Measuring and forecasting financial variability using realised variance with and without a model

OLE E. BARNDORFF-NIELSEN
The Centre for Mathematical Physics and Stochastics (MaPhySto),
University of Aarhus, Ny Munkegade, DK-8000 Aarhus C, Denmark.
oebn@mi.aau.dk

BENT NIELSEN
Nuffield College, University of Oxford, Oxford OX1 1NF, U.K.
bent.nielsen@nuf.ox.ac.uk

NEIL SHEPHARD
Nuffield College, University of Oxford, Oxford OX1 1NF, U.K.
neil.shephard@nuf.ox.ac.uk

CARLA YSUSI
Department of Statistics, University of Oxford,
South Parks Road, Oxford OX1 3TG, U.K.
ysusi@stats.ox.ac.uk

First draft, August. 2002
This draft 7th October, 2002
www.levyprocess.org
Prepared for a volume on "State space and unobserved components models" edited by Andrew C. Harvey, Siem Jan Koopman and Neil Shephard (festschrift for Jim Durbin, F.B.A.). To be published by Cambridge University Press.

Abstract

We use high frequency financial data to proxy, via the realised variance, each day's financial variability. Based on a semiparametric stochastic volatility process, a limit theory shows you can represent the proxy as a true underlying variability plus some measurement noise with known characteristics. Hence filtering, smoothing and forecasting ideas can be used to improve our estimates of variability by exploiting the time series structure of the realised variances. This can be carried out based on a model or without a model. A comparison is made between these two methods.

Keywords: Kalman filter; Mixed Gaussian limit; OU process; Quadratic variation; Realised variance; Realised volatility; Square root process; Stochastic volatility.

1 Introduction

Neil Shephard was fortunate to have Jim Durbin as his supervisor and time series teacher during his first year of graduate studies at the London School of Economics in 1986-87. It was just before Jim retired. Jim was very interested in state space models, having recently written the Harvey and Durbin (1986) influential seat-belt case study on structural time series models. He sent Shephard off to read Kalman (1960) as an interesting place to start research. It was the first research paper Shephard read. Jim thought there was still a considerable amount to be carried through in this area.

Ole Barndorff-Nielsen's main contact to the research work of Jim Durbin has been to his pathbreaking paper Durbin (1980). Together with the papers by Cox (1980) and Hinkley (1980), this was of key import for the discovery of the general form of the p^*-formula for the law of the maximum likelihood estimator and hence the development of the theory that has flown from that formula (see Barndorff-Nielsen and Cox (1994) and the survey paper by Skovgaard (2001)).

Jim's research has had a profound impact on statistics and econometrics. From modelling, estimating and testing time series models to instrumental variables and general estimating equations, through to modern distribution theory, his work has been characterised by energy and inventiveness. He has an original mind. His teaching at the LSE had a profound impact on the course of British econometrics for, with Denis Sargan, he revolutionised the technical standards expected of their students. The current high position of British econometrics is a legacy we largely owe to Denis and Jim.

This paper touches on a number of Jim's interests. It uses continuous time methods, discusses some asymptotic distributional theory and eventually builds towards what might be called a structural time series model.

We use high frequency financial data to proxy each day's financial variability. A limit theory shows you can represent the proxy as a true underlying variability plus some measurement noise with known characteristics. Hence time series filtering, smoothing and forecasting ideas can be used to improve our estimates of variability by exploiting the time series structure of the data. This can be carried out based on a model, which is a particular type of continuous time structural time series model, or without a model. A comparison is made between these two methods.

In Section 2 we review the asymptotic distribution theory of realised variance, linking it to stochastic volatility and quadratic variation. Section 3 uses the distribution theory to derive an optimal filtering, smoothing and forecasting method for integrated variance. We show that this can be implemented in a model free way or based on a parametric model. In Section 4 we discuss how to operationalise the model free approach, while Section 5 discusses the corresponding model based approach. In Section 6 we draw our conclusions. The Appendix contains a proof of a theorem we state in Section 3.

2 Every day is different: historical measures of variability

2.1 The continuous time framework

This paper looks at measuring and forecasting the level of variability of asset prices in a financial market. This theory assumes a flexible stochastic volatility (SV) model for log-prices y^* which follow

$$y^*(t) = \alpha^*(t) + \int_0^t \tau^{1/2}(u)\mathrm{d}w(u), \quad t \geq 0, \tag{1}$$

where the processes $\tau^{1/2}$ and α^* is assumed to be stochastically independent of the standard Brownian motion w. We call $\tau^{1/2}$ the *instantaneous* or *spot volatility*, τ the corresponding *variance* and α^* the *mean* process. A simple example of this is

$$\alpha^*(t) = \mu t + \beta\tau^*(t), \quad \text{where} \quad \tau^*(t) = \int_0^t \tau(u)\mathrm{d}u.$$

The process τ^* is called the *integrated variance.* Throughout we will assume the following conditions hold with probability one:

(C) $\tau > 0$ is càdlàg on $[0, \infty)$, τ^* exists and α^* has the property

$$\hbar^{-3/4} \max_{1 \leq j \leq M} |\alpha^*(j\hbar) - \alpha^*((j-1)\hbar)| = o(1), \tag{2}$$

in $\hbar > 0$ for M a positive integer.

Condition (**C**) implies that the α^* process is continuous and so is predictable, while

$$m(t) = \int_0^t \tau^{1/2}(u)\mathrm{d}w(u),$$

is a continuous local martingale. Hence y^* is a rather flexible continuous semimartingale. Assumption (**C**) also allows the volatility to have, for example, deterministic diurnal effects, jumps, long memory, no unconditional mean or be non-stationary.

Over an interval of time of length $\hbar > 0$, which is here representing a day, returns on the i-th day are defined as

$$y_i = y^*(\hbar i) - y^*\{(i-1)\hbar\}, \qquad i = 1, 2, ..., T, \tag{3}$$

which implies that

$$y_i|\alpha_i, \tau_i \sim N(\alpha_i, \tau_i) \quad \text{where} \quad \alpha_i = \alpha^*(i\hbar) - \alpha^*\{(i-1)\hbar\},$$

while

$$\tau_i = \tau^*(i\hbar) - \tau^*\left\{(i-1)\,\hbar\right\}.$$

Here τ_i is called *actual variance* and α_i is the *actual mean*. Reviews of the literature on the SV topic are given in Taylor (1994), Shephard (1996) and Ghysels, Harvey, and Renault (1996), while statistical and probabilistic aspects are studied in detail in Barndorff-Nielsen and Shephard (2001).

The focus of this paper will eventually be on filtering, smoothing and forecasting τ_i. For shorthand, we call filtering and smoothing "measuring."

2.2 Realised variance

Our econometric approach is motivated by the advent of complete records of quotes or transaction prices for many financial assets. Theoretical and empirical work suggests that the use of such high frequency data is both informative and simplifying for it brings us closer to the theoretical models based on continuous time. However, market microstructure effects (e.g. discreteness of prices, bid/ask bounce, irregular trading etc.) means that there is a mismatch between asset pricing theory based on semimartingales and the data at very fine time intervals. This means that we cannot simply rely on empirical computations based on literally infinitesimal returns, instead we need a distribution theory for these estimators. This theory will reflect the fact that we will use a large but not infinite number of high frequency returns in our empirical work, informing us of the difference between the empirical reality and the theoretical limit of using returns over tiny time intervals.

We suppose there are M intra-$\hbar$ observations during each $\hbar > 0$ time period and that log-price of an asset is written as y^*. Our approach is to think of M as large and increasing. It will drive our limiting theory. Then high frequency observations will be defined as

$$y_{j,i} = y^*\left((i-1)\,\hbar + \frac{\hbar j}{M}\right) - y^*\left((i-1)\,\hbar + \frac{\hbar\,(j-1)}{M}\right), \tag{4}$$

the j-th intra-$\hbar$ return for the i-th period (e.g. if $\hbar$ is a day, $M = 288$, then this is the j-th 5 minute return on the i-th day). This is illustrated in Figure 1 which displays $y^*(t)$ at five minute intervals for the first five days of the Olsen Dollar/DM series. It starts on 1st December 1986 and ignores weekend breaks. This series is constructed every five minutes by the Olsen group from bid and ask quotes which appeared on the Reuters screen (see Dacorogna, Gencay, Muller, Olsen, and Pictet (2001) for details). We have set it up so that $y^*(0) = 0$. Figure 1(b) displays the returns when $M = 1$, which correspond to daily price movements. (c) uses $M = 8$ and shows three hour returns. We can see that the typical variability of each of these higher

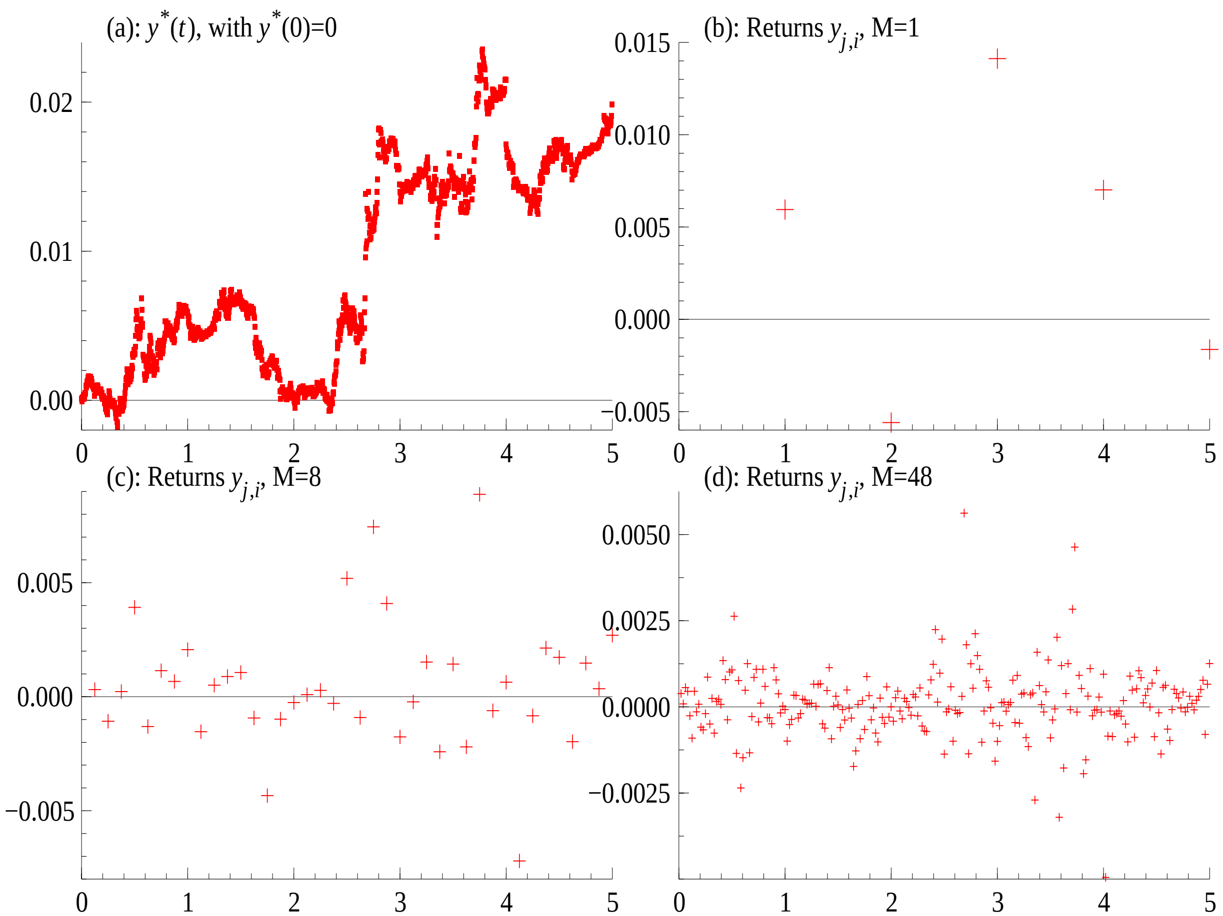

Figure 1: *Log-price and returns at different frequencies for the first five days of the Olsen data. (a): Log price $y^*(t)$ plotted every five minutes with $y^*(0) = 0$. (b): daily returns with $M = 1$. (c) three hour returns with $M = 8$. (d) thirty minute returns with $M = 48$. Code:* `basic_realised.ox`.

frequency observations is smaller than the daily returns. Finally (d) displays the case where $M = 48$, where we are using thirty minute returns.

The basis of our paper is to first work through the historical summery of variability, which can be thought of as estimators of past actual volatility τ_i. These are built using the M intra-$\hbar$ observations. The focus is on the *realised variance*[1]

$$[y^*_M]_i = \sum_{j=1}^{M} y^2_{j,i}. \tag{5}$$

[1]Sums of squared returns are often called realised volatility in econometrics, while we use the name realised variance for that term and realised volatility for the corresponding square root. The use of volatility to denote standard deviations rather than variances is standard in financial economics. See, for example, the literature on volatility and variance swaps, which are derivatives written on realised volatility or variance, which includes Demeterfi, Derman, Kamal, and Zou (1999), Howison, Rafailidis, and Rasmussen (2000) and Chriss and Morokoff (1999). We have chosen to follow this nomenclature rather than the one more familiar in econometrics. Confidence intervals for the realised volatility follow by square rooting the confidence intervals for the realised variance.

Notice this estimator is entirely self-contained, that is it only uses data from the i-th time period to estimate τ_i. Its cousin *realised volatility*

$$\sqrt{\sum_{j=1}^{M} y_{j,i}^2},$$

have been used in financial economics for many years by, for example, Poterba and Summers (1986), Schwert (1989), Taylor and Xu (1997), Christensen and Prabhala (1998), Andersen, Bollerslev, Diebold, and Labys (2001a) and Andersen, Bollerslev, Diebold, and Ebens (2001). However, until recently little theory was known about realised variance outside the Brownian motion case. See the incisive review by Andersen, Bollerslev, and Diebold (2002). Some other pieces on this work we would like to highlight are Meddahi (2002), Andersen, Bollerslev, and Meddahi (2002) and Andreou and Ghysels (2002), although many other interesting papers exist which are discussed by Andersen, Bollerslev, and Diebold (2002).

2.3 Properties of realised variance

It is very well known that the theory of quadratic variation (e.g. Jacod and Shiryaev (1987, p. 55), Protter (1990) and Back (1991)) implies that

$$[y_M^*]_i \xrightarrow{p} \tau_i,$$

as $M \to \infty$. This does not depend upon the exact form of α^* or τ^2.

This consistency result is illustrated in Figure 2 which displays a simulated sample path of integrated variance τ_i from an OU process given by the solution to

$$\mathrm{d}\tau(t) = -\lambda\tau(t)\mathrm{d}t + \mathrm{d}z(\lambda t),$$

where z is a subordinator (a process with independent, stationary and non-negative increments). In this example we construct the process so that $\tau(t)$ has a $\Gamma(4, 8)$ stationary distribution, $\lambda = -\log(0.99)$ and $\hbar = 1$. Also drawn are the sample path of the realised variances $\sum_{j=1}^{M} y_{j,i}^2$ (depicted using crosses) where

$$y^*(t) = \beta\tau^*(t) + \int_0^t \tau^{1/2}(u)\mathrm{d}w(u),$$

and $\beta = 0.5$. The realised variances are computed using a variety of values of M. We see that as M increases the size of $\sum_{j=1}^{M} y_{j,i}^2 - \tau_i$ falls, illustrating the consistency of $\sum_{j=1}^{M} y_{j,i}^2$ for τ_i even though β is not zero.

[2] Indeed the probability limit of realised variance is known under the even weaker assumptions that the price process is a semimartingale.

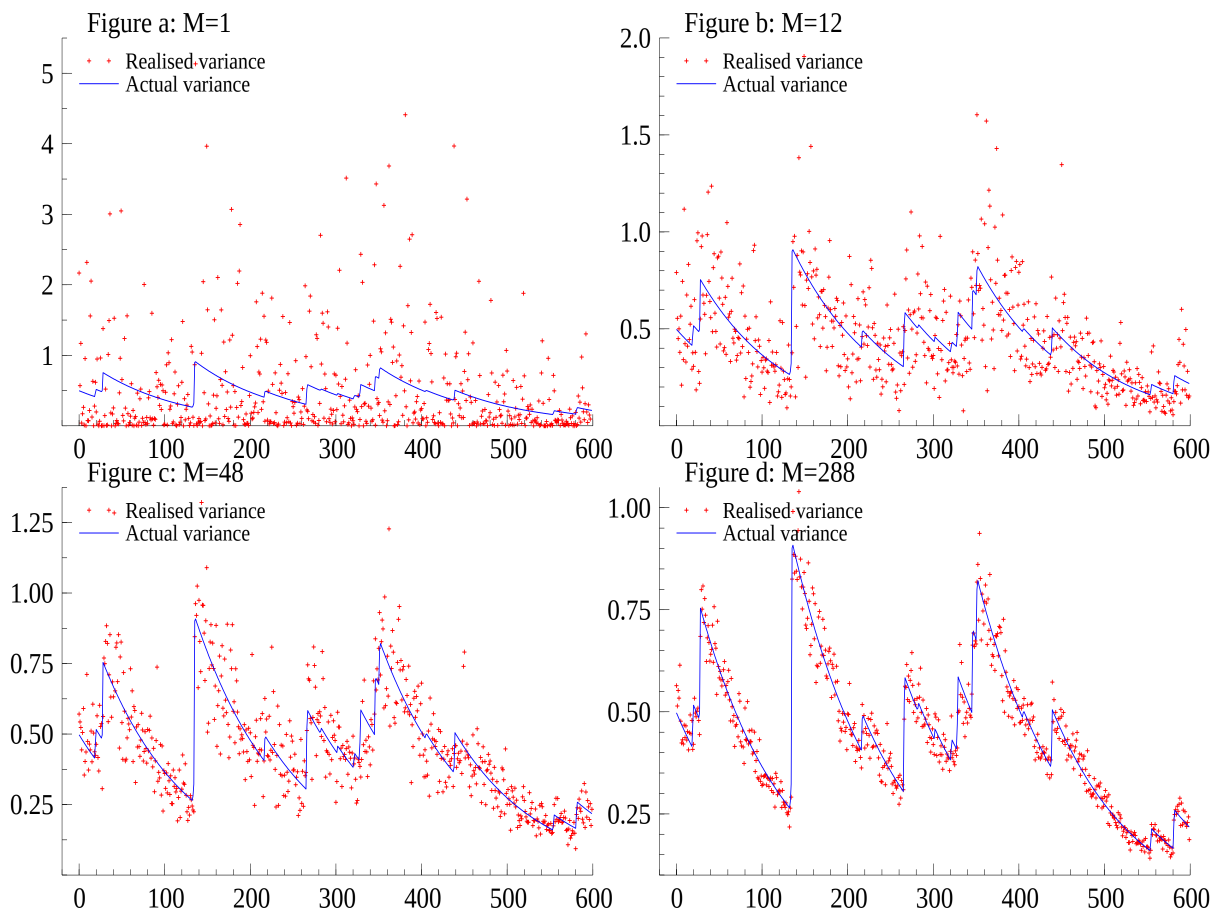

Figure 2: *Actual* τ_i *and realised* $\sum_{j=1}^{M} y_{j,i}^2$ *(with* M varying*) volatility based upon a* $\Gamma(4,8)$-OU process with $\lambda = -\log(0.99)$ and $\hbar = 1$. This implies $\xi = 0.5$ and $\xi\omega^{-2} = 8$. Code: `/code/realised/simple.ox`.

In a recent paper Barndorff-Nielsen and Shephard (2002a), consequently extended in Barndorff-Nielsen and Shephard (2003) and Barndorff-Nielsen and Shephard (2002b), have strengthened the above result considerably. The main result is that:

Theorem 1 *Under assumption* (**C**) *for the SV model in (1), for any positive* $\hbar$ *and* $M \to \infty$

$$\frac{\sqrt{\frac{M}{\hbar}}\left([y_M^*]_i - \tau_i\right)}{\sqrt{2\tau_i^{[2]}}} \overset{d}{\to} N(0,1), \quad \textit{where} \quad \tau_i^{[2]} = \int_{(i-1)\hbar}^{i\hbar} \tau^2(u)\mathrm{d}u. \tag{6}$$

□

We call τ^2 and $\tau_i^{[2]}$ the *spot* and *actual quarticity*, respectively. Of course the problem with this theory is that $\tau_i^{[2]}$ is unknown. This is tackled by using the fact that

$$\frac{M}{\hbar}\sum_{j=1}^{M} y_{j,i}^4 \overset{p}{\to} 3\tau_i^{[2]}.$$

An implication of this is that we can use the feasible limit theory

$$\frac{[y_M^*]_i - \tau_i}{\sqrt{\frac{2}{3}\sum_{j=1}^M y_{j,i}^4}} \xrightarrow{\mathcal{L}} N(0,1), \tag{7}$$

due to Barndorff-Nielsen and Shephard (2002a).

Of course in practice it may make sense to transform the above limit theorem to impose, a priori, positivity on the approximating distribution. In particular it seems natural to work with the logarithmic transformation of the realised variance ratio so that (see Barndorff-Nielsen and Shephard (2002c))

$$\frac{\sqrt{\frac{M}{\hbar}}\left\{\log [y_M^*]_i - \log \tau_i\right\}}{\sqrt{2\tau_i^{[2]}/\left(\tau_i\right)^2}} \xrightarrow{\mathcal{L}} N(0,1) \quad \text{or} \quad \frac{\log [y_M^*]_i - \log \tau_i}{\sqrt{\frac{2}{3[y_M^*]_i^2}\sum_{j=1}^M y_{j,i}^4}} \xrightarrow{\mathcal{L}} N(0,1).$$

The following remarks can be made about these results.

- $\sum_{j=1}^M y_{j,i}^2$ converges to $\int_{\hbar(i-1)}^{\hbar i} \tau(u)\mathrm{d}u$ at rate $\sqrt{M}$.
- The limit theorem is unaffected by the form of the drift process α, smoothness assumption (**C**) is sufficient that its effect becomes negligible.
- Knowledge of the form of the volatility dynamics is not required in order to use this theory.
- The fourth moment of returns need not exist for the asymptotic normality to hold. In such heavy tailed situations, the stochastic denominator $\int_{(i-1)\hbar}^{i\hbar} \tau^2(u)\mathrm{d}u$ loses its unconditional mean.
- The volatility process τ can be non-stationary, exhibit long-memory or include intra-day effects.
- $\sum_{j=1}^M y_{j,i}^2 - \int_{\hbar(i-1)}^{\hbar i} \tau(u)\mathrm{d}u$ has a mixed Gaussian limit implying that marginally it will have heavier tails than a normal.
- The magnitude of the error $\sum_{j=1}^M y_{j,i}^2 - \int_{\hbar(i-1)}^{\hbar i} \tau(u)\mathrm{d}u$ is likely to be large in times of high volatility.

3 Time series of realised variances

3.1 Motivation

So far we have analysed the asymptotics of $\sum_{j=1}^M y_{j,i}^2$ as $M \to \infty$ for a single i. In this section we will explicitly analyse a long time series of realised variances, trying to use the time series

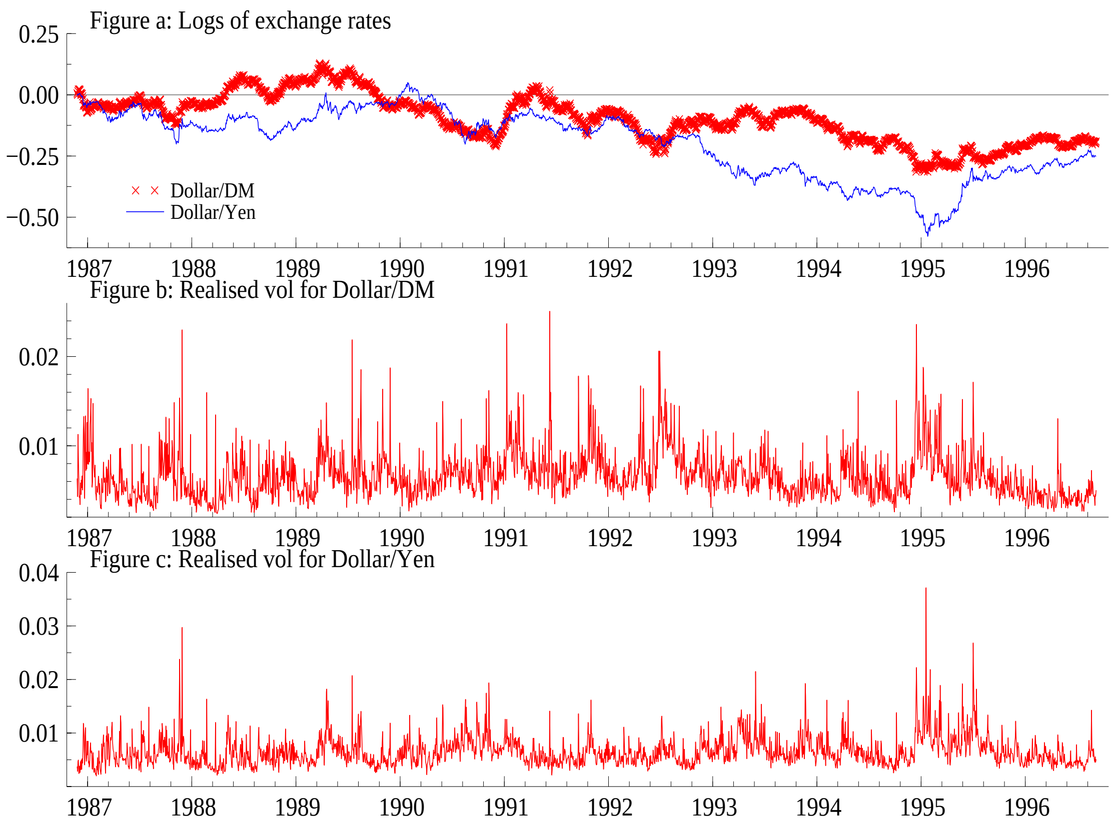

Figure 3: *Long time series of the daily movements in the Dollar against the DM and Yen. Figure (a) the level of the log exchange rates compared to the rate at 1st December 1986. Figure (b) realised volatility each day computed using* $M = 144$ *for the DM series. Figure (c) realised volatility each day computed using* $M = 144$ *for the Yen series. File:* `daily_realised.ox`.

structure to construct more efficient estimators and forecasts of τ_i. To start out we have drawn Figure 3 which displays information on the Olsen data on the DM and Yen against the US Dollar. Figure 3(a) shows the movement of the log prices since 1st December 1986, with the log-prices transformed to be zero at the start of the sample. This is the same series as Figure 1(a) but now the graph is on a very long time scale. Figure 3(b) shows the daily realised volatility $\sqrt{\sum_{j=1}^{M} y_{j,i}^2}$ drawn against i, the day, for the DM series. It is computed using $M = 144$, corresponding to 10 minute returns. It is quite a ragged series but with periods of increased volatility. A similar picture emerges from the corresponding realised volatility for the Yen given in Figure 3(c).

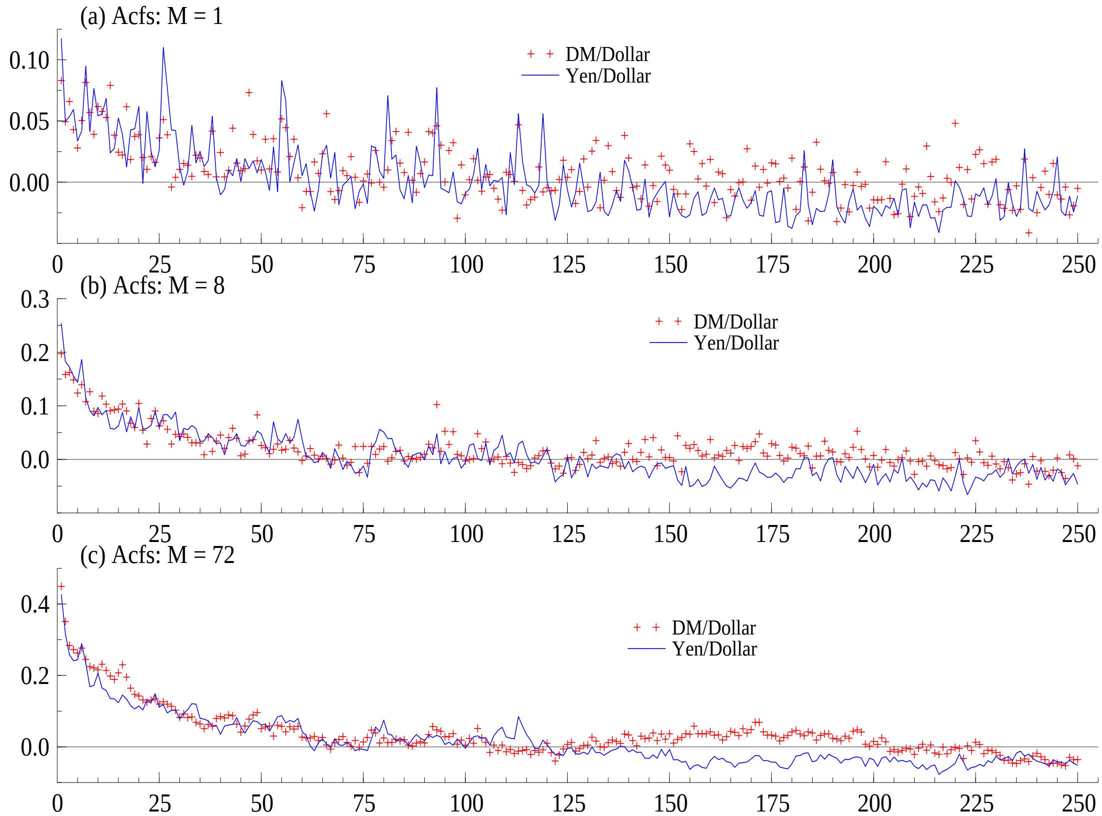

Figure 4: *Autocorrelations of realised variances using a long time series of the movements in the Dollar against the DM and Yen. Figure (a) M=1 case, which corresponds to daily returns. Figure (b)* $M = 8$ *case. Figure (c)* $M = 72$ *for the Yen series. File:* `daily_timeseries.ox`.

3.2 Asymptotics

For each exchange rate we have computed realised variances each day. We can then regard the derived series as a daily time series

$$\sum_{j=1}^{M} y_{j,1}^2, \sum_{j=1}^{M} y_{j,2}^2, ..., \sum_{j=1}^{M} y_{j,T}^2.$$

This new series is of length T, the number of days in the sample.

The correlograms for the daily time series of realised volatilities of these quantities are displayed in Figure 4 for a variety of values of M. 250 lags are used in these figures which correspond to measuring correlations over a one year period. Figure 4(a) shows the results for $M = 1$. In this case the realised variances are simply squared daily returns. The correlogram has the well known slow decay but starting at quite a low level. Figure 4(b) shows the effect of increasing M slightly to 8, now we are computing the realised quantities using 150 minute returns.

Figure 4(c) shows the corresponding results for $M = 72$, which uses 20 minute returns. All the autocorrelations are boosted as M increases from 8, however the broad story is the same. A clear observation is that the autocorrelations are becoming less jagged with the increase in M.

Having observed some of the empirical features of the realised variances we will now set out a theoretical framework for the study of the time series of realised quantities. For the moment we focus on the realised variances.

We define sequences of realised and actual variances for the s-th day to the p-th day

$$[y_M^*]_{s:p} = \left(\sum_{j=1}^{M} y_{j,s}^2, \sum_{j=1}^{M} y_{j,s+1}^2, ..., \sum_{j=1}^{M} y_{j,p}^2 \right)' \qquad \text{and} \qquad \tau_{s:p} = (\tau_s, \tau_{s+1}, ..., \tau_p)',$$

where we recall that $\tau_i = \int_{\hbar(i-1)}^{\hbar i} \tau(u)\mathrm{d}u$. The asymptotic theory of realised variance implies that

$$\sqrt{\frac{M}{\hbar}} \left([y_M^*]_{s:p} - \tau_{s:p} \right) \xrightarrow{\mathcal{L}} N\left\{0, 2diag\left(\tau_{s:p}^{[2]}\right)\right\},$$

where $\tau_i^{[2]} = \int_{\hbar(i-1)}^{\hbar i} \tau^2(u)\mathrm{d}u$.

3.3 Linear estimators

Although estimating $\tau_{s:p}$ by $[y_M^*]_{s:p}$ has attractions, the variance of the error is typically quite large even when M is high. More precise estimators could be obtained by pooling neighbouring time series observations for realised variances tend to be highly correlated through time. This pooling will typically reduce the variance of the estimator, but will induce a bias.

To set up a formal framework for this discussion, abstractly write A as a matrix of non-stochastic weights. Then

$$\sqrt{\frac{M}{\hbar}} \left(A[y_M^*]_{s:p} - A\tau_{s:p} \right) |\tau_{[\nu]}^{[2]} \xrightarrow{\mathcal{L}} N\left\{0, 2Adiag\left(\tau_{s:p}^{[2]}\right)A'\right\}.$$

Now consider the statistic

$$\widehat{\tau_{s:p}} = c\mathrm{E}\left(\tau_{s:p}\right) + A[y_M^*]_{s:p}.$$

We assume that the realised variances constitute a covariance stationary process, which means that

$$\mathrm{E}\left(\tau_{s:p}\right) = \iota \mathrm{E}\left(\tau_t\right),$$

where $\iota = (1, 1, ..., 1)'$. Notice the stationarity is at the daily level, it does not need that the continuous time process τ is stationary.

The population weighted least squares estimator of $\tau_{s:p}$ sets

$$c = (I - A)\,\iota$$

and

$$
\begin{aligned}
A &= \text{Cov}\left(\tau_{s:p}, [y_M^*]_{s:p}\right)\left[\text{Cov}\left([y_M^*]_{s:p}\right)\right]^{-1} \\
&= \text{Cov}\left(\tau_{s:p}\right)\left[\text{Cov}\left([y_M^*]_{s:p}\right)\right]^{-1} \\
&= \text{Cov}\left(\tau_{s:p}\right)\left[\text{Cov}\left(\tau_{s:p}\right) + \frac{2\hbar \text{E}\left(\tau_i^{[2]}\right)}{M} I\right]^{-1}.
\end{aligned}
$$

Of course, as $M \to \infty$ so $A \to I$ and so

$$
\widehat{\tau_{s:p}} \xrightarrow{p} \tau_{s:p}.
$$

Notice that as $M \to \infty$ so $\widehat{A} \to I$ and $\widehat{\tau}_{s:p} \xrightarrow{p} \tau_{s:p}$. Unconditionally $\widehat{\tau_{s:p}}$ has a variance of

$$
(2\hbar/M)\,\text{E}\left(\tau_{s:p}^{[2]}\right) AA' + (I - A)\,\text{Cov}\left(\tau_{s:p}\right)(I - A')
$$

At the end of this Section we will study conditions under which A is guaranteed to be non-negative.

3.4 Implementation

In practice A has to be estimated from the data. Broadly this can be carried out in two ways

1. by estimating A by using empirical averages from the data,
2. implying A from an estimated parametric model.

3.5 Positivity

Before going on to discuss the above issues of implementation issues we will take a moment to give conditions under which all the elements of

$$
A = \text{Cov}\left(\tau_{s:p}\right)\left[\text{Cov}\left(\tau_{s:p}\right) + \frac{2\hbar \text{E}\left(\tau_i^{[2]}\right)}{M} I\right]^{-1}
$$

are non-negative. Such matrices are said to be totally non-negative. The following example shows that A is not necessarily totally non-negative.

Example 2 *Suppose, $|a| < 1$ and we write $u_i = [y_M^*]_i - \tau_i$. Then*

$$
\text{Cov}(\tau_{s:s+1}) = \begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix}, \quad \text{Cov}([y_M^*]_{s:s+1}) = \text{Cov}(\tau_{s:s+1}) + I\text{Var}(u_i),
$$

and so

$$
A = \frac{1}{\{\text{Var}(u_i) + 1\}^2 - a^2}\begin{pmatrix} \text{Var}(u_i) + 1 - a^2 & a\text{Var}(u_i) \\ a\text{Var}(u_i) & \text{Var}(u_i) + 1 - a^2 \end{pmatrix}.
$$

Hence all weights are non-negative iff $a \geq 0$.

The next theorem gives conditions on $\mathrm{Cov}(\tau_{s:p})$ to ensure total non-negativity of A.

Theorem 3 *Assume that* $\mathrm{Cov}(\tau_{s:p})$ *is positive definite. Then the necessary and sufficient condition for all the elements of* A *to be non-negative is that* $\mathrm{Cov}(\tau_{s:p})^{-1}$ *has non-positive off-diagonal elements.*

Proof. Given in the Appendix.

The condition that $\mathrm{Cov}(\tau_{s:p})^{-1}$ has to have non-positive off-diagonal elements has the following straightforward statistical interpretation.

Remark 1 *Suppose* X *is a positive definite covariance matrix. We write the* i,j *element of* X^{-1} *as* $x^{i,j}$*. Then*

$$\frac{-x^{i,j}}{\sqrt{x^{i,i}x^{j,j}}},$$

is the partial correlations between y_i *and* y_j*. That is it is the ordinary correlation between* y_i *and* y_j *conditioning on all the other elements of* y *(see, for example, Cox and Wermuth (1996, p. 69)).*

4 Model free approach

Here we will discuss estimating A by using empirical averages from the data, delaying until the next section a discussion of a model based method.

If we have a large sample from a stationary process of realised variances and the daily process is ergodic then we have that

$$\widehat{\mathrm{E}\left(\tau_i^{[2]}\right)} = \left(\frac{1}{T}\sum_{i=1}^{T}\frac{M}{3\hbar}\sum_{j=1}^{M}y_{j,i}^4\right) \xrightarrow{p} \mathrm{E}\left(\tau_i^{[2]}\right),$$

as T and M go to infinity. Likewise $\mathrm{Cov}\left([y_M^*]_{s:p}\right)$ can be estimated by averages of the time series of realised variances. Hence A can be replaced by

$$\widehat{A} = \left\{\widehat{\mathrm{Cov}\left([y_M^*]_{s:p}\right)} - \widehat{\mathrm{E}\left(\tau_i^{[2]}\right)}\frac{2\hbar}{M}I\right\}\left[\widehat{\mathrm{Cov}\left([y_M^*]_{s:p}\right)}\right]^{-1},$$

which is a feasible weighting matrix. This will imply $\widehat{c} = \left(I - \widehat{A}\right)\iota$ and

$$\widehat{\tau_{s:p}} = \widehat{c}\widehat{\mathrm{E}\left(\tau_{s:p}\right)} + \widehat{A}[y_M^*]_{s:p}.$$

This is a feasible model free, optimal linear estimator of $\tau_{s:p}$ based on $[y_M^*]_{s:p}$.

	DM		Yen	
M	$\widehat{A}$	$\widehat{c}$	$\widehat{A}$	$\widehat{c}$
1	.182	.817	.229	.770
8	.449	.550	.513	.486
72	.778	.221	.789	.210
288	.877	.122	.906	.093

Table 1: *Estimated weights for $\widehat{\tau}_i$, the regression estimator of τ_i which uses only $[y_M^*]_i$ and an intercept. Results for the DM and Yen series against the Dollar. File:* `daily_timeseries.ox`.

4.1 Illustration

Table 1 contains the estimated weights for a single actual variance using a single realised variance sequence, so $s = p = i$, for the DM and Yen series. This is based on the entire time series sample of nearly 2500 days.

We can see the results do not vary very much with the series being used. In particular, for $M = 8$ then the estimator of τ_i for the DM series would be

$$\widehat{\tau}_i = .550\frac{1}{T}\sum_{j=1}^{T}[y_M^*]_j + .449[y_M^*]_i. \tag{8}$$

Thus for small values of M the regression estimator puts a moderate weight on the realised variance and more on the unconditional mean of the variances. As M increases this situation reverses, but even for large values of M the unconditional mean is still quite highly weighted. From now on we will solely focus on the DM series to make the exposition more compact.

In the dynamic case the results are more complicated to present. Here we start by considering estimating three actual variances using three contiguous realised variances — one lag, one lead and the contemporaneous realised variance. Thus

$$s : p = (i-1, i, i+1),$$

and so $\widehat{A}$ will be a 3×3 matrix and $\widehat{c}$ a 3×1 vector. In the case of $M = 8$ we have that

$$\{\text{Cov}([y_M^*]_{1:3})\}^{-1} = \frac{1}{100^2}\begin{pmatrix} 2.07 & -.358 & -.258 \\ -.358 & 2.10 & -.358 \\ -.258 & -.358 & 2.07 \end{pmatrix},$$

while

$$\widehat{A} = \begin{pmatrix} .418 & .100 & .072 \\ .100 & .409 & .100 \\ .072 & .100 & .418 \end{pmatrix}, \quad \widehat{c} = \begin{pmatrix} .408 \\ .388 \\ .408 \end{pmatrix}.$$

Thus the second row of $\widehat{A}$ implies the smoothed estimator of τ_i is

$$\widehat{\tau}_i = .100[y_M^*]_{i-1} + .409[y_M^*]_i + .100[y_M^*]_{i+1} + .388\frac{1}{T}\sum_{j=1}^{T}[y_M^*]_j.$$

The corresponding result for $M = 72$ is

$$\widehat{A} = \begin{pmatrix} .712 & .105 & .053 \\ .105 & .684 & .105 \\ .053 & .105 & .712 \end{pmatrix}, \quad \widehat{c} = \begin{pmatrix} .128 \\ .105 \\ .128 \end{pmatrix}.$$

This shows that the weighting on the diagonal elements of $\widehat{A}$ are much higher, while the size of $\widehat{c}$ has fallen by a factor of around 4. In both cases a lot of weight is put on neighbouring values

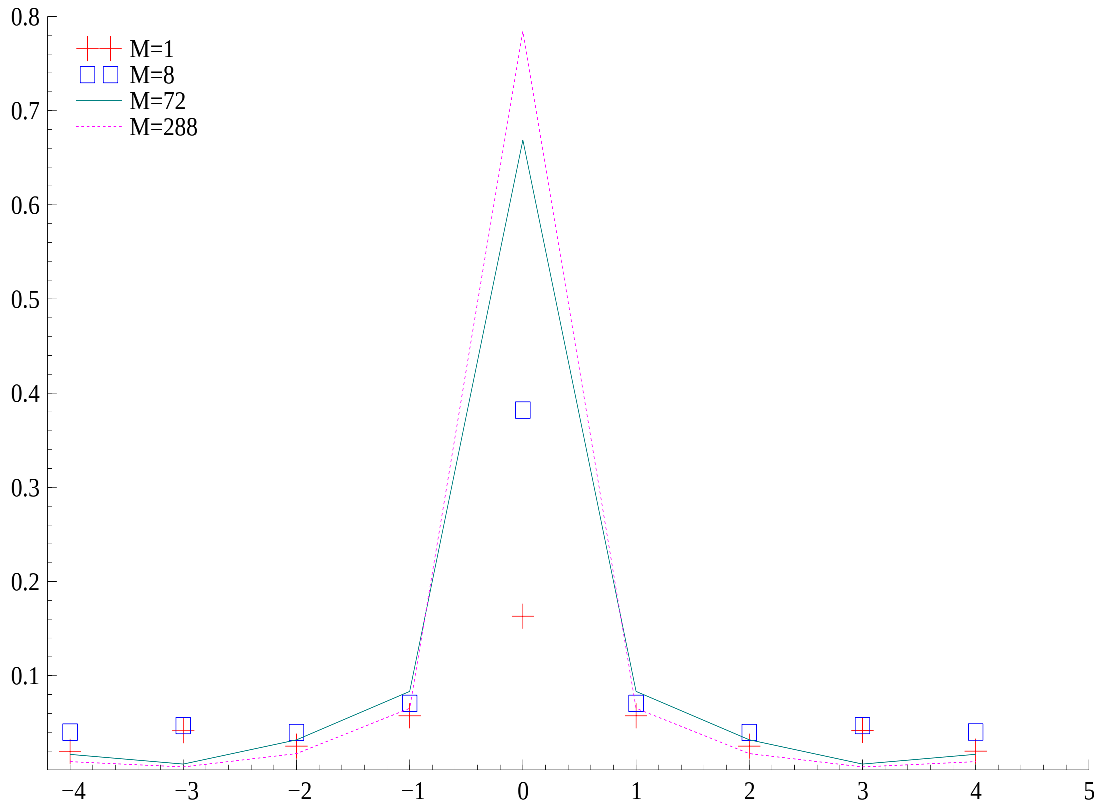

Figure 5: *Estimated weight vector for estimating* τ_i *using* $[y_M^*]_{i-4}, [y_M^*]_{i-3}, \ldots, [y_M^*]_{i+4}$ *drawn against lag length. Computed using the Dollar against the DM. Shows that as* M *increases the weight on* $[y_M^*]_i$ *increases. Corresponding to these results is* $\widehat{c}$*, which moves from* .548, .222, .0553, .026 *as* M *increases through* 1, 8, 72 *to* 288. *File:* `daily_timeseries.ox`.

of the realised variance and on the intercept, although the weight on $[y_M^*]_i$ is not very much smaller than in the univariate case.

The corresponding filtered estimator (which seems a natural competitor to using the raw realised variance $[y_M^*]_i$) is obtained by using the last row of the $\widehat{A}$ matrix. Then we have, for $M = 8$,

$$\widehat{\tau}_i = .072[y_M^*]_{i-2} + .100[y_M^*]_{i-1} + .418[y_M^*]_i + .408 \frac{1}{T} \sum_{j=1}^{T} [y_M^*]_j.$$

Here we see the usual decay in the weight as we go further back in time.

Figure 5 shows middle row of $\widehat{A}$ for the case of estimating τ_i using 9 realised variances, four lags and four leads together with $[y_M^*]_i$. It displays the weights as a function of M indicating how quickly the weights focus on $[y_M^*]_i$ as M increases. The legend of the Figure also gives the value of the weight put on the unconditional mean of the realised variance. For $M = 72$ it is .0553, which is much lower than in the trivariate case of .105 and univariate case of .221.

Figure 6 shows a time series of realised variances for a number of values of M together with the corresponding estimator $\widehat{\tau}_i$ based on nine observations, four leads, the current value and four lags. The smoothed estimator seems to deliver sensible answers, with the results being less sensitive to large values of the realised variances, in particular for small M.

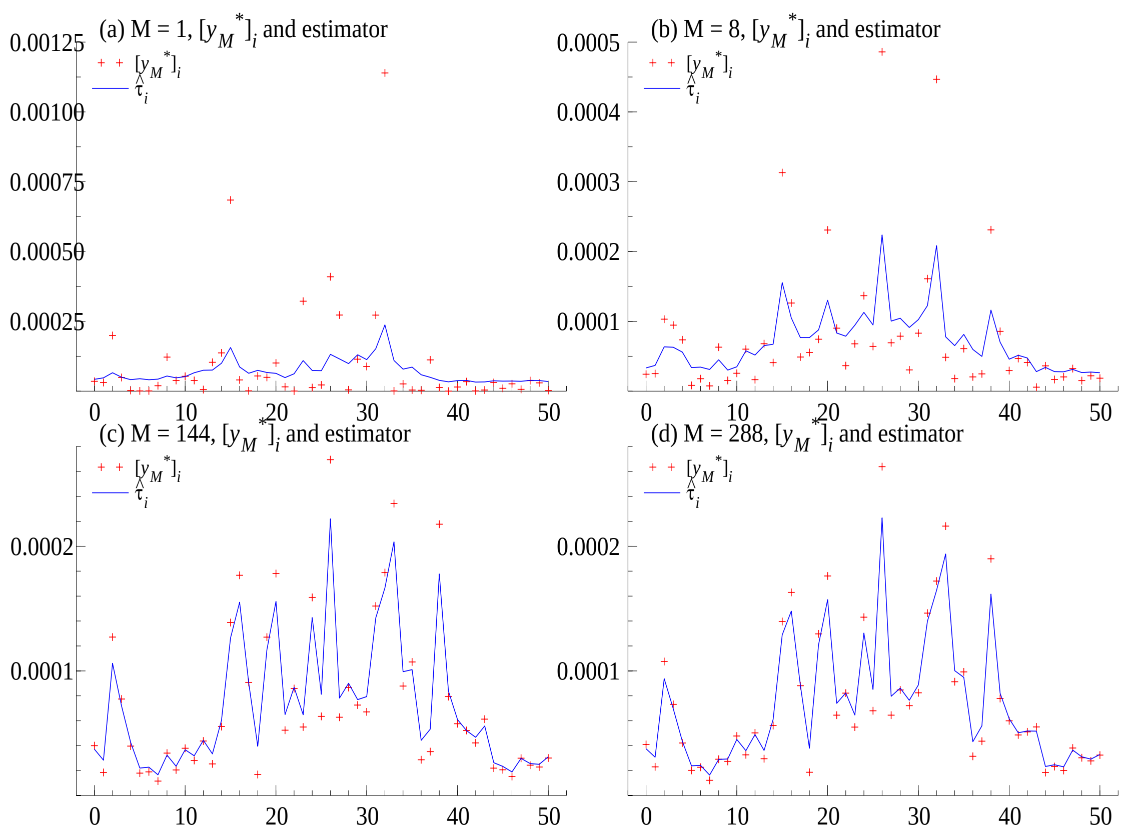

Figure 6: *The estimated* τ_i *using realised variance and weighted version of* $[y_M^*]_{i-4}$,$[y_M^*]_{i-3}$,...,$[y_M^*]_{i+4}$. *Computed using the Dollar against the DM. (a) M=1, (b) M=8, (c) M=144 and (d) M=288. File:* `daily_timeseries.ox`.

Table 2 reports, using the DM data,

$$\frac{1}{T}\sum_{i=1}^{T}\left([y_M^*]_i - [y_{288}^*]_i\right)^2,$$

which is an empirical approximation to the mean square error of the realised variance estimator, using $[y^*_{288}]_i$ as a good proxy for τ_i (the model based estimators would turn out to deliver even more accurate estimators, but this could be interpreted as biasing the results towards the model based approach and so here we use the raw realised variance). The Table shows a rapid decline in the mean square error with M. It also shows the corresponding results for the estimators based on just a regression on a constant and $[y^*_M]_i$, and $\widehat{\tau}_i$, which uses $[y^*_M]_{i-4}$, $[y^*_M]_{i-3}$, ...,$[y^*_M]_{i+3}$,$[y^*_M]_{i+4}$. The results reflect the fact that these adjusted estimators are much more efficient than the realised variance, although the difference between using the time series dynamics and the simple regression estimator is modest.

	DM			Yen		
	$[y^*_M]_i$	$(1-\widehat{A})E(\tau_i)+\widehat{A}[y^*_M]_i$	$\widehat{\tau}_i$	$[y^*_M]_i$	$(1-\widehat{A})E(\tau_i)+\widehat{A}[y^*_M]_i$	$\widehat{\tau}_i$
$M=1$	.822	.175	.145	1.16	.198	.168
$M=8$	.207	.0989	.0769	.186	.117	.0985
$M=72$	.0377	.0345	.0317	.0424	.0406	.0378

Table 2: *Mean square error of the realised variance and the regression estimator and the time series estimators $\widehat{\tau}_i$, which is based on $[y^*_M]_{i-4}$, $[y^*_M]_{i-3}$, ...,$[y^*_M]_{i+3}$,$[y^*_M]_{i+4}$. These are computed using M = 1, 8 and 72. The true value is taken as $[y^*_M]_i$ for 288. File:* `daily_timeseries.ox`.

4.2 Forecasting

Suppose we are interesting in forecasting τ_{p+1} based on the time series of realised variances $[y^*_M]_{s:p}$. Throughout we assume that the integrated and realised variances are second order stationary. The best linear forecast is given by

$$\widehat{\tau_{p+1|s:p}} = c\mathrm{E}(\tau_{p+1}) + A[y^*_M]_{s:p},$$

where

$$c = 1 - A\iota$$

and

$$\begin{aligned} A &= \mathrm{Cov}\left(\tau_{p+1}, [y^*_M]_{s:p}\right)\left[\mathrm{Cov}\left([y^*_M]_{s:p}\right)\right]^{-1} \\ &= \mathrm{Cov}\left([y^*_M]_{p+1}, [y^*_M]_{s:p}\right)\left[\mathrm{Cov}\left([y^*_M]_{s:p}\right)\right]^{-1}. \end{aligned}$$

This is a somewhat surprising result for A can be computed without reference to the details of the asymptotic theory of error. It just falls out from the asymptotic relationship between the realised variances, which can be empirically determined. Hence $\widehat{\tau_{p+1|s:p}}$ is feasible. However, as $M \to \infty$ this is not consistent. Instead

$$A \to \mathrm{Cov}\left(\tau_{p+1}, \tau_{s:p}\right)\left[\mathrm{Cov}\left(\tau_{s:p}\right)\right]^{-1},$$

and so

$$\widehat{\tau_{p+1|s:p}} \xrightarrow{p} c\mathrm{E}\left(\tau_{p+1}\right) + \left\{\mathrm{Cov}\left(\tau_{p+1}, \tau_{s:p}\right)\left[\mathrm{Cov}\left(\tau_{s:p}\right)\right]^{-1}\right\}\tau_{s:p}.$$

Extensions to multistep ahead predictions are straightforward. Importantly the above forecasting framework means that the one-step ahead predictions are generated by a $p-s+1$ order autoregression plus intercept model, where the intercept follows a particularly simple constraint so that the weights on the lagged coefficients plus the intercept add to one. Unconstrained autoregressive forecasting in the context of realised variances has been carried out by Andersen, Bollerslev, Diebold, and Labys (2001b).

The simplest interesting example of the above approach is where $s = p$. Then we are forecasting one-step ahead based on a single realised variance. This produces

$$\widehat{\tau_{p+1|p}} = \mathrm{E}\left(\tau_{p+1}\right) + \mathrm{Cor}\left([y_M^*]_{p+1}, [y_M^*]_p\right)\left\{[y_M^*]_p - \mathrm{E}\left(\tau_{p+1}\right)\right\}.$$

In practice we replace expectations by averages and correlations by empirical correlations. Table 3 provides empirical estimators of A and c for the DM and Yen series for a variety of values of M.

	DM		Yen	
M	$\widehat{A}$	$\widehat{c}$	$\widehat{A}$	$\widehat{c}$
1	.083	.917	.117	.883
8	.197	.803	.254	.746
72	.471	.529	.428	.572
288	.540	.460	.517	.483

Table 3: *Estimated weights for $\widehat{\tau_{p+1|p}}$, the regression estimator of τ_{p+1} which uses only $[y_M^*]_p$ and an intercept. Results for the DM and Yen series against the Dollar. File:* `daily_timeseries.ox`.

We can see again that the results do not vary very much with the series being used. In particular, for $M = 8$ then the estimator of τ_i for the DM series would be

$$\widehat{\tau_{p+1|p}} = .803\frac{1}{T}\sum_{j=1}^{T}[y_M^*]_j + .197[y_M^*]_p.$$

Thus the forecast shrinks much more to the mean than does the corresponding smoother given in (8).

Table 4 provides the weights when we use six lags of realised variances to forecast τ_{p+1}. It shows again that quite a lot of weight is placed on the constant c, while the most recent realised variance is also highly weighted. This results from the fact that the autocorrelation function of the realised variances initially declines very rapidly, followed by a slower decay rate at higher lag lengths.

M	$[y_M^*]_{p-5}$	$[y_M^*]_{p-4}$	$[y_M^*]_{p-3}$	$[y_M^*]_{p-2}$	$[y_M^*]_{p-1}$	$[y_M^*]_p$	$\widehat{c}$
1	.040	.014	.028	.053	.034	.073	.753
8	.074	.046	.074	.089	.083	.138	.493
144	.089	.067	.080	.031	.134	.321	.273
288	.050	.093	.051	.038	.111	.397	.256

Table 4: *Estimated weights for 1-step ahead forecast of integrated variance* τ_{p+1}. *File:* `daily_timeseries.ox`.

4.3 Log-based theory

A similar style of argument could have been used based on the log-realised variances. Here we will write

$$\log[y_M^*]_{s:p} = (\log[y_M^*]_s, ..., \log[y_M^*]_p)'$$

and

$$\log \tau_{s:p} = (\log \tau_s, ..., \log \tau_p)' .$$

The pooled estimator has the asymptotic distribution (see Barndorff-Nielsen and Shephard (2002c))

$$\sqrt{\frac{M}{\hbar}} \left(A \log[y_M^*]_{s:p} - A \log \tau_{s:p} \right) | \tau_{s:p}^{[2]}, \tau_{s:p} \xrightarrow{\mathcal{L}} N \left\{ 0, 2A\mathrm{E} \begin{pmatrix} \tau_s^{[2]} / \left(\tau_s^2\right) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \tau_p^{[2]} / \left(\tau_p^2\right) \end{pmatrix} A' \right\},$$

which would allow us to choose A as a least squares estimator of $\log \tau_{s:p}$ repeating the above argument. Weighting based on the log-realised variances has the advantage that the Monte Carlo evidence suggests that the asymptotics for the log-realised variance is accurate with the errors being approximately homoskedastic which suggests the weighting will be more effective.

The important result that we need to use is that

$$\left\{ \frac{1}{T} \sum_{i=1}^{T} \frac{\frac{M}{3\hbar} \sum_{j=1}^{M} y_{j,i}^4}{\left(\sum_{j=1}^{M} y_{j,i}^2 \right)^2} \right\} \xrightarrow{p} \mathrm{E} \left(\frac{\tau_i^{[2]}}{\tau_i^2} \right),$$

and hence

$$\widehat{\log \tau_{s:p}} = c\mathrm{E} \left(\log \tau_{s:p} \right) + A \log[y_M^*]_{s:p}.$$

Of course

$$\frac{1}{T} \sum_{i=1}^{T} \log[y_M^*]_i \xrightarrow{p} \mathrm{E} \left(\log \tau_i \right),$$

hence we are left with just determining $\widehat{c}$ and $\widehat{A}$. If we assume that the realised variances are a covariance stationary process then the weighted least squares statistic of $\log \tau_{s:p}$ sets

$$c = (I - A) \iota$$

and

$$
\begin{aligned}
A &= \left[\mathrm{Cov}\left(\log \tau_{s:p}\right) + \frac{2\hbar \mathrm{E}\left(\tau_i^{[2]}/\tau_i^2\right)}{M} I\right]^{-1} \mathrm{Cov}\left(\log \tau_{s:p}\right) \\
&= \left[\mathrm{Cov}\left(\log [y_M^*]_{s:p}\right)\right]^{-1} \mathrm{Cov}\left(\log \tau_{s:p}\right).
\end{aligned}
$$

Of course for this statistic

$$
\widehat{\log \tau_{s:p}} \rightarrow \log \tau_{s:p}
$$

as $M \rightarrow \infty$, as expected.

This style of approach extends to the multivariate case where the focus is on estimating the actual covariance matrix (see Barndorff-Nielsen and Shephard (2002b)). Then it makes sense to use these regression approaches based on the logs of the realised variances and the Fisher transformation of the realised correlation. The asymptotic theory of the realised covariation allows this approach to be feasible without specifying a parametric model for the spot covariance matrix.

5 Model based approach

5.1 General discussion and example

Suppose we write (when they exist) ξ, ω^2 and r, respectively, as the mean, variance and the autocorrelation function of the continuous time stationary variance process τ. Here we recall the discussion of Barndorff-Nielsen and Shephard (2002a) on estimating and forecasting τ_i based upon a parametric models for τ and the time series of realised variances . Let us write $u_i = [y_M^*]_i - \tau_i$, then the asymptotic theory tells us that for large M the u_i are approximately uncorrelated with

$$
\mathrm{Var}\left(\sqrt{M} u_i\right) \rightarrow 2\hbar^2\left(\omega^2 + \xi^2\right)
$$

as $M \rightarrow \infty$. Thus the second order properties of $[y_M^*]_i$ can be approximated. In particular $\mathrm{E}\left([y_M^*]_i\right) = \hbar\xi + o(1)$ and for $s > 0$

$$
\begin{aligned}
\mathrm{Var}\left([y_M^*]_i\right) &= 2M^{-1}\hbar^2\left(\omega^2 + \xi^2\right) + \mathrm{Var}(\tau_i) + o(1), \\
\mathrm{Cov}([y_M^*]_i, [y_M^*]_{i+s}) &= \mathrm{Cov}(\tau_i, \tau_{i+s}) + o(1) \\
\mathrm{Cov}([y_M^*]_i, \tau_i) &= \mathrm{Var}(\tau_i) + o(1) \\
\mathrm{Cov}([y_M^*]_i, \tau_{i+s}) &= \mathrm{Cov}(\tau_i, \tau_{i+s}) + o(1).
\end{aligned}
$$

$\mathrm{Var}(\tau_i)$ and $\mathrm{Cov}(\tau_i, \tau_{i+s})$ were given for all covariance stationary processes in Barndorff-Nielsen and Shephard (2001). In particular

$$
\mathrm{Var}\left(\tau_i\right) = 2\omega^2 r^{**}(\hbar) \qquad \text{and} \qquad \mathrm{Cov}\{\tau_i, \tau_{i+s}\} = \omega^2 \diamondsuit r^{**}(\hbar s), \tag{9}
$$

where

$$\diamondsuit r^{**}(s) = r^{**}(s+\hbar) - 2r^{**}(s) + r^{**}(s-\hbar), \tag{10}$$

and

$$r^{**}(t) = \int_0^t r^*(u)\mathrm{d}u \quad \text{where} \quad r^*(t) = \int_0^t r(u)\mathrm{d}u. \tag{11}$$

Thus, for a given model for the covariance stationary process τ we can compute the approximate second order properties of the time series of $[y_M^*]_i$ and τ_i.

The above theory implies we can calculate asymptotically approximate best linear filtered, smoothed and forecast values of τ_i using standard regression theory. This has recently been independently and concurrently studied by Andersen, Bollerslev, and Meddahi (2002) for some diffusion based models for τ. Their results are similar to those we present here.

Suppose we wish to estimate $\tau_{s:p}$ using $[y_M^*]_{s:p}$. Then the best linear estimator is

$$\begin{aligned} \widehat{\tau}_{s:p} &= (I-A)\,\iota E(\tau_i) + A[y_M^*]_i \\ &= A\left\{[y_M^*]_{s:p} - \hbar\xi\iota\right\} + \hbar\xi\iota, \end{aligned}$$

where

$$\begin{aligned} A &= \left\{\mathrm{Cov}([y_M^*]_{s:p})\right\}^{-1}\mathrm{Cov}\left(\tau_{s:p}, [y_M^*]_{s:p}\right) \\ &= \left\{\mathrm{Cov}\left(\tau_{s:p}\right) + 2M^{-1}\hbar^2\left(\omega^2+\xi^2\right)I\right\}^{-1}\mathrm{Cov}\left(\tau_{s:p}\right). \end{aligned}$$

The simplest special case of this is where $s = p = i$, that is we use a single realised variance to estimate actual variance. Then the theory above suggests the efficient linear estimator is constructed using the scalar

$$A = \left\{r^{**}(\hbar) + M^{-1}\hbar^2\left(1+\xi^2/\omega^2\right)\right\}^{-1} r^{**}(\hbar) \in [0,1], \tag{12}$$

which implies $\widehat{\tau}_i \geq 0$. Meddahi (2002) studied this particular regression, which we write as $\widehat{\tau}_i$ and call a Meddahi regression. It is always a consistent estimator of τ_i, but is more efficient than realised variance under the covariance stationarity assumptions.

In practice it is helpful to use the structure of the $\mathrm{Cov}\left(\tau_{s:p}\right)$ in order to carry out the required matrix inverse of $\mathrm{Cov}([y_M^*]_{s:p})$.

5.2 Special case

Suppose τ has the autocorrelation function $r(t) = \exp(-\lambda\,|t|)$. This implies that

$$\mathrm{E}\left(\tau_i\right) = \hbar\xi, \qquad \mathrm{Var}\left(\tau_i\right) = 2\omega^2\lambda^{-2}\left(e^{-\lambda\hbar} - 1 + \lambda\hbar\right),$$

and

$$\text{Cor}\{\tau_i, \tau_{i+s}\} = de^{-\lambda\hbar(s-1)}, \qquad s = 1, 2, ..., \tag{13}$$

where

$$d = \frac{(1 - e^{-\lambda\hbar})^2}{2\,(e^{-\lambda\hbar} - 1 + \lambda\hbar)} \in [0, 1].$$

In this case, in particular, the Meddahi regression has

$$\widehat{A} = \left\{\lambda^{-2}\left(e^{-\lambda\hbar} - 1 + \lambda\hbar\right) + M^{-1}\hbar^2\left(1 + \xi^2/\omega^2\right)\right\}^{-1}\lambda^{-2}\left(e^{-\lambda\hbar} - 1 + \lambda\hbar\right)$$

The above structure implies τ_i has the autocorrelation function of an ARMA$(1,1)$ model

$$\tau_i = \phi\tau_{i-1} + u_i + \theta u_{i-1}, \quad \phi = e^{-\lambda\hbar}.$$

The parameter θ was found numerically in Barndorff-Nielsen and Shephard (2001), however it can be determined analytically as indicated by Meddahi (2002). In particular, write

$$c_i = \tau_i - \phi\tau_{i-1} = u_i + \theta u_{i-1}$$

then

$$\text{Var}\,(c_i) = \left(1 + \phi^2\right)\text{Var}(\tau_i) - 2\phi\text{Cov}(\tau_i, \tau_{i-1})$$

and

$$\begin{aligned}
\text{Cov}(c_i, c_{i-1}) &= \left(1 + \phi^2\right)\text{Cov}(\tau_i, \tau_{i-1}) - \phi\text{Var}(\tau_i) - \phi\text{Cov}(\tau_i, \tau_{i-2}) \\
&= \text{Cov}(\tau_i, \tau_{i-1}) - \phi\text{Var}(\tau_i) \\
&= \text{Var}(\tau_i)\left\{\text{Cor}(\tau_i, \tau_{i-1}) - \phi\right\}.
\end{aligned}$$

Note that $\text{Cor}(\tau_i, \tau_{i-1}) \geq \phi$ as $e^{\lambda\hbar} - e^{-\lambda\hbar} \geq 2\lambda\hbar$. Write

$$\rho_1 = \frac{\text{Cov}(c_i, c_{i-1})}{\text{Var}\,(c_i)} \in \left[0, \frac{1}{2}\right], \quad \text{then} \quad \theta = \frac{1 - \sqrt{1 - 4\rho_1^2}}{2\rho_1} \in [0, 1].$$

This argument extends to the case of a superposition where $r(t) = \sum_{j=1}^{J} w_j \exp(-\lambda_j\,|t|)$, then τ_i can be represented as the sum of J uncorrelated ARMA$(1,1)$ processes, with $\{w_j, \lambda_j\}$ determining the corresponding autoregressive and moving average roots $\{\phi_j, \theta_j\}$.

In calculating $\widehat{\tau}_{s:p}$ Barndorff-Nielsen and Shephard (2001) conveniently placed $[y_M^*]_i$ into a linear state space representation so the filtering, smoothing and forecasting can be carried out using the Kalman filter (see, for example, Harvey (1989) and Durbin and Koopman (2001, Ch. 1)). In particular writing $\alpha_{1i} = (\tau_i - \hbar\xi)$ and $u_i = \sqrt{2M^{-1}\hbar^2\,(\omega^2 + \xi^2)}v_{1i}$, then

$$\begin{aligned}
[y_M^*]_i &= \hbar\xi + (1 \quad 0)\,\alpha_i + u_i, \\
\alpha_{i+1} &= \begin{pmatrix} \phi & 1 \\ 0 & 0 \end{pmatrix}\alpha_i + \begin{pmatrix} \sigma_\sigma \\ \sigma_\sigma\theta \end{pmatrix} v_i,
\end{aligned} \tag{14}$$

where v_i is a zero mean, unit variance, white noise sequence uncorrelated with u_i which has a variance of $2M^{-1}\hbar^2\left(\omega^2+\xi^2\right)$. The parameters ϕ, θ and σ_σ^2 represent the autoregressive root, the moving average root and the variance of the innovation to the ARMA(1,1) representation of the τ_i process. The extension to the superposition case is straightforward. In particular, in the case where $J=2$ this becomes

$$\begin{aligned} [y_M^*]_i &= \hbar\xi + \begin{pmatrix}1 & 0 & 1 & 0\end{pmatrix}\alpha_i + u_i, \\ \alpha_{i+1} &= \begin{pmatrix} \phi_1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \phi_2 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}\alpha_i + \begin{pmatrix} \sigma_{\sigma 1} & 0 \\ \sigma_{\sigma 1}\theta_1 & 0 \\ 0 & \sigma_{\sigma 2} \\ 0 & \sigma_{\sigma 2}\theta_2 \end{pmatrix} v_i, \end{aligned}$$

where again v_i is a zero mean, unit variance, white noise sequence.

M	$\xi=0.5,\ \xi\omega^{-2}=8$			$\xi=0.5,\ \xi\omega^{-2}=4$			$\xi=0.5,\ \xi\omega^{-2}=2$		
$e^{-\hbar\lambda}=0.99$	Smooth	Predict	$[y_M^*]_i$	Smooth	Predict	$[y_M^*]_i$	Smooth	Predict	$[y_M^*]_i$
1	.0134	.0226	.624	.0209	.0369	.749	.0342	.0625	.998
12	.00383	.00792	.0520	.00586	.0126	.0624	.00945	.0211	.0833
48	.00183	.00430	.0130	.00276	.00692	.0156	.00440	.0116	.0208
288	.000660	.00206	.00217	.000967	.00343	.00260	.00149	.00600	.00347
$e^{-\hbar\lambda}=0.9$	Smooth	Predict	$[y_M^*]_i$	Smooth	Predict	$[y_M^*]_i$	Smooth	Predict	$[y_M^*]_i$
1	.0345	.0456	.620	.0569	.0820	.741	.0954	.148	.982
12	.0109	.0233	.0520	.0164	.0396	.0624	.0259	.0697	.0832
48	.00488	.0150	.0130	.00707	.0260	.0156	.0108	.0467	.0208
288	.00144	.00966	.00217	.00195	.0178	.00260	.00280	.0338	.00347

Table 5: *Exact mean square error (steady state) of the estimators of actual volatility. The first two estimators are model based (smoother and 1-step ahead predictor) and the third is* $[y_M^*]_i$. *These measures are calculated for different values of* $\omega^2=\mathrm{Var}(\tau(t))$ *and* λ, *keeping* $\xi=\mathrm{E}(\tau(t))$ *fixed at 0.5. File:* `ssf_mse.ox`.

Table 5 reports the mean square error of the model based one-step ahead predictor and smoother of actual variance, as well as the corresponding result for $[y_M^*]_i$. The results in the left hand block of the Table corresponds to the model which was simulated in Figure 2, while the other blocks represent other choices of the ratio of ξ to ω^2. The exercise is repeated for two values of λ.

The main conclusion from the results in Table 5 is that model based approaches can potentially lead to very significant reductions in mean square error, with the reductions being highest for persistent (low value of λ) variance processes with high values of $\xi\omega^{-2}$. Even for moderately large values of M the model based predictor can be more accurate than realised variance, sometimes by a considerable amount. This is an important result from a forecasting viewpoint. However, when there is not much persistence and M is very large, this result is reversed and realised variance can be moderately more accurate. The smoother is always substantially more accurate than realised variance, even when M is very large and there is not much memory in variance.

Estimating the parameters of continuous time stochastic volatility models is known to be difficult due to our inability to compute the appropriate likelihood function. This has prompted the development of a sizable collection of methods to deal with this problem (e.g. Kim, Shephard, and Chib (1998) and Gallant, Hsieh, and Tauchen (1997)). Barndorff-Nielsen and Shephard (2002a) used quasi-likelihood estimation methods based on the time series of realised variance. The quasi-likelihood is constructed using the output of the Kalman filter. It is suboptimal for it does not exploit the non-Gaussian nature of the variance dynamics, but it provides a consistent and asymptotically normal set of estimators. Monte Carlo results reported in Barndorff-Nielsen and Shephard (2002a) indicate that the finite sample behaviour of this approach is quite good. Further the estimation takes only a few seconds on a modern computer.

5.3 Empirical illustration

To illustrate some of these results we have fitted a set of superposition based models to the realised variance time series constructed from the five minute US/DM exchange rate return data discussed above. Here we use the quasi-likelihood method to estimate the parameters of the model — ξ, ω^2, $\lambda_1, ..., \lambda_J$ and $w_1, ..., w_J$. We do this for a variety of values of M, starting with $M = 6$, which corresponds to working with four hour returns. The resulting parameter estimates are given in Table 6. For the moment we will focus on this case.

M	J	ξ	ω^2	λ_1	λ_2	λ_3	w_1	w_2	Quasi-L	BP
6	3	0.4783	0.376	0.0370	1.61	246	0.212	0.180	-113,258	11.2
6	2	0.4785	0.310	0.0383	3.76	—	0.262	—	-113,261	11.3
6	1	0.4907	0.358	1.37	—	—	—	—	-117,397	302
18	3	0.460	0.373	0.0145	0.0587	3.27	0.0560	0.190	-101,864	26.4
18	2	0.460	0.533	0.0448	4.17	—	0.170	—	-101,876	26.5
18	1	0.465	0.497	1.83	—	—	—	—	-107,076	443
144	3	0.508	4.79	0.0331	0.973	268	0.0183	0.0180	-68,377	15.3
144	2	0.509	0.461	0.0429	3.74	—	0.212	—	-68,586	23.3
144	1	0.513	0.374	1.44	—	—	—	—	-76,953	765

Table 6: *Fit of the superposition of J volatility processes for a SV model based on realised variance computed using $M = 6$, $M = 18$ and $M = 144$. We do not record w_J as this is 1 minus the sum of the other weights. Estimation method: quasi-likelihood using output from a Kalman filter. BP denotes Box–Pierce statistic, based on 20 lags, which is a test of serial dependence in the scaled residuals. File:* `ssf_empirical.ox`.

The fitted parameters suggests a dramatic shift in the fitted model as we go from $J = 1$ to $J = 2$ or 3. The more flexible models allow for a factor which has quite a large degree of memory, as well as a more rapidly decaying component or two. A simple measure of fit of the model is the Box–Pierce statistic, which shows a large jump from a massive 302 when $J = 1$, down to a more acceptable number for a superposition model.

To provide a more detailed assessment of the fit of the model we have drawn a series of graphs in Figure 7 based on $M = 8$ and $M = 144$. Figure 7(a) draws the computed realised variance $[y_M^*]$, together with the corresponding smoothed estimate (based on $J = 3$) of actual variance using the model. These are based on the $M = 8$ case. We can see that realised variance is much more jagged than the smoothed quantity. These are quite close to the semi-parametric estimator given in Figure 6. Figure 7(b) shows the corresponding autocorrelation function for

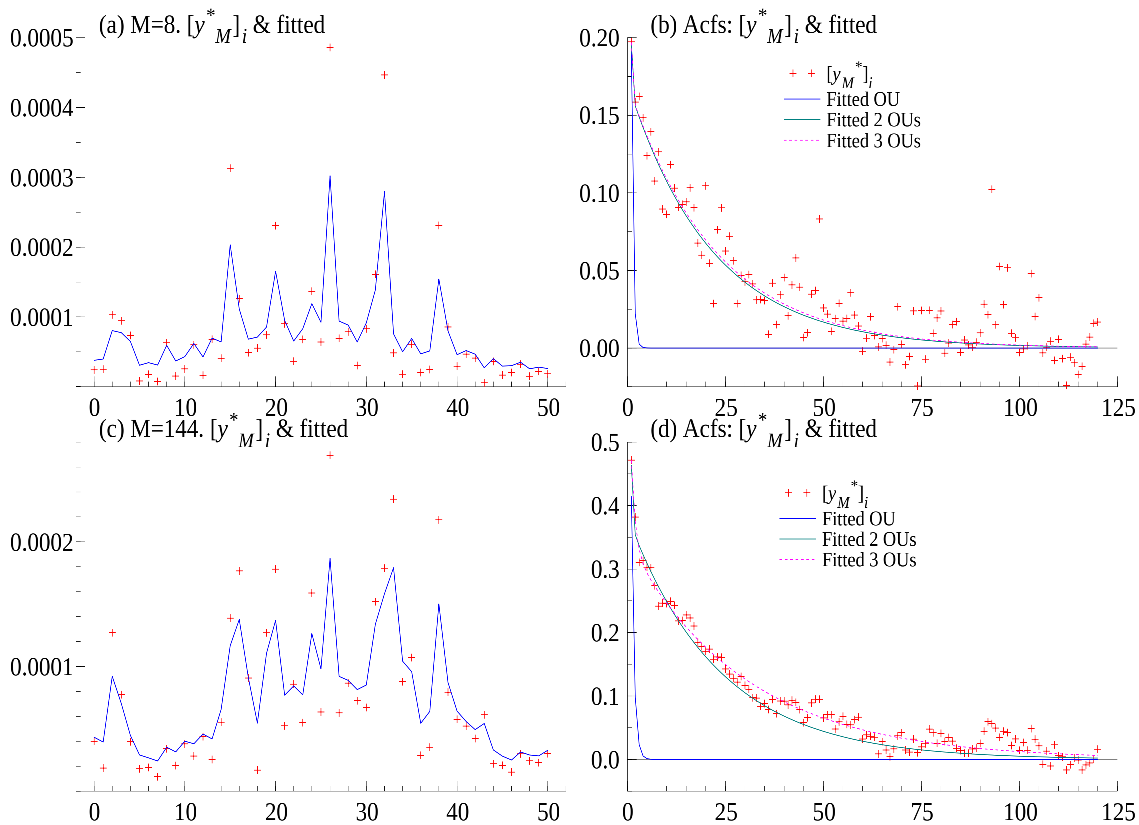

Figure 7: *Results from $M = 8$ and $M = 144$. (a) Using M=8, first 50 observations of $[y_M^*]_i$ & smoother. (b) Using M=8, Acf of $[y_M^*]_i$ and the fitted version for various values of J. (c) Using M=144, first 50 observations of $[y_M^*]_i$ & smoother. (d) Using M=144, Acf of $[y_M^*]_i$ and the fitted version for various values of J.* File: `daily_timeseries.ox`.

the realised variance series together with the corresponding empirical correlogram. We see from this figure that when $J = 1$ we are entirely unable to fit the data, as its autocorrelation function starts at around 0.6 and then decays to zero in a couple of days. A superposition of two processes is much better, picking up the longer-range dependence in the data. The superposition of two and three processes give very similar fits, indeed in the graph they are indistinguishable.

We next ask how these results vary as M increases. We reanalyse the situation when $M = 144$, which corresponds to working with ten minute returns. Figure 7(c) and (d) gives the corresponding results. Broadly the smoother has not produced very different results, while the $J = 3$ case now gives a slightly different fit to the Acf than the $J = 2$. The latter result is of importance, for as M increases the correlogram becomes more informative, allowing us to discriminate between different models more easily.

5.4 Comparison

We can compare the fit of the smoothers from the model free and model based approaches. In Figure 8 we display, using crosses, the time series of the model free smoother, based on 4 leads and 4 lags. This is drawn, for a variety of values of M, as the square root of the estimate, so it is estimating the square root of integrated variance. The corresponding model based approach is drawn using a line and it shows a close connection with the model free estimator. Table 7 gives the correlations between the two estimators as a function of M and the number of leads and lags in the model free approach. As the number of leads and lags increases the connection between the two estimators becomes stronger. Likewise, as M increase the two estimators become more closely correlated.

M	RV	1 lead, 1 lag	4 leads, 4 lags
6	.702	.849	.929
48	.903	.924	.932
144	.961	.985	.989
288	.984	.997	.998

Table 7: *Correlations between the model free and model based smoothers based on the Dollar/DM data. We vary M and the number of leads and lags.*

6 Conclusion

In this paper we have shown how we can use a time series of realised variances to measure and forecast integrated variances. These high frequency financial data statistics allow either model based or model free approaches to the problem. We have spent some time comparing the two smoothed estimators, which tend to be quite similar when M is large and we employ quite a few leads and lags.

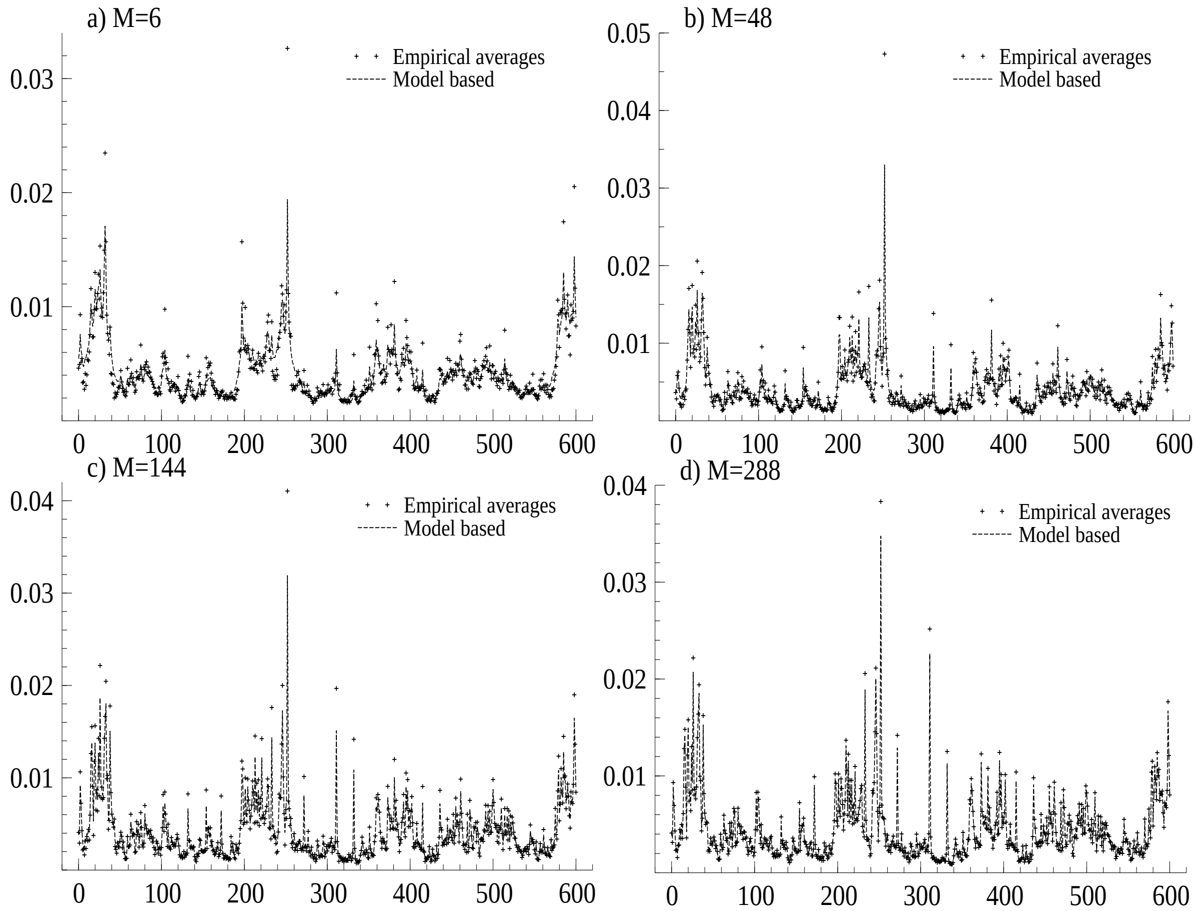

Figure 8: *Shows a comparison of the model free smoother based on 4 leads and lags and the model based approach. We show the estimators for the first 600 days in the sample, using a variety of values of M.*

7 Acknowledgments

Ole E. Barndorff-Nielsen's work is supported by CAF (`www.caf.dk`), which is funded by the Danish Social Science Research Council, and by MaPhySto (`www.maphysto.dk`), which is funded by the Danish National Research Foundation. Neil Shephard's research is supported by the UK's ESRC through the grant "Econometrics of trade-by-trade price dynamics," which is coded R00023839. All the calculations made in this paper are based on software written by the third author using the Ox language of Doornik (2001). We would like to thank Torben Andersen, Tim Bollerslev and Nour Meddahi for helpful conversations on this topic and Eric Zivot for his comments on our initial draft. We thank Michel M. Dacorogna for allowing us to use Olsen's high frequency exchange rate data in our study. Full details of this type of data is available in Dacorogna, Gencay, Muller, Olsen, and Pictet (2001).

8 Proof of the Theorem

We split the proof into two sections, dealing with the diagonal and non-diagonal elements of the matrix

$$A = (X + \sigma I)^{-1} X.$$

Here X is positive semi-definite and $\sigma > 0$.

(a) Diagonal elements of A.

Since X and I commute then A is positive definite, implying that A has positive diagonal. To be more explicit write $X = V\Lambda V'$ where $I = VV'$ and Λ is diagonal. From

$$A = \left\{V\left(\Lambda + \sigma I\right)V'\right\}^{-1} V\Lambda V' = V\left\{\left(\Lambda + \sigma I\right)^{-1}\Lambda\right\}V',$$

it is seen that A is symmetric and positive definite since $(\Lambda + \sigma I)^{-1}\Lambda$ is diagonal with positive diagonal elements. [3]

(b) Off-diagonal elements of A.

Rewrite

$$A = (X + \sigma I)^{-1} X = \left(I + \sigma X^{-1}\right)^{-1} = \eta\left(\eta I + X^{-1}\right)^{-1}, \qquad \eta = 1/\sigma.$$

It suffices to consider off diagonal elements of

$$N = \left(\eta I + X^{-1}\right)^{-1}.$$

The proof follows by induction. We use subscripts to denote the size of matrices, and superscripts to denote the elements of the inverse of a matrix.

Dimension 2. It holds that $N_2^{-1} = N_2^{\#}/\det(N_2)$, where

$$N_2^{\#} = \begin{pmatrix} \eta + X^{22} & -X^{12} \\ -X^{21} & \eta + X^{11} \end{pmatrix}.$$

Therefore the off-diagonal element is non-negative, $N_2^{12} \geq 0$, for all σ if and only if $X^{12} \leq 0$.

Dimension $k+1$. Simultaneous permutation of the i-th and j-th column and the i-th and j-th row preserves the positive definiteness of matrix. Thus we can look at an arbitrary off diagonal element to establish this result. Thus, look at the upper right element of N_{k+1}. This is given as

$$N_{k+1}^{1,k+1} = \frac{(-1)^k}{\det(N_{k+1})} \det\begin{pmatrix} X^{2,1} & \eta + X^{2,2} & \cdots & X^{2,k} \\ \vdots & & & \vdots \\ X^{k,1} & X^{k,2} & \cdots & \eta + X^{k,k} \\ X^{k+1,1} & X^{k+1,2} & \cdots & X^{k+1,k} \end{pmatrix}$$

[3]"Notice that all of the eigenvalues are strictly less than one." deleted as it seems unnecessary.

Expanding the latter determinant along the last row it follows that

$$\begin{aligned} N_{k+1}^{1,k+1} &= \frac{(-1)^k}{\det(N_{k+1})} \sum_{j=1}^{k} (-1)^{k-j}\, X^{k+1,j} (-1)^{j+1} N_k^{1,j} \det(N_k) \\ &= (-1)\, \frac{\det(N_k)}{\det(N_{k+1})} \sum_{j=1}^{k} X^{k+1,j} N_k^{1,j}. \end{aligned}$$

By induction it holds $N_k^{1,j} \geq 0$ for all j, and therefore a sufficient condition for $N_{k+1}^{1,k+1} \geq 0$ is that $X^{k+1,j} \leq 0$ for all $j \leq k$.

To prove necessity note that $N_k^{1,j} \det(N_k)$ is a polynomial in η of order $k-1$ if $j = 1$ and $j \neq 1$. Thus for large η

$$N_{k+1}^{1,k+1} \approx (-1) \frac{\det(N_k)}{\det(N_{k+1})} X^{k+1,1} N_k^{1,1},$$

so if $N_{k+1}^{1,k+1}$ is non-negative for large η then $X^{k+1,1}$ must be non-positive.

This completes the proof.

References

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2002). Parametric and nonparametric measurement of volatility. In Y. Ait-Sahalia and L. P. Hansen (Eds.), *Handbook of Financial Econometrics*. Amsterdam: North Holland. Forthcoming.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics 61*, 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001a). The distribution of exchange rate volatility. *Journal of the American Statistical Association 96*, 42–55.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001b). Modeling and forecasting realized volatility. Unpublished paper: Department of Economics, Duke University.

Andersen, T. G., T. Bollerslev, and N. Meddahi (2002). Analytic evaluation of volatility forecasts. Unpublished paper: Department of Economics, Duke University.

Andreou, E. and E. Ghysels (2002). Rolling-sampling volatility estimators: some new theoretical, simulation and empirical results. *Journal of Business and Economic Statistics 20*, 363–376.

Back, K. (1991). Asset pricing for general processes. *Journal of Mathematical Economics 20*, 371–395.

Barndorff-Nielsen, O. E. and D. R. Cox (1994). *Inference and Asymptotics*. London: Chapman & Hall.

Barndorff-Nielsen, O. E. and N. Shephard (2001). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics (with discussion). *Journal of the Royal Statistical Society, Series B 63*, 167–241.

Barndorff-Nielsen, O. E. and N. Shephard (2002b). Econometric analysis of realised covariation: high frequency covariance, regression and correlation in financial economics. Unpublished paper: Nuffield College, Oxford.

Barndorff-Nielsen, O. E. and N. Shephard (2002a). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B 64*, 253–280.

Barndorff-Nielsen, O. E. and N. Shephard (2002c). Estimating quadratic variation using realised variance. *Journal of Applied Econometrics*. Forthcoming.

Barndorff-Nielsen, O. E. and N. Shephard (2003). Realised power variation and stochastic volatility. *Bernoulli*. Forthcoming.

Chriss, N. and W. Morokoff (1999). Volatility and variance swaps. *Risk 12*, 55–59.

Christensen, B. J. and N. R. Prabhala (1998). The relation between implied and realized volatility. *Journal of Financial Economics 37*, 125–150.

Cox, D. R. (1980). Local ancillarity. *Biometrika 67*, 279–86.

Cox, D. R. and N. Wermuth (1996). *Multivariate dependencies.* London: Chapman and Hall.

Dacorogna, M. M., R. Gencay, U. A. Muller, R. B. Olsen, and O. V. Pictet (2001). *An Introduction to High-Frequency Finance.* San Diego: Academic Press.

Demeterfi, K., E. Derman, M. Kamal, and J. Zou (1999). A guide to volatility and variance swaps. *Journal of Derivatives 6*, 9–32.

Doornik, J. A. (2001). *Ox: Object Oriented Matrix Programming, 3.0.* London: Timberlake Consultants Press.

Durbin, J. (1980). Approximations for densities of sufficient estimators. *Biometrika 67*, 311–333.

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods.* Oxford: Oxford University Press.

Gallant, A. R., D. Hsieh, and G. Tauchen (1997). Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics 81*, 159–192.

Ghysels, E., A. C. Harvey, and E. Renault (1996). Stochastic volatility. In C. R. Rao and G. S. Maddala (Eds.), *Statistical Methods in Finance*, pp. 119–191. Amsterdam: North-Holland.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge: Cambridge University Press.

Harvey, A. C. and J. Durbin (1986). The effects of seat belt legislation on British road casualties: A case study in structural time series modelling. *Journal of the Royal Statistical Society, Series B 149*, 187–227.

Hinkley, D. V. (1980). Likelihood as approximate pivotal distribution. *Biometrika 67*, 287–292.

Howison, S. D., A. Rafailidis, and H. O. Rasmussen (2000). A note on the pricing and hedging of volatility derivatives. Unpublished paper: Mathematical Institute, University of Oxford.

Jacod, J. and A. N. Shiryaev (1987). *Limit Theorems for Stochastic Processes.* Springer-Verlag: Berlin.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *J. Basic Engineering, Transactions ASMA, Series D 82*, 35–45.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*, 361–393.

Meddahi, N. (2002). A theoretical comparison between integrated and realized volatilities. *Journal of Applied Econometrics*. Forthcoming.

Poterba, J. and L. Summers (1986). The persistence of volatility and stock market fluctuations. *American Economic Review 76*, 1124–1141.

Protter, P. (1990). *Stochastic Integration and Differential Equations: A New Approach*. New York: Springer-Verlag.

Schwert, G. W. (1989). Why does stock market volatility change over time? *Journal of Finance 44*, 1115–1153.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielsen (Eds.), *Time Series Models in Econometrics, Finance and Other Fields*, pp. 1–67. London: Chapman & Hall.

Skovgaard, I. M. (2001). Likelihood asymptotics. *Scandinavian Journal of Statistics 28*, 3–32.

Taylor, S. J. (1994). Modelling stochastic volatility. *Mathematical Finance 4*, 183–204.

Taylor, S. J. and X. Xu (1997). The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance 4*, 317–340.