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
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.

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.

Simulations by Andersen and Bollerslev (1998) and Androul 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 all frictions, as a group, cannot be.

Essentially, three strands of microstructure noise in realize remove the noise with a simple Andersen, Bollerslev, Diebolt Müller, and Dacorogna (2001) based on a volatility signature. In contrast, Russell and Bai microstructure noise. Rather they determine an optimal sampling they construct a mean-square precision against the corresponding frequency increases.

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-return inate the noise in size. In co-estimator in which the weight during periods of high volatility.

Our argument has three parts in terms of market microstructure filter can be used to remove attention to how the variability high-frequency volatility. This removing the noise.

To formalize, consider a sequence 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 $\eta_t$ denotes microstructure noise. The observed return is
\[ \tilde{r}_t = r_t + \epsilon_t \]  

where \( \tilde{r}_t = p_t - p_{t-1} \) is the latent (true) return and \( \epsilon_t = \eta_t - \eta_{t-1} \) 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,

\[ p_t = \int_0^t \nu_s \, dw_s \]

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

\[ V_r = \int_0^r \nu_s \, ds < \infty \]

with probability one for all \( r \).

Assumptions about noise dynamics must be selected with care. Close study of microstructure noise reveals strong positive correlation at high frequency. 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. (Microstructure noise)** A random i.i.d. sequence of random variables \( \sigma_i^2 < \infty \) and independent.

We do not make any assumption about the dependent features, and we cannot assume that \( \sigma_i^2 < \infty \) and independent for all \( i \).

Under Assumption 2, we assume that the latent realizations of \( \epsilon_t \) are independent.

**Lemma 1.** If \( \epsilon_t \) is an i.i.d. sequence of random variables \( \epsilon_i^2 < \infty \), then obeys

\[ \text{Var}(\epsilon_t) \]

Moreover, the first-order correlation

\[ \text{Cor}(\epsilon_t) \]

For each day, which conta measures.

1. The integrated volatility
2. An infeasible estimator
3. A feasible estimator, \( \tilde{V} \)

To motivate the form of the 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 \( \sigma_n^2 < \infty \) 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 \( \sigma_n^2 \), then the autocovariance function of observed returns obeys

\[
Cov(\tilde{r}_t, \tilde{r}_{t-k}) = \begin{cases} 
\sigma_n^2 + 2\sigma_n^2 & \text{if } k = 0 \\
-\sigma_n^2 & \text{if } k = 1 \\
0 & \text{if } k > 1
\end{cases}
\]

Moreover, the first-order autocorrelation is given by

\[
\rho = \frac{-\sigma_n^2}{\sigma_n^2 + 2\sigma_n^2}
\]

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

1. The integrated volatility, \( V = \sum_{t=1}^{n} \sigma_t^2 \).
2. An infeasible estimator, constructed from latent returns, \( \bar{V} = \sum_{t=1}^{n} \tilde{r}_t^2 \).
3. A feasible estimator, \( \tilde{V} = \sum_{t=1}^{n} \tilde{r}_t^2 \), where \( \tilde{r}_t = \tilde{r}_t \) in the absence of noise.

To motivate the form of the feasible estimator, decompose the estimation error as
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 \( E(\overline{V} - \overline{V})^2 \), where \( \hat{r} = (\hat{r}_l, \ldots, \hat{r}_r) \) 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

\[
\hat{V} = E(\overline{V}|\hat{r}) = E\left( \sum_{i=1}^{n} r_i^2 | \hat{r} \right) = \sum_{i=1}^{n} E(\hat{r}_i^2 | \hat{r})
\]

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

\[
\hat{r}_t = H^T \xi_t
\]

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

\[
\xi_{t+1} = F \xi_t + R \eta_{t+1}
\]

where \( \eta_t = (r_t, \eta_t) \) with covariance matrix

\[
Q_t = \begin{pmatrix}
\sigma_r^2 & 0 \\
0 & \sigma_\eta^2
\end{pmatrix}
\]

and the coefficient matrix

\[
F = \begin{pmatrix}
0 \\
0 \\
0
\end{pmatrix}
\]

The Kalman filter delivers the sequence of observed returns the corresponding likelihoods. While these linear projects have larger MSE's than non-linear projects, the conditional expectation \( \{\hat{r}_t, \hat{r}_{t-1}, \ldots, \hat{r}_1, 1\} \). Let \( \xi_t \) represent the mean squared latent variable, the one-step-ahead matrix with the first diagonal \( \sigma_n^2 \). The third diagonal element is determined through a direct step-ahead prediction error variance of \( \hat{u}_t \), which is \( \sigma_r^2 + \sigma_\eta^2 + c_t \).

The projections from the
and the coefficient matrices are

\[
F = \begin{pmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 1 & 0
\end{pmatrix} \quad \text{and} \quad R = \begin{pmatrix}
1 & 0 \\
0 & 1 \\
0 & 0
\end{pmatrix}
\]

The Kalman filter delivers the linear projection of the state vector \( \xi_t \), given the sequence of observations \( \{ \tilde{r}_t, \ldots, \tilde{r}_1 \} \). The Kalman smoother delivers the corresponding linear projection onto the extended sequence \( \hat{r} \). 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 \( \hat{\xi}_t \) represent linear projection onto \( F_t = (\hat{r}_t, \hat{r}_{t-1}, \ldots, \hat{r}_1, 1) \). Let \( \hat{\xi}_t = \hat{E}_t(\hat{\xi}_t) \) and let

\[
P_{tt} = E \left[ (\xi_t - \hat{\xi}_t)(\xi_t - \hat{\xi}_t)^T \right]
\]

(5)

represent the mean squared error matrices of these projections. For example, the one-step-ahead mean squared error matrix \( P_{tt-1} \) is a diagonal matrix with the first diagonal element equal to \( \sigma^2 \) and the second equal to \( \sigma^2 \). 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 \( M_t = H^T P_{tt-1} H = \sigma^2 + \sigma^2 + c_t \).

The projections from the Kalman filter are given by the recursion

\[
\hat{r}_{tt} \sigma^2 = \frac{\sigma^2}{M_t} (\hat{r}_t + \hat{\eta}_{t-1|t-1})
\]

(6)

\[
\hat{\eta}_{tt} = \frac{\sigma^2}{M_t} (\hat{r}_t + \hat{\eta}_{t-1|t-1})
\]

(7)

\[
c_{t+1} = \frac{\sigma^2 (\sigma^2 + c_t)}{M_t}
\]

(8)
The recursion and the boundary conditions $\hat{\eta}_{\tau T} = 0$ and $c_i = \sigma^2_{\eta}$ determine the sequence of filtered returns and filtered noise.

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

$$\hat{r}_{\tau T} \left( \sigma^2_{\tau} \right) = \hat{r}_{\tau T} \left( \sigma^2_{\eta} \right) - \frac{\sigma^2_{\eta}}{\sigma^2_{\eta} + c_i} \left( \hat{\eta}_{\tau T} - \hat{\eta}_{\tau \tau} \right)$$  \hspace{1cm} (9)

$$\hat{\eta}_{\tau -1:T} = \frac{c_i}{\sigma^2_{\eta} + c_i} \left( \hat{\eta}_{\tau T} - \hat{\eta}_{\tau \tau} \right)$$  \hspace{1cm} (10)

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

$$d_{T} = c_i \left( \frac{\sigma^2_{\eta} + \sigma^2_{\tau}}{\sigma^2_{\eta} + \sigma^2_{\tau} + c_i} \right) - \left( \frac{c_i}{\sigma^2_{\eta} + c_i} \right)^2 (c_{i + 1} - d_{T + 1})$$  \hspace{1cm} (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 $d_{i} < c_{T}$ for $i = 1, 2, \ldots, T$, establishing that $\text{Var}(\hat{\eta}_{T + 1}) < \text{Var}(\hat{\eta}_{T})$ 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 $\sigma^2_{\tau} = \sigma^2_{\eta}$ for all $t$, then the weights are nearly the same for all smoothed returns. To see the point clearly, consider a numerical example. If $\sigma^2_{\tau} = 10, \sigma^2_{\eta} = 1$, and $T = 7$, then, ignoring weights less than 0.001, we have that

$$\hat{r}_{\eta T} = 0.006 \hat{r}_{T} + 0.0709 \hat{r}_{T - 1} + 0.8452 \hat{r}_{T - 2} + 0.0709 \hat{r}_{T - 3} + 0.006 \hat{r}_{T - 4}$$

while

$$\hat{r}_{\eta T} = 0.0059 \hat{r}_{T} + 0.0709 \hat{r}_{T - 1} + 0.8452 \hat{r}_{T - 2} + 0.0709 \hat{r}_{T - 3} + 0.006 \hat{r}_{T - 4}.$$

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 realized volatility by $\sum_{t=1}^{n} \hat{r}_{\tau T}^2$ filtering is a linear transform of the squared bias estimator of $E(r^2_t | \hat{r}_T)$. I determined in the normal case that $E \left( \left( r_t - \hat{r}_{\tau T} \right)^2 \right)$, the pr

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

We find that the bias is

$$b_{\eta}(\sigma^2_{\tau}) = b_{\eta}^i$$

where $b_{\eta}^i(\sigma^2_{\tau})$ is the bias of the smoothed estimator. As is smaller than the bias of the filtered estimator. As is smaller than the bias of the smoothed estimator in what is necessary.

To determine the magnitude $\sigma^2_{\tau} = \sigma^2_{\eta}$ (constant volatility)

In accord with intuition, the variance and an increasing bias is 15 percent of the expected squared noise for $\sigma^2_{\tau} / \sigma^2_{\eta} \approx 0$, then the bias is near
2.2. Bias

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

\[
E\left[ \left(r_i^2 - \hat{r}_{nT}^2\right)^2 \right] = E\left[ \hat{E}_T\left(r_i^2 - 2\hat{r}_{nT}r_i + \hat{r}_{nT}^2\right) \right] = E\left[r_i^2 - \hat{r}_{nT}^2\right] \tag{12}
\]

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

We find that the bias is

\[
h_i(\sigma_i^2) = b'_i(\sigma_i^2) - \left(\frac{\sigma_i^2}{\sigma_i^2 + c_i}\right) (c_{i+1} - d_{i+1}) \tag{13}
\]

where \( b'_i(\sigma_i^2) \) (the bias of the filtered return \( \hat{r}_{nT} \)) is

\[
b'_i(\sigma_i^2) = \sigma_i^2 \left(\frac{\sigma_i^2 + c_i}{\sigma_i^2 + \sigma_i^2 + c_{i-1}}\right) \tag{14}
\]

Recall that \( c_i \) is the element of the variance for the filtered prediction of \( \eta_{i-1} \) and \( d_i \) is the corresponding variance element for the smoothed prediction.

As shown in (11), for \( i < T \) 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 \( \sigma_i^2 = \sigma_T^2 \) (constant volatility). The bias, \( h_i \), is well approximated by

\[
\sigma_T^2 \left(1 - \frac{1}{\sqrt{1 + 4\sigma_i^2/\sigma_T^2}}\right)
\]

In accord with intuition, the bias is a decreasing function of the return variance and an increasing function of the noise variance. If \( \sigma_i^2/\sigma_T^2 = .1 \), 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 \( \sigma_i^2/\sigma_T^2 \approx 0 \), then the bias is approximately \( \sigma_T^2 \). 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 \( \mathbf{r} = (r_1, r_2, \ldots, r_T)' \) and \( \eta = (\eta_0, \eta_1, \ldots, \eta_T)' \) so that

\[
\tilde{\mathbf{r}} = \mathbf{r} + B \eta
\]

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

\[
\tilde{\mathbf{r}} = \Lambda \left( \Lambda + \sigma_i^2 B B' \right)^{-1} \tilde{\mathbf{r}} \quad \text{and} \quad \Sigma = \sigma_i^2 B \left( I + \sigma_i^2 B' \Lambda^{-1} B \right)^{-1} B'
\]

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

\[
\begin{pmatrix} \mathbf{r} \\ \eta \end{pmatrix} \sim N_T \left( \begin{pmatrix} \Lambda & 0 \\ 0 & \sigma_i^2 I \end{pmatrix} \right)
\]

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

\[
\tilde{\mathbf{r}} | \mathbf{r} \sim N_T (\tilde{\mathbf{r}}, \Sigma)
\]

where \( \tilde{\mathbf{r}} \) and \( \Sigma \) are identical to the quantities from the Kalman smoother (15).

Under the assumption of joint normality \( \tilde{\mathbf{r}} = \text{E}(\mathbf{r} | \tilde{\mathbf{r}}) \), so, the smoothed estimator is the conditional expectation rather than simply the optimal linear projection. Similarly, \( \Sigma = \text{Var}(\mathbf{r} | \tilde{\mathbf{r}}) \) 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

\[
\text{Var}(r_i | \tilde{r}) = \text{E}(r_i^2 | \tilde{r}) - \text{E}^2(r_i | \tilde{r})
\]

The optimal estimator \( \text{E}(r_i^2 | \tilde{r}) \) exceeds the square of the optimal estimator for latent returns \( \text{E}^2(r_i | \tilde{r}) \). The correction term \( \text{Var}(r_i | \tilde{r}) \) 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

\[
\text{E}(r_i^2 | \tilde{r}) = \text{E}^2(r_i | \tilde{r}) + \text{Var}(r_i | \tilde{r})
\]

4. IM

The above analysis, in which suggests the use of the bias-corrected estimator \( \tilde{\mathbf{r}} = \mathbf{r} + B \eta \)

To implement the method, w

If the latent return variance bias-corrected estimator \( \tilde{\mathbf{r}} \), ma 1, the first two autocovariance for determining the var the true returns. (If one wishes then ML estimators may be estimators.) Andersen et al. (2) similar to \( \tilde{\mathbf{r}} \), does not cc correction.

For the case in which the re \( \tilde{\mathbf{r}}_m^2 (\tilde{\sigma}_m^2) \) and \( b_i (\tilde{\sigma}_m^2) \). We then taking window.'

\[
\tilde{\sigma}_m^2 = \frac{1}{25}
\]

The estimated time-varying \( \tilde{\sigma}_m^2 (\tilde{\sigma}_m^2) \) from (9) and \( b_i (\tilde{\sigma}_m^2) \)

For the case of constant va

5. F

To test the performance of th

use a model for the simulated behavior of the S&P 500 station 1 is
4. IMPLEMENTATION

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

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

To implement the method, we need estimators of \( \{\sigma_t^2\} \) and \( \sigma_n^2 \).

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 \( \sigma_r^2 \) and \( \sigma_n^2 \). 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 \( \hat{r}_{i|T}^2(\hat{\sigma}_t^2) \) and \( b_i(\hat{\sigma}_t^2) \). We then estimate the time-varying variance with a rolling window:

\[
\hat{\sigma}_n^2 = \frac{1}{25} \sum_{k=1}^{12} \left( \hat{r}_{i|T}^2(\hat{\sigma}_t^2) + b_i(\hat{\sigma}_t^2) \right)
\]

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

For the case of constant variance, laws of large numbers ensure the consistency of \( \hat{\sigma}_t^2 \) and \( \hat{\sigma}_r^2 \). 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
\[ dp_t = \sigma_t dw_t, \]
\[ d\sigma_t^2 = \theta(\omega - \sigma_t)dt + (2\lambda^2)\frac{1}{2}dw_t, \]  
(19)

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

\[ p_t - p_{t-1/m} = \gamma_{(m),t} \]
\[ \sigma_{(m),t}^2 = \phi_{(m)} + \alpha_{(m)}y_{(m),t}^2 + \beta_{(m)}\sigma_{(m),t-1/m}^2 \]  
(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

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

These parameters imply an unconditional return variance of \(\sigma^2 = 7.9\) 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 \(\eta_t\) is generated as an i.i.d. \(N(0, \sigma^2)\) random variable. To determine the noise variance, we invert the formula for \(\rho\) in Lemma 1 to obtain

\[ \sigma^2 = \frac{-\rho}{1 + 2\rho} \sigma_r^2. \]

Hasbrouck and Ho (1987) report estimates of \(\rho\) between \(-.4\) and \(-.1\), so we allow \(\rho\) to take the values \([-0.4, -0.3, -0.2, -0.1]\). As decreasing the value of \(\rho\) increases \(\sigma^2\), the resultant values of noise variance vary from \(\sigma^2 = 1\) (for \(\rho = -0.1\)) to \(\sigma^2 = 15.8\) (for \(\rho = -0.4\)).

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

\[ \bar{V} = \sum_{i=1+178}^{78} r_{ij}^2, \]

the feasible bias-corrected realize

\[ \hat{\mathbb{V}}_j(\hat{\sigma}_{ij}^2) = \sum_{i=4}^{78} \hat{\mathbb{V}}_j(\hat{\sigma}_{ij}^2) \]

and the infeasible bias-corrected

To compare this filter to that such as the MA(1) filter mention the gains from smoothing filtered (rather than smoothed)

\[ \hat{\mathbb{V}}_j(\hat{\sigma}_{ij}^2) = \sum_{i=4}^{78} \hat{\mathbb{V}}_j(\hat{\sigma}_{ij}^2) \]

where \(\hat{\sigma}_{ij}^2\) is obtained from (18) and \(b_h(\hat{\sigma}_{ij}^2)\), respectively. Finally, To judge the quality of the \(\alpha\) mean squared error (MSE) of \(\alpha\) estimator) estimator. For example,

\[ \text{MSE} \left[ \hat{\mathbb{V}}(\hat{\sigma}_{ij}^2) \right] \]

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

<table>
<thead>
<tr>
<th>(\sigma^2)</th>
<th>Constant Varianc</th>
</tr>
</thead>
<tbody>
<tr>
<td>(15.8)</td>
<td>8.4</td>
</tr>
<tr>
<td>5.9</td>
<td>7.6</td>
</tr>
<tr>
<td>2.6</td>
<td>6.1</td>
</tr>
<tr>
<td>1.0</td>
<td>3.9</td>
</tr>
</tbody>
</table>
the feasible bias-corrected realized volatility estimator
\[ \hat{V}_f(\hat{\sigma}_n^2) = \sum_{j=1}^{784} \left[ \hat{r}_{ij}^2 \left( \hat{\sigma}_n^2 \right) + b_t(\hat{\sigma}_n^2) \right] \]

and the infeasible bias-corrected estimator \( \hat{V}_f(\sigma_n^2) \).

To compare this filter to methods that assume constant return variance, such as the MA(1) filter mentioned above, we construct \( \hat{V}_f(\sigma_n^2) \). To determine the gains from smoothing, we also construct the estimator based on filtered (rather than smoothed) quantities
\[ \hat{V}'_f(\hat{\sigma}_n^2) = \sum_{j=1}^{784} \left[ \hat{r}_{ij}^2 \left( \hat{\sigma}_n^2 \right) + b'_t(\hat{\sigma}_n^2) \right] \]

where \( \hat{\sigma}_n^2 \) is obtained from (18) with \( \hat{r}_{ij}^2(\hat{\sigma}_n^2) \) and \( b'_t(\hat{\sigma}_n^2) \) in place of \( \hat{r}_{ij}^2(\hat{\sigma}_n^2) \)
and \( b_t(\hat{\sigma}_n^2) \), respectively. Finally, for completeness, we construct \( \hat{V}'_f(\sigma_n^2) \).

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 \( \hat{V}(\hat{\sigma}_n^2) \) is
\[ \frac{\text{MSE}\left[ \hat{V}(\hat{\sigma}_n^2) \right]}{\text{MSE}\left[ \hat{V}(\sigma_n^2) \right]} = \frac{10000}{\text{MSE}\left[ \hat{V}(\sigma_n^2) \right]} \left( \frac{\sum_{j=1}^{10000} (\hat{V}_j - \hat{V}_f(\hat{\sigma}_n^2))^2}{\sum_{j=1}^{10000} (\hat{V}_j - \hat{V}_f(\sigma_n^2))^2} \right) \]

In Table 1, we present the relative efficiency calculations. Regardless of the degree of noise variance, or indeed 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,
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 \( \{ \sigma_{l_{m,l}}^2 \} \):

\[
\sigma_{l_{m,l}}^2 = \sigma_{l_{m,l}}^2 \left( 1 + \frac{1}{3} \cos \left( \frac{2\pi}{78} I \right) \right)
\]

where \( \sigma_{l_{m,l}}^2 \) 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 \( \sigma_n^2 = 1.0 \)), 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 \( \sigma_n^2 = 15.8 \)), the filtered-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.**

<table>
<thead>
<tr>
<th>( \sigma_n^2 )</th>
<th>Constant Variance</th>
<th>Time-Varying Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\sigma}_n^4 )</td>
<td>( \hat{\sigma}_n^4 )</td>
<td>( \hat{\sigma}_n^4 )</td>
</tr>
<tr>
<td>( \hat{\sigma}_n^4 )</td>
<td>( \hat{\sigma}_n^4 )</td>
<td>( \hat{\sigma}_n^4 )</td>
</tr>
<tr>
<td>( \hat{\sigma}_n^4 )</td>
<td>( \hat{\sigma}_n^4 )</td>
<td>( \hat{\sigma}_n^4 )</td>
</tr>
<tr>
<td>15.8</td>
<td>8.1</td>
<td>6.5</td>
</tr>
<tr>
<td>5.9</td>
<td>7.4</td>
<td>5.9</td>
</tr>
<tr>
<td>2.6</td>
<td>6.0</td>
<td>5.0</td>
</tr>
<tr>
<td>1.0</td>
<td>3.9</td>
<td>3.5</td>
</tr>
</tbody>
</table>

Although currently we exploit the gains available frequency volatility proc filtered naïve estimators are naive based on our proposed

This article applies main noise from a popt Kalman smoother can are required for deter counterparts. However, edge of a latent stochastic. We show that a simple rolling regressions realized volatility estimators hide the mean square presence of strong diurnal volatility estimators can instead of merely filter correcting and using volatility.

1. Andersen, Bollerslev (1997), and Madhavan (2
2. Noise outcomes of a (1997), and Madhavan (2
3. Noise outcomes of a when no transaction inter m
4. Brockwell and Davi
5. If \( t = T \), then the stn
6. The bias of the filter
7. The rolling window and variance in the present
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 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.
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 $-2 \rho \sigma^2$ (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


