's theorem yields (5). This completes the proof of Proposition 2.1. □

B Examples of distributions satisfying Assumption 3.1

Lemma B.1 *For any random variable X and any nonnegative integer r, we have $\mathbb{E}|X^2 - 1|^r \leq \mathbb{E}X^{2r} + 1$.*

Proof. For any real x, we have

$$|x^2 - 1|^r \leq \max\{x^2, 1\}^r = \max\{x^{2r}, 1\} \leq x^{2r} + 1.$$

Using the above for X and applying the expectation operator on both sides yields the result. □

Lemma B.2 *For an integer $r \geq 3$, we have*

$$(2r)! \leq \frac{10}{3}(r!)^3.$$

Proof. Denote $a_r = (2r)!/(r!)^3$. We have

$$\frac{a_{r+1}}{a_r} = \frac{(2r+2)(2r+1)}{(r+1)^3} < 1 \quad \text{for} \quad r \geq 3,$$

and thus a_r is a decreasing sequence for $r \geq 3$. Noting that $a_3 = 10/3$, we get the result.□

Lemma B.3 *Let Z_t be a standard normal random variable. Then, Assumption 3.1 is satisfied with $\gamma = 0$.*

Proof. Applying Lemma B.1 and noting the standard result for normal variables

$$\mathbb{E}Z_t^{2n} = \frac{\Gamma(n+1/2)2^n}{\sqrt{\pi}},$$

we obtain

$$\mathbb{E}|Z_t^2 - 1|^n \leq \mathbb{E}Z_t^{2n} + 1 = \frac{\Gamma(n+1/2)2^n}{\sqrt{\pi}} + 1.$$

Recall that $v = \mathbb{E}|Z_t^2 - 1|^2 = 2$. Setting $\gamma = 0$, Assumption 3.1 will be satisfied if there exists $c > 0$ such that

$$\frac{\Gamma(n+1/2)2^n}{\sqrt{\pi}} + 1 \leq 2c^{n-2}n!, \quad \text{for all} \quad n \geq 3.$$

But this is true as by elementary algebra, the sequence

$$\left(\frac{\frac{\Gamma(n+1/2)2^n}{\sqrt{\pi}} + 1}{2n!} \right)^{\frac{1}{n-2}}$$

is bounded from above. □

Lemma B.4 *Let Z_t be a Laplace (double exponential) variable, standardised in such a way that $\mathbb{E}Z_t^2 = 1$. Then, Assumption 3.1 is satisfied with $\gamma = 2$.*

Proof. Z_t has density

$$f(z) = \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2}|z|\right),$$

and, by standard results for the Laplace distribution (see, e.g., Johnson & Kotz, 1970, Chapter 23), we have $\mathbb{E}Z_t^{2n} = (2n)!2^{-n}$. Applying Lemma B.1 and then Lemma B.2, we obtain

$$\mathbb{E}|Z_t^2 - 1|^n \leq \mathbb{E}Z_t^{2n} + 1 = (2n)!2^{-n} + 1 \leq \frac{10}{3}(n!)^3 2^{-n} + 1.$$

Recall that $v = \mathbb{E}|Z_t^2 - 1|^2 = 5$. Setting $\gamma = 2$, Assumption 3.1 will be satisfied if there exists $c > 0$ such that

$$\frac{10}{3}(n!)^3 2^{-n} + 1 \leq c^{n-2} 5(n!)^3, \quad \text{for all} \quad n \geq 3.$$

This is true as by simple algebra, the sequence

$$\left(\frac{\frac{10}{3}(n!)^3 2^{-n} + 1}{5(n!)^3} \right)^{\frac{1}{n-2}}$$

is bounded from above. □

C Properties of the Haar-Fisz estimator

Proof of Theorem 3.1. The first of the two equalities in (16) is due to the orthonormality of the discrete Haar transform. We now show the second equality. For notational clarity, denote

$$\begin{aligned} d_{1,j,k} &= s_{j+1,2k-1}/\sqrt{2}, \\ d_{2,j,k} &= s_{j+1,2k}/\sqrt{2}, \end{aligned}$$

so that $d_{j,k} = d_{1,j,k} - d_{2,j,k}$ and $s_{j,k} = d_{1,j,k} + d_{2,j,k}$. Denote further $\mu_{i,j,k} = \mathbb{E}\, d_{i,j,k}$ and $w_{i,j,k}^2 = \operatorname{Var} d_{i,j,k}$ for $i = 1, 2$. Finally, denote $w_{j,k}^2 = \operatorname{Var} d_{j,k}$.
For the reader's convenience, we now give explicit formulae for $d_{1,j,k}$ and $d_{2,j,k}$:

$$\begin{aligned} d_{1,j,k} &= 2^{\frac{j-J}{2}} \sum_{i=2^{J-j}(k-1)+1}^{2^{J-j}(k-1/2)} X_i^2, \\ d_{2,j,k} &= 2^{\frac{j-J}{2}} \sum_{i=2^{J-j}(k-1/2)+1}^{2^{J-j}k} X_i^2. \end{aligned}$$

We now compute the risk of $\hat{\mu}_{j,k}^{(h)}$ for $(j,k) \in \mathcal{I}_N$.

1. Case $\sigma^2(i/N) :=$ constant $:= \sigma^2$, for $i = 2^{J-j}(k-1)+1, \ldots, 2^{J-j}k$ (so that $\mu_{1,j,k} =$

$\mu_{2,j,k}$). Without loss of generality, consider $k = 1$ to shorten the notation.

$$\mathbb{E}\left\{(d_{1,j,1} - d_{2,j,1})\mathbb{I}\left(\frac{|d_{1,j,1} - d_{2,j,1}|}{d_{1,j,1} + d_{2,j,1}} > t_j\right) - (\mu_{1,j,1} - \mu_{2,j,1})\right\}^2 =$$
$$\mathbb{E}\left\{(d_{1,j,1} - d_{2,j,1})\mathbb{I}\left(\frac{|d_{1,j,1} - d_{2,j,1}|}{d_{1,j,1} + d_{2,j,1}} > t_j\right)\right\}^2 =$$
$$2^{j-J}\sigma^4\,\mathbb{E}\left\{\left(\sum_{i=1}^{2^{J-j-1}} Z_i^2 - Z_{i+2^{J-j-1}}^2\right)^2 \mathbb{I}\left(\frac{|d_{1,j,1} - d_{2,j,1}|}{d_{1,j,1} + d_{2,j,1}} > t_j\right)\right\} =$$
$$2^{j-J}\sigma^4\,\mathbb{E}\left\{\left(\sum_{i=1}^{2^{J-j-1}} (Z_i^2 - Z_{i+2^{J-j-1}}^2)^2 + \sum_{\substack{i,l=1\\ i\neq l}}^{2^{J-j-1}} (Z_i^2 - Z_{i+2^{J-j-1}}^2)(Z_l^2 - Z_{l+2^{J-j-1}}^2)\right) \times \right.$$
$$\left. \times \quad \mathbb{I}\left(\frac{|d_{1,j,1} - d_{2,j,1}|}{d_{1,j,1} + d_{2,j,1}} > t_j\right)\right\}. \tag{26}$$

Note that by symmetry arguments, for any $i \neq l$, we have

$$\mathbb{E}\left\{(Z_i^2 - Z_{i+2^{J-j-1}}^2)(Z_l^2 - Z_{l+2^{J-j-1}}^2)\mathbb{I}\left(\frac{|d_{1,j,1} - d_{2,j,1}|}{d_{1,j,1} + d_{2,j,1}} > t_j\right)\right\} = 0,$$

which simplifies (26) to

$$2^{j-J}\sigma^4 \sum_{i=1}^{2^{J-j-1}} \mathbb{E}\left\{(Z_i^2 - Z_{i+2^{J-j-1}}^2)^2\mathbb{I}\left(\frac{|d_{1,j,1} - d_{2,j,1}|}{d_{1,j,1} + d_{2,j,1}} > t_j\right)\right\} =$$
$$\frac{\sigma^4}{2}\,\mathbb{E}\left\{(Z_1^2 - Z_{1+2^{J-j-1}}^2)^2\mathbb{I}\left(\frac{|d_{1,j,1} - d_{2,j,1}|}{d_{1,j,1} + d_{2,j,1}} > t_j\right)\right\} \leq$$
$$\frac{\sigma^4}{2}\left(\mathbb{E}\left\{(Z_1^2 - Z_{1+2^{J-j-1}}^2)^{2r}\right\}\right)^{1/r}\mathbb{P}\left(\frac{|d_{1,j,1} - d_{2,j,1}|}{d_{1,j,1} + d_{2,j,1}} > t_j\right)^{1-1/r} :=$$
$$\frac{\sigma^4}{2}C_r\mathbb{P}\left(\frac{|d_{1,j,1} - d_{2,j,1}|}{d_{1,j,1} + d_{2,j,1}} > t_j\right)^{1-1/r}, \tag{27}$$

where $C_r = (\mathbb{E}\{(Z_1^2 - Z_{1+2^{J-j-1}}^2)^{2r}\})^{1/r}$, and the last but one step above uses Hölder's inequality with $r > 1$ but otherwise arbitrary. Simple algebra gives

$$\mathbb{P}\left(\frac{|d_{1,j,1} - d_{2,j,1}|}{d_{1,j,1} + d_{2,j,1}} > t_j\right) = 2\,\mathbb{P}\left(\frac{d_{1,j,1} - d_{2,j,1}}{d_{1,j,1} + d_{2,j,1}} > t_j\right) =$$
$$2\,\mathbb{P}\left(\frac{2^{\frac{j-J}{2}}}{\sqrt{v(1+t_j^2)}}\left\{\sum_{i=1}^{2^{J-j-1}} (Z_i^2 - 1)(1 - t_j) - \sum_{i=2^{J-j-1}+1}^{2^{J-j}} (Z_i^2 - 1)(1 + t_j)\right\} > \frac{t_j 2^{\frac{J-j}{2}}}{\sqrt{v(1+t_j^2)}}\right). \tag{28}$$

Since the condition of Theorem 1 from Rudzkis *et al.* (1978) holds due to our Assumption 3.1, we are able to apply the Corollary to Theorem 1 from Rudzkis *et al.* (1978). Recalling that $t_j = 2^{\frac{-J+j+1}{2}}\sqrt{2\log(N)}$, that $t_j \to 0$ on $\mathcal{I}_N$, and that $2^j < 2^{J^*} = O(N^{1-\delta})$, it is easy to see that

$$\frac{t_j 2^{\frac{J-j}{2}}}{\sqrt{v(1+t_j^2)}} = o\left(\left\{\frac{2^{\frac{J-j}{2}}\frac{\sqrt{v(1+t_j^2)}}{1+t_j}}{2\max\{c,\sqrt{v}\}}\right\}^{\nu}\right), \quad \text{as} \quad N \to \infty,$$

for any positive ν. Therefore, by the Corollary to Theorem 1 from Rudzkis *et al.* (1978), we bound (28) from above by

$$2C\,\mathbb{P}\left(N(0,1) > \frac{t_j 2^{\frac{J-j}{2}}}{\sqrt{v(1+t_j^2)}}\right). \tag{29}$$

Recalling again that $t_j = 2^{\frac{-J+j+1}{2}}\sqrt{2\log(N)}$ and denoting by $\Phi(\cdot)$ the cumulative distribution function (cdf) of a $N(0,1)$ random variable, we now bound (29) from above using Mill's ratio inequality (see, e.g., Shorack & Wellner, 1986, p. 850):

$$2C\left(1-\Phi\left(\frac{2\sqrt{\log(N)}}{\sqrt{v(1+t_j^2)}}\right)\right) \le C\,\exp\left\{-\frac{4\log(N)}{2v(1+t_j^2)}\right\} = C\,N^{-\frac{2}{v(1+t_j^2)}}. \tag{30}$$

Plugging (30) into (27), we obtain

$$\frac{\sigma^4}{2}C_r\,\mathbb{P}\left(\frac{|d_{1,j,1}-d_{2,j,1}|}{d_{1,j,1}+d_{2,j,1}} > t_j\right)^{1-1/r} \le \sigma^4\tilde{C}_r N^{-\frac{2(1-1/r)}{v(1+t_j^2)}}, \tag{31}$$

for some appropriate positive $\tilde{C}_r$. Noting that $t_j \le t_{J^*} = O(\sqrt{\log(N)}N^{-\delta/2})$ uniformly on $\mathcal{I}_N$, it is easy to show by direct comparison that

$$N^{-\frac{2(1-1/r)}{v(1+t_j^2)}} = O\left(N^{-\frac{2}{v}(1-1/r)}\right), \quad \text{as} \quad N \to \infty.$$

Upon choosing $r = 2(v\delta)^{-1}$, we obtain the final bound for (31) as $\tilde{C}_\delta \sup_z \sigma^4(z) N^{-\frac{2}{v}+\delta}$.

2. Case $\sigma^2(i/N) \ne$ constant, for $i = 2^{J-j}(k-1)+1,\dots,2^{J-j}k$ (so that possibly $\mu_{1,j,k} \ne \mu_{2,j,k}$).

$$\begin{aligned}
&\mathbb{E}\left\{(d_{1,j,k}-d_{2,j,k})\mathbb{I}\left(\frac{|d_{1,j,k}-d_{2,j,k}|}{d_{1,j,k}+d_{2,j,k}} > t_j\right) - (\mu_{1,j,k}-\mu_{2,j,k})\right\}^2 \le \\
&2\,\mathbb{E}\left\{(d_{1,j,k}-d_{2,j,k}-(\mu_{1,j,k}-\mu_{2,j,k}))\mathbb{I}\left(\frac{|d_{1,j,k}-d_{2,j,k}|}{d_{1,j,k}+d_{2,j,k}} > t_j\right)\right\}^2 + \\
&2\,(\mu_{1,j,k}-\mu_{2,j,k})^2\mathbb{P}\left(\frac{|d_{1,j,k}-d_{2,j,k}|}{d_{1,j,k}+d_{2,j,k}} < t_j\right) \le \\
&2w_{j,k}^2 + 2(\mu_{1,j,k}-\mu_{2,j,k})^2\mathbb{P}\left(\frac{|d_{1,j,k}-d_{2,j,k}|}{d_{1,j,k}+d_{2,j,k}} < t_j\right).
\end{aligned} \tag{32}$$

If $\mu_{1,j,k} = \mu_{2,j,k}$ then the second summand disappears. Assume, without loss of generality, that $\mu_{1,j,k} > \mu_{2,j,k}$. Noting that $w^2_{i,j,k} \le v \sup_z \sigma^4(z)/2$, we bound (32) from above by

$$
\begin{aligned}
&2v \sup_z \sigma^4(z) + 2(\mu_{1,j,k} - \mu_{2,j,k})^2 \mathbb{P}\left(\frac{d_{1,j,k} - d_{2,j,k}}{d_{1,j,k} + d_{2,j,k}} < t_j\right) = \\
&\quad 2v \sup_z \sigma^4(z) + 2(\mu_{1,j,k} - \mu_{2,j,k})^2 \mathbb{P}((d_{1,j,k} - \mu_{1,j,k})(t_j - 1) + (d_{2,j,k} - \mu_{2,j,k})(t_j + 1) + \\
&\quad + 2\,\mu_{1,j,k} t_j > (1 + t_j)(\mu_{1,j,k} - \mu_{2,j,k})) \le \quad \text{[Markov's inequality]} \\
&\quad 2v \sup_z \sigma^4(z) + \frac{2}{(1+t_j)^2}\left((1-t_j)^2 w^2_{1,j,k} + (1+t_j^2) w^2_{2,j,k} + 4\mu^2_{1,j,k} t_j^2\right) \le \\
&\quad 4v \sup_z \sigma^4(z) + 8 \sup_z \sigma^4(z) \log(N) = \\
&\quad 4 \sup_z \sigma^4(z)\, (v + 2\log(N))\,.
\end{aligned}
$$

We now move on to the last step of the proof: evaluation of the full L_2 risk. Define the set

$$\mathcal{N}_N = \{(j,k) : \sigma^2(i/N) := \text{constant, for } i = 2^{J-j}(k-1)+1, \ldots, 2^{J-j}k\}$$

(see Case 1 above). Denote by M the number of jumps in $\sigma^2(z)$. At each scale j, at most M indices (j,k) are in $\mathcal{N}_N^c$. We have

$$
\begin{aligned}
&\frac{1}{N}\sum_{j=0}^{J-1}\sum_{k=1}^{2^j} \mathbb{E}\left\{\mu_{j,k} - \hat{\mu}^{(h)}_{j,k}\right\}^2 = \\
&\quad \frac{1}{N}\sum_{(j,k)\in\mathcal{I}_N\cap\mathcal{N}_N} \mathbb{E}\left\{\mu_{j,k} - \hat{\mu}^{(h)}_{j,k}\right\}^2 + \frac{1}{N}\sum_{(j,k)\in\mathcal{I}_N\cap\mathcal{N}_N^c} \mathbb{E}\left\{\mu_{j,k} - \hat{\mu}^{(h)}_{j,k}\right\}^2 + \frac{1}{N}\sum_{(j,k)\in\mathcal{I}_N^c} \mu^2_{j,k} \le \\
&\quad \frac{1}{N}\sum_{(j,k)\in\mathcal{I}_N\cap\mathcal{N}_N} \tilde{C}_\delta \sup_z \sigma^4(z) N^{-\frac{2}{v}+\delta} + \frac{MJ^*}{N} 4 \sup_z \sigma^4(z)\,(v + 2\log(N)) + \\
&\quad + \frac{M}{N}\sum_{j=J^*}^{J-1} \frac{2^{J-j}}{4}\left\{\sup_z \sigma^2(z) - \inf_z \sigma^2(z)\right\}^2 \le \\
&\quad N^{-1} 2^{J^*} \tilde{C}_\delta \sup_z \sigma^4(z) N^{-\frac{2}{v}+\delta} + \frac{MJ^*}{N} 4 \sup_z \sigma^4(z)\,(v + 2\log(N)) + \\
&\quad + \frac{M}{2}\left\{\sup_z \sigma^2(z) - \inf_z \sigma^2(z)\right\}^2 \frac{2^{J-J^*} - 1}{N} = \\
&\quad O\left(N^{-\frac{2}{v}}\right) + O\left(N^{-1}\log^2(N)\right) + O\left(N^{\delta-1}\right) = O\left(N^{-\min\left(1-\delta, \frac{2}{v}\right)}\right), \quad \text{as} \quad N \to \infty.
\end{aligned}
$$

This completes the proof of Theorem 3.1. □

Proof of Theorem 3.2. The first of the two equalities in (18) is due to the orthonormality of the discrete Haar transform. We now show the second equality. Throughout the proof, we use the notation from the proof of Theorem 3.1. As in the latter, we first assume that $(j,k) \in \mathcal{I}_N$ and consider two cases:

1. Case $\sigma^2(i/N) :=$ constant $:= \sigma^2$, for $i = 2^{J-j}(k-1)+1, \ldots, 2^{J-j}k$ (so that $\mu_{j,k} = 0$). Without loss of generality, consider $k = 1$ to shorten the notation.

Noting that $\left(\hat{\mu}_{j,k}^{(s)}\right)^2 \le \left(\hat{\mu}_{j,k}^{(h)}\right)^2$, we have

$$\mathbb{E}\left\{\hat{\mu}_{j,k}^{(s)} - \mu_{j,k}\right\}^2 = \mathbb{E}\left\{\hat{\mu}_{j,k}^{(s)}\right\}^2 \le \mathbb{E}\left\{\hat{\mu}_{j,k}^{(h)}\right\}^2 \le \tilde{C}_\delta \sup_z \sigma^4(z) N^{-\frac{2}{v}+\delta},$$

where the last step uses the bound obtained in the proof of Theorem 3.1.

2. Case $\sigma^2(i/N) \neq$ constant, for $i = 2^{J-j}(k-1)+1, \ldots, 2^{J-j}k$ (so that possibly $\mu_{j,k} \neq 0$). Using the bound from the proof of Theorem 3.1, we obtain

$$\begin{aligned}\mathbb{E}\left\{\hat{\mu}_{j,k}^{(s)} - \mu_{j,k}\right\}^2 = \mathbb{E}\left\{\hat{\mu}_{j,k}^{(s)} - \hat{\mu}_{j,k}^{(h)} + \hat{\mu}_{j,k}^{(h)} - \mu_{j,k}\right\}^2 \le \\ 2\,\mathbb{E}\left\{\hat{\mu}_{j,k}^{(s)} - \hat{\mu}_{j,k}^{(h)}\right\}^2 + 2\,\mathbb{E}\left\{\hat{\mu}_{j,k}^{(h)} - \mu_{j,k}\right\}^2 \le 2\,\mathbb{E}\left\{\hat{\mu}_{j,k}^{(s)} - \hat{\mu}_{j,k}^{(h)}\right\}^2 + 8 \sup_z \sigma^4(z)(v + \log(N)).\end{aligned}$$

We now bound $2\,\mathbb{E}\left\{\hat{\mu}_{j,k}^{(s)} - \hat{\mu}_{j,k}^{(h)}\right\}^2$. Using the representation

$$\hat{\mu}_{j,k}^{(h)} = s_{j,k} f_{j,k} \mathbb{I}(|f_{j,k}| > t_j),$$

it is easy to see by direct comparison that

$$\left|\hat{\mu}_{j,k}^{(s)} - \hat{\mu}_{j,k}^{(h)}\right| \le s_{j,k} t_j,$$

which, using the explicit definition of $s_{j,k}$, the formula for t_j, and the bounds for $w_{i,j,k}^2$ and $\mu_{i,j,k}^2$, leads to

$$\begin{aligned} 2\,\mathbb{E}\left\{\hat{\mu}_{j,k}^{(s)} - \hat{\mu}_{j,k}^{(h)}\right\}^2 &\le 2\,t_j^2 \mathbb{E}\left\{s_{j,k}\right\}^2 \\ &= 2\,t_j^2 \left\{w_{1,j,k}^2 + w_{2,j,k}^2 + (\mu_{1,j,k} + \mu_{2,j,k})^2\right\} \\ &\le 8\,\log(N) \sup_z \sigma^4(z) \left\{v 2^{j-J} + 1\right\} \\ &\le 8\,\log(N) \sup_z \sigma^4(z) \left\{v + 1\right\}. \end{aligned}$$

This yields

$$\mathbb{E}\left\{\hat{\mu}_{j,k}^{(s)} - \mu_{j,k}\right\}^2 \le 8 \sup_z \sigma^4(z) \left\{\log(N)(v+2) + v\right\}.$$

The remaining part of the proof is completely analogous to the last part of the proof of Theorem 3.1, leading to the same rate. We omit the details. This completes the proof of Theorem 3.2. □

Proof of Lemma 3.1. Denote $U = (X^{(1)} - X^{(2)})/(X^{(1)} + X^{(2)})$. We have

$$F_U(t) = P(U < t) = P\left(\frac{X^{(1)}}{X^{(2)}} < \frac{1+t}{1-t}\right) = P\left(Z < \frac{1+t}{1-t}\right) = F_Z\left(\frac{1+t}{1-t}\right),$$

where $Z \sim F(m, m)$. Therefore, after some simple algebra,

$$\begin{aligned} f_U(t) &= f_Z\left(\frac{1+t}{1-t}\right) \frac{2}{(1-t)^2} \\ &= 2^{1-m} \frac{\Gamma(m)}{\Gamma(m/2)\Gamma(m/2)} (1+t)^{m/2-1}(1-t)^{m/2-1}, \end{aligned}$$

which, as can easily be verified, is the desired density. This completes the proof of Lemma 3.1. □

Acknowledgements

Piotr Fryzlewicz gratefully acknowledges the financial support of the Department of Mathematics at Imperial College London, and David Hand. The work of Theofanis Sapatinas was partly supported by a 'University of Cyprus – 2004 Research Grant'. The work of Suhasini Subba Rao was supported by DFG (DA 187/12-2) grant. Piotr Fryzlewicz would like to thank Theofanis Sapatinas for financial support and warm hospitality whilst visiting Nicosia to carry out part of this work. The authors also thank Guy Nason and Rainer von Sachs for their helpful comments.

References

[1] Abramovich, F., Bailey, T.C. & Sapatinas, T. (2000). Wavelet analysis and its statistical applications. *The Statistician*, **49**, 1–29.

[2] Antoniadis, A. & Fryzlewicz, P. (2004). Parametric modelling of thresholds across scales in wavelet regression. *Preprint.*

[3] Bera, A.K. & Higgins, M.L. (1993). ARCH models: properties, estimation and testing. *Journal of Economic Surveys*, **7**, 305–366.

[4] Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, **31**, 307–327.

[5] Clémençon, S. & Slim, S. (2004). Statistical analysis of financial time series under the assumption of local stationarity. *Quantitative Finance*, **4**, 208–220.

[6] Coifman, R.R. & Donoho, D.L. (1995). Translation-invariant de-noising. In *Wavelets and Statistics*, Antoniadis, A. & Oppenheim, G. (Eds.), Lecture Notes in Statistics, **103**, pp. 125–150, New York: Springer-Verlag.

[7] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics*, **16**, 1–37.

[8] Dahlhaus, R. & Subba Rao, S. (2003). Statistical inference for time-varying ARCH processes. *Preprint.*

[9] Donoho, D., Mallat, S., von Sachs, R. & Samuelides, Y. (2003). Locally stationary covariance and signal estimation with macrotiles. *IEEE Transactions on Signal Processing*, **53**, 614–627.

[10] Drees, H. & Stărică, C. (2003). A simple non-stationary model for stock returns. *Preprint.*

[11] Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1008.

[12] Fan, J. & Yao, Q. (2003). *Nonlinear Time Series.* New York: Springer-Verlag.

[13] Fisz, M. (1955). The limiting distribution of a function of two independent random variables and its statistical application. *Colloquium Mathematicum*, **3**, 138–146.

[14] Fryzlewicz, P. (2004). Modelling and forecasting financial log-returns as locally stationary wavelet processes. To appear in the *Journal of Applied Statistics.*

[15] Fryzlewicz, P. & Nason, G.P. (2003). Denoising the wavelet periodogram using the Haar-Fisz transform. *Preprint.*

[16] Fryzlewicz, P. & Nason, G.P. (2004). A Haar-Fisz algorithm for Poisson intensity estimation. *Journal of Computational and Graphical Statistics*, **13**, 621–638.

[17] Gao, H.-Y. (1997). Choice of thresholds for wavelet shrinkage estimate of the spectrum. *Journal of Time Series Analysis*, **18**, 231–252.

[18] Giraitis, L., Leipus, R. & Surgailis, D. (2003). Recent advances in ARCH modelling. *Preprint.*

[19] Hull, J.C. (1997). *Options, Futures, and Other Derivatives.* 3rd edition, London: Prentice-Hall.

[20] Johnson, N.L. & Kotz, S. (1970). *Continuous Univariate Distributions*, Vol. **1**. Boston: Houghton Mifflin Company.

[21] Kim, W. (1998). Econometric analysis of locally stationary time series models. *Unpublished manuscript, Yale University.*

[22] Kokoszka, P. & Leipus, R. (2000). Change-point estimation in ARCH models. *Bernoulli*, **6**, 513–539.

[23] Mikosch, T. & Stărică, C. (2000). Is it really long memory we see in financial returns? In *Extremes and Integrated Risk Management*, (Ed. P. Embrechts), pp. 149–168, London: Risk Books.

[24] Mikosch, T. & Stărică, C. (2003). Long-range dependence effects and ARCH modelling. In *Theory and Applications of Long Range Dependence*, (Eds. P. Doukhan, G. Oppenheim and M.S. Taqqu), pp. 439–459, Boston: Birkhäuser.

[25] Mikosch, T. & Stărică, C. (2004). Non-stationarities in financial time series, the long-range dependence, and the IGARCH effects. *The Review of Economics and Statistics*, **86**, 378-390.

[26] Moulin, P. (1994). Wavelet thresholding techniques for power spectrum estimation. *IEEE Transactions on Signal Processing*, **42**, 3126–3136.

[27] Nason, G.P., von Sachs, R. & Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary spectrum. *Journal of the Royal Statistical Society*, Series B, **62**, 271–292.

[28] Nason, G.P. & Silverman, B.W. (1995). The stationary wavelet transform and some statistical applications. In *Wavelets and Statistics*, Antoniadis, A. & Oppenheim, G. (Eds.), Lecture Notes in Statistics, **103**, pp. 281–300, New York: Springer-Verlag.

[29] Neumann, M. & von Sachs, R. (1995). Wavelet thresholding: beyond the Gaussian i.i.d. situation. In *Wavelets and Statistics*, Antoniadis, A. & Oppenheim, G. (Eds.), Lecture Notes in Statistics, **103**, pp. 301–329, New York: Springer-Verlag.

[30] Pensky, M. & Vidakovic, B. (2004). Bayesian decision theoretic scale adaptive estimation of a log-spectral density. *Preprint*.

[31] Rudzkis, R., Saulis, L. & Statulevičius, V. (1978). A general lemma on probabilities of large deviations. *Lithuanian Mathematical Journal*, **18**, 226–238.

[32] Shorack, G.R. & Wellner, J.A. (1986) *Empirical Processes with Applications to Statistics.* New York: John Wiley & Sons.

[33] Stărică, C. (2003). Is GARCH(1,1) as good a model as the accolades of the Nobel prize would imply? *Preprint*.

[34] Stărică, C. & Granger, C.W.J. (2003). Non-stationarities in stock returns. To appear in *The Review of Economics and Statistics*.

[35] Taylor, S.J. (1986). *Modelling Financial Time Series*. Chichester: John Wiley & Sons.

[36] Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. New York: John Wiley & Sons.

[37] Wahba, G. (1980). Automatic smoothing of the log-periodogram. *Journal of the American Statistical Association*, **75**, 122–132.

[38] Walden, A.T., Percival, D.B. & McCoy, E.J. (1998). Spectrum estimation by wavelet thresholding of multitaper estimators. *IEEE Transactions on Signal Processing*, **46**, 3153–3165.