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.

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

Econometric Analysis of Financial and Economic Time Series/Part A
Advances in Econometrics, Volume 20, 211–227

ISSN: 0731-9053/doi:10.1016/S0731-9053(05)20008-7

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 Ebens (2001) and Barndorff-Nielsen and Shephard (2002a,b) 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.

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 (1998) and Andreou and Ghysels (2002), among many 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 th
frictions, as a group, cannot b

Essentially, three strands c
microstructure noise in realize
remove the noise with a simp
Andersen, Bollerslev, Diebol
Müller, and Dacorogna (200
based on a volatility signature
In contrast, Russell and Bar
microstructure noise. Rather t
determine an optimal samplin
they construct a mean-squarec
precision against the correspo
pling frequency increases. Alth
for a given asset, the optimal
assets. Ait-Sahalia, Mykland,
similar treatments. Finally, H
West (1987) type correction to
returns.

Theory suggests that noise
ever, it is known that return
contributions of noise and a
During periods of high-retur
inate the noise in size. In co
estimator in which the weight
during periods of high volatil

Our argument has three par
in terms of market microstruc
filter can be used to remove
attention to how the variabil
high-frequency volatility. Thi
removing the noise.

To formalize, consider a sequ
example) indexed by t. The
where p_t denotes true price
observed return is

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, Müller, and Dacorogna (2001) select a sample frequency of five minutes based on a volatility signature plots, and then apply a moving-average filter. In contrast, Russell and Bandi (2004) 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, Mykland, and 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 toward 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 formalize, consider a sequence of fixed intervals (five-minute periods, for example) indexed by t. The log of the observed price at t is $\tilde{p}_t = p_t + \eta_t$, where p_t denotes true price and η_t denotes microstructure noise. The observed return is

$$\tilde{r}_t = r_t + \varepsilon_t \tag{1}$$

where $\tilde{r}_t = p_t - p_{t-1}$ is the latent (true) return and $\varepsilon_t = \eta_t - \eta_{t-1}$ is the return noise.

We employ assumptions typical of the realized volatility literature. Our first assumption concerns the latent (true) price process.

Assumption 1. (*Latent price process*). The log true price process is a continuous local martingale. Specifically,

$$p_\tau = \int_0^\tau v_s \mathrm{d}w_s$$

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

$$V_\tau = \int_0^\tau v_s \mathrm{d}s < \infty$$

with probability one for all τ.

Assumptions about noise dynamics must be selected with care. Close study of microstructure noise reveals strong positive correlation at high frequency.[3] The correlation declines sharply with the sampling frequency, due to intervening transactions. To understand these effects, we discuss three prominent sources of noise.

The bid–ask bounce, discussed by Roll (1984), arises because transactions cluster at quotes rather than the true price. Hasbrouck and Ho (1987) show that this source of noise may be positively correlated as a result of clustered trade at one quote (due to the break up of large block trades). However, the positive noise correlation due to trade clustering nearly vanishes between trades more than a few transactions apart. In a similar fashion, positive noise correlation arising from the common rounding of adjacent transactions, vanishes at lower sampling frequencies.

The nonsynchronous trading effect, discussed by Lo and MacKinlay (1990), arises when transactions are relatively infrequent. If transactions are infrequent relative to the measurement of prices at regular intervals, then multiple price measurements refer to the same transaction, inducing positive noise correlation. Again, the positive noise correlation vanishes as the sampling frequency declines.

To determine the sampling frequency at which noise is uncorrelated, Hasbrouck and Ho (1987) study a large sample of NYSE stocks. They find no significant correlation for observations sampled more than 10 transactions

apart. Hansen and Lund
study of Dow Jones Indu
that the sampling interval
microstructure noise as an

Assumption 2. (*Microst*
i.i.d. sequence of rand
$\sigma_\eta^2 < \infty$ and independen

We do not make any
noise. However, as the n
dependent features, and
assumption of normally di
Consequently, we consid
Section 3.

Under Assumption 2, it
with a unit root. To deter
we assume that latent ret

Lemma 1. If in additio
weakly stationary proc
tional variance σ_r^2, then
obeys

$Cov(\tilde{r}_t,$

Moreover, the first-order

For each day, which conta
measures.

1. The integrated volatilit
2. An infeasible estimator
3. A feasible estimator, $\hat{V}$

To motivate the form o
error as

apart. Hansen and Lunde (2004) find supporting evidence in their recent study of Dow Jones Industrial Average stocks. In consequence, we assume that the sampling interval contains at least 10 transactions to justify treating microstructure noise as an i.i.d. sequence.

Assumption 2. (*Microstructure noise*). The microstructure noise forms an i.i.d. sequence of random variables each with mean zero and variance $\sigma_\eta^2 < \infty$ and independent of the latent 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 normally distributed noise may be a plausible approximation. Consequently, we consider normally distributed microstructure noise in Section 3.

Under Assumption 2, it is clear that return noise forms an MA(1) process with a unit root. To determine the covariance structure of observed returns, we assume that latent returns form a weakly stationary martingale.

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

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

Moreover, the first-order autocorrelation is given by

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

For each day, which contains n intervals, define the following three volatility measures.

1. The integrated volatility, $V = \Sigma_{t=1}^n \sigma_t^2$.
2. An infeasible estimator, constructed from latent returns, $\overline{V} = \Sigma_{t=1}^n r_t^2$.
3. A feasible estimator, $\hat{V} = \Sigma_{t=1}^n \hat{r}_t^2$, where $\hat{r}_t = \tilde{r}_t$ in the absence of noise.

To motivate the form of the feasible estimator, decompose the estimation error as

$$V - \hat{V} = V - \overline{V} + \overline{V} - \hat{V} \tag{2}$$

The behavior of $V - \overline{V}$ as a function of step length and the underlying volatility process has been studied by Barndorff-Nielsen and Shephard (2002a). If the step length is chosen (hence n is fixed), then this part of the error is beyond the control of the researcher. Therefore, we focus on minimizing the mean squared error $\mathrm{E}(\overline{V} - \hat{V})^2$, where $\tilde{r}' = (\tilde{r}_1, \cdots, \tilde{r}_T)$ and $T = n \cdot J$ (J is the number of days in the sample). It is well known that the mean squared error is minimized by choosing

$$\hat{V} = \mathrm{E}(\overline{V}|\tilde{r}) = \mathrm{E}\left(\sum_{t=1}^{n} r_t^2 \middle| \tilde{r}\right) = \sum_{t=1}^{n} \mathrm{E}(r_t^2|\tilde{r})$$

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

2.1. *Kalman Filter and Smoother*

The Kalman filter provides a technique to separate (observed) contaminated returns into two components: the first corresponds to (latent) true returns and the second to microstructure noise. To construct the filter, we follow the notation in Hamilton (1994) (Harvey, 1989 also provides textbook treatment). The state vector consists of latent variables, $\xi_t' = (r_t, \eta_t, \eta_{t-1})$. The observation equation relates the state vector to observed returns

$$\tilde{r}_t = \mathbf{H}'\xi_t \tag{3}$$

where $H' = (1, 1, -1)$. The state equation describes the dynamic evolution of the state vector

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

where $v_t' = (r_t, \eta_t)$ with covariance matrix

$$Q_t = \begin{pmatrix} \sigma_t^2 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix}$$

and the coefficient matrice

$$\mathbf{F} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

The Kalman filter deliv
given the sequence of obse
livers the corresponding li
While these linear projecti
have larger MSE's than nc
need not be linear project
and conditional expectatio
$\{\tilde{r}_t, \tilde{r}_{t-1}, \ldots, \tilde{r}_1, 1\}$. Let $\hat{\xi}_{\tau|}$

$$\mathbf{P}_{\tau|t} =$$

represent the mean squar
ample, the one-step-ahead
matrix with the first diago
σ_η^2. The third diagonal ele
determined though a recu
step-ahead prediction erro
the variance of u_t, which
$\sigma_t^2 + \sigma_\eta^2 + c_t$.

The projections from th

$\hat{r}_{t|}$

and the coefficient matrices are

$$\mathbf{F} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{R} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

The Kalman filter delivers the linear projection of the state vector ξ_t, given the sequence of observations $\{\tilde{r}_1, \cdots, \tilde{r}_t\}$. The Kalman smoother delivers the corresponding linear projection onto the extended sequence $\tilde{r}$. While these linear projections make efficient use of information, they may have larger MSE's than nonlinear projections. As conditional expectations need not be linear projections, we distinguish between linear projections and conditional expectations. Let $\hat{\mathrm{E}}_t$ represent linear projection onto $F_t = \{\tilde{r}_t, \tilde{r}_{t-1}, \ldots, \tilde{r}_1, 1\}$. Let $\hat{\xi}_{\tau|t} = \hat{\mathrm{E}}_t(\xi_\tau)$ and let

$$\mathbf{P}_{\tau|t} = \mathrm{E}\left[\left(\xi_\tau - \hat{\xi}_{\tau|t}\right)\left(\xi_\tau - \hat{\xi}_{\tau|t}\right)'\right] \tag{5}$$

represent the mean squared error matrices of these projections. For example, the one-step-ahead mean squared error matrix $\mathbf{P}_{t|t-1}$ is a diagonal matrix with the first diagonal element equal to σ_t^2 and the second equal to σ_η^2. The third diagonal element we define as $c_t = Var(\hat{\eta}_{t-1|t-1})$. The c_t are determined though a recursion described below. Let u_t denote the one-step-ahead prediction error for the observed returns. Then it follows that the variance of u_t, which we denote by M_t, is given by $M_t = H' P_{t|t-1} H = \sigma_t^2 + \sigma_\eta^2 + c_t$.

The projections from the Kalman filter are given by the recursion

$$\hat{r}_{t|t}(\sigma_t^2) = \frac{\sigma_t^2}{M_t}\left(\tilde{r}_t + \hat{\eta}_{t-1|t-1}\right) \tag{6}$$

$$\hat{\eta}_{t|t} = \frac{\sigma_\eta^2}{M_t}\left(\tilde{r}_t + \hat{\eta}_{t-1|t-1}\right) \tag{7}$$

$$c_{t+1} = \frac{\sigma_\eta^2(\sigma_t^2 + c_t)}{M_t} \tag{8}$$

The recursion and the boundary conditions $\hat{\eta}_{t|t} = 0$ and $c_1 = \sigma_\eta^2$ determine the sequence of filtered returns and filtered noise.

The projections from the Kalman smoother, which are the elements of $\hat{\xi}_{t|T} = \hat{E}_T(\xi_t)$, are

$$\hat{r}_{t|T}(\sigma_t^2) = \hat{r}_{t|t}(\sigma_t^2) - \frac{\sigma_t^2}{\sigma_t^2 + c_t}(\hat{\eta}_{t|T} - \hat{\eta}_{t|t}) \tag{9}$$

$$\hat{\eta}_{t-1|T} = \hat{\eta}_{t-1|t} + \frac{c_t}{\sigma_t^2 + c_t}(\hat{\eta}_{t|T} - \hat{\eta}_{t|t}) \tag{10}$$

Smoothed quantities exhibit smaller variances than their filtered counterparts. For example, if we let $d_{t+1} = Var(\hat{\eta}_{t|T})$, then it can be shown that

$$d_t = c_t\left(\frac{\sigma_t^2 + \sigma_\eta^2}{\sigma_t^2 + \sigma_\eta^2 + c_t}\right) - \left(\frac{c_t}{\sigma_t^2 + c_t}\right)^2 (c_{t+1} - d_{t+1}) \tag{11}$$

As is easily verified from their definitions, $d_{T+1} = c_{T+1}$. Consequently, $d_T < c_T$, and, by an induction argument, it follows that $d_t < c_t$ for $t = 1, 2, \ldots, T$, establishing that $Var(\hat{\eta}_{t|T}) < Var(\hat{\eta}_{t|t})$ for $t = 1, 2, \ldots, T-1$.

The smoother estimates the latent returns as weighted averages of contemporaneous, lagged, and future observed returns. If variance is constant, so that $\sigma_t^2 = \sigma_r^2$ for all t, then the weights are nearly the same for all smoothed returns. To see the point clearly, consider a numerical example. If $\sigma_r^2 = 10$, $\sigma_\eta^2 = 1$, and $T = 7$, then, ignoring weights less than 0.001, we have that

$$\tilde{r}_{4|7} = 0.006\tilde{r}_2 + 0.0709\tilde{r}_3 + 0.8452\tilde{r}_4 + 0.0709\tilde{r}_5 + 0.006\tilde{r}_6$$

while

$$\tilde{r}_{3|7} = 0.0059\tilde{r}_1 + 0.0709\tilde{r}_2 + 0.8452\tilde{r}_3 + 0.0709\tilde{r}_4 + 0.006\tilde{r}_5.$$

Thus, an almost identical weighting scheme determines the third and fourth optimally estimated latent returns. 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 optimal weights for estimating latent returns in (6)–(8) and (9)–(10) are not constant. Instead, during periods of high volatility, the optimal weights are larger for the currently observed return, and lower for the other returns.

With the estimated latent ret
realized volatility by $\Sigma_{t=1}^{n} \hat{r}_t^2$
filtering is a linear transform
biased estimator of $E(r_t^2|\tilde{r})$. I
determined in the normal co
the bias $E\left(r_t^2 - \hat{r}_{t|T}^2\right)$, the pr

$$E\left[\left(r_t - \hat{r}_{t|T}^2\right)^2\right] = E$$

Thus, the bias equals the (1,
$\hat{r}_{t|T}$.

We find that the bias is

$$b_t(\sigma_t^2) = b_t^f$$

where $b_t^f(\sigma_t^2)$ (the bias of th

$$b_t^f(\sigma_t^2$$

Recall that c_t is the element
and d_t is the corresponding
As shown in (11), for $t < T$ t
the filtered estimator.[5] As
smaller than the bias of the
smoothed estimator in what

To determine the magnitu
$\sigma_t^2 = \sigma_r^2$ (constant volatility)

$$\sigma_r^2$$

In accord with intuition, t
variance and an increasing f
the bias is 15 percent of the
expected squared noise ter
$\sigma_r^2/\sigma_\eta^2 \simeq 0$, then the bias is
inates, then the bias is near

2.2. Bias

With the estimated latent returns $\hat{r}_{t|T} = \hat{\mathrm{E}}(r_t|\tilde{r})$, it seems natural to estimate realized volatility by $\Sigma_{t=1}^{n} \hat{r}_{t|T}^2$ (note, $\hat{r}_{t|T}^2$ stands for $(\hat{r}_{t|T})^2$). Yet, because filtering is a linear transformation, while squaring is not, $\hat{r}_t^2$ is a downward biased estimator of $\mathrm{E}(r_t^2|\tilde{r})$. Fortunately, the size and direction of the bias is determined in the normal course of constructing the Kalman smoother. For the bias $\mathrm{E}\left(r_t^2 - \hat{r}_{t|T}^2\right)$, the properties of projection mappings imply[4]

$$\mathrm{E}\left[\left(r_t - \hat{r}_{t|T}^2\right)^2\right] = \mathrm{E}\left[\hat{\mathrm{E}}_T\left(r_t^2 - 2\hat{r}_{t|T} r_t + \hat{r}_{t|T}^2\right)\right] = \mathrm{E}\left[r_t^2 - \hat{r}_{t|T}^2\right] \tag{12}$$

Thus, the bias equals the (1, 1) element of the mean squared error matrix for $\hat{r}_{t|T}$.

We find that the bias is

$$b_t(\sigma_t^2) = b_t^f(\sigma_t^2) - \left(\frac{\sigma_t^2}{\sigma_t^2 + c_t}\right)^2 (c_{t+1} - d_{t+1}) \tag{13}$$

where $b_t^f(\sigma_t^2)$ (the bias of the filtered return $\hat{r}_{t|t}$) is

$$b_t^f(\sigma_t^2) = \sigma_t^2 \left(\frac{\sigma_\eta^2 + c_t}{\sigma_t^2 + \sigma_\eta^2 + c_{t-1}}\right) \tag{14}$$

Recall that c_t is the element of the variance for the filtered prediction of η_{t-1} and d_t is the corresponding variance element for the smoothed prediction. As shown in (11), for $t < T$ the smoothed estimator has lower variance than the filtered estimator.[5] As a result, the bias of the smoothed estimator is smaller than the bias of the filtered estimator, and so we concentrate on the smoothed estimator in what follows.

To determine the magnitude of the bias, consider the simple case in which $\sigma_t^2 = \sigma_r^2$ (constant volatility). The bias, b_t, is well approximated by

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

In accord with intuition, the bias is a decreasing function of the return variance and an increasing function of the noise variance. If $\sigma_\eta^2/\sigma_r^2 = .1$, then the bias is 15 percent of the return variance and 50 percent larger than the expected squared noise term.[6] If the noise variance dominates, so that $\sigma_r^2/\sigma_\eta^2 \simeq 0$, then the bias is approximately σ_r^2. If the return variance dominates, then the bias is near zero.

3. MULTIVARIATE NORMAL APPROACH

To analyze the multivariate normal case, it is convenient to work in vector form. Let $\mathbf{r} = (r_1, r_2, \cdots, r_T)'$ and $\eta = (\eta_0, \eta_1, \ldots, \eta_T)'$ so that

$$\tilde{\mathbf{r}} = \mathbf{r} + \mathbf{B}\eta$$

Here $\mathbf{B}$ is a selection matrix with first row $[-1, -1, 0, \cdots, 0]$. The covariance matrix of r is $\Lambda = diag(\sigma_t^2)$. The Kalman smoother equations (in vector form) are

$$\hat{\mathbf{r}} = \Lambda\left(\Lambda + \sigma_\eta^2 \mathbf{B}\mathbf{B}'\right)^{-1}\tilde{\mathbf{r}} \text{ and } \Sigma = \sigma_\eta^2 \mathbf{B}\left(\mathbf{I} + \sigma_\eta^2 \mathbf{B}'\Lambda^{-1}\mathbf{B}\right)^{-1}\mathbf{B}' \tag{15}$$

From Assumption 1, it follows that $r_t|\sigma_t^2 \sim N(0, \sigma_t^2)$. If we extend the assumption to

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

then it is simple to derive the conditional distribution of $\tilde{r}|r$. Specifically,

$$\mathbf{r}|\tilde{\mathbf{r}} \sim N_T(\hat{\mathbf{r}}, \Sigma) \tag{16}$$

where $\hat{\mathbf{r}}$ and Σ are identical to the quantites from the Kalman smoother (15).

Under the assumption of joint normality $\hat{r} = \mathrm{E}(r|\tilde{r})$, so, the smoothed estimator is the conditional expectation rather than simply the optimal linear projection. Similarly, $\Sigma = Var(r|\hat{r})$ rather than simply the MSE matrix of the linear projection. This is especially useful for understanding the source of the bias that arises from squaring filtered returns. Here

$$Var(r_t|\tilde{\mathbf{r}}) = \mathrm{E}\left(r_t^2|\tilde{\mathbf{r}}\right) - \mathrm{E}^2(r_t|\tilde{\mathbf{r}}) \tag{17}$$

The optimal estimator $\mathrm{E}\left(r_t^2|\tilde{\mathbf{r}}\right)$ exceeds the square of the optimal estimator for latent returns $\mathrm{E}^2(r_t|\tilde{\mathbf{r}})$. The correction term $Var(r_t|\tilde{\mathbf{r}})$ does more than simply correct for the bias. Because the correction term corresponds to the conditional covariance matrix of r given the observed returns, the correction delivers the conditional expectation of squared returns. In consequence, we are able to form an optimal nonlinear estimator from an optimal linear estimator as

$$\mathrm{E}\left(r_t^2|\tilde{\mathbf{r}}\right) = \mathrm{E}^2(r_t|\tilde{\mathbf{r}}) + Var(r_t|\tilde{\mathbf{r}})$$

4. IM

The above analysis, in which suggests the use of the bias-c

$$\hat{V}\left(\sigma_t^2\right) =$$

To implement the method, w

If the latent return varianc bias-corrected estimator $\hat{V}(\sigma_t^2$ ma 1, the first two autocovar ficient for determining the var the true returns. (If one wishe then ML estimators may be timators.) Andersen et al. (2 similar to $\hat{V}\left(\sigma_r^2\right)$, does not cc correction.

For the case in which the re $\hat{r}_{t|T}^2\left(\hat{\sigma}_r^2\right)$ and $b_t\left(\hat{\sigma}_r^2\right)$. We then ing window.[7]

$$\hat{\sigma}_{t|T}^2 = \frac{1}{25}$$

The estimated time-varying $\hat{r}_{t|T}^2\left(\hat{\sigma}_t^2\right)$ from (9) and $b_t\left(\hat{\sigma}_t^2\right)$ f

For the case of constant va sistency of $\hat{\sigma}_r^2$ and $\hat{\sigma}_\eta^2$. Similar Sahalia et al. (2003). To esta constant, it seems natural to than focus on this problem, w a general purpose filter that r

5. P

To test the performance of th use a model for the simulated behavior of the S&P 500 sto tion 1 is

4. IMPLEMENTATION

The above analysis, in which it is assumed that $\{\sigma_t^2\}$ and σ_η^2 are known, suggests the use of the bias-corrected estimator

$$\hat{V}(\sigma_t^2) = \sum_{t=1}^{n} \left[\hat{r}_{t|T}^2(\sigma_t^2) + b_t(\sigma_t^2)\right]$$

To implement the method, we need estimators of $\{\sigma_t^2\}$ and σ_η^2.

If the latent return variance is assumed constant, then $\sigma_t^2 = \sigma_r^2$ and the bias-corrected estimator $\hat{V}(\sigma_r^2)$ is a function only of σ_r^2 and σ_η^2. From Lemma 1, 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. (If one wishes to make further distributional assumptions, then ML estimators may be used in place of the method of moments estimators.) Andersen et al. (2001) employ an MA(1) estimator that, while similar to $\hat{V}(\sigma_r^2)$, does not contain smoothed estimates and makes no bias correction.

For the case in which the return variances are not constant, we begin with $\hat{r}_{t|T}^2(\hat{\sigma}_r^2)$ and $b_t(\hat{\sigma}_r^2)$. We then estimate the time-varying variance with a rolling window.[7]

$$\hat{\sigma}_{t|T}^2 = \frac{1}{25} \sum_{k=t-12}^{t+12} \left(\hat{r}_{k|T}^2(\hat{\sigma}_r^2) + b_k(\hat{\sigma}_r^2)\right) \tag{18}$$

The estimated time-varying variances from (18) together with $\hat{\sigma}_\eta^2$, yield $\hat{r}_{t|T}^2(\hat{\sigma}_t^2)$ from (9) and $b_t(\hat{\sigma}_t^2)$ from (13).

For the case of constant variance, laws of large numbers ensure the consistency of $\hat{\sigma}_r^2$ and $\hat{\sigma}_\eta^2$. Similar results are derived for ML estimators in Ait-Sahalia et al. (2003). To establish consistency if the return variance is not constant, it seems natural to specify a dynamic structure for $\{\sigma_t^2\}$. Rather than focus on this problem, we seek to recover latent realized volatility with a general purpose filter that minimizes mean squared error.

5. PERFORMANCE

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. A popular special case of Assumption 1 is

$$\begin{aligned} dp_t &= \sigma_t dw_t \\ d\sigma_t^2 &= \theta(\omega - \sigma_t)dt + (2\lambda\theta)^{1/2} dw_{\sigma_t} \end{aligned} \tag{19}$$

where w_t and w_{σ_t} are independent Brownian motion. Drost and Werker (1996, Corollary 3.2) provide the map between (19) and the (discrete) GARCH(1,1),

$$\begin{aligned} p_t - p_{t-1/m} &= r_{(m),t} = \sigma_{(m),t}\, z_{(m),t} \\ \sigma^2_{(m),t} &= \phi_{(m)} + \alpha_{(m)} r^2_{(m),t} + \beta_{(m)} \sigma^2_{(m),t-1/m} \end{aligned} \tag{20}$$

where $z_{(m),t}$ is (for the purposes of simulation) i.i.d. $N(0,1)$. Andreou and Ghysels (2002) find that 5-minute returns from the S&P 500 index are well approximated by the values

$$\phi_{(m)} = 0.0004, \quad \alpha_{(m)} = 0.0037, \quad \beta_{(m)} = 0.9963. \tag{21}$$

These parameters imply an unconditional return variance of $\sigma_r^2 = 7.9$ basis points over the 5-minute interval. While this unconditional variance is high (daily estimates of return variance are roughly eight basis points), an appropriate rescaling by multiplying by 1/78 results in such small parameter values that simulation is difficult. As the relative mean squared error measurements that we report are invariant to such scaling, we follow Andreou and Ghysels and use the values in (21).

From (20) and (21) we simulate latent returns, r_t. We construct observed returns as $\tilde{r}_t = r_t + \eta_t - \eta_{t-1}$, where η_t is generated as an i.i.d. $N\left(0, \sigma_\eta^2\right)$ random variable. To determine the noise variance, we invert the formula for ρ in Lemma 1 to obtain

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

Hasbrouck and Ho (1987) report estimates of ρ between $-.4$ and $-.1$, so we allow ρ to take the values $[-.4, -.3, -.2, -.1]$. As decreasing the value of ρ increases σ_η^2, the resultant values of noise variance vary from $\sigma_\eta^2 = 1$ (for $\rho = -.1$) to $\sigma_\eta^2 = 15.8$ (for $\rho = -.4$).

To mirror trading days on the NYSE, which are 6.5 hours long, each simulated day contains 78 5-minute returns. We generate 10,000 trading days, a span that roughly corresponds to 50 years. For each day, j, we construct the latent realized volatility

$$\overline{V} = \sum_{t=(j-1)78+1}^{78j} r_t^2,$$

the feasible bias-corrected realize

$$\hat{V}_j\left(\hat{\sigma}^2_{t|T}\right) = \sum_{t=(j-1}^{78}$$

and the infeasible bias-corrected

To compare this filter to meth such as the MA(1) filter mentio mine the gains from smoothing filtered (rather than smoothed)

$$\hat{V}^f_j\left(\hat{\sigma}^2_{t|t}\right) = \sum_{t=(j-}$$

where $\hat{\sigma}^2_{t|t}$ is obtained from (18) and $b_k\left(\hat{\sigma}_r^2\right)$, respectively. Finally

To judge the quality of the re mean squared error (MSE) of e timal) estimator. For example,

$$\frac{MSE\left[\hat{V}\left(\hat{\sigma}^2_{t|T}\right)\right]}{MSE\left[\hat{V}\left(\sigma_t^2\right)\right]}$$

In Table 1, we present the r the degree of noise variance, o from estimating a time-varying with the smallest noise varian mator is reduced by more than

Table 1.

σ_η^2	Constant Varian
	$\hat{V}^f(\hat{\sigma}_r^2)$
15.8	8.4
5.9	7.6
2.6	6.1
1.0	3.9

the feasible bias-corrected realized volatility estimator

$$\hat{V}_j\left(\hat{\sigma}^2_{t|T}\right) = \sum_{t=(j-1)78+1}^{78j}\left[\hat{r}^2_{t|T}\left(\hat{\sigma}^2_{t|T}\right) + b_t\left(\hat{\sigma}^2_{t|T}\right)\right]$$

and the infeasible bias-corrected estimator $\hat{V}_j(\sigma_t^2)$.

To compare this filter to methods that assume constant return variance, such as the MA(1) filter mentioned above, we construct $\hat{V}_j(\hat{\sigma}_r^2)$. To determine the gains from smoothing, we also construct the estimator based on filtered (rather than smoothed) quantites

$$\hat{V}^f_j\left(\hat{\sigma}^2_{t|t}\right) = \sum_{t=(j-1)78+1}^{78j}\left[\hat{r}^2_{t|t}\left(\hat{\sigma}^2_{t|t}\right) + b^f_t\left(\hat{\sigma}^2_{t|t}\right)\right]$$

where $\hat{\sigma}^2_{t|t}$ is obtained from (18) with $\hat{r}^2_{k|t}(\hat{\sigma}_r^2)$ and $b^f_k(\hat{\sigma}_r^2)$ in place of $\hat{r}^2_{k|T}(\hat{\sigma}_r^2)$ and $b_k(\hat{\sigma}_r^2)$, respectively. Finally, for completeness, we construct $\hat{V}^f_j(\hat{\sigma}_r^2)$.

To judge the quality of the realized volatility estimators, we measure the mean squared error (MSE) of each estimator relative to the infeasible (optimal) estimator. For example, the relative MSE for $\hat{V}\left(\hat{\sigma}^2_{t|T}\right)$ is

$$\frac{MSE\left[\hat{V}\left(\hat{\sigma}^2_{t|T}\right)\right]}{MSE\left[\hat{V}\left(\sigma_t^2\right)\right]} = \frac{\sum_{j=1}^{10000}\left(\overline{V}_j - \hat{V}_j\left(\hat{\sigma}^2_{t|T}\right)\right)^2}{\sum_{j=1}^{10000}\left(\overline{V}_j - \hat{V}_j\left(\sigma_t^2\right)\right)^2}$$

In Table 1, we present the relative efficiency calculations. Regardless of the degree of noise variance, or indeed of the decision to smooth, the gain from estimating a time-varying return variance is substantial. For the case with the smallest noise variance, the relative MSE for the smoothed estimator is reduced by more than half (from 3.5 to 1.4). As one would expect,

Table 1. Relative Efficiency.

σ_η^2	Constant Variance		Time-Varying Variance	
	$\hat{V}^f(\hat{\sigma}_r^2)$	$\hat{V}(\hat{\sigma}_r^2)$	$\hat{V}^f(\hat{\sigma}^2_{t\|t})$	$\hat{V}(\hat{\sigma}^2_{t\|T})$
15.8	8.4	6.7	5.7	4.7
5.9	7.6	6.0	3.9	3.4
2.6	6.1	5.1	2.4	2.2
1.0	3.9	3.5	1.5	1.4

increasing the noise variance renders the estimation problem more difficult, yet even for the highest noise variance the relative MSE for the smoothed estimator is substantially reduced (from 6.7 to 4.7). Moreover, while smoothing always leads to an efficiency gain, the magnitude of the efficiency gain resulting from smoothing is dominated by efficiency gain from allowing for time-varying volatility.

To determine the impact of diurnal patterns, we generate time-varying volatility that mirrors the U-shape pattern often observed in empirical returns. To do so, we construct a new sequence of return variances $\left\{\sigma^{2*}_{(m),t}\right\}$:

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

where $\sigma^{2}_{(m),t}$ is obtained from (20). Note that with the cyclic component, the expected variance doubles between the diurnal peak and trough. This process mimics the U-shape pattern as the maximum of the cosine term to corresponds to the beginning and ending of each day.

In Table 2, we find that the relative MSE measurements are surprisingly robust to the presence of diurnal patterns. When noise variance is about an order of magnitude smaller than the expected innovation variance (when $\sigma^2_\eta = 1.0$), the MSE of the realized volatility estimator is about 4 percent larger when based on filtered returns. When noise variance is roughly twice as large as the expected innovation variance (when $\sigma^2_\eta = 15.8$), 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 160 percent larger than corresponding mean squared errors based on the volatility proxy. The improvements from smoothing, relative to filtering, are shown in the last column.

Table 2. Relative Efficiency with a Diurnal Pattern.

σ^2_η	Constant Variance		Time-Varying Variance	
	$\hat{V}^{-1}(\hat{\sigma}^2_r)$	$\hat{V}(\hat{\sigma}^2_r)$	$\hat{V}^{-1}(\hat{\sigma}^2_{\eta,t})$	$\hat{V}(\hat{\sigma}^2_{\eta,t})$
15.8	8.1	6.5	5.5	4.6
5.9	7.4	5.9	3.8	3.3
2.6	6.0	5.0	2.4	2.2
1.0	3.9	3.5	1.5	1.4

Although currently
exploit the gains availal
frequency volatility pro
filtered naïve estimator
naïve estimators are m
based on our proposed

This article applies ma
ing noise from a popu
Kalman smoother can
are required for determ
counterparts. However
edge of a latent stocha
sible. We show that a
simple rolling regressic
realized volatility estim
duces the mean square
presence of strong diu
volatility estimators ca
instead of merely filte
correcting and using
volatility.

1. Andersen, Bollerslev
empirics for realized vola
2. Surveys by O'Hara
(1997), and Madhavan (2
3. Noise outcomes of a
when no transaction inter
measured price remains c
price changes through tim
4. Brockwell and Davis
5. If $t = T$, then the sn
6. The bias of the filter
7. The rolling window
and variance in the prese

Although currently used filters vary widely, we are aware of none that exploit the gains available from either smoothing or from the used of a high-frequency volatility proxy. Most filtering methods in uses are similar to the filtered naïve estimator. Notice that the mean squared errors of the filtered naïve 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 removing noise from 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.

NOTES

1. Andersen, Bollerslev, and Diebold (2005) provides a survey of both theory and empirics for realized volatility.
2. Surveys by O'Hara (1995), Hasbrouck (1996), Campbell, Lo, & MacKinlay (1997), and Madhavan (2000) document these and other microstructure frictions.
3. 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).
4. Brockwell and Davis (1987, Proposition 2.3.2).
5. If $t = T$, then the smoothed estimator is identical to the filtered estimator.
6. The bias of the filtered estimator is approximately $-2\rho\sigma_r^2$ (recall $\rho < 0$).
7. The rolling window width of 24 corresponds to two hours, which balances bias and variance in the presence of diurnal features.

REFERENCES

Ait-Sahalia, Y., Mykland, P. & Zhang, L. (2003). *How often to sample a continuous-time process in the presence of market microstructure noise.* Working Paper w9611, NBER.

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review, 39,* 885–905.

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2005). Parametric and nonparametric volatility measurement. In: Y. Aït-Sahalia & L. P. Hansen (Eds), *Handbook of financial econometrics.* Amsterdam: Elsevier Science B.V.

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

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

Atchison, M., Butler, K., & Simonds, R. (1987). Nonsynchronous security trading and market index autocorrelation. *Journal of Finance, 42,* 111–118.

Ball, C. (1988). Estimation bias induced by discrete security prices. *Journal of Finance, 43,* 841–865.

Barndorff-Nielsen, O. E. & Shephard, N. (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., & Shephard, N. (2002b). Estimating quadratic variation using realised variance. *Journal of Applied Econometrics, 17,* 457–477.

Brockwell, P., & Davis, R. (1987). *Time series: Theory and methods.* New York: Springer.

Campbell, A., Lo, W., & MacKinlay, A. (1997). *The econometrics of financial markets.* Princeton: Princeton University Press.

Cho, D., & Frees, E. (1988). Estimating the volatility of discrete stock prices. *Journal of Finance, 43,* 451–466.

Cohen, K., Maier, S., Schwartz, R., & Whitcomb, D. (1978). The returns generation process, returns variance, and the effect of thinness in securities markets. *Journal of Finance, 33,* 149–167.

Cohen, K., Maier, S., Schwartz, R., & Whitcomb, D. (1979). On the existence of serial correlation in an efficient securities market. *TIMS Studies in the Management Sciences, 11,* 151–168.

Corsi, F., Zumbach, G., Müller, U., & Dacorogna, M. M. (2001). Consistent high-precision volatility from high-frequency data. *Economic Notes, 30,* 183–204.

Drost, F., & Werker, B. (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., & Kalay, A. (1985). Implications of the discreteness of observed stock prices. *Journal of Finance, 40,* 135–154.

Hamilton, J. (1994). *Time series analysis.* Princeton: Princeton University Press.

Hansen, P., & Lunde, A. (2004). *An unbiased measure of realized variance.* Manuscript, Department of Economics, Brown University.

Harvey, A. (1989). *Forecasting, structural time series models and the Kalman Filter.* Cambridge: Cambridge University Press.

Hasbrouck, J. (1996). Modelin
Rao (Eds), *Handbook o*
Holland.

Hasbrouck, J., & Ho, T. (19
process. *Journal of Fina*

Lo, A., & MacKinlay, A. C. (1
from a simple specificat

Lo, A., & MacKinlay, A. C.
Journal of Econometrics

Madhavan, A. (2000). Marke
205–258.

Newey, W., & West, K. (1987
correlation consistent c

O'Hara, M. (1995). *Market m*

Oomen, R. (2004). *Properties*
time. Manuscript. Wary

Roll, R. (1984). A simple meas
of Finance, 39, 1127–11

Russell, J., & Bandi, F. (2004).
Econometric Society 20

Zhou, B. (1996). High-frequen
ness and Economic Stat

Hasbrouck, J. (1996). Modeling market microstructure time series. In: G. S. Maddala & C. R. Rao (Eds), *Handbook of Statistics*. (Vol. 14, pp. 647–692). Amsterdam: Elsevier-North-Holland.

Hasbrouck, J., & Ho, T. (1987). Order arrival, quote behaviour, and the return-generating process. *Journal of Finance*, *42*, 1035–1048.

Lo, A., & MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, *1*, 41–66.

Lo, A., & MacKinlay, A. C. (1990). An econometric analysis of nonsynchronous-trading. *Journal of Econometrics*, *45*, 181–212.

Madhavan, A. (2000). Market microstructure: A survey. *Journal of Financial Markets*, *3*, 205–258.

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

O'Hara, M. (1995). *Market microstructure theory*. Cambridge: Blackwell Publishers.

Oomen, R. (2004). *Properties of bias corrected realized variance in calendar time and business time*. Manuscript. Warwick Business School.

Roll, R. (1984). A simple measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, *39*, 1127–1139.

Russell, J., & Bandi, F. (2004). *Microstructure Noise, Realized Volatility, and Optimal Sampling*. Econometric Society 2004 Latin American Meetings 220, Econometric Society.

Zhou, B. (1996). High-frequency data and volatility in foreign exchange rates. *Journal of Business and Economic Statistics*, *14*, 45–52.