

Noise Reduced Realized Volatility: A Kalman Filter Approach

John P. Owens * and Douglas G. Steigerwald[†]

Abstract

Microstructure noise contaminates high-frequency estimates of asset price volatility. Recent work has determined a preferred sampling frequency under the assumption that the properties of noise are constant. Given the sampling frequency, the high-frequency observations are given equal weight. While convenient, constant weights are not necessarily efficient. We use the Kalman filter to derive more efficient weights, for any given sampling frequency. We demonstrate the efficacy of the procedure through an extensive simulation exercise, showing that our filter compares favorably to more traditional methods.

Keywords: Realized Volatility, Microstructure Noise, Kalman Filter

JEL Classification: C32, C52, G00

*School of Economics and Finance, Faculty of Commerce and Administration, Victoria University of Wellington, PO Box 600, Wellington, New Zealand, phone: +64-4-463 5451, fax: +64-4-463 5014, e-mail: John.Owens@vuw.ac.nz.

[†]Department of Economics, University of California, Santa Barbara

1 Introduction

Long-standing interest in asset price volatility, combined with recent developments in its estimation with high-frequency data, has provoked research on the correct use of such data. In this paper we offer a framework for high-frequency measurement of asset returns that provides a means of clarifying the impact of microstructure noise. Additionally, we provide Kalman filter based techniques for the efficient removal of such noise.

In a series of widely cited articles, Andersen, Bollerslev, Diebold, and Labys (2001a,b) and Barndorff-Nielsen and Shephard (2002a,b,c) lay out a theory of volatility estimation from high-frequency sample variances. According to the theory, realized volatility estimators can recover the volatility defined by the quadratic variation of the semimartingale for prices. Realized volatility estimators are constructed as the sums of squared returns, where each return is measured over a short interval of time.¹

Realized volatility differs markedly from model-based estimation of volatility. The widely used class of volatility models derived from the ARCH specification of Engle (1982) place constraints on the parameters that correspond to the interval over which returns are measured. Empirical analyses of these models rarely support the constraints. In contrast, realized volatility estimators do not require a specified volatility model.

¹Andersen (2002) provides a survey of both theory and empirics for realized volatility.

The asymptotic theory underpinning realized volatility estimators suggests that the estimators should be constructed from the highest frequency data available. One would then sum the squares of these high-frequency returns, giving each squared return equal weight. In practice, however, very high frequency data is contaminated by noise arising from the microstructure of asset markets.

By now, it is widely accepted that market microstructure contamination obscures high-frequency returns through several channels. For example, transaction returns exhibit negative serial correlation due to what Roll (1984) terms the bid-ask bounce. When prices are observed at only regular intervals, or are treated as if this were the case, measured returns exhibit nonsynchronous trading biases as described in Cohen, Maier, Schwartz, and Whitcomb (1978, 1979), and Atchison, Butler, and Simonds (1987), and Lo and MacKinlay (1988, 1990). Because transaction prices are discrete and tend to cluster at certain fractional values, prices exhibit rounding distortions as described in Gottlieb and Kalay (1985), Ball (1988), and Cho and Frees (1988). Noise cannot be removed simply by working with the middle of specialist quotes; while mid-quotes are less impacted by asynchronous trade and the bid-ask bounce, mid-quotes are distorted by the inventory needs of specialists and by the regulatory requirements that they face.²

Simulations by Andersen and Bollerslev (1998a), Andersen, Bollerslev and Lange (1999), Andreou and Ghysels (2002), Bai, Russell and Tiao (2000), Barndorff-Nielsen

²Surveys by O'Hara (1995), Hasbrouck (1996), Campbell, Lo, and MacKinlay (1997), and Madhaven (2000) document these and other microstructure frictions.

and Shephard (2002c), Barucci and Renò (2002), Bollerslev and Zhou (2002), and Zumbach, Corsi and Trapletti (2002), among others, illustrate the effects of finite sampling and microstructure noise on volatility estimates under a variety of specifications. Differences in model formulation and assumed frictions make drawing robust conclusions about the effects of specific microstructure features difficult. Nevertheless, from the cited work, it is clear that microstructure frictions, as a group, cannot be safely ignored.

Essentially, three strands of research exist that treat the problem of microstructure noise in realized volatility estimation. The first attempts to remove the noise with a simple moving-average filter as in Zhou (1996). Andersen, Bollerslev, Diebold and Ebens (2001) and Corsi, Zumbach, Muller and Dacorogna (2001) who select a sample frequency of five minutes based on a volatility signature plots and then apply a moving-average filter. In contrast, Bandi and Russell (2003) work with an explicit model of microstructure noise. Rather than filtering the data to reduce the noise, they determine an optimal sampling frequency in the presence of noise. To do so, they construct a mean-squared error criterion that trades off the increase in precision against the corresponding increase in noise that arises as the sampling frequency increases. Although squared returns are given equal weight for a given asset, the optimal sampling interval that arises can vary across assets. Ait-Sahalia, Y., P. A. Mykland, and L. Zhang (2003) and Oomen (2004) offer similar treatments. Finally, Hansen and Lunde (2004) derive a Newey and West (1987) type correction

to account for spurious correlations in observed returns.

Theory suggests that noise volatility remains relatively constant. However, it is known that return volatility varies markedly. Thus, the relative contributions of noise and actual returns towards observed returns vary. During periods of high return volatility, return innovations tend to dominate the noise in size. In consequence we propose a somewhat different estimator in which the weight given to each return varies. Observed returns during periods of high volatility are given larger weight.

Our argument has three parts. First, we frame a precise definition of noise in terms of market microstructure theory. Second, we show how the Kalman filter can be used to remove the microstructure noise. We pay particular attention to how the variability of the optimal return weights depends on high-frequency volatility. Third, we demonstrate the efficacy of the filter in removing the noise.

2 Model

To be consistent with previous work, we employ assumptions typical of the realized volatility literature. Our first assumption concerns the true price process.

Assumption 1. (The true price process.)

The log true price process is a continuous local martingale. Specifically,

$$p_\tau = \int_0^\tau v_s dw_s$$

where w_τ is standard Brownian motion and where the spot volatility process v_τ is a strictly positive cadlag process such that the quadratic variation (or integrated volatility) process, denoted by V_τ , obeys

$$V_\tau = \int_0^\tau v_s ds < \infty$$

with probability one for all τ .

Let the step length be indexed by m , so that each period is divided into subintervals of equal length. From a well developed theory of quadratic variation (see Andersen (2002) for example), Assumption 1 implies that

$$\lim_{m \rightarrow \infty} \sum_{j=1, \dots, [m\tau]} (p_{\tau+j/m} - p_{\tau+(j-1)/m})^2 \rightarrow V_\tau \quad (1)$$

almost surely for all τ . Moreover, Assumption 1 implies that true returns, conditional on the path of integrated volatility, are normally distributed. Specifically, if we write $\sigma_t^2 = V_t - V_{t-1}$, then

$$r_t | \sigma_t^2 \sim N(0, \sigma_t^2).$$

Thus, the underlying integrated volatility can, in principle, be recovered from a continuous record of the true price process. But, of course, in practice the usefulness of the asymptotic result is limited by microstructure features that prevent the accurate measurement of true price at the required sampling frequencies. Only a finite number of noisy proxies of the required true prices are available.

To formalize the issues, consider a sequence of fixed intervals (5-minute periods, for example) indexed by t . Write the log of the observed price at t as $\tilde{p}_t = p_t + \eta_t$, where p_t denotes true price and η_t denotes microstructure noise. Let $r_t = p_t - p_{t-1}$ denote the true returns and let $\varepsilon_t = \eta_t - \eta_{t-1}$ denote the return noise. Write the observed returns as

$$\tilde{r}_t = r_t + \varepsilon_t. \quad (2)$$

For a given period (a day, for example), normalized to begin at $t = 0$ and assumed to last an integer multiple d of the chosen fixed interval, define the following three volatility measures.

1. The period's integrated volatility denoted by $V = \int_0^d \sigma_s^2 ds$.
2. The period's latent realized volatility, constructed from unobserved true returns, denoted by $\bar{V} = \sum_{t=1}^d r_t^2$.
3. The period's estimated integrated volatility denoted by \hat{V} .

The estimation error can be decomposed as

$$V - \hat{V} = V - \bar{V} + \bar{V} - \hat{V}. \quad (3)$$

The first difference on the right side of (3) has already received significant attention in the literature. Its properties are a function of the step length and the properties of the underlying volatility processes (Barndorff-Nielsen and Shephard (2002a) study the issues for a specific parameterization). If the step length is chosen, then this part of the error is beyond the control of the researcher. Therefore, we focus on minimizing the second difference. Specifically, we seek to minimize the mean squared error of \hat{V} as an estimate of \bar{V} ($MSE = E(\bar{V} - \hat{V})^2$) given the information contained in the sequence of observed returns. It is well known that the mean squared error is minimized by choosing the estimate according to

$$\hat{V} = E(\bar{V} | \{\tilde{r}_t\}_1^d) = E\left(\sum_{t=1}^d r_t^2 \middle| \{\tilde{r}_t\}_1^d\right) = \sum_{t=1}^d E(r_t^2 | \{\tilde{r}_t\}_1^d).$$

Thus, in order to minimize the effects of microstructure noise, we must extract the expected squared returns from the observed returns. The effectiveness with which the extraction can be achieved depends on the correct treatment of the microstructure noise.

Assumptions about noise dynamics must be selected with care. On the one hand, noise must exhibit strong positive serial correlation when prices are sampled

at extremely high frequencies. Noise outcomes of adjacent price measurements are almost perfectly correlated when no transaction intervenes (they are not perfectly correlated because, although measured price remains constant in the absence of new transactions, the latent true price changes through time). On the other hand, intervening transactions sharply curtail correlation among noise outcomes when prices are measured at lower frequencies.

Take the issue of bid-ask bounce for example. Roll (1984) explains that transaction prices are noisy proxies for true price because transactions tend to occur at the quotes instead of at the true price. A transaction at the ask reflects the true price as well as the premium charged by suppliers of liquidity. For similar reasons, transactions at the bid induce noise as well. Roll assumes that trade directions are independently determined. Under this assumption, he shows that the differencing operation that converts prices to returns causes the returns to exhibit spurious negative first-order serial correlation.

However, as discussed in Hasbrouck and Ho (1987), trade directions are not independently determined. Instead trade direction exhibits positive serial correlation as a result of, among other things, traders breaking up large transactions. When traders need to buy a large number of shares, they tend to distribute the purchases over time. However, trade direction is only weakly correlated between trades more than a few transactions apart. Thus, noise outcomes separated this distance are similarly weakly correlated.

Rounding errors are another source of noise. At the highest sampling frequencies, these errors exhibit strong temporal interdependence. This is because adjacent price measurements tend to round to the same value. However, if prices are sampled infrequently enough so that price measurements tend to round to different values, then only weak correlation between the corresponding rounding errors results.

If, as is typical, the random timing of trades is ignored in favor of measuring prices at regular intervals, then Lo and MacKinlay (1990) show that a so-called nonsynchronous trading effect leads to another source of microstructure noise. Nonsynchronous trading noise is serially correlated when transactions are sufficiently rare, relative to the sampling frequency, so that price measurements often refer to the same previous transaction. However, if each sampling interval typically contains several transactions, then this type of noise is a far less significant source of serial correlation in noise.

Clearly, there exist sampling frequencies for which noise exhibits temporal interdependence. However, for each type of microstructure noise, there exist time scales over which separate noise outcomes are essentially independent. Thus, the question of how long microstructure noise remains serially correlated, it seems, is an empirical one. Hasbrouck and Ho (1987) provide perhaps the most comprehensive treatment of the subject. Using transaction data from a large sample of stocks of the NYSE, they find that statistically significant autocorrelations do not persist

beyond 10 lags (trades) for either returns based on transaction prices or for returns based on midquotes. Although the authors' theory suggests that an ARMA(2,2) model is appropriate for observed trade-by-trade returns, their fitted ARMA models predict extremely small serial correlations after 10 lags. These results are supported by recent work. Using a different approach, Hansen and Lunde (2004) find that noise has time dependence lasting only two minutes in their sample of recent data from the stocks of the Dow Industrial Average.

We assume that prices are sampled infrequently enough to justify treating microstructure noise as an iid sequence. As a practical matter, we assume that the sampling intervals contain at least 10 transactions (or at least 10 quote revisions if prices are measured as midquotes) the vast majority of the time. By sampling in this way, theory provides clear guidance about how to best recover the true returns. We formalize our ideas about noise with the second assumption.

Assumption 2. (The microstructure noise.)

The microstructure noise forms an iid sequence of random variables each with mean zero and variance $\sigma_\eta^2 < \infty$ and independent of the true return process.

We do not make any distributional assumptions about microstructure noise. However, as the noise is composed of a sum of several largely independent features, and because these features tend to be symmetric, the assumption of nor-

mally distributed noise seems to be a plausible approximation. Consequently, we will consider normally distributed microstructure noise as a special case.

Under Assumption 2, it is clear that return noise forms an MA(1) process with a unit root. If true returns also form a weakly stationary martingale, then we can say more.

Lemma 1 *If in addition to Assumption 1 and Assumption 2, r_t forms a weakly stationary process with unconditional mean zero and unconditional variance σ_r^2 , then the autocovariance function of observed returns obeys*

$$Cov(\tilde{r}_t, \tilde{r}_{t-k}) = \begin{cases} \sigma_r^2 + 2\sigma_\eta^2 & \text{if } k = 0 \\ -\sigma_\eta^2 & \text{if } k = 1 \\ 0 & \text{if } k > 1 \end{cases} .$$

Moreover, the first-order autocorrelation is given by

$$\rho = -\frac{\sigma_\eta^2}{\sigma_r^2 + 2\sigma_\eta^2}$$

Thus, the first two autocovariances of the observed returns series are sufficient for determining the variance of the noise and the expected variance of the true returns. These two parameters are required for the Kalman filter that is used to optimally extract true returns from the observed returns.

2.1 Kalman filter

Kalman filter techniques are well known to the econometrics literature. Harvey (1989) and Hamilton (1994) provide textbook treatments of the subject, laying out the mechanics and many of the optimal characteristics of the technique. We follow Hamilton in our notation. To infer the (latent) true returns, we separate (observed) contaminated returns into two components: the first corresponding to true returns and the second to microstructure noise.

The State Equation (4) describes the dynamics of the latent variables:

$$\xi_{t+1} = \mathbf{F}\xi_t + \mathbf{R}\mathbf{v}_{t+1} \quad (4)$$

where

$$\xi_t = \begin{pmatrix} r_t \\ \eta_t \\ \eta_{t-1} \end{pmatrix}; \mathbf{F} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}; \mathbf{R} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}; \mathbf{v}_t = \begin{pmatrix} r_t \\ \eta_t \\ 0 \end{pmatrix}.$$

The Observation Equation (5) describes the dynamics of the observed returns

$$\tilde{r}_t = \mathbf{H}^T \xi_t \quad (5)$$

where $\mathbf{H}^T = (1, 1, -1)$.

The Kalman filter allows one to efficiently determine linear projections of la-

tent returns and latent microstructure noise on the history of observed returns.

Of course, linear projections are optimal (in the mean squared error sense) among all linear estimators that make use of this history. In order to explain how the technique works, we establish the following notation. Let $\hat{\mathbf{E}}_t$ represent linear projection on to the history of observed returns and a constant. Denote the history by

$\mathcal{F}_t = \{\tilde{r}_t, \tilde{r}_{t-1}, \dots, \tilde{r}_1, 1\}$. Let $\hat{\boldsymbol{\xi}}_{\tau|t} = \hat{\mathbf{E}}_t(\boldsymbol{\xi}_\tau)$ and let

$$\mathbf{P}_{\tau|t} = \mathbb{E} \left[(\boldsymbol{\xi}_\tau - \hat{\boldsymbol{\xi}}_{\tau|t}) (\boldsymbol{\xi}_\tau - \hat{\boldsymbol{\xi}}_{\tau|t})^T \right] \quad (6)$$

represent the mean squared error matrices of these projections. Let u_t denote the one-step-ahead prediction error for the observed returns and let M_t denote its variance:

$$u_t = \tilde{r}_t - \mathbf{H}^T \hat{\boldsymbol{\xi}}_{t|t-1}$$

$$M_t = \mathbf{H}^T \mathbf{P}_{t|t-1} \mathbf{H}.$$

We first describe an infeasible filter based on knowing the true sequence of integrated volatilities $\{\sigma_t^2\}$. If the sequence $\{\sigma_t^2\}$ is known, then the Kalman filter possesses the previously described optimal properties, and the construction of the filter is straightforward. We examine the practical consequences of the discrepancies between the infeasible filter and a feasible filter, which we propose below, through simulations.

Let \mathbf{Q}_t denote the sequence of state variable innovation variances where

$$\mathbf{Q}_t = E\mathbf{v}_t\mathbf{v}_t^T = \begin{pmatrix} \sigma_t^2 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix}.$$

Using standard results, we obtain the sequence of linear projections through the recursion

$$\hat{\boldsymbol{\xi}}_{t|t} = \hat{\boldsymbol{\xi}}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{H} M_t^{-1} u_t \quad (7)$$

$$\hat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{F} \hat{\boldsymbol{\xi}}_{t|t} \quad (8)$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{H} \mathbf{H}^T \mathbf{P}_{t|t-1} M_t^{-1} \quad (9)$$

$$\mathbf{P}_{t+1|t} = \mathbf{F} \mathbf{P}_{t|t} \mathbf{F}^T + \mathbf{R} \mathbf{Q}_t \mathbf{R}^T \quad (10)$$

and the boundary conditions

$$\hat{\boldsymbol{\xi}}_{1|0}^T = (0, 0, 0) \quad (11)$$

$$vec(\mathbf{P}_{1|0}) = [\mathbf{I} - (\mathbf{F} \otimes \mathbf{F})]^{-1} vec(\mathbf{R} \mathbf{Q}_0 \mathbf{R}') \quad (12)$$

where we assume that the unconditional variance of the innovations exists and is

given by \mathbf{Q} . The recursion (10) implies that

$$\mathbf{P}_{t+1|t} = \begin{pmatrix} \sigma_{t+1}^2 & 0 & 0 \\ 0 & \sigma_\eta^2 & 0 \\ 0 & 0 & \frac{\sigma_\eta^2(\sigma_t^2 + c_t)}{\sigma_t^2 + \sigma_\eta^2 + c_t} \end{pmatrix}$$

where c_t denotes the (3,3) element of $\mathbf{P}_{t|t-1}$. The boundary condition (12) implies that $c_1 = \sigma_\eta^2$. Thus, the dynamics of the mean squared error matrices are completely determined by the difference equation obeyed by their third diagonal elements, c_t .

Although the recursion (7) determines the filtered returns, as presented, the recursions are difficult to interpret. After some tedious simplifications, we find that an equivalent way to express (7) is through

$$\hat{r}_{t|t} = \frac{\sigma_t^2}{M_t} (\tilde{r}_t + \hat{\eta}_{t-1|t-1}) \quad (13)$$

$$\hat{\eta}_{t|t} = \frac{\sigma_\eta^2}{M_t} (\tilde{r}_t + \hat{\eta}_{t-1|t-1}) \quad (14)$$

$$\hat{\eta}_{t-1|t} = \hat{\eta}_{t-1|t-1} - \frac{c_t}{M_t} (\tilde{r}_t + \hat{\eta}_{t-1|t-1}). \quad (15)$$

As represented in (13)-(15), the filtering algorithm is seen to work in two steps. First, the observed return is adjusted according to what is known about the error terms at time $t - 1$. Notice that at that time the best guess of η_t is zero. The adjusted quantity is then apportioned to the filtered values of the latent return and to the

filtered values of the current and the previous noise terms. Notice that

$$\hat{r}_{t|t} + \hat{\eta}_{t|t} - \hat{\eta}_{t-1|t} = \tilde{r}_t \quad \forall t.$$

Thus, the filter simply finds the optimal allocation of the observed return across its constituent parts. Finally, notice that the ratio of filtered latent returns to filtered contemporary noise equals the signal-to-noise ratio σ_t^2/σ_η^2 .

In the special case of constant return volatility, which corresponds to the uniform weighting of much previous research, a solution to this difference equation can be found; however, it is too complicated to be shown here. The asymptotic behavior is much more simply expressed. If $\sigma_t = \sigma_r$ for all t , then it holds that

$$\lim_{t \rightarrow \infty} c_t = \frac{1}{2} \sigma_r^2 \left(\sqrt{1 + 4 \frac{\sigma_\eta^2}{\sigma_r^2}} - 1 \right). \quad (16)$$

This can be verified directly by taking the limit of the solution to the difference equations obeyed by c_t ; also, notice that the limit is a stationary point of the recursion.

The term c_t represents the MSE of filtered lagged noise. The filtered noise terms are known with greater precision than their unfiltered counterparts. Thus, it is consistent with intuition that the recursive definition of c_t implies that $c_t \leq \sigma_\eta^2$. It is also intuitive that the limit (16) converges to σ_η^2 as the volatility becomes unboundedly

large:

$$\lim_{\sigma_r^2 \rightarrow \infty} \frac{1}{2} \sigma_r^2 \left(\sqrt{1 + 4 \frac{\sigma_\eta^2}{\sigma_r^2}} - 1 \right) = \sigma_\eta^2.$$

Which is easily verified with L'Hopital's rule. As volatility becomes large, we lose the ability to learn anything about the noise terms from the observable data.

If constant return volatility is assumed, so that $\sigma_t^2 = \sigma_r^2$ for all t , then fractions (13) and (14) are essentially constant (as a practical matter, c_t converges to a fixed value after the first few observations). With constant fractions the weights are constant across observations. Of course, return volatility likely varies over time. Simple calculus arguments show that the observed return allocated to the filtered return increases with the integrated volatility of the period. In essence, a large observed return is more likely to have come from a large latent return during a period of high volatility. Notice that the signal-to-noise ratio is larger when volatility is high. This highlights a potential source of improvements for estimators of volatility. An algorithm based on a proxy for high-frequency volatility may be able to improve on a filter that assumes a constant volatility by allowing the weights to be sensitive to the changing signal-to-noise ratio.

2.2 Kalman smoother

Although filtered returns make optimal use of the history of observed returns, a superior estimator is available if we make use of future outcomes of observed returns. By extending the methods described above, the projections onto the entire

sample of observed returns can be obtained, yielding a new estimator $\hat{r}_{t|T} = \hat{\mathbf{E}}_T(r_t)$.

Using standard results, we obtain the sequence of linear projections through the recursion

$$\mathbf{J}_t = \mathbf{P}_{t|t} \mathbf{F}' \mathbf{P}_{t+1|t}^{-1} \quad (17)$$

$$\hat{\boldsymbol{\xi}}_{t|T} = \hat{\boldsymbol{\xi}}_{t|t} + \mathbf{J}_t (\hat{\boldsymbol{\xi}}_{t+1|T} - \hat{\boldsymbol{\xi}}_{t+1|t}) \quad (18)$$

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t} + \mathbf{J}_t (\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t}) \mathbf{J}_t^T \quad (19)$$

which simplifies as

$$\hat{r}_{t|T} = \hat{r}_{t|t} - \frac{\sigma_t^2}{\sigma_t^2 + c_t} (\hat{\eta}_{t|T} - \hat{\eta}_{t|t}) \quad (20)$$

$$\hat{\eta}_{t|T} = \hat{\eta}_{t|t} + (\hat{\eta}_{t|T} - \hat{\eta}_{t|t}) \quad (21)$$

$$\hat{\eta}_{t-1|T} = \hat{\eta}_{t-1|t} + \frac{c_t}{\sigma_t^2 + c_t} (\hat{\eta}_{t|T} - \hat{\eta}_{t|t}). \quad (22)$$

Smoothing is seen to be a matter of adjustment to the filtered quantities. Working backward from the end of the series, we improve our understanding of the noise terms. The change in the estimated noise term that results from the information contained in the future observations is then used to adjust the filtered returns and the filtered previous noise term. Here again, the correct adjustment to the estimated latent returns depends on current volatility.

As with the filtered case, it holds that

$$\hat{r}_{t|T} + \hat{\eta}_{t|T} - \hat{\eta}_{t-1|T} = \tilde{r}_t \quad \forall t$$

Thus, the information contained in future observations simply improves the optimal distribution of observed returns across its constituents.

The recursions of (13)-(22) provide useful insight into the filtering and smoothing process. However, they also provide a second benefit. Notice that each step of the recursions requires information about only adjacent returns. Thus, the data can be smoothed with only two passes over the data. A forward pass accomplishing (13)-(15) delivers the filtered returns. Then, a backward pass accomplishing (20)-(22) determines the smoothed values.

2.3 Bias

Once the estimated latent returns are obtained, it seems natural to estimate realized volatility by the sum of their squares. However, there are reasons for caution. Squaring returns is a nonlinear transformation whereas filtering is a linear one. There is no reason to expect that filtered squared returns equal squared filtered returns. In fact, using squared filtered returns necessarily biases estimated realized volatility towards zero. Fortunately, however, the size and direction of the bias is determined in the normal course of constructing the Kalman filter. Denote the bias

as $b_t^F = \mathbb{E} \left(r_t^2 - \hat{r}_{t|t}^2 \right)$. We have that

$$\mathbb{E} \left[(r_t - \hat{r}_{t|t})^2 \right] = \mathbb{E} \left[\hat{\mathbb{E}}_t \left(r_t^2 - 2\hat{r}_{t|t}r_t + \hat{r}_{t|t}^2 \right) \right] = \mathbb{E} \left[r_t^2 - \hat{r}_{t|t}^2 \right] \quad (23)$$

where we have used the appropriate law of iterated expectations.³ Thus, the bias equals the (1, 1) element of the mean squared error matrix $P_{t|t}$. This suggests a simple correction. An unbiased estimate of squared latent returns is given by $\hat{r}_{t|t}^2 + \hat{b}_t^F$.

The recursions (9) and (10) imply that the bias is

$$b_t^F = \sigma_t^2 \left(\frac{\sigma_\eta^2 + c_t}{M_t} \right). \quad (24)$$

Thus, the bias equals the period's integrated volatility down weighted by the ratio of the one-step-ahead prediction variance of the noise terms to the one-step-ahead prediction variance of observed returns.

In the special case of constant volatility, the bias converges quickly to a limiting value and can effectively be considered a constant. Substituting the limit of c_t as defined in (16) into (24) suggests a simple bias correction based solely on the parameters of the model. However, the resulting expression is complex and provides little intuition. To give some idea of the bias' magnitude and its relationship to underlying model parameters, notice that if the return variance dominates noise

³The iterated conditional expectation result is based on properties of projection mappings. See, for example, Brockwell and Davis (1987) Proposition 2.3.2.

variance, then c_t converges to a value near σ_η^2 . In this case, the bias is well approximated by replacing c_t with σ_η^2 , which yields $b_t^F/\sigma_r^2 \simeq -2\rho$. As unadjusted high-frequency returns often exhibit substantial first-order serial correlation, the bias can easily reach unacceptable levels.

Of course, the analysis of the bias depends on the properties of the Kalman filter. We cannot determine what bias results if an estimator is used that does not deliver the optimal linear filter of the latent returns. For example, if constant return volatility is assumed when it is not appropriate, then the analysis does not apply. However, as the essential reason for the bias is that projections (filters) are generally less noisy than their actual counterparts, we suppose that the same sort of bias affects commonly used filters (such as simple MA filters). We verify this supposition in the Simulation section.

Basing realized volatility on squared smoothed returns results in the same sort of bias as obtains when squared filtered returns are used. Denote this bias as b_t^S where

$$b_t^S = \mathbb{E} (r_t^2 - \hat{r}_{t|T}^2).$$

After tedious simplification of the MSE matrices of the smoother (see (19) above), we have

$$b_t^S = b_t^F - \left(\frac{\sigma_t^2}{\sigma_t^2 + c_t} \right)^2 (c_{t+1} - d_{t+1})$$

where the d_t is given by

$$d_t = c_t \left(\frac{\sigma_t^2 + \sigma_\eta^2}{\sigma_t^2 + \sigma_\eta^2 + c_t} \right) - \left(\frac{c_t}{\sigma_t^2 + c_t} \right)^2 (c_{t+1} - d_{t+1}).$$

The recursions make clear that the smoothed bias is smaller than the filtered bias. This is the expected result; smoothed returns are more precisely known. Beyond this, the interpretation of the smoothed bias is difficult. However, if volatility is constant, then the bias is well approximated by

$$b_t^S \simeq \sigma_r^2 \left(1 - \frac{1}{\sqrt{1 + 4\sigma_\eta^2/\sigma_r^2}} \right).$$

Here, the bias is a simple function of the constant integrated volatility and the variance of the noise term. If the noise term dominates, so that $\sigma_\eta^2/\sigma_r^2 \simeq 0$, then the bias is approximately σ_r^2 . If the integrated volatility term dominates, then the bias is near zero. Moreover, we see that the bias is monotonically increasing in noise volatility.

3 Multivariate Normal Approach

We gain additional insight into the optimal filter if we are willing to assume the normality of microstructure noise. Define the vectors $\mathbf{r} = (r_1, r_2, \dots, r_T)'$ and $\boldsymbol{\eta} = (\eta_0, \eta_1, \dots, \eta_T)'$. According to the assumptions, the vector of latent returns are

normally distributed, if we condition on integrated volatility. With the assumption that the vector of microstructure noise terms are normally distributed as well, we have

$$\begin{pmatrix} \mathbf{r} \\ \boldsymbol{\eta} \end{pmatrix} \sim MVN \left(\mathbf{0}, \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \sigma_\eta^2 \mathbf{I} \end{pmatrix} \right)$$

where $\Lambda = diag(\sigma_t^2)$. We may express the observed returns as $\tilde{\mathbf{r}} = (\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_T)'$, which is related to latent returns and noise though $\tilde{\mathbf{r}} = \mathbf{r} + \mathbf{B}\boldsymbol{\eta}$, where

$$\mathbf{B} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}.$$

It follows that, $\tilde{\mathbf{r}}$ and \mathbf{r} are jointly normally distributed. Using basic facts about linear functions of multivariate normal distributions, it is straightforward to determine the conditional distribution of the latent returns given the observed returns by applying well known results about conditional distributions. We have

$$\mathbf{r} | \tilde{\mathbf{r}} \sim MVN(\hat{\mathbf{r}}, \Sigma) \tag{25}$$

where $\hat{\mathbf{r}} = \Lambda (\Lambda + \sigma_\eta^2 \mathbf{B} \mathbf{B}')^{-1} \tilde{\mathbf{r}}$ and $\Sigma = \sigma_\eta^2 \mathbf{B} (\mathbf{I} + \sigma_\eta^2 \mathbf{B}' \Lambda^{-1} \mathbf{B})^{-1} \mathbf{B}'$.

Although the approach is different here, the end result is identical to the Kalman smoother defined above. The vector $\hat{\mathbf{r}}$ equals the smoothed returns and the matrix

Σ contains the corresponding MSE for the smoothed returns. Thus, even if the normality assumptions do not hold, then the indicated matrix calculations still deliver optimal linear estimates of the latent returns. If the normality assumption does hold, then the smoothed returns are conditional expectations of latent returns in addition to being linear projections.

The matrix form of the smoother allows us to directly examine how the constant volatility assumption effects the weights for the observed returns. If volatility is constant, then the latent integrated volatility matrix simplifies as $\Lambda = \sigma_r^2 \mathbf{I}$ where σ_r^2 is the common integrated volatility over the chosen step length. The optimal estimator of latent returns becomes

$$\hat{\mathbf{r}} = \left(\mathbf{I} + \frac{\sigma_\eta^2}{\sigma_r^2} \mathbf{B} \mathbf{B}' \right)^{-1} \tilde{\mathbf{r}}. \quad (26)$$

Thus, if the strength of the signal (σ_r^2) is much stronger than the strength of the noise (σ_η^2), then estimated latent returns and observed returns are nearly identical, as the matrix multiplying observed returns in (26) is nearly the identity matrix. As the relative size of the noise increases, the off-diagonal elements of the matrix become more important and neighboring observed returns become relevant for estimating latent returns. To see this point clearly, consider when $\sigma_r^2 = 10$, $\sigma_\eta^2 = 1$, and $T = 7$. Under these conditions, the matrix multiplying observed returns in

(26) becomes

$$\begin{pmatrix} 0.8392 & 0.0704 & 0.0059 & 0.0005 & 0 & 0 & 0 \\ 0.0704 & 0.8451 & 0.0709 & 0.006 & 0.0005 & 0 & 0 \\ 0.0059 & 0.0709 & 0.8452 & 0.0709 & 0.006 & 0.0005 & 0 \\ \mathbf{0.0005} & \mathbf{0.006} & \mathbf{0.0709} & \mathbf{0.8452} & \mathbf{0.0709} & \mathbf{0.006} & \mathbf{0.0005} \\ 0 & 0.0005 & 0.006 & 0.0709 & 0.8452 & 0.0709 & 0.0059 \\ 0 & 0 & 0.0005 & 0.006 & 0.0709 & 0.8451 & 0.0704 \\ 0 & 0 & 0 & 0.0005 & 0.0059 & 0.0704 & 0.8392 \end{pmatrix}.$$

The fourth row (in bold) indicates that the optimally estimated fourth latent return equals 85 percent of the fourth observed returns plus 7 percent of the adjacent observed returns plus something less than one percent of the other observed returns. An almost identical weighting scheme determines the third and fifth optimally estimated latent return. For large samples, the weights are even more consistent. Except for a few returns at the beginning and end of the sample, the assumption of constant volatility leads to estimates of latent returns that are essentially a weighted average of the observed returns where the weights, for all practical purposes, are constants.

If, as is almost certainly the case in practice, latent returns do not exhibit constant volatility, then the mean of the conditional distribution of (25) determines the optimal weights for estimating latent returns. Instead of constant weights, during

periods of high volatility the optimal weights are higher for the current observed return and lower for the other returns. During periods of low volatility the opposite is true.

The multivariate approach is especially useful for understanding the source of the bias. To see how the bias comes about, recall the following fact about conditional expectations.

$$\text{var}(r_t | \tilde{\mathbf{r}}) = \mathbb{E}(r_t^2 | \tilde{\mathbf{r}}) - \mathbb{E}^2(r_t | \tilde{\mathbf{r}}) \quad (27)$$

Thus, the optimal estimator for the square of the latent returns that we want, $\mathbb{E}(r_t^2 | \tilde{\mathbf{r}})$, exceeds by a positive amount the square of the optimal estimator for latent returns that we have in hand, $\mathbb{E}^2(r_t | \tilde{\mathbf{r}}) = \hat{r}_{t|T}^2$. However, as mentioned, the bias can be recovered through the MSE matrices which are given in Σ . The MSE bias correction, which is given by tt element of Σ , equals $\text{var}(r_t | \tilde{\mathbf{r}})$. However, here, the correction does more than simply correct for the bias. Because, in the multivariate normal case, the MSE matrices correspond to the conditional covariance matrix of the latent returns given the observed returns, the correction delivers the conditional expectation of the squared returns. Notice that

$$\mathbb{E}(r_t^2 | \tilde{\mathbf{r}}) = \mathbb{E}^2(r_t | \tilde{\mathbf{r}}) + \text{var}(r_t | \tilde{\mathbf{r}}).$$

Thus, if noise is normally distributed, then the suggested correction creates an optimal non-linear estimate out of a linear one.

To see that the bias can be substantial, consider the constant variance example given above. With $\sigma_r^2 = 10$ and $\sigma_\eta^2 = 1$, the diagonal elements of the covariance matrix Σ all exceed 1.5. Thus, the bias is more than 15 percent of the expected latent squared return and 50 percent larger than the expected squared noise term. This reinforces the fact that squared linearly extracted returns may be a poor proxy for the non-linearly extracted squared return.

4 Suggested Estimator

From the analysis, it is clear that simple MA and AR type filters have a number of shortcomings when data exhibit stochastic volatility. Implicitly, these simple filters assume constant volatility and therefore weigh all observed returns equally in their construction of filtered latent returns. Optimal filters weigh the observed returns according to the dynamics of the signal-to-noise ratio between return innovations and microstructure noise. The analysis shows that if the squares of linearly filtered latent returns are used as proxies for squared latent returns, then realized volatility estimators based on these proxies will exhibit downward biases. To redress both of these issues, we suggest the use of bias-corrected squares of smoothed returns and the use of a rolling average of these as a proxy for latent return volatility when specifying the Kalman filter.

The potential benefits of the approach are clear. The proxy may be able to account for the time varying signal-to-noise relationship. However, the potential

problems of miss-specifying volatility are equally clear. We use simulations to demonstrate that the suggested filter does improve realized volatility measurements relative to commonly used filters.

5 Simulations

To test the performance of the suggested filter against realistic scenarios, we use a model for the simulated latent returns that is consistent with the return behavior of the S&P 500 stock index. The following continuous time model is our starting point.

$$\begin{aligned} dp_t &= \sigma_t dw_t \\ d\sigma_t^2 &= \theta (\omega - \sigma_t) dt + (2\lambda\theta)^{1/2} dw_{\sigma_t}. \end{aligned} \tag{28}$$

This specification is a popular special case of Assumption 1. Drost and Werker (1996) show that this, so-called GARCH diffusion, discretizes as an exact GARCH(1,1):

$$\begin{aligned} p_t - p_{t-1/m} &= r_{(m),t} = \sigma_{(m),t} z_{(m),t} \\ \sigma_{(m),t} &= \phi_{(m)} + \alpha_{(m)} r_{(m),t}^2 + \beta_{(m)} \sigma_{(m),t-1/m} \end{aligned} \tag{29}$$

where $z_{(m),t}$ is (for the purposes of simulation) iid normal (0,1) and $1/m$ is the sampling frequency. Drost and Werker (1996, corollary 3.2) provides the map between the parameters of the diffusion and the GARCH parameters. This is the same scheme as is applied in Andreou and Ghysels (2002). According to those authors, the following parameters are consistent with 5-minute returns from the S&P 500

index.

$$\phi_{(m)} = 0.000426, \alpha_{(m)} = 0.003670, \beta_{(m)} = 0.996276. \quad (30)$$

These parameters imply an unconditional return variance of $\sigma_r^2 = 7.88889$, where returns are measured in basis points.

Using model (29) with the parameters (30), we simulate 10,000, 6.5 hour days worth of data (78,000 5-minute intervals).

We add to the returns noise generated by $\varepsilon_t = \eta_t - \eta_{t-1}$ where the η_t are iid normal distributed with mean zero. Because noise is inherently unobservable, it is difficult to obtain guidance from the literature about how to best choose its variance for the purposes of the study. However, first-order serial correlation of observed returns are observable and are reported in the literature. The first-order serial correlation is related to the variance of the microstructure noise through the equation $\rho = -\frac{\sigma_\eta^2}{\sigma_r^2 + 2\sigma_\eta^2}$. Therefore, we repeat the simulation under a variety of assumed first-order serial correlations for the observed returns. Specifically, we let $\rho = -0.4, -0.3, -0.2$, and -0.1 , in each case setting the variance of microstructure noise equal to

$$\sigma_\eta^2 = \frac{-\rho}{1 + 2\rho} \sigma_r^2.$$

These first-order serial correlations are chosen to be consistent with observed correlations in stock market data (as reported in Hasbrouck and Ho (1987) for example).

For each of the 10,000 simulated days, we record true realized volatility and three different types of estimated realized volatility. First we consider the optimal (though infeasible) return estimates based on knowledge of the true sequence of return variances. The optimal estimates form a benchmark against which other feasible estimators are compared. Next, we consider naive return estimates based on the assumption of constant return variance. The estimated (smoothed) returns are then used to construct a rolling average estimator for the sequence of return volatilities. Specifically, we take the arithmetic average of the current smoothed squared return plus the bias correction and the previous 12 and subsequent 12 of these. The resulting sequence of return volatility estimates are then used in the Kalman filter as if they are the correct return variances. Let \bar{V}_j denote the actual realized volatility and let \hat{V}_j denote the one of the Kalman filter based estimators of it. Thus,

$$\begin{aligned}\bar{V}_j &= \sum_{t=(j-1)78+1}^{78j} r_t^2 \\ \hat{V}_j &= \sum_{t=(j-1)78+1}^{78j} (\hat{r}_t^2 + \hat{b}_t).\end{aligned}$$

As a statistical criteria for appraising the quality of various realized volatility estimators, we measure the mean squared error of each type of estimator by

$$MSE_{xy} = \frac{1}{10000} \sum_{j=1}^{10000} (\bar{V}_j - \hat{V}_j)^2$$

where $x = s$ or f depending on whether the estimated returns and bias corrections are based on smoothed estimates or filtered estimates, respectively, and $y = n, r$, or o if the estimated returns and bias corrections are based on the naive assumption of constant volatility, the rolling volatility estimators, or the infeasible optimal estimators, respectively. We then measure the relative efficiency of the various types of estimators relative to infeasible benchmarks that require knowledge of instantaneous return volatility. These results are listed in Table (1).

Next, in order to check for the robustness of the various estimators to diurnal patterns, we modify the simulation scheme as follows. We first generate a sequence of return variances, $\sigma_{(m),t}^2$, as in Equation (29). Next, we construct a new sequence of return variance $\sigma_{(m),t}^{2*}$ where

$$\sigma_{(m),t}^{2*} = \sigma_{(m),t}^2 \left(1 + 1/3 \cos \left(\frac{2\pi}{78} t \right) \right).$$

Then, the latent returns are generated using $r_{(m),t} = \sigma_{(m),t}^{*} z_{(m),t}$ and noise is generated as before. Notice that the expected variance is twice as high when the diurnal term is largest compared to when it is smallest. The diurnal patterns mimic the U-shaped pattern observed in real markets as we have arranged for the maximum of the cosine term to correspond to the beginning and ending of the days. These results are listed in Table (2).

The relative efficiencies are surprisingly robust to the presence of diurnal patterns. The improvements from smoothing, relative to filtering, are shown in the

		MSE_{fn}	MSE_{sn}	MSE_{fr}	MSE_{sr}	MSE_{fo}
$\rho = -0.4$	MSE_{fo}	7.6381	6.1246	5.1509	4.3131	1.0000
$\sigma_\eta^2 = 15.778$	MSE_{so}	8.3937	6.7306	5.6605	4.7398	1.0989
$\rho = -0.3$	MSE_{fo}	6.9550	5.5434	3.5496	3.0532	1.0000
$\sigma_\eta^2 = 5.917$	MSE_{so}	7.5725	6.0356	3.8648	3.3243	1.0888
$\rho = -0.2$	MSE_{fo}	5.7041	4.7485	2.2763	2.0704	1.0000
$\sigma_\eta^2 = 2.630$	MSE_{so}	6.0819	5.0630	2.4270	2.2075	1.0662
$\rho = -0.1$	MSE_{fo}	3.7725	3.4057	1.4410	1.3819	1.0000
$\sigma_\eta^2 = 0.986$	MSE_{so}	3.9121	3.5317	1.4943	1.4330	1.0370

Table 1: Relative efficiency with no diurnal pattern.

		MSE_{fn}	MSE_{sn}	MSE_{fr}	MSE_{sr}	MSE_{fo}
$\rho = -0.4$	MSE_{fo}	7.4039	5.9493	4.9939	4.1897	1.0000
$\sigma_\eta^2 = 15.778$	MSE_{so}	8.1353	6.5371	5.4872	4.6036	1.0988
$\rho = -0.3$	MSE_{fo}	6.8300	5.4567	3.4688	2.9925	1.0000
$\sigma_\eta^2 = 5.917$	MSE_{so}	7.4268	5.9335	3.7719	3.2539	1.0874
$\rho = -0.2$	MSE_{fo}	5.6596	4.7231	2.2449	2.0461	1.0000
$\sigma_\eta^2 = 2.630$	MSE_{so}	6.0287	5.0311	2.3913	2.1795	1.0652
$\rho = -0.1$	MSE_{fo}	3.7774	3.4187	1.4406	1.3814	1.0000
$\sigma_\eta^2 = 0.986$	MSE_{so}	3.9151	3.5433	1.4931	1.4317	1.0364

Table 2: Relative efficiency with a diurnal pattern.

last column. When noise variance is about an order of magnitude smaller than the expected innovation variance (when $\rho = -0.1$), the mean squared error of the realized volatility estimate is about 4 percent larger when based on filtered returns. When noise variance is roughly twice as large as the expected innovation variance (when $\rho = -0.4$), the filter based mean squared error is about 10 percent larger. Larger gains are achieved by the estimator based on the rolling volatility proxy, especially when noise volatility is relatively small. The mean squared errors based on the naive estimators are between 40??? and 140???? percent larger than corresponding mean squared errors based on the volatility proxy.

Although currently used filters vary widely, we are aware of none that exploit the gains available from either smoothing or from the use of a high-frequency volatility proxy. Most filtering methods in use are similar to the filtered naive estimator. Notice that the mean squared errors of the filtered naive estimators are more than double those of the smoothed estimators based on our proposed smoothed estimator based on the volatility proxy.

6 Conclusions

This article applies market microstructure theory to the problem of denoising a popular volatility estimate. The theory suggests that a Kalman smoother can optimally extract the latent squared returns, which are required for determining realized volatility from their noisy observable counterparts. However, the correct

specification of the filter requires knowledge of a latent stochastic volatility state variable, and is therefore infeasible. We show that a feasible Kalman smoothing algorithm based on a simple rolling regression proxy for high-frequency volatility can improve realized volatility estimates. In simulations, the algorithm substantially reduces the mean squared error of realized volatility estimators even in the presence of strong diurnal patterns. The broad conclusion is that realized volatility estimators can be improved in an obvious way, by smoothing instead of merely filtering the data, and in a less obvious way, by bias correcting and using a straightforward proxy of latent high-frequency volatility.

References

- Ait-Sahalia, Y., P. Mykland, and L. Zhang (2003), How often to sample a continuous-time process in the presence of market microstructure noise, Working Paper w9611, NBER.
- Andersen, T.G. and T. Bollerslev (1998), "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39, 885-905.
- Andersen, T.G., T. Bollerslev and F.X. Diebold (2002), "Parametric and Nonparametric Measurements of Volatility," In Y. Ait-Sahalia and L.P. Hansen (Eds.) *Handbook of Financial Econometrics*. Elsvier 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 (2000), "Exchange Rate Returns Standardized by Realized Volatility are (Nearly) Gaussian," *Multinational Finance Journal*, 4, 159-179.
- Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001a), "The Distribution of Realized 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," NBER Working Paper No. 8160.

Andersen, T.G., T. Bollerslev and S. Lange (1999), "Forecasting Financial Market Volatility Sample Frequency vis-a-vis Forecast Horizon," Journal of Empirical Finance, 6, 457-477.

Andreou, E. and E. Ghysels (2002), "Rolling Sample Volatility Estimators: Some New Theoretical, Simulation and Empirical Results," Journal of Business and Economic Statistics, 20, 363-376.

Atchison, M., K. Butler, and R. Simonds (1987), "Nonsynchronous Security Trading and Market Index Autocorrelation," Journal of Finance, 42, 111-118.

Bai, X., J. R. Russell and G. C. Tiao (2000), "Beyond Merton's Utopia: Effects of Non-Normality and Dependence on the Precision of Variance Estimates Using High-Frequency Financial Data," Manuscript, Graduate School of Business, University of Chicago.

Ball, C. (1988) "Estimation Bias Induced by Discrete Security Prices," Journal of Finance, 43, 841-865.

Bandi, F., and J. Russell (2003) "Microstructure Noise, Realized Volatility, and Optimal Sampling" Manuscript, The University of Chicago.

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 (2002b), "Estimating Quadratic Variation Using Realised Variance," *Journal of Applied Econometrics*, forthcoming.

Barndorff-Nielsen, O. E. and N. Shephard (2002c), "How Accurate is the Asymptotic Approximation to the Distribution of Realized Volatility?" in D. W. F. Andrews, J. L.

Bollerslev, T. and H. Zhou (2002), "Estimating Stochastic Volatility Diffusion Using Conditional Moments of Integrated Volatility," *Journal of Econometrics*, 109, 33-65.

Brockwell, P. and R. Davis (1987), "Time series : theory and methods," New York ; Berlin: Springer-Verlag.

Campbell, A., W. Lo, A. MacKinlay (1997), "The Econometrics of Financial Markets," Princeton University Press.

Cho, D., and E. Frees (1988) "Estimating the Volatility of Discrete Stock Prices," *Journal of Finance*, 43, 451-466.

Cohen, K., S. Maier, R. Schwartz, and D. Whitcomb (1978), "The Returns Generation Process, Returns Variance, and the Effect of Thinness in Securities Markets," *Journal of Finance*, 33, 149-167.

Cohen, K., S. Maier, R. Schwartz, and D. Whitcomb (1979), "On the Existence of Serial Correlation in an Efficient Securities Market," *TIMS Studies in the Management Sciences*, 11, 151-168.

Corsi, F., G. Zumbach, U. Müller, and M. M. Dacorogna (2001), "Consistent High-Precision Volatility from High-Frequency Data," *Economic Notes*, 30, 183-204.

Drost, F. and B. Werker (1996) "Closing the GARCH Gap: Continuous Time GARCH Modeling," *Journal of Econometrics* 74, 31-57.

Engle, R. (1982) "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1007.

Gottlieb, G., and A. Kalay (1985) "Implications of the Discreteness of Observed Stock Prices," *Journal of Finance*, 40, 135-154.

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

Hamilton, J. (1994) *Time Series Analysis*. Princeton: Princeton University Press.

Hansen, P., and A. Lunde, 2004, "An Unbiased Measure of Realized Variance," manuscript Department of Economics, Brown University.

Hasbrouck, J. (1996), "Modeling Market Microstructure Time Series," in G. S. Maddala and C. R. Rao (eds.) *Handbook of Statistics*, Vol.14, 647-692. Amsterdam: North-Holland.

Hasbrouck, J. and T. Ho. (1987), "Order Arrival, Quote Behaviour, and the Return-Generating Process," *Journal of Finance*, 42, 1035-1048.

Lo, A., and A. C. MacKinlay (1988), "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test," *Review of Financial Studies*, 1, 41-66.

Lo, A., and A. C. MacKinlay (1990), "An Econometric Analysis of Nonsynchronous-Trading," *Journal of Econometrics*, 45, 181-212.

Madhavan, A. (2000) "Market Microstructure: A Survey." Working paper, Marshall School of Business, USC.

Newey,W. and K. West (1987), "A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica* 55, 703–708.

O'Hara, M. (1995) *Market Microstructure Theory*, Blackwell Publishers, Cambridge, MA.

Roll, R. (1984), "A Simple Measure of the Effective Bid-Ask Spread in an Efficient Market." *Journal of Finance*, 39, 1127-1139.

Oomen, R., 2004, "Properties of Bias Corrected Realized Variance in Calendar Time and Business Time," Manuscript Warwick Business School.

Zhang, L., P. Mykland, and Y. Ait-Sahalia, 2003, "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High Frequency Data," manuscript Princeton University.

Zhou, B. (1996), "High-Frequency Data and Volatility in Foreign Exchange Rates," Journal of Business and Economic Statistics, 14, 45-52.

Zumbach, G., F. Corsi and A. Trapletti (2002), "Efficient Estimation of Volatility using High Frequency Data," Manuscript, Olsen & Associates, Zürich, Switzerland.