A Haar-Fisz technique for locally stationary volatility estimation

By PIOTR FRYZLEWICZ
Department of Mathematics, University of Bristol,
University Walk, Bristol BS8 1TW, U.K.
p.z.fryzlewicz@bris.ac.uk

THEOFANIS SAPATINAS
Department of Mathematics and Statistics, University of Cyprus,
P.O. Box 20537, CY 1678 Nicosia, Cyprus.
t.sapatinas@ucy.ac.cy

AND SUHASINI SUBBA RAO
Department of Mathematics, University of Bristol,
University Walk, Bristol BS8 1TW, U.K.
s.subbarao@bris.ac.uk

Summary

We consider 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 the model explains well the common stylised facts of log-returns. We propose a new wavelet thresholding algorithm for volatility estimation in this model, where Haar wavelets are 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. We show that our approach both gives a very good fit to selected currency exchange datasets, and achieves accurate long- and short-term volatility forecasts in comparison to the GARCH(1,1) and moving window techniques.

Some key words: 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 is roughly symmetric (or slightly skewed), has a peak at zero, and is heavy-tailed.
3. The sample autocorrelations are "small" at almost all lags; however, the sample autocorrelations of the absolute values and squares are significant for a large number of lags.
4. Volatility is "clustered", i.e. days of either large or small movements are likely to be followed by days with similar characteristics.

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 ARCH (Engle, 1982) and GARCH (Bollerslev, 1986; Taylor, 1986) models, 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. (2005).

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 for example Mikosch & Stărică (2004). Indeed, Kokoszka & Leipus (2000) considered the problem of change-point detection in the ARCH model, whilst Dahlhaus & Subba Rao (2006) 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 relaxes 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 (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 for example Clémençon & Slim (2004), who apply the locally stationary covariance estimation methodology of Donoho et al. (2003) to log-returns, or Fryzlewicz (2005), who provides an exploratory analysis of log-returns in the framework of Nason et al. (2000).

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 show that the proposed piecewise-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.

2 The non-stationary model and motivation

In this section, we introduce the proposed 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 $\{P_{t,N}\}_{t=1}^{N}$ (for example, a stock index, a currency exchange rate, a share price, etc.), our object of interest is the log-return series $X_{t,N} := \log(P_{t,N}) - \log(P_{t-1,N})$. We propose the following non-stationary "stochastic triangular array" model for $\{X_{t,N}\}_{t=1}^{N}$:

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

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

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 $\{X_{t,N}\}$. 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 process $\{X_{t,N}\}_{t=1}^N$) rather than $\sigma(z)$ itself, we now specify the smoothness assumption for $\sigma^2(z)$ which will be used throughout the paper.

Assumption 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 $\{Z_t\}_t$ will be specified later.

Assumption 1 is not merely a "technical" assumption introduced in order that the theoretical results of this paper (for example, mean-square consistency of the proposed estimators) may hold. Piecewise stationarity is arguably the simplest type of departure from stationarity. In the paper, we demonstrate that this simple type of non-stationarity is already flexible and powerful enough to enable the successful modelling and forecasting of volatilities. We achieve this both by theoretically verifying that the piecewise constant model captures the commonly observed stylised facts (§ 2.2), and by demonstrating its good practical performance (§§ 4 and 5). We note that the piecewise constant modelling of volatilities has also been considered by Mercurio & Spokoiny (2004) and Polzehl & Spokoiny (2004).

In contrast to the ARCH/GARCH modelling approach where the volatility is modelled as a stationary stochastic process (more specifically, a function of the volatility 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 Stărică & Granger (2005), 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 function.

We also observe that our approach is different to the use of a piecewise-constant noise variance in threshold autoregressive models of Tong (1990). In the latter approach, different autoregressive regimes are followed above and below a certain threshold, which introduces nonlinear dependence in the process. In contrast to that approach, our model is linear (indeed, it is merely a sequence of independent variables) with a locally constant unconditional variance.

2.2 Explanation of the common stylised facts

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

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

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

The following proposition holds.

Proposition 1 *Suppose that $\{X_{t,N}\}_{t=1}^N$ follows model (1), and that Assumption 1 holds. Assume further that $E(Z_t^8) < \infty$. Suppose $\{h_N\}_N$ is a sequence such that $h_N > 0$ and for some $\beta \geq 0$, $h_N/N \to \beta$ as $N \to \infty$. Then we have*

$$\bar{X}_N^1 \quad \to \quad 0, \tag{2}$$

in mean-square, as $N \to \infty$,

$$\frac{\bar{X}_N^4}{\left(\bar{X}_N^2\right)^2} \quad \to \quad E(Z_t^4)\frac{\int_0^1 \sigma^4(z)dz}{\left\{\int_0^1 \sigma^2(z)dz\right\}^2}, \tag{3}$$

in probability, as $N \to \infty$,

$$\gamma_{X^1}^N(h) \quad \to \quad 0, \text{ for a fixed } h > 0, \tag{4}$$

in mean-square, as $N \to \infty$,

$$\gamma_{X^2}^N(h_N) \quad \to \quad \int_0^{1-\beta} \sigma^2(z)\sigma^2(z+\beta)dz - \left\{\int_0^1 \sigma^2(z)dz\right\}^2, \tag{5}$$

in mean-square, as $N \to \infty$.

Typically, upon assuming that our observations come from a stationary (not necessarily linear) process, which is indeed often done in log-return analysis, we would use the quantities on the left-hand sides of formulae (2) – (5) as measures of the mean, the kurtosis, the autocovariance, and the autocovariance of the squares of the data, respectively. As mentioned in § 1, we would then in most cases observe (a) the sample mean of the data being close to 0, (b) the sample kurtosis being greater than 3 (3 being the kurtosis of a Gaussian variable), (c) the sample autocovariance of the data at lag $h > 0$ being close to 0, (d) the sample autocovariances of the squares of the data not decaying to 0, or decaying to 0 only very slowly. Proposition 1 shows that these stylised facts can be well explained in our model: indeed, formulae (2) – (5) provide a heuristic explanation for the above stylised facts (a) – (d), respectively, even in the case of Z_t being Gaussian. In particular, note that the ratio on the right-hand side of formula (3) is always greater that 1, unless $\sigma^2(z)$ is constant when it is equal to 1. Similarly, if $h_N = h > 0$, then the integral on the right-hand side of formula (5) 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 1 demonstrates that the estimated sample autocovariance evaluated under the premise that the process is stationary gives a misleading view as to the true dependence structure of the underlying process. More precisely, it is clear that the true correlations of the process $\{X_{t,N}^2\}_{t=1}^N$ are zero. However for all $h > 0$, it is straightforward to see from (5) that the autocovariance estimator $\gamma_{X^2}^N(h)$ does not necessarily converge to zero as $N \to 0$ and the correlations appear to persist for all values of h, thus giving the wrong impression that $\{X_{t,N}^2\}_{t=1}^N$ may be correlated, or even have the long-memory property, when in fact they are independent.

We finally note that our model naturally captures the often observed clustering of volatility. Indeed, the piecewise-constant form of $\sigma^2(z)$ means that the local variance remains at the same level for a number of time units, thus modelling the volatility clustering.

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 $\{X_{t,N}\}_{t=1}^N$ 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; in other words, if the regularity of $\sigma^2(z)$ varies from one region to another. For an overview of wavelet methods in statistics, we refer the reader to the monograph of Vidakovic (1999).

The starting point for these considerations is a reformulation of (1):

$$X_{t,N}^2 = \sigma^2(t/N)\, Z_t^2, \quad t = 1, 2, \ldots, N. \tag{6}$$

Note that $X_{t,N}^2$ is an unbiased but inconsistent estimate of $\sigma^2(t/N)$, and thus needs to be smoothed to achieve consistency. Obviously, (6) can be rewritten as:

$$X_{t,N}^2 = \sigma^2(t/N) + \sigma^2(t/N)(Z_t^2 - 1), \quad t = 1, 2, \ldots, N, \tag{7}$$

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

Neumann & von Sachs (1995) used a nonlinear wavelet estimation technique in a setting similar to (7). However, their method involved finding an estimate of the local variance of the "noise" (in our case: $\sigma^2(t/N)(Z_t^2 - 1)$), which in our case would amount to finding a pre-estimate of $\sigma^2(t/N)$ itself. This is an obvious drawback of the estimation procedure, and can hamper the practical performance of the method (Fryzlewicz, 2005).

In order to avoid having to find a pre-estimate of $\sigma^2(t/N)$, an obvious step would be to take the logarithmic transformation of (6):

$$\log X_{t,N}^2 = \log \sigma^2(t/N) + \log Z_t^2, \quad t = 1, 2, \ldots, N. \tag{8}$$

The logarithmic transformation transforms model (6) from multiplicative 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 $\{Z_t\}_t$ being a sequence of independent and identically distributed $N(0, 1)$ 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 (for example Moulin, 1994; Gao, 1997), and those techniques could be adapted to our framework. However, any wavelet estimator in the setting specified by (8) 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). Exponentiating this estimate would yield *an* estimate of $\sigma^2(t/N)$; however, the statistical properties of the latter, such as mean-square consistency, would not be easy to establish (note that, generally, the existence of the second moment of a random variable Y does not imply the existence of the second moment of $\exp(Y)$).

- Any wavelet estimator in the model (8) would suffer from a bias of order $E(\log Z_t^2)$. Since, as mentioned before, we do not assume any specific distributional form of the innovation process $\{Z_t\}_t$, 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) and it yields an asymptotically unbiased, mean-square consistent estimate of $\sigma^2(t/N)$, as opposed to $\log \sigma^2(t/N)$ (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 $\{X_{t,N}^2\}_{t=1}^N$: here, we assume that N is an integer power of two; techniques for adapting wavelet transforms to non-dyadic sample sizes are described in Wickerhauser (1994). To simplify the notation, we drop the subscript N and consider the sequence $X_t^2 := X_{t,N}^2$. We denote $J = \log_2 N$. The estimation algorithm proceeds as follows:

1. Compute the Haar decomposition of $\{X_t^2\}_{t=1}^N$ using the following algorithm:
   (a) Let $s_{J,k} := X_k^2$, $k = 1, 2, \ldots, 2^J$.
   (b) For each $j = J-1, J-2, \ldots, 0$, recursively form vectors s_j, d_j and f_j with elements:
$$\begin{aligned} s_{j,k} &= \frac{s_{j+1,2k-1} + s_{j+1,2k}}{\sqrt{2}} \\ d_{j,k} &= \frac{s_{j+1,2k-1} - s_{j+1,2k}}{\sqrt{2}} \\ f_{j,k} &= \frac{d_{j,k}}{s_{j,k}}, \end{aligned}$$
   where $k = 1, \ldots, 2^j$.

2. For each $j = J-1, J-2, \ldots, 0$ and $k = 1, 2, \ldots, 2^j$, denote $\mu_{j,k} := E(d_{j,k})$. For most levels j (in a sense to be made precise later), estimate $\mu_{j,k}$ by
$$\begin{aligned} \hat{\mu}_{j,k}^{(h)} &= s_{j,k} f_{j,k}\, I(|f_{j,k}| > t_j) \\ &= d_{j,k}\, I(|f_{j,k}| > t_j) \quad \text{(hard thresholding)}, \end{aligned} \tag{9}$$

or by

$$\hat{\mu}^{(s)}_{j,k} = s_{j,k}\,\mathrm{sgn}(f_{j,k})\,(|f_{j,k}| - t_j)_+ \quad \text{(soft thresholding)}, \tag{10}$$

where $I(\cdot)$ and $\mathrm{sgn}(\cdot)$ are the indicator and signum functions, respectively, and $(x)_+ = \max(0, x)$. 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 thresholded 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). Explicit formulae for these two estimators are given later in this section.

Asymptotic Gaussianity 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 $f_{J-1,1} = (Z_1^2 - Z_2^2)/(Z_1^2 + Z_2^2)$, and the variance of $f_{J-1,1}$ does not depend on $\sigma^2(z)$. Thus, the thresholds t_j in (9) and (10) also do not need to depend on $\sigma^2(z)$, and can therefore be selected more easily.

In the above example, if $\sigma^2(1/N)$ was not equal to $\sigma^2(2/N)$ (if a jump occurred between times $1/N$ and $2/N$), then the distribution of $f_{J-1,1}$ would depend on $\sigma^2(1/N)$ and $\sigma^2(2/N)$ in a non-trivial way. In particular, we could expect $f_{J-1,1}$ to be significantly deviated from zero, if the value of $\sigma^2(1/N)$ was much different from that of $\sigma^2(2/N)$. In that case, hopefully, the corresponding coefficient $d_{J-1,1}$ would "survive" the process of thresholding.

Note that the Haar-Fisz *transform* for Poisson data, an algorithmic device for stabilizing the variance of Poisson data and bringing their distribution closer to normality, was introduced by Fryzlewicz & Nason (2004).

We now give precise and explicit definitions of $\hat{\sigma}^2_{(h)}(t/N)$ and $\hat{\sigma}^2_{(s)}(t/N)$ in terms of Haar wavelet vectors. For $j = 0, \ldots, J - 1$ and $k = 1, \ldots, 2^j$, define the Haar wavelet vectors $\{\psi_{j,k}(t)\}_{t=1}^{2^J}$ as

$$\begin{aligned}\psi_{j,k}(t) &= 2^{(j-J)/2}\, I[t \in \{(k-1)2^{J-j} + 1, \ldots, (k - \frac{1}{2})2^{J-j}\}] \\ &- 2^{(j-J)/2}\, I[t \in \{(k - \frac{1}{2})2^{J-j} + 1, \ldots, k2^{J-j}\}].\end{aligned}$$

Fix $\delta \in (0,1)$. For each $N = 2^J$, define the set $\mathcal{I}_N = \{(j,k) : j < J^*\}$, with $2^{J^*} = O\left(N^{1-\delta}\right)$. The estimators $\hat{\sigma}^2_{(h)}(t/N)$ and $\hat{\sigma}^2_{(s)}(t/N)$ are defined as

$$\begin{aligned} \hat{\sigma}^2_{(h)}(t/N) &= \bar{X}^2_N + \sum_{(j,k)\in\mathcal{I}_N} \hat{\mu}^{(h)}_{j,k}\psi_{j,k}(t) && (11) \\ \hat{\sigma}^2_{(s)}(t/N) &= \bar{X}^2_N + \sum_{(j,k)\in\mathcal{I}_N} \hat{\mu}^{(s)}_{j,k}\psi_{j,k}(t), && (12) \end{aligned}$$

where $\hat{\mu}^{(h)}_{j,k}$ and $\hat{\mu}^{(s)}_{j,k}$ are as in formulae (9) and (10), respectively, with

$$t_j = 2^{-\frac{J-j-1}{2}}\sqrt{(2\log N)}. \tag{13}$$

Outside the set $\mathcal{I}_N$, we simply define $\hat{\mu}^{(h)}_{j,k} = \hat{\mu}^{(s)}_{j,k} = 0$. Denote

$$v := E(|Z_t^2 - 1|^2), \tag{14}$$

and consider the following assumption:

Assumption 2. The law of the random variable Z_t^2 has no atom at 0, and there exist $c > 0$, $\gamma \geq 0$ such that

$$E(|Z_t^2 - 1|^n) \leq c^{n-2}(n!)^{1+\gamma}v, \quad \text{for all} \quad n \geq 3. \tag{15}$$

Remark 1. By elementary properties of the Gaussian and Laplace distributions (Johnson & Kotz, 1970), Assumption 2 is satisfied, in particular, if Z_t is standard Gaussian (with $v = 2$, $\gamma = 0$) or standard Laplace (with $v = 5$, $\gamma = 2$). Assumption 2 can also accommodate other distributions which are leptokurtic or possess a degree of skewness.

The following theorem demonstrates the mean-square consistency of $\hat{\sigma}^2_{(h)}(z)$ and $\hat{\sigma}^2_{(s)}(z)$.

THEOREM **1** *Suppose that* $\{X_{t,N}\}_{t=1}^N$ *follows model (1), and that Assumptions 1 – 2 hold. Let* (e) *be either one of:* (h) *and* (s)*. Then we have*

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

where v *is defined as in formula (14).*

Remark 2. Note that, in the particular case of Z_t being standard Gaussian (so that $v = 2$), the mean-square error rate in (16) reduces to $O(N^{-1+\delta})$, which is arbitrarily close to the parametric rate $O(N^{-1})$. Intuitively, this is not surprising as our problem is "almost parametric" in the sense that our target function is piecewise constant with a finite number of jumps, but the exact number, locations or magnitudes of the jumps are not known. It is clear from the proof of Theorem 1 that the exact parametric rate is impossible to obtain for our estimation procedure, due to the fact that we only use non-trivial estimators of $\mu_{j,k}$ in the set $\mathcal{I}_N$, which is essential for a certain asymptotic normality effect to hold. This effect holds for any choice of $\delta > 0$, and thus, in theory, it is beneficial to choose δ to be "as small as possible". Employing the asymptotic normality as a tool, the rate is shown to be $O(N^{-1+\delta})$ using the fact that the target function is piecewise constant.

3.2 Noise-free reconstruction

In this section, we consider the case when the errors Z_t in (1) are standard Gaussian. 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 $\{X_{t,N}^2\}_{t=1}^N$.

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 $N \to \infty$, or more precisely:

$$\mathrm{pr}\left\{\cup_{j=0}^{J-1} \cup_{k=1}^{2^j} (|f_{j,k}| > \tilde{t}_j)\right\} \to 0, \quad \text{as} \quad N \to \infty \quad (J = \log_2 N), \tag{17}$$

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 § 3.1.

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

Lemma 1 *Let $\{X_i\}_{i=1}^{2m}$ be a sequence of independent and identically distributed χ_1^2 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 $m/2$ (in other words, $Y \sim \beta(m/2, m/2)$).*

We now derive the thresholds $\tilde{t}_j$. As the distribution of $f_{j,k}$ does not depend on k, we can define $\alpha_j(N) = \mathrm{pr}(|f_{j,k}| < \tilde{t}_j)$. We have the following bound for (17) (note the use of the Bonferroni inequality):

$$\mathrm{pr}\left\{\cup_{j=0}^{J-1} \cup_{k=1}^{2^j} (|f_{j,k}| > \tilde{t}_j)\right\} \leq \sum_{j=0}^{J-1} 2^j \{1 - \alpha_j(N)\}. \tag{18}$$

Our objective is to choose $\{\alpha_j(N)\}_{j=0}^{J-1}$ such that $\sum_{j=0}^{J-1} 2^j\{1 - \alpha_j(N)\} \to 0$, as $N \to \infty$. The choice of $\{\alpha_j(N)\}_{j=0}^{J-1}$ will determine how fast $\sum_{j=0}^{J-1} 2^j\{1 - \alpha_j(N)\}$ approaches zero and thus the rate of convergence. Probably the simplest option is to mimick standard universal thresholding in a classical independent and identically distributed Gaussian nonparametric regression setting for wavelets, where the analogue of $\alpha_j(N)$ is constant across

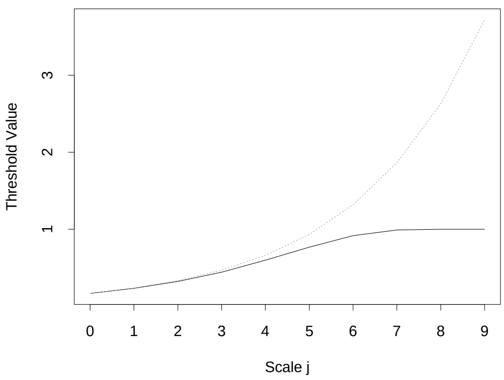

Figure 1: Thresholds $\tilde{t}_j$ (solid line) and t_j (dotted line) for $J = 10$ and $j = 0, \dots, J-1$. See discussion in §3.2.

scales ($\alpha_j(N) = \alpha(N)$) and the rate of convergence to 0 of the probability corresponding to (17) is equal to $(\pi J \log 2)^{-1/2}$. To guarantee such a rate we choose $\alpha(N)$ such that:

$$\mathrm{pr}\left\{\cup_{j=0}^{J-1} \cup_{k=1}^{2^j} \left(|f_{j,k}| > \tilde{t}_j\right)\right\} \leq \sum_{j=0}^{J-1} 2^j\{1 - \alpha(N)\} = \frac{1}{\sqrt{(\pi J \log 2)}},$$

which is uniquely solved by

$$\alpha^*(N) = 1 - (2^J - 1)^{-1}(\pi J \log 2)^{-1/2}. \tag{19}$$

In the light of what was said above, $\alpha^*(N)$ guarantees the convergence of (17) to 0 at a rate of *at least* $(\pi J \log(2))^{-1/2}$.

By Lemma 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 $\alpha^*(N) = P(|f_{j,k}| < \tilde{t}_j)$. 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 $X^{(2)} = \sum_{i=m+1}^{2m} Z_i^2$.

Fig. 1 compares the thresholds t_j and $\tilde{t}_j$ for $J = 10$ and $j = 0, \dots, J-1$. 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 (Antoniadis & Fryzlewicz, 2006). For simplicity, we consider a linear dependence of $\alpha_j(N)$ on j:

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

The equation for $(\alpha_0(N), \alpha_{J-1}(N))$ is:

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

where $c_J \downarrow 0$ is the desired rate of convergence. This simplifies to

$$\alpha_{J-1}(N)\{2^J(J-2)+2\} + \alpha_0(N)(2^J - J - 1) = (2^J - 1 - c_J)(J-1).$$

One possibility is to set $\alpha_{J-1}(N) = \alpha^*(N)$ and then solve for $\alpha_0(N)$. As a special case, note that setting $\alpha_{J-1}(N) = \alpha^*(N)$ and $c_J = (\pi J \log 2)^{-1/2}$ gives the solution $\alpha_0(N) = \alpha^*(N)$, which implies that $\alpha_j(N) = \alpha^*(N)$ for all j; in this case, $\alpha_j(N)$ does not depend on j.

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 Fig. 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 (13)) which guarantee mean-square consistency. We take $J^* = J - 1$. See § 3.1 for details.

NF-p-H(-TI): 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 (19)) and $\alpha_0 = \frac{p}{100}\alpha_{J-1}$, where $p \leq 100$. See § 3.2 for details. The suffix TI denotes the translation invariant version: in TI versions of wavelet-based denoising algorithms, the final estimator is obtained as the average of the estimators obtained for all circular shifts of the data. This is common practice in wavelet regression. The fast $O(N \log N)$ implementation of TI wavelet thresholding algorithms uses the Non-Decimated Wavelet Transform (NDWT) (Nason & Silverman, 1995).

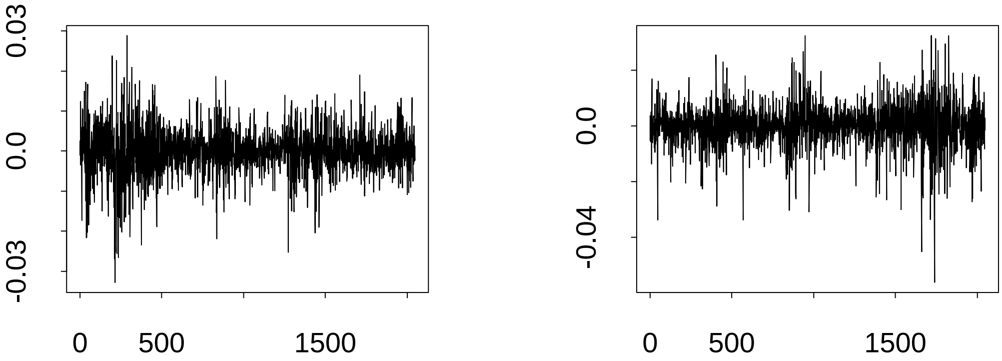

Figure 2: The last $N = 2048$ observations of the USD/GBP (left) and USD/JPY (right) exchange rate series, running from 8 November 1991 to 31 December 1999.

NF-p-S(-TI): The same as above, but with soft thresholding.

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 autocorrelation in the squares, that is the p-value of the Ljung-Box test for lack of serial correlation in $\hat{Z}_t^2$ is above a pre-specified threshold λ (in all of the examples in the paper, we use $\lambda = 0.05$).

In an extensive empirical study which compared several parameter configurations for several currency exchange datasets, we found that for $p = 100$, the corresponding estimators NF-100-H, NF-100-S(-TI) often oversmoothed, in the sense that the empirical residuals displayed significant dependence in the squares. The suitable value of p was then chosen as follows: we decreased p over the grid $100, 99, 98, \ldots$ until the Ljung-Box test indicated no significant correlation in the squared empirical residuals. We found that either of the two estimators: NF-98-S or NF-97-S, as well as their TI versions, performed well for most of the datasets considered. On the other hand, the estimators NF-p-H for $p < 100$ were often extremely "spiky". Typically, NF-100-H-TI produced correctly behaved empirical residuals, although the reconstructions were also often spiky. The price for using any 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 MS-H estimator typically gave slightly oversmoothed reconstructions. This was due to the fact that, as mentioned in §3.2, 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.

To summarise, NF-p-S, NF-p-S-TI and NF-p-H-TI were the preferred estimators. For the USD/GBP series, the corresponding values of p for those estimators, selected by the automatic procedure desribed above, were: $p = 97$, $p = 97$ and $p = 100$, respectively. The respective p-values of the Ljung-Box test were 0.09, 0.06 and 0.82. For the USD/JPY series, the selected values of p for the above three estimators were: $p = 97$, $p = 98$ and $p = 100$, respectively. The respective p-values of the Ljung-Box test were 0.19, 0.18 and 0.94.

The left column of Fig. 3 shows the results for the USD/GBP series. The top plot shows the NF-97-S-TI estimate and the middle plot shows the NF-100-H-TI estimate. The NF-97-S estimate is a piecewise constant function whose breakpoints can be loosely interpreted as "significant changes" in the volatility. The bottom plot shows the squared returns and the locations of the breakpoints of NF-97-S. For clarity, we only plot the first 250 observations, which roughly corresponds to one business year starting 8 November 1991. The right plot shows the corresponding results for the USD/JPY series except that the top plot shows NF-98-S-TI and the bottom plot shows the last 250 observations (which roughly corresponds to one business years starting 1 January 1999). Note that TI estimates are not useful for breakpoint detection as they are continuous.

Although the soft-thresholding TI estimates were smoother and thus more visually appealing than the hard-thresholding TI estimates, it was far from obvious that they should be preferred, as the p-values for the latter ones were much higher. This might imply that some of the spikes observed in the hard-thresholding estimates were not merely artefacts from hard thresholding but served to explain significant transient features of the volatility function. On the other hand, the low p-values produced by the soft-thresholding estimates might be due to the occasional bias introduced by soft thresholding, which is a well-known phenomenon in the classical Gaussian regression context.

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. Suppose that we observe $X_{1,N}^2, X_{2,N}^2, \ldots, X_{t,N}^2$ from model (1) and want to forecast the volatility at times $t+1, \ldots, t+h$, where $t+h \leq N$. The Mean-Square-optimal forecasts are given by:

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

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 (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 and NF-100-S. Both

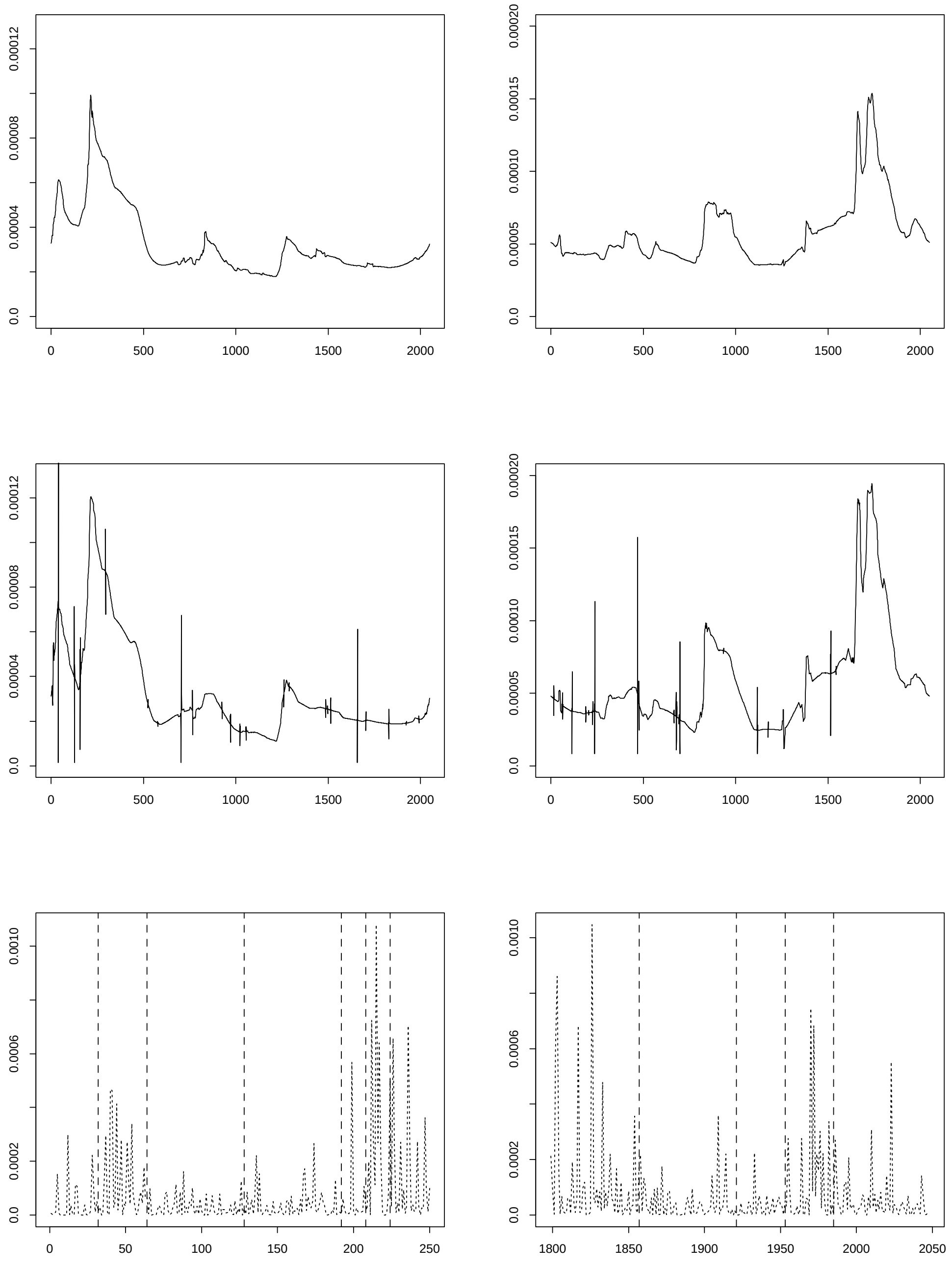

Figure 3: Empirical results for the USD/GBP (left column) and the USD/JPY (right column) series. Left column, top: NF-97-S-TI estimate; middle: NF-100-H-TI estimate; bottom: 1st 250 observations of the squared series (dotted) and the corresponding breakpoints of the NF-97-S estimate (dashed). Right column, top: NF-98-S-TI estimate; middle: NF-100-H-TI estimate; bottom: last 250 observations of the squared series (dotted) and the corresponding breakpoints of NF-97-S (dashed).

estimates are computed on $X^2_{t-1023,N}, \ldots, X^2_{t,N}$. Then, for all h, the forecast $\sigma^{2,\mathrm{HF}}_{t|t+h,N}$ is simply obtained as the value of our estimate at the "last" time point t/N.

For GARCH-based forecasts, we forecast the volatility at time $t+1, \ldots, t+h$ from $X^2_1, \ldots, X^2_t$ using the following methods:

G-SCROLL: We fit the stationary GARCH(1,1) model with standard 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`.

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 $(X^2_{t-h+1}, \ldots, X^2_t)$. 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 it is to be applied.

Let $\sigma^2_{t|t+h}$ 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 $\bar{\sigma}^2_{t|t+250} = \sum_{h=1}^{250} \sigma^2_{t|t+h}$, and compare it to the "realised" volatility $\bar{X}^2_{t|t+250} = \sum_{h=1}^{250} X^2_{t+h}$, using the Average Squared Error

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

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 § 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. This was due to the fact that the numerical maximiser of the likelihood in the S-Plus routine `garch` failed to converge. Our NF-100-S method and the simple MW technique performed the best, or nearly 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.

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 §5.

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 stationary 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 worse 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.

Appendix 1
Explanation of the stylised facts

Proof of Proposition 1. We first show (5). Denote $K_\sigma = \max_z \sigma^2(z)$, and let $\mathrm{TV}(\sigma^2)$ be the total variation of $\sigma^2(z)$. We have

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

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

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

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

$$E\left[\left(\bar{X}_N^2\right)^2 - \left\{\int_0^1 \sigma^2(z)dz\right\}^2\right]^2 \leq$$
$$2\,E\left[\left(\bar{X}_N^2\right)^2 - \left\{\frac{1}{N}\sum_{t=1}^{N}\sigma^2\left(\frac{t}{N}\right)\right\}^2\right]^2 + 2\left[\left\{\frac{1}{N}\sum_{t=1}^{N}\sigma^2\left(\frac{t}{N}\right)\right\}^2 - \left\{\int_0^1 \sigma^2(z)dz\right\}^2\right]^2$$
$$=: III + IIII.$$

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

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

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

$$III_1 =$$

$$\frac{1}{N^4} E\left[\left\{\sum_{t=1}^{N} \sigma^4\left(\frac{t}{N}\right)(Z_t^2-1)^2\right\}^2 + 4\sum_{t=1}^{N}\sum_{s\neq t}\sigma^4\left(\frac{t}{N}\right)\sigma^4\left(\frac{s}{N}\right)(Z_t^2-1)^2(Z_s^2-1)^2\right]$$
$$= O(N^{-2}).$$

Hence (5) follows.

We now show (2), (3) and (4). Since the proofs of (2) and (4) use exactly the same technique, we omit the details. For (3), note that, using the same technique as above, it is easy to show that

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

in mean-square as $N \to \infty$, which also imply convergence in probability. Thus, Slutsky's theorem yields (3). This completes the proof of Proposition 1. $\square$

Appendix 2
Properties of the Haar-Fisz estimator

Proof of Theorem 1 (for (e) = (h)). The first of the two equalities in (16) is due to the orthonormality of the discrete Haar transform. Note that the term $vN^{-2}\sum_{t=1}^{N}\sigma^4(t/N)$ arises because of the inclusion of the term $\bar{X}_N^2$ in the estimator (11). We now show the second equality. For notational clarity, denote $d_{1,j,k} = s_{j+1,2k-1}/\sqrt{2}$ and $d_{2,j,k} = s_{j+1,2k}/\sqrt{2}$, 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 $\mu_{i,j,k} = E(d_{i,j,k})$ and $w_{i,j,k}^2 = \text{var}(d_{i,j,k})$ for $i = 1, 2$. Finally, denote $w_{j,k}^2 = \text{var}(d_{j,k})$.

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 $(j,k) \in \mathfrak{I}_N$.

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.

$$E\left\{(d_{1,j,1} - d_{2,j,1}) I\left(\frac{|d_{1,j,1} - d_{2,j,1}|}{d_{1,j,1} + d_{2,j,1}} > t_j\right) - (\mu_{1,j,1} - \mu_{2,j,1})\right\}^2 =$$
$$E\left\{(d_{1,j,1} - d_{2,j,1}) I\left(\frac{|d_{1,j,1} - d_{2,j,1}|}{d_{1,j,1} + d_{2,j,1}} > t_j\right)\right\}^2 =$$
$$2^{j-J}\sigma^4\, E\left\{\left(\sum_{i=1}^{2^{J-j-1}} Z_i^2 - Z_{i+2^{J-j-1}}^2\right)^2 I\left(\frac{|d_{1,j,1} - d_{2,j,1}|}{d_{1,j,1} + d_{2,j,1}} > t_j\right)\right\} =$$

$$2^{j-J}\sigma^4\, E\left[\left\{\sum_{i=1}^{2^{J-j-1}}(Z_i^2-Z_{i+2^{J-j-1}}^2)^2+\sum_{\substack{i,l=1\\ i\neq l}}^{2^{J-j-1}}(Z_i^2-Z_{i+2^{J-j-1}}^2)(Z_l^2-Z_{l+2^{J-j-1}}^2)\right\}\times\right.$$
$$\left.\times\quad I\left(\frac{|d_{1,j,1}-d_{2,j,1}|}{d_{1,j,1}+d_{2,j,1}}>t_j\right)\right]. \tag{20}$$

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

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

which simplifies (20) to

$$\begin{aligned}
&2^{j-J}\sigma^4\sum_{i=1}^{2^{J-j-1}}E\left\{(Z_i^2-Z_{i+2^{J-j-1}}^2)^2 I\left(\frac{|d_{1,j,1}-d_{2,j,1}|}{d_{1,j,1}+d_{2,j,1}}>t_j\right)\right\}=\\
&\frac{\sigma^4}{2}\,E\left\{(Z_1^2-Z_{1+2^{J-j-1}}^2)^2 I\left(\frac{|d_{1,j,1}-d_{2,j,1}|}{d_{1,j,1}+d_{2,j,1}}>t_j\right)\right\}\le\\
&\frac{\sigma^4}{2}\left[E\left\{(Z_1^2-Z_{1+2^{J-j-1}}^2)^{2r}\right\}\right]^{1/r}\mathrm{pr}\left(\frac{|d_{1,j,1}-d_{2,j,1}|}{d_{1,j,1}+d_{2,j,1}}>t_j\right)^{1-1/r}:=\\
&\frac{\sigma^4}{2}C_r\mathrm{pr}\left(\frac{|d_{1,j,1}-d_{2,j,1}|}{d_{1,j,1}+d_{2,j,1}}>t_j\right)^{1-1/r},
\end{aligned}\tag{21}$$

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

$$\begin{aligned}
&\mathrm{pr}\left(\frac{|d_{1,j,1}-d_{2,j,1}|}{d_{1,j,1}+d_{2,j,1}}>t_j\right)=2\,\mathrm{pr}\left(\frac{d_{1,j,1}-d_{2,j,1}}{d_{1,j,1}+d_{2,j,1}}>t_j\right)=\\
&2\,\mathrm{pr}\left[\frac{2^{\frac{j-J}{2}}}{\surd\{v(1+t_j^2)\}}\left\{\sum_{i=1}^{2^{J-j-1}}(Z_i^2-1)(1-t_j)-\sum_{i=2^{J-j-1}+1}^{2^{J-j}}(Z_i^2-1)(1+t_j)\right\}>\frac{t_j2^{\frac{J-j}{2}}}{\surd\{v(1+t_j^2)\}}\right].
\end{aligned}\tag{22}$$

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

$$\frac{t_j2^{\frac{J-j}{2}}}{\surd\{v(1+t_j^2)\}}=o\left(\left\{\frac{2^{\frac{J-j}{2}}\frac{\surd\{v(1+t_j^2)\}}{1+t_j}}{2\max\{c,\surd v\}}\right\}^{\nu}\right),\quad \text{as}\quad N\to\infty,$$

for any positive ν. Therefore, by the Corollary to Theorem 1 from Rudzkis *et al.* (1978), we bound (22) from above by

$$2C\,\mathrm{pr}\left\{N(0,1)>\frac{t_j2^{\frac{J-j}{2}}}{\surd\{v(1+t_j^2)\}}\right\}. \tag{23}$$

Recalling again that $t_j = 2^{\frac{-J+j+1}{2}}\sqrt{(2\log N)}$ and denoting by $\Phi(\cdot)$ the cumulative distribution function of a $N(0,1)$ random variable, we now bound (23) from above using Mill's ratio inequality (Shorack & Wellner, 1986, p. 850):

$$2C\left[1-\Phi\left\{\frac{2\sqrt{(\log N)}}{\sqrt{\{v(1+t_j^2)\}}}\right\}\right] \leq C\,\exp\left\{-\frac{4\log N}{2v(1+t_j^2)}\right\} = C\,N^{-\frac{2}{v(1+t_j^2)}}. \tag{24}$$

Plugging (24) into (21), we obtain

$$\frac{\sigma^4}{2}C_r\,\mathrm{pr}\left(\frac{|d_{1,j,1}-d_{2,j,1}|}{d_{1,j,1}+d_{2,j,1}} > t_j\right)^{1-1/r} \leq \sigma^4\tilde{C}_r N^{-\frac{2(1-1/r)}{v(1+t_j^2)}}, \tag{25}$$

for some appropriate positive $\tilde{C}_r$. Noting that $t_j \leq t_{J^*} = O(N^{-\delta/2}\sqrt{(\log N)})$ uniformly on $\mathcal{I}_N$, it is easy to show by direct comparison that $N^{-2(1-1/r)/(v+vt_j^2)} = O(N^{-2(1-1/r)/v})$ as $N \to \infty$. Upon choosing $r = 2(v\delta)^{-1}$, we obtain the final bound for (25) as $\tilde{C}_\delta \sup_z \sigma^4(z)N^{-\frac{2}{v}+\delta}$.

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

$$\begin{aligned}
&E\left\{(d_{1,j,k}-d_{2,j,k})I\left(\frac{|d_{1,j,k}-d_{2,j,k}|}{d_{1,j,k}+d_{2,j,k}} > t_j\right) - (\mu_{1,j,k}-\mu_{2,j,k})\right\}^2 \leq \\
&\quad 2\,E\left\{(d_{1,j,k}-d_{2,j,k}-(\mu_{1,j,k}-\mu_{2,j,k}))I\left(\frac{|d_{1,j,k}-d_{2,j,k}|}{d_{1,j,k}+d_{2,j,k}} > t_j\right)\right\}^2 + \\
&\quad 2\,(\mu_{1,j,k}-\mu_{2,j,k})^2\mathrm{pr}\left(\frac{|d_{1,j,k}-d_{2,j,k}|}{d_{1,j,k}+d_{2,j,k}} < t_j\right) \leq \\
&\quad 2w_{j,k}^2 + 2(\mu_{1,j,k}-\mu_{2,j,k})^2\mathrm{pr}\left(\frac{|d_{1,j,k}-d_{2,j,k}|}{d_{1,j,k}+d_{2,j,k}} < t_j\right).
\end{aligned} \tag{26}$$

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 $w_{i,j,k}^2 \leq v\sup_z \sigma^4(z)/2$, we bound (26) from above by

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

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

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

(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

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

This completes the proof of Theorem 1 for (e) = (h). □

Proof of Theorem 1 (for (e) = (s).) The first of the two equalities in (16) is due to the orthonormality of the discrete Haar transform. Note that the term $vN^{-2}\sum_{t=1}^{N}\sigma^4(t/N)$ arises because of the inclusion of the term $\bar{X}_N^2$ in the estimator (12). We now show the second equality. Throughout the proof, we use the notation from the proof of Theorem 1 for (e) = (h). 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

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

where the last step uses the bound obtained in the proof of Theorem 1 for (e) = (h).

2. Case $\sigma^2(i/N) \ne$ constant, for $i = 2^{J-j}(k-1)+1,\ldots,2^{J-j}k$ (so that possibly $\mu_{j,k}\ne 0$).

Using the bound from the proof of Theorem 1 for (e) = (h), we obtain

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

We now bound $2\,E\left(\hat{\mu}_{j,k}^{(s)}-\hat{\mu}_{j,k}^{(h)}\right)^2$. Using the representation $\hat{\mu}_{j,k}^{(h)} = s_{j,k}f_{j,k}I(|f_{j,k}|>t_j)$, it is easy to see by direct comparison that $|\hat{\mu}_{j,k}^{(s)}-\hat{\mu}_{j,k}^{(h)}| \le s_{j,k}t_j$, which, using the explicit definition of $s_{j,k}$, the formula for t_j, and the bounds for $w^2_{i,j,k}$ and $\mu^2_{i,j,k}$, leads to

$$\begin{aligned} 2\,E\left(\hat{\mu}^{(s)}_{j,k} - \hat{\mu}^{(h)}_{j,k}\right)^2 &\leq 2\,t_j^2 E\left(s_{j,k}\right)^2 = 2\,t_j^2\left\{w^2_{1,j,k} + w^2_{2,j,k} + \left(\mu_{1,j,k} + \mu_{2,j,k}\right)^2\right\} \\ &\leq 8\,\log N \sup_z \sigma^4(z)\left(v2^{j-J} + 1\right) \leq 8\,\log N \sup_z \sigma^4(z)\,(v+1)\,. \end{aligned}$$

This yields $E(\hat{\mu}^{(s)}_{j,k} - \mu_{j,k})^2 \leq 8 \sup_z \sigma^4(z)\,\{(v+2)\log N + v\}$. The remaining part of the proof is completely analogous to the last part of the proof of Theorem 1 for (e) = (h), leading to the same rate. We omit the details. This completes the proof of Theorem 1 for (e) = (s). □

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

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

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

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

which is the desired density. This completes the proof of Lemma 1. □

Acknowledgement

Piotr Fryzlewicz thanks Theofanis Sapatinas and David Hand for financial support and Theofanis Sapatinas for his hospitality during a visit to Nicosia. Piotr Fryzlewicz is also grateful to the Department of Mathematics, Imperial College London, U.K. (where most of the work for the current version of this paper was done) for the excellent research environment. Theofanis Sapatinas was partly supported by a 'University of Cyprus – 2004 Research Grant'. Suhasini Subba Rao was supported by DFG (DA 187/12-2) grant. We thank the editor, an associate editor, a referee, Guy Nason, and Rainer von Sachs for their helpful comments.

References

Antoniadis, A. & Fryzlewicz, P. (2006). Parametric modelling of thresholds across scales in wavelet regression. *Biometrika* **93** (to appear).

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

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

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

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

Dahlhaus, R. & Subba Rao, S. (2006). Statistical inference for time-varying ARCH processes. *Annals of Statistics* **34** (to appear).

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.

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

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

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

Fryzlewicz, P. (2005). Modelling and forecasting financial log-returns as locally stationary wavelet processes. *Journal of Applied Statistics* **32**, 503–28.

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

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

Giraitis, L., Leipus, R. & Surgailis, D. (2005). Recent advances in ARCH modelling. In *Long memory in Economics*, Ed. A Kirman and G. Teyssiere, pp. 3–39, Berlin: Springer.

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

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

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

Mercurio, D. & Spokoiny, V. (2004). Statistical inference for time-inhomogeneous volatility models. *Annals of Statistics* **32**, 577–602.

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.

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

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.

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

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

Polzehl, J. & Spokoiny, V. (2004). Varying coefficient GARCH versus local constant volatility modelling. Comparison of the predictive power. *WIAS Preprint 997.*

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

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

Stărică, C. & Granger, C.W.J. (2005). Non-stationarities in stock returns. *The Review of Economics and Statistics* **87**, 503–22.

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

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* Oxford: Oxford University Press.

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

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

Wickerhauser, M.V. (1994). *Adapted Wavelet Analysis from Theory to Software.* Wellesley, MA: AK Peters, Ltd.