Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations*

Federico M. Bandi† and Jeffrey R. Russell‡
Booth School of Business, University of Chicago

First draft: May, 2005
This draft: June, 2009

Abstract

A growing literature has advocated consistent kernel estimation of integrated variance in the presence of financial market microstructure noise. We find that, for realistic sample sizes encountered in practice, the asymptotic results derived for these estimators may provide unsatisfactory representations of their finite sample properties. In addition, the existing asymptotic results might not offer sufficient guidance for practical implementations. We show how to optimize the finite sample properties of kernel-based integrated variance estimators. Empirically, we find that their suboptimal implementation can, in some cases, lead to little or no finite sample gains when compared to the classical realized variance estimator. Significant statistical and economic gains can, however, be recovered by using our proposed finite sample methods.

Keywords: Integrated variance, Realized variance, Microstructure noise, Kernel-based estimators

JEL Classification: C13, C14, C22

---

*This paper was previously circulated under the title "On the finite sample properties of kernel-based integrated variance estimators." We are grateful to Peter Hansen, Yixiao Sun, the Co-Editors (Nour Meddahi, Per Mykland, and Neil Shephard), two anonymous referees, and seminar participants at Chicago Booth, Yale, Rice, UC Los Angeles, UC San Diego, Singapore Management University, the CIREQ conference on realized volatility (Montreal, April 22-23, 2006), and the "International Conference on Financial Econometrics, Finance, and Risk" (Perth, June 29-July 1, 2006) for useful comments and discussions. The work of Federico Bandi was supported by the William S. Fishman Faculty Research Fund at the Booth School of Business of the University of Chicago.

†Address: Office 408, 5807 South Woodlawn Avenue, Chicago, IL 60637. E-mail: federico.bandi@chicagobooth.edu.

‡Address: Office 452, 5807 South Woodlawn Avenue, Chicago, IL 60637. E-mail: jeffrey.russell@chicagobooth.edu.

1 Introduction

The asymptotic consistency of kernel-based or HAC-type variance estimators relies on a limiting condition requiring the number of autocovariances to diverge to infinity as the ratio (ϕ, say) of the number of autocovariances over the number of observations goes to zero. As noticed as early as Neave (1970), while this condition is "mathematically convenient," it might lead to inaccurate asymptotic approximations to the estimators' finite sample properties. In effect, the ratio ϕ is fixed in any given sample. For a fixed ϕ, the magnitude of the finite sample mean-squared error (MSE) of HAC-type variance estimators can differ substantially from asymptotic approximations relying on a vanishing ϕ. Importantly, some influential recent contributions on integrated variance estimation by virtue of noisy high-frequency asset price data rely on a similar asymptotic condition for "near-consistency" or consistency (see, e.g., Barndorff-Nielsen et al., 2005, 2008, Hansen and Lunde, 2006, Zhang, 2006, and Zhang et al., 2005). These contributions are subject to the same observation: applied researchers are necessarily forced to select a value for ϕ.

This paper shows that, for a given ϕ, the finite sample properties of HAC-type variance estimators might not conform closely with existing asymptotic approximations. However, the ratio ϕ can be chosen *optimally* on the basis of a finite sample MSE criterion. In other words, the finite sample properties of HAC-type integrated variance estimators can be optimized. Our approach relates to the optimal MSE approach to integrated variance estimation by virtue of *realized variance* of Bandi and Russell (2003, 2008). As in Bandi and Russell (2003, 2008), we focus on finite sample performance and study an MSE-based method to optimize such a performance. Bandi and Russell (2003, 2008) write the conditional (on the volatility path of the underlying price process) MSE of the classical realized variance estimator (Andersen et al., 2003, and Barndorff-Nielsen and Shephard, 2002) as a function of the sampling frequency and select an optimal sampling frequency which minimizes the MSE. Here, the conditional MSEs of alternative integrated variance estimators are written as a function of ϕ and selection of an optimal ϕ is conducted for a *given* number of intra-daily observations.

Interestingly, Kiefer and Vogelsang (2005) have also recently highlighted the importance of treating the ratio ϕ as fixed in deriving asymptotic approximations to the properties of HAC estimators. Differently from Kiefer and Vogelsang (2005), however, we do not aim to derive asymptotic approximations for HAC estimators (and corresponding test statistics) for any value of ϕ. Rather, we study selection of ϕ in order to optimize the estimators' finite sample performance as summarized by their conditional MSEs. Importantly, we can do so because, contrary to the more classical HAC literature, the existing work on integrated variance estimation in the presence of market microstructure noise has relied on a class of price formation models (discussed in Section 2 below) which readily lends itself to finite sample investigations.

Using midpoints of bid-ask quotes for a sample of S&P100 stocks, we find that the root MSEs of HAC-type integrated variance estimators at the optimal ϕ value imply precise estimation of the integrated price variance over the period. In the case of *biased* (but even consistent) kernel estimators, estimation accuracy deteriorates quickly with suboptimal choices of ϕ. While the optimal finite sample MSE values of these estimators are smaller than the optimal finite sample MSE values of the classical realized variance estimator, the gains that biased estimators provide over the realized variance estimator can be either reduced or lost by suboptimal choices of ϕ. Importantly, asymptotic selection criteria for ϕ do not perform well. They imply a large finite sample bias component. We show how to choose ϕ in practise. Our optimal choice of ϕ yields considerable finite sample gains by reducing the impact of the bias term. The case of *unbiased* (or roughly unbiased) and consistent kernel estimators is somewhat different. Asymptotic selection criteria for ϕ perform, in general, better than in the biased case. Additionally, the MSEs (variances) of these estimators are fairly flat, albeit still convex, in ϕ. Hence, suboptimal choices of ϕ do not lead to drastic losses. Even though these estimators have the potential to be substantially more accurate than biased kernel estimators, as in the biased case, the asymptotic approximations to the estimators' finite sample variances may overestimate their finite sample precision. We quantify the difference between their asymptotic and finite sample accuracy in practise. Finally, we provide some evidence about the economic gains yielded by our finite sample (MSE-based) procedures in the context of a classical portfolio-choice problem. We refer the interested reader to a companion paper (Bandi et al., 2008) for broader applications of our methods to variance forecasting for the purpose of option pricing and trading.

Our previous work on realized variance focused on the finite sample performance of the classical realized variance estimator. If realized variance is used to identify integrated variance in the presence of market microstructure noise, the number of observations ought to be chosen optimally based on finite sample criteria. Bandi and Russell (2003, 2008) provide one such (MSE-based) criterion, but other metrics (possibly dictated by economic theory) may be used. This paper looks at the finite sample performance of the recently-proposed integrated variance kernel estimators. Again, if HAC-type estimators are used to identify integrated variance over a period, the number of auto-covariances may be chosen optimally (given a kernel function and a certain number of intra-daily observations) to optimize a finite sample criterion. We, again, provide a statistically meaningful (MSE-based) procedure to do so.

While one could rank alternative approaches to integrated variance estimation in the presence of microstructure noise based solely on asymptotic properties, this paper (and our previous work on realized variance) shows that the resulting ranking can be misleading. In effect, un-optimized consistent kernel estimators can perform substantially worse than optimized realized variance (see Subsection 4.1.), despite the well-known inconsistency of realized variance (Bandi and Russell, 2003, and Zhang et al., 2005). In addition, as a further example, consistent kernel estimators with the same asymptotic distribution can have drastically different finite sample properties (see

Subsection 4.2.).

The analysis in this paper sheds light on the relative performance of several recent approaches to integrated variance estimation (including realized variance, the two-scale estimator of Zhang et al., 2005, in its traditional and bias-corrected form, and the class of flat top, unbiased kernel estimators proposed by Barndorff-Nielsen et al., 2005, 2008). Our intent is not to advocate a specific method. However, *regardless of the estimator used*, we recommend explicit optimization of its finite sample properties, when possible. Our goal is to facilitate this approach and provide directions for practical implementations.

The paper proceeds as follows. Section 2 discusses the model and the class of HAC-type estimators which are the focus of our work. Section 3 presents the finite sample MSEs (as a function of ϕ) of some recently-proposed HAC estimators of the integrated price variance and discusses choice of ϕ. In Section 4 we apply the methods to three representative stocks, i.e., Goldman Sachs, SBC Communications, and EXXON Mobile Corporation. Section 5 evaluates the usefulness of our finite sample methods in the context of a classical portfolio-choice problem. Section 6 concludes. The Appendix contains the proofs.

2 The framework

Following the notation in Bandi and Russell (2008), *inter alia*, denote a trading day by $h = [0, 1]$. The trading day is divided into m equispaced sub-periods $t_i - t_{i-1} = \frac{1}{m}$ with $i = 1, ..., m$ so that $t_0 = 0$ and $t_m = 1$.[1] Now define

$$\underbrace{p(t_i) - p(t_{i-1})}_{r_i} = \underbrace{p^e(t_i) - p^e(t_{i-1})}_{r_i^e} + \underbrace{\eta(t_i) - \eta(t_{i-1})}_{\varepsilon_i}, \tag{1}$$

where r_i is an observed continuously-compounded intra-daily return, r_i^e is an equilibrium continuously-compounded intra-daily return,[2] and ε_i is a market microstructure contamination in the intra-daily return process. As in Bandi and Russell (2006a), Barndorff-Nielsen et al. (2005, 2008), Zhang et al. (2005), and Zhang (2006), among others, we make the following assumptions:

Assumption 1. *The equilibrium price process p^e is a stochastic volatility local martingale, namely,*

$$p^e(t) = \int_0^t \sigma_s dW_s, \tag{2}$$

[1] Extensions to deterministic non-equispaced arrival times are immediate. Allowing for stochastically spaced observations possibly dependent on the price process is more involved but may be empirically important. We refer to the work of Renault and Werker (2008) for discussions.

[2] As is customary in this literature, we are purposely unspecific about the nature of the equilibrium.

where $\{W_t : t \geq 0\}$ *is a standard Brownian motion assumed to be independent of the càdlàg spot volatility process* σ_t *for all* t*. Furthermore,*

$$Q(t) = \int_0^t \sigma_s^4 ds < \infty \tag{3}$$

for all t*.*

Assumption 2. *The logarithmic price contaminations* $\eta(t)$ *are i.i.d. mean zero with* $E(\eta^2) = \sigma_\eta^2$ *and* $E(\eta^4) = \theta\sigma_\eta^4 < \infty$*.*[3] *The* $\eta(t)'s$ *are independent of* $p^e(t)$*.*

These conditions are standard. They provide a uniform framework in which the asymptotic properties of the recently-proposed kernel estimators have been derived. Since we compare finite sample performance to asymptotic performance for each estimator (and across estimators), we use assumptions under which limiting results have been derived for *all* estimators studied in this paper. The empirical validity of these assumptions depends on the market structure (centralized versus decentralized markets), the nature of the price measurements (transaction prices versus midpoints of bid-ask spreads, for instance), and the sampling method (calendar time sampling versus event time sampling). We refer the reader to Bandi and Russell (2006b) for discussions. Hansen and Lunde (2006) study the empirical features of the noise for a sample of NYSE and NASDAQ stocks. Awartani et al. (2009) propose hypothesis tests on the noise properties.

The object of econometric interest is the integrated price variance over the trading day, namely $V = \int_0^1 \sigma_s^2 ds$. To this extent, consider the asymmetric kernel estimator

$$\widehat{V} = w_0\widehat{\gamma}_0 + 2\sum_{s=1}^{q} w_s\widehat{\gamma}_s, \tag{4}$$

where $\widehat{\gamma}_s = \sum_{i=1}^{m-s} r_i r_{i+s}$ and the w_s's are generic weights. This class of integrated variance estimators is in the tradition of zero frequency nonparametric spectral density estimators, or HAC estimators (Andrews, 1991, Andrews and Monahan, 1992, and Newey and West, 1987, among others). Hansen and Lunde (2006) study the finite sample MSE properties of $\widehat{V}$ for the case $q = 1$, $w_0 = 1$, and $w_1 = \frac{m}{m-1}$ (see Zhou, 1996, for an introduction to this estimator).[4] Hansen and Lunde (2006) also discuss the finite sample bias properties of $\widehat{V}$ for the more general case of

[3] In what follows, for convenience, we set $\theta = 3$ (i.e., the Gaussian case). This simplification allows us to only use estimates of σ_η^2, rather than estimates of both σ_η^2 and $E(\eta^4)$, in the empirical evaluation of the MSEs. The simplification is conceptually unimportant. The optimal number of autocovariances (and the MSEs at the optimum) of the biased estimators (in Theorem 1 and Theorem 2) are hardly affected by the properties of the noise (see Subsection 4.1.). This is because the bias term of these estimators plays a fundamental role in finite samples and the bias does not depend on the properties of the noise. In order to account for $\theta \neq 3$ in the case of the class of unbiased estimators in Theorem 3, the variance term of these estimators can be modified by simply writing $\Omega_2[1,1] = \theta, \Omega_2[1,2] = \Omega_2[2,1] = -1-\theta, \Omega_2[2,2] = 4+\theta$, and $\Omega_3[1,1] = (-\theta+1)/2, \Omega_3[1,2] = \Omega_3[2,1] = (\theta-1)/2+1, \Omega_3[2,2] = (-\theta+1)/2 - 7/2$. Similar modifications may be introduced in Corollary 1 through 3.

[4] The estimator's finite sample MSE properties in the context of a pure jump process of finite variation for the equilibrium price are studied by Oomen (2006).

an unrestricted q with $w_0 = 1$ and $w_s = \frac{m}{m-s}$. The limiting features of $\widehat{V}$ as a "near-consistent" estimator of the integrated variance of the equilibrium price process V are examined in Barndorff-Nielsen et al. (2005). Under Assumptions 1 and 2, Barndorff-Nielsen et al. (2005) show that, when using Bartlett-type kernel weights (i.e., when $w_0 = \frac{m-1}{m}\frac{q-1}{q}$ and $w_s = \frac{q-s}{q}$ for $s = 1, ..., q$), the asymptotic variance of $\widehat{V}$ coincides with the theoretical lower bound of the limiting variance of asymmetric kernel estimators in the class represented by Eq. (4), namely $4\left(\mathbf{E}\left(\eta^2\right)\right)^2$. For an average stock, $\left(\mathbf{E}\left(\eta^2\right)\right)^2$ is very small relative to $V = \int_0^1 \sigma_s^2 ds$ (see Section 4), hence the "near-consistency" of the Bartlett-type kernel-based estimator. Importantly, "near-consistency" requires $q, m \to \infty$ with $\frac{q}{m} \to 0$ and $\frac{q^2}{m} \to \infty$.

Asymmetric kernels are inconsistent, unless appropriately modified. The two-scale estimator proposed by Zhang et al. (2005) (see Eq. (12) below) is a "modified" Bartlett-type kernel estimator. Its consistency and asymptotic mixed normality (at rate $m^{1/6}$) are derived under $q = cm^{2/3}$ (where c is a constant to be chosen appropriately), thereby implying similar limiting conditions on q and m as those leading to the "near-consistency" of the Bartlett kernel estimator. The Bartlett kernel estimator and its modified version belong to the class of quadratic estimators. An interesting discussion of quadratic estimators and their use in integrated variance estimation is contained in Sun (2006).

Barndorff-Nielsen et al. (2005, 2008) have recently advocated (unbiased) flat-top symmetric kernels of the type

$$\widehat{V}^{BNHLS} = \widehat{\gamma}_0 + \sum_{s=1}^{q} w_s \left(\widehat{\gamma}_s + \widehat{\gamma}_{-s}\right), \tag{5}$$

where $\widehat{\gamma}_s = \sum_{i=1}^{m} r_i r_{i-s}$ with $s = -q, ..., q$, $w_s = k\left(\frac{s-1}{q}\right)$ and k is a function defined on $[0, 1]$ satisfying $k(0) = 0$ and $k(1) = 0$.[5] When $q = cm^{2/3}$, this class of estimators has an asymptotic mixed normal distribution. This distribution is the same as that of the two-scale estimator when using the flat-top Bartlett kernel, i.e., $k(x) = 1 - x$. The additional requirements $k^{'}(0) = 0$ and $k^{'}(1) = 0$ yield a faster rate of convergence of the estimators ($m^{1/4}$) to their mixed normal distribution. When $k(x) = 1 - 3x^2 + 2x^2$, the estimator has the same limiting distribution as the multi-scale estimator of Zhang (2006).

In all cases, the number of autocovariances is assumed to diverge to infinity with the sample size at a certain rate. This condition may lead to imprecise asymptotic representations of the estimators' finite sample properties and equally inaccurate (asymptotic) choices of the number

[5] Barndorff-Nielsen et al. (2008) re-express the logarithmic price end-points, $p(0)$ and $p(1)$, as an average of M observations in $\left(0 \pm \frac{1}{m}\right)$ and $\left(1 \pm \frac{1}{m}\right)$. This averaging (provided $M \to \infty$) simplifies the look of the limiting variances of the estimators in certain cases. Naturally, M is fixed in practise and recommended to be small (see Barndorff-Nielsen et al., 2008). In what follows, we set it equal to 1. This choice is designed to make the comparison between finite sample and asymptotic findings more straightforward. Because a diverging M will sometimes decrease the limiting variance of the corresponding estimator, this choice is also favorable to asymptotic approaches when compared to their finite sample counterparts.

of autocovariances. In practise, $q = \lfloor \phi m \rfloor$, where, as is customary, $\lfloor x \rfloor$ denotes the largest integer that is smaller than x or equal to x, with $0 < \phi \leq 1$. Next, we show how to select ϕ optimally on the basis of a finite sample MSE criterion. We study (*1*) the asymmetric Bartlett-type kernel estimator in its traditional and bias-corrected form, (*2*) the *modified* Bartlett-type kernel estimator, i.e., the two-scale estimator of Zhang et al. (2005), again in its traditional and bias-corrected form, and (*3*) the general class of flat-top symmetric kernel estimators proposed by Barndorff-Nielsen et al. (2005). We focus on (*2*) since the two-scale estimator is, to the best of our knowledge, the first integrated variance estimator found to be consistent in the presence of market microstructure noise. We study (*1*) because the asymmetric Bartlett-type estimator has, as we show below, very similar finite sample properties as the two-scale estimator, despite being theoretically inconsistent. Finally, we analyze the general class of flat-top symmetric kernel estimators because, contrary to (*1*) and (*2*), these estimators are unbiased by construction. In addition, this class includes estimators which, while unbiased, have the same limiting properties as the two-scale estimator and the multi-scale estimator, under appropriate assumptions. We do not explicitly consider the multi-scale estimator but expect it, when suitably bias-adjusted, to behave similarly to the bias-corrected two-scale estimator and the flat-top symmetric kernel estimators.

3 Choosing ϕ

3.1 The asymmetric Bartlett kernel estimator

We start with the estimator in Eq. (4) computed using Bartlett-type kernel weights. In what follows, $Q = \int_0^1 \sigma_s^4 ds$. Theorem 1 contains the conditional (on the volatility path - as in Bandi and Russell, 2003, 2008) MSE of the estimator expressed as a function of the ratio $\phi = \frac{q}{m}$. The optimal ϕ, ϕ^*, is defined as the *argmin* of the conditional MSE. The MSE in Theorem 1 and in the subsequent theorems should be interpreted as "nearly" exact. We solely replace quantities like $\frac{m}{3} E \sum_{i=1}^{m} (r_i^e)^4$ with the integrated quarticity Q for ease of interpretation. This is justifiable in that $\frac{m}{3} \sum_{i=1}^{m} (r_i^e)^4$ estimates Q consistently as $m \to \infty$ (Barndorff-Nielsen and Shephard, 2002). For sufficiently liquid stocks, in fact, the available number of intra-daily observations is large enough (c.f., Section 4) for the representation to be empirically meaningful (see, e.g., Barndorff-Nielsen and Shephard, 2005, for further discussions).

Theorem 1. *Consider*

$$\widehat{V}^{Bar} = w_0 \widehat{\gamma}_0 + 2 \sum_{s=1}^{q} w_s \widehat{\gamma}_s, \tag{6}$$

with $\widehat{\gamma}_s = \sum_{i=1}^{m-s} r_i r_{i+s}$. *Assume* $w_0 = \left(\frac{m-1}{m}\right)\left(\frac{q-1}{q}\right)$ *and* $w_s = \frac{q-s}{q}$ *for* $s = 1, ..., q$. *The optimal (in a conditional MSE sense)* ϕ *is defined as*

$$\phi^*_{Bar} = \underset{0<\phi\le 1}{\arg\min}\left[(bias(\phi))^2 + var(\phi)\right], \tag{7}$$

where

$$(bias(\phi))^2 = \frac{V^2}{m^2} + \left(\frac{2}{m^2} - \frac{2}{m^3}\right)\frac{V^2}{\phi} + \left(\frac{1}{m^2} - \frac{2}{m^3} + \frac{1}{m^4}\right)\frac{V^2}{\phi^2} \tag{8}$$

and

$$\begin{aligned}
var(\phi) &= K^{Bar} - \frac{Q}{3}\phi^2 + \left(\frac{8}{3}\sigma_\eta^2 V + \frac{4}{3}Q\right)\phi + \\
&\left[-\frac{4Q}{m^4} + \frac{\left(4\sigma_\eta^4 + 8\sigma_\eta^2 V\right)}{m} + \frac{\left(8\sigma_\eta^4 + 16\sigma_\eta^2 V + 8Q\right)}{m^3} + \frac{\left(-\frac{56}{3}\sigma_\eta^2 V - \frac{10}{3}Q - 24\sigma_\eta^4\right)}{m^2}\right]\frac{1}{\phi} + \\
&\left[8\frac{\sigma_\eta^4}{m} + \frac{2Q}{m^5} + \frac{\left(-24\sigma_\eta^4 - 8\sigma_\eta^2 V\right)}{m^2} + \frac{\left(20\sigma_\eta^4 + 16\sigma_\eta^2 V + 2Q\right)}{m^3} + \frac{\left(-4\sigma_\eta^4 - 8\sigma_\eta^2 V - 4Q\right)}{m^4}\right]\frac{1}{\phi^2},
\end{aligned} \tag{9}$$

with

$$K^{Bar} = 4\sigma_\eta^4 + \frac{4}{m}\sigma_\eta^4 + \left(-\frac{4}{m^2}\sigma_\eta^4 - \frac{8}{m^2}\sigma_\eta^2 V - \frac{11}{3}\frac{1}{m^2}Q\right) + \frac{2}{m^3}Q. \tag{10}$$

Proof. *The proof follows the same lines as that of Theorem 2 - see Appendix.*

When $q = \lfloor \phi m \rfloor$ for $0 < \phi \le 1$, as is the case in practise, a traditional bias-variance trade-off arises. For a given number of intra-daily observations m, larger values of ϕ lead to larger values of q and hence a smaller bias. However, since the variance term contains terms of order $O\left(\frac{q}{m}\right)$ and $O\left(\frac{q^2}{m^2}\right)$, higher values of ϕ translate into a larger variance. Interestingly, the variance itself is a convex function of ϕ. This result mirrors a similar result in the context of integrated variance estimation by virtue of realized variance ($\widehat{\gamma}_0$). There, Bandi and Russell (2003, 2008) show that, when market microstructure noise plays a role, the variance of the classical realized variance estimator is a convex function of the number of intra-daily observations used to compute the estimator.

Corollary 1. (The bias-corrected Bartlett estimator.) *Consider* $\widehat{V}^{Bar\_adj} = \left(\frac{\phi m^2 - m - \phi m + 1}{\phi m^2}\right)^{-1}\widehat{V}^{Bar}$. *Then,* $(bias(\phi))^2 = 0$ *and*

$$var_{\widehat{V}^{Bar\_adj}}(\phi) = \left(\frac{\phi m^2 - m - \phi m + 1}{\phi m^2}\right)^{-2} var(\phi), \tag{11}$$

where $var(\phi)$ *is defined in Eq.* (9).

The closed-form bias expression in Theorem 1 readily lends itself to a simple procedure for correcting the finite sample distortion of the original estimator. The finite sample variance of the bias-corrected estimator (defined in Eq. (11)) can therefore be used to select ϕ optimally in this case. The relative performance of the Bartlett estimator in its corrected and uncorrected form will be one of the subjects of our subsequent analysis. We now turn to the approach advocated by Zhang et al. (2005).

3.2 The two-scale estimator

Given the initial sampling grid $\Phi := \{t_0 = 0, t_1, ..., t_m = 1\}$, consider non-overlapping subgrids $\Phi_u^q := \{t_{u-1}, t_{u-1+q}, ..., t_{u-1+c_u q}\}$ with $c_u = \left\lfloor \frac{m-u+1}{q} \right\rfloor$ for $u = 1, ..., q$. The two-scale (or subsampling) estimator of Zhang et al. (2005) is defined as

$$\widehat{V}^{ZMA} = \frac{1}{q}\sum_{u=1}^{q}\left(\sum_{t_i \in \Phi_u^q}(p(t_{i+q}) - p(t_i))^2\right) - \frac{m-q+1}{mq}\sum_{i=1}^{m} r_i^2. \tag{12}$$

Barndorff-Nielsen et al. (2005) show that this estimator can be rewritten as a "modified" Bartlett-type kernel estimator. Specifically,

$$\widehat{V}^{ZMA} = \left(1 - \frac{m-q+1}{mq}\right)\widehat{\gamma}_0 + 2\sum_{s=1}^{q}\left(\frac{q-s}{q}\right)\widehat{\gamma}_s - \frac{1}{q}\vartheta_q \tag{13}$$

with $\vartheta_1 = 0$ and $\vartheta_q = \vartheta_{q-1} + (r_1 + ... + r_{q-1})^2 + (r_{m-q+2} + ... + r_m)^2$ for $q \geq 2$. The addition of the term $\frac{1}{q}\vartheta_q$, which subsampling yields by construction, is what makes the estimator consistent. Bartlett (1950) motivated the Bartlett kernel with subsampling.

Theorem 2. *Assume $\sigma_i^2 = \int_{t_{i-1}}^{t_i}\sigma_s^2 ds = \frac{V}{m}$ for all i. The optimal (in a conditional MSE sense) ϕ of the subsampling estimator in Eq. (13) is defined as*

$$\phi^*_{ZMA} = \underset{0<\phi\leq 1/2}{\arg\min}\left[(bias(\phi))^2 + var(\phi)\right], \tag{14}$$

where

$$(bias(\phi))^2 = \left(\frac{6V^2}{m^2} + \frac{2V^2}{m}\right) + V^2\phi^2 - \frac{4V^2}{m}\phi + \left(-4\frac{V^2}{m^2} - 4\frac{V^2}{m^3}\right)\frac{1}{\phi} + \left(\frac{V^2}{m^2} + \frac{2V^2}{m^3} + \frac{V^2}{m^4}\right)\frac{1}{\phi^2}, \tag{15}$$

and, if $\phi \leq 1/2$,

$$\begin{aligned} var(\phi) = {} & K^{ZMA} - \frac{1}{3}(Q+V^2)\phi^2 + \left(-\frac{1}{3}V^2\frac{1}{m} - 4V^2\frac{1}{m^2} + \frac{4}{3}Q\right)\phi + \\ & \left[-\frac{4\left(Q+V^2\right)}{m^4} + \left(\frac{8\sigma_\eta^4 + 16\sigma_\eta^2 V - 8Q - \frac{56}{3}V^2}{m^3}\right) + \left(\frac{24\sigma_\eta^2 V - \frac{10}{3}Q + 8\sigma_\eta^4}{m^2}\right) + \left(\frac{-8\sigma_\eta^4 + 8\sigma_\eta^2 V}{m}\right)\right]\frac{1}{\phi} + \\ & \left[\frac{2Q}{m^5} + \left(\frac{-4\sigma_\eta^4 - 8\sigma_\eta^2 V + 4Q - 8V^2}{m^4}\right) + \left(\frac{-4\sigma_\eta^4 - 16\sigma_\eta^2 V + 2Q}{m^3}\right) + \left(\frac{8\sigma_\eta^4 - 8\sigma_\eta^2 V}{m^2}\right) + \frac{8\sigma_\eta^4}{m}\right]\frac{1}{\phi^2}, \end{aligned} \tag{16}$$

with

$$K^{ZMA} = (-4\sigma_\eta^4 - 8V\sigma_\eta^2)\frac{1}{m} + \left(-4\sigma_\eta^4 - 8\sigma_\eta^2 V + \frac{13}{3}Q + \frac{79}{3}V^2\right)\frac{1}{m^2} + \frac{1}{m^3}\left(2Q + 8V^2\right). \tag{17}$$

Proof. *See Appendix.*

The addition of the term $\frac{1}{q}\vartheta_q$ is what makes the quantity $4\sigma_\eta^4$, which appears in K^{Bar}, not appear in K^{ZMA}. Under standard asymptotic conditions, i.e, as $q, m \to \infty$ with $\frac{q}{m} \to 0$ and $\frac{q^2}{m} \to \infty$, this simple modification is sufficient for the consistency of Zhang et al.'s two-scale estimator. In practise, though, $q = \lfloor \phi m \rfloor$ and the overall contribution of $4\sigma_\eta^4$ to the finite sample conditional MSE of the estimator is small, as we show in the next section. As earlier in the case of $\widehat{V}^{Bar}$, for a given number of intra-daily observations m, the choice of the number of subsamples or, equivalently, the choice of ϕ, induces a bias-variance trade-off which can be optimized. Also, as before, a bias-corrected version of the estimator $\left(\widehat{V}^{ZMA\_adj}\right)$ may be easily defined. The corresponding finite sample variance can therefore be readily optimized.

Corollary 2. (The bias-corrected ZMA estimator.) *Consider* $\widehat{V}^{ZMA\_adj} = \left(\frac{\phi m^2-1+2\phi m-\phi^2m^2-m}{\phi m^2}\right)^{-1}\widehat{V}^{ZMA}$. *Then,* $(bias(\phi))^2 = 0$ *and*

$$var_{\widehat{V}^{ZMA\_adj}}(\phi) = \left(\frac{\phi m^2 - 1 + 2\phi m - \phi^2 m^2 - m}{\phi m^2}\right)^{-2} var(\phi), \tag{18}$$

where $var(\phi)$ *is defined in Eq.* (16).

Remark 1. The additional assumption $\sigma_j^2 = \int_{t_{j-1}}^{t_j} \sigma_s^2 ds = \frac{V}{m}\ \forall j$ is exact under business time sampling. We refer the reader to Oomen (2005, 2006) for a thorough approach to business time sampling. The expression is of course not exact, but is empirically sensible, if volatility is fairly stable within the day. Importantly, we use it here solely to handle the term $\frac{1}{q}\vartheta_q$ and express the MSE in a way that is empirically fully tractable. It is noted, in fact, that the assumption is not required in the case of the inconsistent Bartlett estimator (in Theorem 1) in light of the absence there of the term $\frac{1}{q}\vartheta_q$. To get a feel for this point, we may focus on the bias term but the same reasoning applies to the full MSE, of course. Note that the exact finite sample bias of the two-scale estimator is

$$\left(\frac{-m+q-1}{qm}\right)V - \frac{1}{q}\sum_{s=1}^{q-1}(q-s)(\sigma_s^2 + \sigma_{m+1-s}^2),$$

thereby including terms (resulting from $\frac{1}{q}\vartheta_q$, in the summation above) which are hard to handle empirically. Under the assumption, the bias reduces to $\left(\frac{-m+2q-q^2-1}{qm}\right)V$ (see Theorem 2), but this term can now be easily evaluated in applied work (using an estimate of V and given ϕ and m). Importantly, since the assumption plays a role *only* through the term $\frac{1}{q}\vartheta_q$, and the impact of this term is minimal in finite samples (as we discuss in Section 4), the empirical consequences of the assumption may be considered immaterial.

Remark 2. Zhang et al. (2005) also propose a bias correction. The form of our correction is however slightly different from that contained in Zhang et al. (2005) where $\widehat{V}^{ZMA\_adj} = \left(\frac{qm-1+q-m}{qm}\right)^{-1}\widehat{V}^{ZMA}$. Under $\sigma_j^2 = \int_{t_{j-1}}^{t_j} \sigma_s^2 ds = \frac{V}{m}\ \forall j$, the correction in Corollary 2 is exact. While Zhang et al. (2005) do not adjust for the term $\frac{1}{q}\sum_{s=1}^{q-1}(q-s)(\sigma_s^2+\sigma_{m+1-s}^2)$ (or $\left(\frac{q^2-q}{qm}\right)V$ under $\sigma_j^2 = \int_{t_{j-1}}^{t_j} \sigma_s^2 ds = \frac{V}{m}\ \forall j$) in the bias expansion, this term is of order $\frac{q}{m}-\frac{1}{m}$. Hence, the difference between the exact correction discussed in this paper and the approximate correction in Zhang et al. (2005) is asymptotically negligible and may be empirically small. Section 4 below shows that this is indeed the case by evaluating the finite sample MSE of the two-scale estimator corrected as in Zhang et al. (2005). This MSE is easily derived using Theorem 2 above and is provided in Corollary 3. Below we show that the guidance provided by the asymptotics for selecting the optimal number of subsamples is considerably better in the case of bias-corrected versions of the two-scale estimator than for the unadjusted counterpart. Even though they may be improved upon, asymptotic approximations to the estimator's MSE are also superior in the presence of bias corrections.

Corollary 3. (The bias-correction in Zhang et al., 2005) *Consider* $\widehat{V}^{ZMA\_adj2} = \left(\frac{\phi m^2-1+\phi m-m}{\phi m^2}\right)^{-1}\widehat{V}^{ZMA}$, *where* $\left(\frac{\phi m^2-1+\phi m-m}{\phi m^2}\right)^{-1}$ *is the correction in Zhang et al. (2005). Then,* $(bias(\phi))^2 = \left(\frac{\phi m-\phi^2 m^2}{\phi m^2-1+\phi m-m}\right)^2 V^2$ *and*

$$var_{\widehat{V}^{ZMA\_adj2}}(\phi) = \left(\frac{\phi m^2-1+\phi m-m}{\phi m^2}\right)^{-2} var(\phi), \tag{19}$$

where $var(\phi)$ *is defined in Eq.* (16).

3.3 Flat-top symmetric kernels

Finally, we study the MSE (variance) of the symmetric flat-top kernels proposed by Barndorff-Nielsen et al. (2005, 2008). Suitable modifications of the results in Barndorff-Nielsen et al. (2005, 2008), as outlined in the Appendix, lead to the expression in Theorem 3.

Theorem 3. *The optimal (in a conditional MSE sense)* ϕ *of the estimator in Eq. (5) is defined as*

$$\phi^*_{BNHLS} = \underset{0<\phi\leq 1}{\arg\min}\left[(bias(\phi))^2 + var(\phi)\right], \tag{20}$$

where

$$bias(\phi) = 0 \tag{21}$$

and

$$Var\,(\phi) = \frac{Q}{m} w^{\intercal}\Omega_1 w + 4\sigma_\eta^4 m(w^{\intercal}\Omega_2 w) + 4\sigma_\eta^4(w^{\intercal}\Omega_3 w) + (2\sigma_\eta^2 V)4(w^{\intercal}\Omega_4 w), \tag{22}$$

with

$$w = \left(1, 1, k\left(\frac{1}{\phi m}\right), ..., k\left(\frac{\phi m - 1}{\phi m}\right)\right)^{\intercal}, \tag{23}$$

and Ω_a $a = 1, ..., 4$ are $(\phi m + 1, \phi m + 1)$ square matrices. For $j \leq \phi m$, the matrices Ω_1 and Ω_4 are defined as follows:

$$\begin{aligned} \Omega_1[1,1] &= 2,\ \Omega_1[1+j, 1+j] = 4, \\ \Omega_4[1,1] &= 1,\ \Omega_4[2,1] = -1,\ \Omega_4[1,2] = -1,\ \Omega_4[2,2] = 2,\ \Omega_4[1+j, 1+j] = 2, \\ \Omega_4[1+j, j] &= -1,\ \Omega_4[j, j+1] = -1, \end{aligned} \tag{24}$$

and zeros everywhere else. For $j \leq \phi m - 1$, the matrices Ω_2 and Ω_3 are defined as follows:

$$\begin{aligned} \Omega_2[1,1] &= 3,\ \Omega_2[1,2] = -4,\ \Omega_2[2,1] = -4,\ \Omega_2[2,2] = 7, \\ \Omega_2[2+j, 2+j] &= 6,\ \Omega_2[2+j, 1+j] = -4,\ \Omega_2[1+j, 2+j] = -4, \\ \Omega_2[2+j, j] &= 1,\ \Omega_2[j, 2+j] = 1, \\ \Omega_3[1,1] &= -1,\ \Omega_3[1,2] = 2,\ \Omega_3[2,1] = 2,\ \Omega_3[2,2] = -4.5, \\ \Omega_3[j+2, j+2] &= -3(j+1) - 1, \\ \Omega_3[2+j, 1+j] &= 2(j+1),\ \Omega_3[1+j, 2+j] = 2(j+1), \\ \Omega_3[2+j, j] &= -(j+1)/2,\ \Omega_3[j, 2+j] = -(j+1)/2, \end{aligned} \tag{25}$$

and zeros everywhere else.

Proof. *See Appendix.*

The MSE (variance) of the class of flat-top symmetric estimators is defined implicitly as a function of ϕ. The presence of potentially highly nonlinear kernels $k(.)$ renders the formulation of a closed-form expression unnecessarily cumbersome and kernel-specific. The expression in Theorem 3, on the other hand, can be easily evaluated for *all* kernels suggested by Barndorff-Nielsen et al. (2005, 2008). Importantly, the variance in Eq. (22) is a convex function of ϕ. Hence, even though this class of estimators is unbiased, there is still important scope for choosing the number of autocovariances optimally.

The next Section applies our methods. In the context of three representative stocks, we discuss optimal choice of the number of autocovariances as well as the validity of the existing asymptotic approximations to the estimators' finite sample properties.

4 Three representative stocks: GS, SBC, and XOM.

We consider Goldman Sachs (GS), SBC Communications (SBC), and EXXON Mobile Corporation (XOM). The data come from the TAQ data set. They are midpoints of bid-ask quotes posted on two exchanges, the NYSE and the MIDWEST, over the month of February 2002. The relevant parameter values for σ_η^2, V, Q, and m come from Table 1 in Bandi and Russell (2006a).[6] More generally, the interested reader is referred to Bandi and Russell (2006a) for simple methods to evaluate σ_η^2, V, and Q. We choose GS, SBC, and XOM since they represent median and extreme features of the S&P100 stocks as summarized by the ratio between the second moment of the noise returns, σ_ε^2, and the integrated variance of the underlying equilibrium price process, V. Specifically, GS, SBC, and XOM correspond to the first, the fifth, and the ninth decile of the cross-sectional distribution of the ratios, respectively. The relevant parameters are in Table 1.

4.1 The Bartlett kernel estimator and the two-scale estimator

We use Theorem 1 and Theorem 2 to derive the finite sample optimal ϕ and q values (ϕ^* and $q^* = \lfloor \phi^* m \rfloor$). In both cases, q^* is between 13 and 15 (Table 2, rows 1 and 2). The MSEs of the two estimators are also very similar. This is, of course, to be expected given the theoretical relation between the two approaches. The MSE values at the optimal (in finite samples) number of autocovariances are in Table 3 (rows 2 and 3). These values are rather small. Consider $\widehat{V}^{Bar}$ and SBC, for instance. The root MSE value at the optimum is equal to $5.3e-05$. The corresponding integrated variance over the day is $4.1e-04$.

It is useful to notice that, in light of the empirically-relevant magnitudes of the price and noise moments, the dominating bias and variance terms are $\frac{V^2}{m^2}\frac{1}{\phi^2}$ and $\frac{4}{3}Q\phi$ in the case of both estimators. Hence, the expression

$$\phi^*_{BR} = \phi^*_{Bar,ZMA} \approx \left(\frac{3}{2} \frac{\frac{V^2}{m^2}}{Q} \right)^{1/3} \tag{26}$$

provides a convenient *rule-of-thumb* to choose ϕ in practise. The approximate $\phi^*_{Bar,ZMA}$ optimally balances bias and variance. It is the presence of a large bias component relative to the variance component, and the necessity to reduce it, which leads to a rather large number of autocovariances,

[6] The second moments of the noise σ_η^2 are obtained by dividing by two the corresponding σ_ε^2 values contained in the column labelled "Mid. Var." The daily integrated variances V are in the column labelled "V^*." The integrated quarticities Q are obtained by using the approximate optimal frequencies in the column labelled "D^a" as follows. Since,

$$\frac{6.5 \times 60}{D^a} = \left(\frac{Q}{\sigma_\varepsilon^4} \right)^{1/3},$$

(c.f., Bandi and Russell, 2006b, Proposition 4), then

$$Q = \left(\frac{6.5 \times 60}{D^a} \right)^3 \sigma_\varepsilon^4.$$

Finally, m is obtained by using the average durations d (in seconds) in the column labelled "Avg. Dur." as follows:

$$m = \frac{6.5 \times 60 \times 60}{d}.$$

In the same fashion, the interested reader can obtain representative values for all S&P100 stocks by referring to Table 1 in Bandi and Russell (2006a).

as reported earlier. In agreement with this observation, we show below that unbiased (or roughly unbiased) versions of these estimators require a smaller (optimal) number of autocovariances.

The ratio in Eq. (26) closely resembles the approximate optimal sampling frequency of the classical realized variance estimator $\widehat{\gamma}_0$ derived in Bandi and Russell (2003, 2008), i.e.,

$$\frac{1}{m^*_{\widehat{\gamma}_0}} \approx \left(\frac{(2\sigma_\eta^2)^2}{Q} \right)^{1/3}. \tag{27}$$

As before, the larger the bias component $2\sigma_\eta^2$ relative to the variance component Q, the smaller the optimal number of intra-daily observations m^* required to optimize the finite sample MSE properties of the realized variance estimator. Interestingly, contrary to the realized variance case, the finite sample biases of the kernel estimators discussed here do not depend on the moments of the noise (c.f., Eq. (8) and Eq. (15)). Similarly, the approximate rule to select their optimal number of autocovariances only depends on the moments of the underlying equilibrium price process.

We are now in a position to evaluate how the optimal MSE values derived above (Table 3, rows 2 and 3) fare as compared to the optimal MSE values of the realized variance estimator (Table 3, row 1). In the latter case, the optimal values are achieved by sampling continuously-compounded returns in order to minimize the estimator's finite sample MSE under Assumptions 1 and 2 (Bandi and Russell, 2006a). The use of the approximate rule in Eq. (27) would yield very similar results (Bandi and Russell, 2006a). We find that there are substantial MSE gains to be obtained by employing kernel-based integrated variance estimators. Hansen and Lunde (2006) reach an analogous conclusion by comparing the MSE of the classical realized variance estimator to the MSE of $\widehat{V}$ for the case $q = 1$, $w_0 = 1$, and $w_1 = \frac{m}{m-1}$.

Importantly, Zhang et al. (2005) provide a distribution theory for their proposed estimator. It is therefore interesting to evaluate the finite sample performance of the asymptotically-optimal q ($\widetilde{q}_{ZMA}$, say) values implied by their limiting results, i.e.,

$$\widetilde{q}_{ZMA} = \left(\frac{16\left(\mathbf{E}(\eta^2)\right)^2}{\frac{4}{3}Q} \right)^{1/3} m^{2/3} \tag{28}$$

(Table 4, row 1).[7] By comparing Table 2 (row 2) and Table 4 (row 1) we note that the selected $\widetilde{q}_{ZMA}$ values are excessively small in finite samples. We can now plug these figures into the finite sample MSE expansion in Theorem 2 to obtain the corresponding finite sample MSE values (Table 3, row 7). In light of the sub-optimal (in finite samples) choice of the number of autocovariances

[7] In Zhang et al.'s case, $\widetilde{q}_{ZMA} = cm^{2/3}$, where c is a constant (defined in the main text) depending, of course, only on moments of the noise and price process, not on the number of observations m. In our case, $q = \phi m$, where ϕ is a function of m. The fact that the asymptotically-optimal rule requires $q \propto m^{2/3}$ does not imply that our rule is suboptimal. From a finite sample standpoint, our rule is, in fact, "nearly" optimal and selects the correct number of autocovariances *given* the available number of observations (and the model's parameter).

implied by the asymptotic criterion, the MSE values are larger than those implied by finite sample optimal choices of q. The difference is striking. When appropriately optimized, Zhang et al.'s two-scale estimator can provide significant MSE gains over the classical realized variance estimator (c.f., Table 3, rows 1 and 3). However, choosing the optimal number of autocovariances using asymptotic criteria may induce biases which have the potential to reduce and/or lose these gains (c.f., Table 3, rows 1 and 7). The reason for this outcome is simple. The limiting results do not take into account the finite sample bias of the estimator, i.e., the bias does not appear in the asymptotic distribution. Hence, the asymptotic criterion for choosing ϕ downplays the relevance of the finite sample bias term and leads to estimates which are (potentially severely) affected by it. The same considerations apply to the inconsistent, but otherwise similar, Bartlett kernel estimator.

We can also examine Zhang et al.'s implied asymptotic approximation to the finite sample MSE of $\widehat{V}^{ZMA}$, namely

$$asyMSE\left(\widetilde{\phi}_{ZMA}\right) = \frac{8}{m^{1/3}}\left(\left(\frac{16\left(\mathbf{E}(\eta^2)\right)^2}{\frac{4}{3}Q}\right)^{1/3}\right)^{-2}\left(\mathbf{E}(\eta^2)\right)^2 + \frac{1}{m^{1/3}}\left(\left(\frac{16\left(\mathbf{E}(\eta^2)\right)^2}{\frac{4}{3}Q}\right)^{1/3}\right)\frac{4}{3}Q$$

(Table 5, row 1). Table 3, row 3, and Table 5, row 1, reveal that the finite sample MSE of the estimator at its (finite sample) optimal ϕ value is considerably larger than the asymptotic MSE value at its corresponding asymptotically-optimal value. Hence, using asymptotic approximations relying on a vanishing ϕ might severely underestimate the estimator's true sampling error.

We now turn to the bias-corrected versions of the two estimators. As expected, the corrections reduce the need for a large number of autocovariances. The optimal q^* values (in Table 2, rows 3 and 4) are now equal to 3, 4, and 8 (for both estimators) and closer to the optimal asymptotic choice in the two-scale case (which, of course, is not affected by the correction since $q, m \to \infty$ with $\frac{q}{m} \to 0$). Importantly, the absence of sizeable bias components reduces the magnitude of the finite sample MSE values (Table 3, rows 4 and 5). We can now focus on the consistent, two-scale estimator. Even though the MSE reduction implied by the bias-correction is considerable, the difference between the estimator's finite sample MSE at the optimal q^* value (Table 3, row 5) and its asymptotic MSE at the optimal asymptotic value (Tables 5, row 2) is substantial. For all stocks, the former is at least twice as large as the former implying, again, that asymptotic arguments may overstate the precision of the estimates. We will show that this result applies to the unbiased kernel estimators in the flat-top symmetric class as well. As expected, the empirical difference between the exact bias correction discussed in Corollary 2 and the approximate bias correction in Zhang et al. (2005) (c.f. Corollary 3) is minimal (c.f., Table 3, rows 5 and 6).

In sum, much can be learned from using Bartlett-type kernel estimators of integrated variance. However, one should exercise care when bringing these tools to the data. We find that:

[1] The finite sample MSEs of $\widehat{V}^{Bar}$ and $\widehat{V}^{ZMA}$ are very similar.

[2] The root MSEs at the optimal ϕ value imply precise estimation of V. However, estimation accuracy deteriorates quickly when choosing ϕ suboptimally. Due to the presence of a potentially large bias term, this is particularly evident when the selected ϕ value is excessively small.

[3] The asymptotically-optimal choices of the number of autocovariances are drastically different from (smaller than) the finite sample optimal choices. Importantly for empirical purposes, the implied finite sample MSEs are considerably larger in the former case than in the latter case.

[4] The finite sample (optimal) MSE values are much larger than those implied by asymptotic approximations.

[5] The optimal finite sample MSE values of $\widehat{V}^{Bar}$ and $\widehat{V}^{ZMA}$ are smaller than those of the classical realized variance estimator.

[6] The gains that $\widehat{V}^{Bar}$ and $\widehat{V}^{ZMA}$ provide over the classical realized variance estimator may be either reduced or lost by suboptimally choosing the number of autocovariances (or, equivalently, the number of subsamples in the case of Zhang et al.'s estimator).

[7] Finite sample bias-corrections are helpful. Their implied (finite sample) optimal MSE values are drastically lower than those obtained in the presence of a bias term. Asymptotic approximations to the estimators' estimation errors are also more accurate in this case. However, these approximations continue to drastically overstate the estimators' precision.

4.2 Flat-top symmetric kernel estimators

We evaluate flat-top kernel estimators obtained by using three kernels, namely the Barlett kernel, $k(x) = 1 - x$, the cubic kernel, $k(x) = 1 - 3x^2 + 2x^3$, and the modified Tukey-Hanning kernel, $\left(1 - \cos\left(\pi(1-x)^2\right)\right)/2$. The Bartlett kernel delivers an estimator with the same asymptotic properties as the two-scale estimator of Zhang et al. (2005). Provided $q = cm^{1/2}$, the cubic kernel yields a more efficient estimator converging at rate $m^{1/4}$, the fastest convergence rate for this problem. The limiting distribution of the cubic flat-top kernel estimator is the same as that of the multi-scale estimator of Zhang (2006). Finally, the modified Tukey-Hanning kernel estimator is asymptotically more efficient than the cubic estimator.

This family of estimators is unbiased by construction. As shown, this is an important property since a large component of the MSEs of the asymmetric Bartlett kernel and of the two-scale estimator is bias-induced. As in the case of bias-corrected versions of those estimators, we expect a smaller number of autocovariances to be needed for MSE optimization. In addition, we expect the resulting MSEs to be smaller than the MSEs of the previously-analyzed estimators when considered in their traditional (unadjusted) versions. However, we anticipate that they will be similar to the MSE values of bias-corrected versions of those estimators. Our subsequent results reveal that this is, indeed, the case.

Having made these points, we stress that the use of a small number of autocovariances might render the asymptotic approximations to the distributions of the estimators in this class inaccurate. To gain some intuition, notice that, in the Bartlett case, one of the quadratic forms constituting the variance of the estimator, $w^\intercal \Omega_3 w$, can be approximated by $-\left(\frac{1}{q}\right)\left(-1+\frac{1}{2}\right)+O\left(\frac{1}{q^2}\right)$. If $q = 3$, for example, $w^\intercal \Omega_3 w = -0.0555$. However, $-\left(\frac{1}{q}\right)\left(-1+\frac{1}{2}\right) = 0.166$. If $q = 100$, $w^\intercal \Omega_3 w = 0.0048$ and $-\left(\frac{1}{q}\right)\left(-1+\frac{1}{2}\right) = 0.0050$. Since approximations of this sort are used to compute the asymptotic variances of the flat-top symmetric kernel estimators, and the optimal number of autocovariances will be shown to be small, the use of asymptotic approximations may, again, lead to understated estimation errors.

Table 2 (row 6 through 8) reports the (finite sample) optimal number of autocovariances of the flat-top kernel estimators for our three choices of kernels and our three representative stocks. As conjectured, these figures are smaller than in the case of the biased estimators discussed above (Table 2, rows 1 and 2) but similar to values which would be chosen, optimally, for bias-corrected versions of those estimators (Table 2, rows 3 and 4). Similarly, the corresponding MSEs (Table 3, row 8 through 10) are similar to those of the bias-corrected Bartlett and two-scale estimator (Table 3, rows 4 and 5) but are considerably smaller than those of their uncorrected versions (Table 3, row 2 and 3).

We find that the cubic flat-top kernel estimator, which is asymptotically distributed as the multi-scale estimator of Zhang (2006), does not seem to improve on the finite sample performance of the flat-top Bartlett kernel estimator. In addition, the modified Tukey-Hanning kernel estimator does only marginally better than both the Bartlett kernel estimator and the cubic kernel estimator.

We now evaluate the accuracy of the asymptotic variance of the flat-top kernel estimators. When $k(x) = 1 - x$, i.e., the Bartlett case, the asymptotically-optimal choice of q, as well as the lower bound on the asymptotic variance, coincide with the corresponding values in the two-scale case. Asymptotic criteria tend to select an excessively small number of autocovariances. The implied asymptotic variances tend to overestimate the finite sample precision of the estimator. The finite sample variances are small. However, they are more than twice as large as those implied by the asymptotics (c.f., Table 3, row 8, and Table 5, row 3). When $k(x) = 1 - 3x^2 + 2x^3$, the asymptotically-optimal choice of the number of autocovariances is $\widetilde{q}_Z = cm^{\frac{1}{2}}$, where c is selected to minimize

$$asyMSE\,(c) = \left(4c(0.371)Q - \frac{8}{c}(-1.2)\left(\left(\mathbf{E}(\eta^2)\right)V + \frac{\left(\mathbf{E}(\eta^2)\right)^2}{2}\right) + 4\left(\mathbf{E}(\eta^2)\right)^2 \frac{1}{c^3}12\right)\frac{1}{m^{1/2}}$$

We report $\widetilde{q}_Z$ in Table 4, row 4. The limiting variance selects virtually the same number of autocovariances as the finite sample criterion in Theorem 3 (c.f., Table 2, row 7, and Table 4, row 4). The implied asymptotic variances are in Table 5, row 4. Again, the asymptotic approximations tend to overstate the precision of the estimator. This overstatement is rather marginal in the case of XOM, but somewhat substantial in the case of GS (c.f., Table 3, row 9, and Table 5, row 4). We summarize our findings as follows:

[8] In a finite sample, the symmetric flat-top kernel estimators are preferable to biased versions of the Bartlett-type kernel estimators (even when their limiting distributions are identical). However, they perform very similarly to bias-corrected versions of those estimators.

[9] The use of asymptotic criteria to select the optimal number of autocovariances can be more or less satisfactory depending on the kernel function used. We found it to be inaccurate when employing a Bartlett kernel and accurate when using a cubic kernel, for instance.

[10] In light of the absence of a bias term (by construction), suboptimal bandwidth choices do not lead to extremely large losses due to the flatness of the variance term as a function of ϕ.

[11] Even though the cubic kernel implies a faster rate of convergence ($m^{1/4}$) than the Bartlett kernel ($m^{1/6}$), the finite sample performance of these two estimators is virtually identical.

[12] The asymptotic approximations to the flat-top kernel estimators' finite sample dispersions can be imprecise. As earlier in the case of corrected and uncorrected Bartlett-type kernel estimators, a careful assessment of the accuracy of these estimators requires a closer look at their finite sample properties.

5 An economic metric: portfolio choice

This section examines the economic benefits of our finite sample methods by way of a dynamic portfolio allocation experiment in the spirit of Fleming et al. (2001, 2003) and West et al. (1993). Bandi and Russell (2006b) have emphasized the importance of economic metrics to evaluate methods intended to "purge" high-frequency variance measures of their market microstructure noise contaminations. In a companion paper (Bandi et al., 2008), we evaluate the usefulness of the finite sample approach studied in the present paper for the purpose of option pricing.

We consider the utility of a mean-variance optimizing agent who chooses between a risky asset (a market index) and a risk-less asset. The agent selects time-varying portfolio weights obtained from conditional variance forecasts. Expected returns are assumed fixed. Hence, the agent's strategy is effectively a volatility-timing trading strategy. We generate 12 time-series of daily variance estimates using the finite sample optimal (inconsistent) Bartlett estimator, the traditional two-scale estimator in two versions (asymptotically optimal and optimal in a finite sample, as implied by Theorem 3), the bias-corrected two-scale estimator in two versions (asymptotically-optimal and optimal in finite sample, as implied by Corollary 2), the flat-top Bartlett, cubic, and Tukey-Hanning kernel estimators in two versions (asymptotically-optimal and optimal in finite samples, as implied by Theorem 3), and the VIX.[8] Given the 12 time-series of daily variance estimates, we generate out-of-sample, one-day ahead, variance forecasts using an ARFIMA (2,d,2) model.[9]

The agent selects optimal portfolio weights for each forecast. Specifically, the optimal (mean-variance) weight at day t is given by

$$w_t = \frac{E_t(R_{t,t+1} - R^f_{t,t+1})}{\lambda V_t(R_{t,t+1})},$$

where $R_{t,t+1}$ is the $t, t+1$ holding period return on the risky asset, $R^f_{t,t+1}$ is the corresponding return on the risk-less asset, λ is a risk-aversion parameter, and $V_t(R_{t,t+1})$ is a time t variance forecast. We set λ equal to 7. For each series of portfolio weights w_t, we construct the sequence of portfolio returns

$$R^p_{t,t+1} = R^f_{t,t+1} + w_t(R_{t,t+1} - R^f_{t,t+1}).$$

Finally, we evaluate the agent's long-run utility associated with specific variance estimates (and variance forecasts) by computing

$$AU = \overline{R}^p - \frac{\lambda}{2}\frac{1}{n}\sum_{t=1}^{n}\left(R^p_{t,t+1} - \overline{R}^p\right)^2,$$

where $\overline{R}^p = \frac{1}{n}\sum_{t=1}^{n} R^p_{t,t+1}$ and n is the number of days over which the prediction is conducted. The difference between the average utilities for two different sequences of variance estimates, variance forecasts, and portfolio weights is the difference in certainty equivalents and forms the relevant metric to measure the economic benefit of two different variance estimates/forecasts.[10]

[8] We find it interesting to evaluate how forecasts based on high-frequency variance estimates fare as compared to a well-known (and readily-available) market-based forecast, such as the VIX. We therefore consider today's VIX as a forecast for tomorrow's variance. We handle volatility risk premia as in Bandi et al. (2008).

[9] Preliminary investigations suggest that this specification is sufficiently parsimonious and empirically-reasonable for all variance series. We leave the evaluation of competing forecasting models for future work. MIDAS (Ghysels et al., 2006) and HAR-RV (Corsi, 2009) regressions, for example, are recent, succesful approaches in this literature.

[10] Alternatively, one could set a target return for each period, minimize variance under the target return, and simply compare long-run variances. This procedure has the advantage of avoiding noisy first moment estimation and potential contaminations in the resulting ranking of the forecasts.

We consider an investor who chooses between the S&P500 index and a risk-less security and use SPIDERS mid-quotes (on the NYSE) to compute the daily market variances. Our sample extends from January 2, 1998 to March 31, 2006. We remove quotes whose associated price changes and/or spreads are larger than 10%. The average duration between quote updates in our sample is 11.53 seconds. The average spread and the average price level are 0.0015 and 117.27, respectively.

The asymptotically-optimal and finite sample optimal number of autocovariances are re-computed for each day in the sample. In other words, the variance estimates are optimized daily, thereby requiring daily computation of the optimization inputs (the noise second moment σ_η^2, the integrated quarticity Q, and a preliminary estimate of the daily integrated variance V). The noise second moment is estimated by calculating the sample second moment of the quote-to-quote continuously-compounded returns. Hence, we use the highest quote arrival frequency.[11] The integrated variance and the integrated quarticity are computed using realized variance and the (realized) quarticity estimator in Barndorff-Nielsen and Shephard (2002). These measures are constructed using fixed calendar-time intervals and the prevailing quote method. They both rely on 15-minute returns.[12]

Virtually 6 year's worth of data are used to construct the first forecast (starting on January 20, 2004). The total number of out-of-sample forecasts n is equal to 553. The conditional mean of the risky asset $E_t(R_{t,t+1})$ is set equal to the unconditional mean of the daily S&P500 returns over the forecasting horizon. The risk-less asset return is set to an annual value of 4% (consistent with short-term interest rate values over the period) and converted to a daily value by dividing by 365. Since all variance measures are obtained over a 6.5-hour period, we derive daily measures (before forecasting) by multiplying each 6.5-hour variance estimate $\widehat{V}_{t,t+1}$ by $\zeta = \left(\sum_{t=1}^{n} R_{t,t+1}^2 / \sum_{t=1}^{n} \widehat{V}_{t,t+1} \right)$. This well-known and simple method ensures that the average of the corrected daily realized variances coincides with the variance of the daily returns.[13] We use the GPH estimator of Geweke and Porter-Hudak (1983) to estimate the d parameters.[14]

[11] Since the SPIDERS' variance signature plots are very upward sloping at high sampling frequencies in our sample, this identification procedure is empirically meaningful (see, e.g., Bandi and Russell, 2008, for discussions).

[12] Bandi and Russell (2008) thoroughly evaluate, by simulation, the empirical validity of this simple (from an applied standpoint) method to estimate the price moments of interest. They simulate an underlying price process allowing for extreme variance dynamics to show that, even when the equilibrium price moments are not estimated very accurately, the impact of their estimation error on the resulting optimization procedures (for realized variance) is minimal. Having made this point, the study of efficient estimation procedures for integrated quarticity is an important topic for future research.

[13] See Hansen and Lunde (2005) for further discussions.

[14] The estimated d values are equal to 0.426 (Bartlett kernel estimator), 0.489 (asymptotically-optimal two-scale estimator), 0.429 (finite sample optimal two-scale estimator), 0.493 (asymptotically-optimal bias-corrected two-scale estimator), 0.491 (finite sample optimal bias-corrected two-scale estimator), 0.479 (asymptotically-optimal flat-top Bartlett kernel estimator), 0.461 (finite sample optimal flat-top Bartlett kernel estimator), 0.435 (asymptotically-optimal flat-top cubic kernel estimator), 0.461 (finite sample optimal flat-top cubic kernel estimator), 0.447 (asymptotically-optimal flat-top Tukey-Hanning kernel estimator), and 0.452 (finite sample optimal flat-top Tukey-Hanning kernel estimator).

Our findings (in Tables 6 and 7) are largely consistent with our previous observations based on analytical representations of the estimators' MSEs. The optimized (in finite samples) flat-top symmetric kernels yield the highest utility (with relatively smaller standard errors, with the exception of the VIX). In particular, the finite-sample optimal flat-top Bartlett and cubic kernel estimators perform very similarly and, for our sample, slightly better than the optimized (in finite samples) flat-top Tukey-Hanning estimator. The finite sample optimal Bartlett and two-scale estimator fare well and, as expected, very similarly to each other in terms of average utility. However, while the use of a large number of autocovariances in these cases translates into small biases, it leads to some variability in the variance estimates/forecasts and in the corresponding utilities, as testified by the corresponding (larger) standard errors. Again, asymptotically-optimal versions of the unadjusted two-scale estimator translate into biased and rather variable variance estimates resulting in low average utilities and high variability of the utility values. Importantly for our purposes, pairwise t-statistics of the null of equal mean utilities across estimators optimized in finite samples and estimators optimized asymptotically clearly favor the former. The t-statistics are always positive, thereby signifying higher average utilities delivered by the finite-sample optimal estimators. In addition, these statistics are significantly different from zero at the 5% level in three out of five cases, i.e., the flat-top Bartlett kernel estimator and the two-scale estimator in its two versions (adjusted and unadjusted).

Even though these results are fully consistent with the broader set of findings obtained by Bandi et al. (2008) using profits and Sharpe ratios from option trading as a way to evaluate the accuracy of high-frequency variance estimates/forecasts, they should only be considered as suggestive of the potential of carefully-crafted finite sample methods. In general, of course, the optimization of forecast quality (given a forecasting model and, possibly, well-posed economic metrics) might require trading-off biases and variances, for instance, differently from what is implied by an MSE criterion. Direct optimization of forecast quality in the context of meaningful economic metrics is an important research topic better left for future work.

6 Conclusions

Nonparametric price variance estimation in the presence of market microstructure noise contaminations is a difficult task. While the classical realized variance estimator of Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2002) has been shown to be inconsistent for integrated variance under realistic price formation mechanisms with noise (Bandi and Russell, 2003, 2008, and Zhang et al., 2005), stimulating recent work has proposed HAC-type estimators which are consistent for the object of interest (Barndorff-Nielsen et al., 2005, 2008, Zhang et al., 2005, and Zhang, 2006).

We argue, and show, that asymptotic representations relying on limiting conditions on the number of autocovariances q and the number of observations m may constitute unsatisfactory approximations to the finite sample properties of these estimators. In light of this observation, we treat the ratio between q and m as fixed, as it is in practise, and choose it optimally in order to optimize the finite sample MSE of the estimators as done by Bandi and Russell (2003, 2008) in the context of realized variance. Differently put, relying on the data generating process used to derive asymptotic results in the literature, we choose the number of autocovariances optimally as a function of the model parameters and the number of intra-period observations.

The work on HAC-type integrated variance estimation is developing in interesting directions. We feel that any approach used must be serious about evaluating, and optimizing, the finite sample properties of the proposed estimator(s). Complete reliance on asymptotic approximations (and rules) has the potential to be unnecessarily detrimental both from a statistical and from an economic standpoint. In this paper, and in our previous work on the traditional realized variance estimator, finite sample optimality is assessed based on an MSE criterion. As pointed out earlier, it is now of interest to directly optimize relevant economic metrics in addition to purely statistical criteria. In the same vein, the economic benefits of alternative statistical methods can, and should, be evaluated in the context of relevant economic metrics. The importance of finite sample performance in high-frequency variance estimation, and the use of economic metrics to judge this performance, has been recently emphasized by Bandi and Russell (2006b). Some work has been conducted along these lines, largely in the context of optimally-sampled realized variance (Bandi and Russell, 2006a, 2008). This paper and Bandi et al. (2008) provide initial investigations in the context of "optimal" kernel estimators. Much remains to be done in future research.

7 Appendix

Proof of Theorem 2. Define $x_{i,s} = r_i r_{i+s}$,

$$z_1 = x_{1,0}, \qquad z_2 = x_{2,0} + 2x_{1,1}, \qquad z_3 = x_{3,0} + 2x_{2,1} + 2x_{1,2}, \qquad \ldots$$

and

$$\widetilde{z}_1 = x_{m,0}, \qquad \widetilde{z}_2 = x_{m-1,0} + 2x_{m-1,1}, \qquad \widetilde{z}_3 = x_{m-2,0} + 2x_{m-2,1} + 2x_{m-2,2}, \qquad \ldots$$

Write

$$\widehat{V}^{ZMA} = \underbrace{\left(1 - \frac{m-q+1}{mq}\right)\widehat{\gamma}_0 + 2\sum_{s=1}^{q}\left(\frac{q-s}{q}\right)\widehat{\gamma}_s}_{\alpha} \underbrace{-\frac{1}{q}\vartheta_q}_{\beta}$$

with $\frac{1}{q}\vartheta_q = \frac{1}{q}\sum_{s=1}^{q-1}(q-s)z_s + \frac{1}{q}\sum_{s=1}^{q-1}(q-s)\widetilde{z}_s$. We start with the variance of α. Using the appropriate kernel weights (i.e., $w_0 = 1 - \frac{m-q+1}{mq}$ and $w_s = \frac{q-s}{q}$ for $s \geq 1$) and Corollary 2 of Barndorff-Nielsen et al. (2005), we obtain

$$Var(\alpha) = \Lambda_1^{\alpha}\sigma_\eta^4 m + \Lambda_0^{\alpha}\sigma_\eta^2 + \Lambda_{-1}^{\alpha}\frac{1}{m},$$

where

$$
\begin{aligned}
\Lambda_1^\alpha &= 12\left(\frac{mq-m+q-1}{mq}\right)^2+4\left(\frac{m-1}{m}\right)\left(\frac{q-1}{q}\right)\left(7\left(\frac{q-1}{q}\right)-8\left(\frac{mq-m+q-1}{mq}\right)\right)\\
&+\left(\frac{m-2}{m}\right)\left(\frac{q-2}{q}\right)8\left(3\frac{q-2}{q}-4\frac{q-1}{q}+\left(\frac{mq-m+q-1}{mq}\right)\right)\\
&+\sum_{j=3}^{q}\left(\frac{m-j}{m}\right)\left(\frac{q-j}{q}\right)8\left(3\left(\frac{q-j}{q}\right)-4\left(\frac{q-(j-1)}{q}\right)+\left(\frac{q-(j-2)}{q}\right)\right)\\
&-\frac{4}{m}\left(\frac{mq-m+q-1}{mq}\right)^2-\frac{8}{m}\sum_{j=1}^{q}\left(\frac{q-j}{q}\right)^2,
\end{aligned}
$$

$$
\begin{aligned}
\Lambda_0^\alpha &= 8V\left(\frac{mq-m+q-1}{mq}\right)^2+16V\left(\frac{m-1}{m}\right)\left(\frac{q-1}{q}\right)\left(\left(\frac{q-1}{q}\right)-\left(\frac{mq-m+q-1}{mq}\right)\right)\\
&+\sum_{j=2}^{q}16V_{j/m}\left(\frac{q-j}{q}\right)\left(\left(\frac{q-j}{q}\right)-\left(\frac{q-(j-1)}{q}\right)\right),
\end{aligned}
$$

and

$$\Lambda_{-1}^\alpha = 2Q\left(\frac{mq-m+q-1}{mq}\right)^2+\sum_{j=1}^{q}4Q_{j/m}\left(\frac{q-j}{q}\right)^2.$$

In light of the fact that

$$\sum_{j=1}^{q} j=\frac{q^2+q}{2},\quad \sum_{j=1}^{q} j^2=\frac{(q+1)(2q+1)q}{6},\quad \sum_{j=1}^{q} j^3=\frac{q^4+2q^3+q^2}{4},$$

and since $V_{j/m}=V\left(\frac{m-j}{m}\right)$ and $Q_{j/m}=Q\left(\frac{m-j}{m}\right)$, simple algebra gives

$$\Lambda_1^\alpha=\frac{4}{m}+\frac{8}{q^2}+\frac{20}{m^2}+\frac{4}{m^2q^2}-\frac{12}{mq}-\frac{8}{mq^2}-\frac{24}{m^2q}-\frac{4}{m^3}-\frac{4}{m^3q^2}+\frac{8}{m^3q},$$

$$\Lambda_0^\alpha=\left(\frac{8}{3}\frac{q}{m}-\frac{8}{q^2}+\frac{24}{m^2}+\frac{24}{m^2q^2}+\frac{8}{q}-\frac{56}{3}\frac{1}{mq}+\frac{16}{mq^2}-\frac{48}{m^2q}\right)V,$$

and

$$\Lambda_{-1}^\alpha=\left(\frac{2}{q^2}+\frac{2}{m^2}-\frac{10}{3}\frac{1}{q}+\frac{13}{3}\frac{1}{m}-\frac{8}{mq}+\frac{4}{mq^2}+\frac{4}{3}q+\frac{2}{m^2q^2}-\frac{1}{3}\frac{q^2}{m}-\frac{4}{m^2q}\right)Q.$$

Hence,

$$
\begin{aligned}
&Var(\alpha)=\left(\frac{4}{m}+\frac{8}{q^2}+\frac{20}{m^2}+\frac{4}{m^2q^2}-\frac{12}{qm}-\frac{8}{mq^2}-\frac{24}{m^2q}-\frac{4}{m^3}-\frac{4}{m^3q^2}+\frac{8}{m^3q}\right)\sigma_\eta^4 m\\
&+\left(\frac{8}{3}\frac{q}{m}-\frac{8}{q^2}+\frac{24}{m^2}+\frac{24}{m^2q^2}+\frac{8}{q}-\frac{56}{3mq}+\frac{16}{mq^2}-\frac{48}{m^2q}\right)\sigma_\eta^2 V\\
&+\left(\frac{2}{q^2}+\frac{2}{m^2}-\frac{10}{3}\frac{1}{q}+\frac{13}{3}\frac{1}{m}-\frac{8}{mq}+\frac{4}{mq^2}+\frac{4}{3}q+\frac{2}{m^2q^2}-\frac{1}{3}\frac{q^2}{m}-\frac{4}{m^2q}\right)\frac{1}{m}Q\\
=\;&\left(4+\frac{8m}{q^2}+\frac{20}{m}+\frac{4}{mq^2}-\frac{12}{q}-\frac{8}{q^2}-\frac{24}{mq}-\frac{4}{m^2}-\frac{4}{m^2q^2}+\frac{8}{m^2q}\right)\sigma_\eta^4\\
&+\left(\frac{8}{3}\frac{q}{m}-\frac{8}{q^2}+\frac{24}{m^2}+\frac{24}{m^2q^2}+\frac{8}{q}-\frac{56}{3mq}+\frac{16}{mq^2}-\frac{48}{m^2q}\right)\sigma_\eta^2 V\\
&+\left(\frac{2}{mq^2}+\frac{2}{m^3}-\frac{10}{3}\frac{1}{qm}+\frac{13}{3}\frac{1}{m^2}-\frac{8}{m^2q}+\frac{4}{m^2q^2}+\frac{4}{3}\frac{q}{m}+\frac{2}{m^3q^2}-\frac{1}{3}\frac{q^2}{m^2}-\frac{4}{m^3q}\right)Q,
\end{aligned}
$$

which yields

$$\begin{aligned}
Var(\alpha) &= 4\sigma_\eta^4 + \frac{20}{m}\sigma_\eta^4 + \left(-\frac{4}{m^2}\sigma_\eta^4 + \frac{24}{m^2}\sigma_\eta^2 V + \frac{13}{3}\frac{1}{m^2}Q\right) + \frac{2}{m^3}Q \\
&\quad + 8\frac{m}{q^2}\sigma_\eta^4 + \frac{2}{m^3 q^2}Q - \frac{1}{3}\frac{q^2}{m^2}Q - \frac{4}{m^3 q}Q + \left(-12\sigma_\eta^4 + 8\sigma_\eta^2 V\right)\frac{1}{q} + \left(-8\sigma_\eta^4 - 8\sigma_\eta^2 V\right)\frac{1}{q^2} \\
&\quad + \left(\frac{8}{3}\sigma_\eta^2 V + \frac{4}{3}Q\right)\frac{q}{m} + \left(4\sigma_\eta^4 + 16\sigma_\eta^2 V + 2Q\right)\frac{1}{mq^2} \\
&\quad + \left(8\sigma_\eta^4 - 48\sigma_\eta^2 V - 8Q\right)\frac{1}{m^2 q} + \left(-4\sigma_\eta^4 + 24\sigma_\eta^2 V + 4Q\right)\frac{1}{m^2 q^2} + \left(-\frac{56}{3}\sigma_\eta^2 V - \frac{10}{3}Q - 24\sigma_\eta^4\right)\frac{1}{mq}.
\end{aligned}$$

Consider now $Var(\beta)$, namely,

$$Var(\beta) = Var\left(-\frac{1}{q}\vartheta_q\right) = \frac{1}{q^2}Var\left(\sum_{s=1}^{q-1}(q-s)z_s\right) + \frac{1}{q^2}Var\left(\sum_{s=1}^{q-1}(q-s)\widetilde{z}_s\right)$$

for $q \leq m/2$. By virtue of Lemma A.3 and Lemma 7 of Barndorff-Nielsen et al. (2005), we can write

$$\begin{aligned}
& Var\left(\sum_{s=1}^{q-1}(q-s)z_s\right) = \sum_{s=1}^{q-1}(q-s)^2 Var(z_s) + 2\sum_{s=1}^{q-2}(q-s)(q-s-1)\,cov(z_s, z_{s+1}) \\
=\; & \sum_{s=1}^{q-1}(q-s)^2\left[12\sigma_\eta^4 + 8\sigma_\eta^2(\sigma_1^2 + ... + \sigma_s^2) + \sigma_s^2\left(4\sigma_1^2 + ... + 4\sigma_{s-1}^2 + 2\sigma_s^2\right)\right] \\
& + 2\sum_{s=1}^{q-2}(q-s)(q-s-1)\left[-6\sigma_\eta^4 - 4\sigma_\eta^2(\sigma_1^2 + ... + \sigma_s^2)\right] - \left[(q-1)^2 4\sigma_\eta^4\right] \\
=\; & A + B + C.
\end{aligned}$$

We start with A. Write

$$\begin{aligned}
A &= \sum_{s=1}^{q-1}(q-s)^2\left[12\sigma_\eta^4 + 8\sigma_\eta^2(\sigma_1^2 + ... + \sigma_s^2) + \sigma_s^2\left(4\sigma_1^2 + ... + 4\sigma_{s-1}^2 + 2\sigma_s^2\right)\right] \\
&= \sum_{s=1}^{q-1}(q-s)^2\left[12\sigma_\eta^4 + 8\sigma_\eta^2 s\frac{V}{m} + 4(s-1)\frac{V^2}{m^2} + 2\frac{V^2}{m^2}\right] \\
&= 2\sigma_\eta^4 q - 6\sigma_\eta^4 q^2 + 4\sigma_\eta^4 q^3 - \frac{1}{3}V^2\frac{q}{m^2} + \frac{2}{3}V^2\frac{q^2}{m^2} - \frac{2}{3}V^2\frac{q^3}{m^2} + \frac{1}{3}V^2\frac{q^4}{m^2} - \frac{2}{3}\sigma_\eta^2 V\frac{q^2}{m} + \frac{2}{3}\sigma_\eta^2 V\frac{q^4}{m}.
\end{aligned}$$

We now turn to B. Write

$$\begin{aligned}
B &= 2\sum_{s=1}^{q-2}(q-s)(q-s-1)\left[-6\sigma_\eta^4 - 4\sigma_\eta^2(\sigma_1^2 + ... + \sigma_s^2)\right] \\
&= q(-8\sigma_\eta^4) + q^2\left(12\sigma_\eta^4\right) + q^3(-4\sigma_\eta^4) + \frac{q}{m}\left(-\frac{4}{3}\sigma_\eta^2 V\right) + \frac{q^2}{m}\left(\frac{2}{3}\sigma_\eta^2 V\right) + \frac{q^3}{m}\left(\frac{4}{3}\sigma_\eta^2 V\right) + \frac{q^4}{m}\left(-\frac{2}{3}\sigma_\eta^2 V\right).
\end{aligned}$$

Finally,

$$C = -(q-1)^2 4\sigma_\eta^4 = -q^2(4\sigma_\eta^4) - q(-8\sigma_\eta^4) - 4\sigma_\eta^4.$$

Thus,

$$\frac{1}{q^2}Var\left(\sum_{s=1}^{q-1}(q-s)z_s\right) = \frac{1}{q^2}(A+B+C)$$

$$= 2\sigma_\eta^4 + \frac{1}{q}\left(2\sigma_\eta^4\right) + \frac{1}{q^2}(-4\sigma_\eta^4) + \frac{1}{m^2}\left(\frac{2}{3}V^2\right) + \frac{1}{mq}\left(-\frac{4}{3}V\sigma_\eta^2\right) + \frac{q}{m}\left(\frac{4}{3}\sigma_\eta^2 V\right)$$
$$+\frac{q^2}{m^2}\left(\frac{1}{3}V^2\right) + \frac{1}{qm^2}\left(-\frac{1}{3}V^2\right) - \frac{2}{3}\frac{q}{m^2}V^2.$$

Since

$$Var(\widetilde{z}_1) = 8\sigma_\eta^4 + 8\sigma_\eta^2\sigma_m^2 + 2\sigma_m^4,$$

$$Var(\widetilde{z}_j) = 12\sigma_\eta^4 + 8\sigma_\eta^2\left(\sigma_m^2 + ... + \sigma_{m-j+1}^2\right) + \sigma_{m-j+1}^2\left(4\sigma_m^2 + ... + 4\sigma_{m-j+2}^2 + 2\sigma_{m-j+1}^2\right)$$

for $j = 2, ...,$

$$Cov(\widetilde{z}_1, \widetilde{z}_2) = -6\sigma_\eta^4 - 4\sigma_\eta^2\sigma_m^2,$$

and

$$Cov(\widetilde{z}_j, \widetilde{z}_{j+1}) = -6\sigma_\eta^4 - 4\sigma_\eta^2\left[\sigma_m^2 + ... + \sigma_{m-j+1}^2\right] \qquad j = 2, ...,$$

then $\frac{1}{q^2}Var\left(\sum_{s=1}^{q-1}(q-s)\widetilde{z}_s\right)$ can be represented similarly. Hence,

$$\frac{1}{q^2}Var\left(\sum_{s=1}^{q-1}(q-s)z_s\right) + \frac{1}{q^2}Var\left(\sum_{s=1}^{q-1}(q-s)\widetilde{z}_s\right)$$
$$= 4\sigma_\eta^4 + \frac{1}{q}\left(4\sigma_\eta^4\right) + \frac{1}{q^2}(-8\sigma_\eta^4) + \frac{1}{m^2}\left(\frac{4}{3}V^2\right) + \frac{1}{mq}\left(-\frac{8}{3}V\sigma_\eta^2\right) + \frac{q}{m}\left(\frac{8}{3}\sigma_\eta^2 V\right)$$
$$+\frac{q^2}{m^2}\left(\frac{2}{3}V^2\right) + \frac{1}{qm^2}\left(-\frac{2}{3}V^2\right) - \frac{4}{3}\frac{q}{m^2}V^2.$$

We now turn to the covariance between α and β. Start with

$$2\frac{1}{q}\sum_{s=1}^{q-1}\left(\frac{mq-m+q-1}{mq}\right)(q-s)cov\left(\gamma_0, z_s\right)$$
$$= 2\frac{1}{q}\left(\frac{m+1}{m}\right)\left(\frac{q-1}{q}\right)(q-1)[10\sigma_\eta^4 + 8\sigma_\eta^2\sigma_1^2 + 2\sigma_1^4]$$
$$+2\frac{1}{q}\left(\frac{m+1}{m}\right)\left(\frac{q-1}{q}\right)(q-2)[-4\sigma_\eta^4 + 4\sigma_\eta^2(\sigma_2^2 - \sigma_1^2) + 2\sigma_2^4]$$
$$+2\frac{1}{q}\left(\frac{m+1}{m}\right)\left(\frac{q-1}{q}\right)\sum_{s=3}^{q-1}(q-s)\left[4\sigma_\eta^2(\sigma_s^2 - \sigma_{s-1}^2) + 2\sigma_s^4\right]$$
$$= 12\sigma_\eta^4 - \frac{16}{q}\sigma_\eta^4 + \frac{4}{q^2}\sigma_\eta^4 + \frac{1}{m}(12\sigma_\eta^4 + 16\sigma_\eta^2 V) + \frac{1}{m^2}(16\sigma_\eta^2 V - 4V^2)$$
$$+2\frac{V^2}{m^3}q + 2\frac{V^2}{m^2}q + \frac{1}{qm}\left(-16\sigma_\eta^4 - 32\sigma_\eta^2 V\right) + \frac{1}{qm^2}\left(2V^2 - 32\sigma_\eta^2 V\right)$$
$$+\frac{1}{qm^3}\left(2V^2\right) + \frac{1}{q^2m^2}\left(16\sigma_\eta^2 V\right) + \frac{1}{mq^2}\left(4\sigma_\eta^4 + 16\sigma_\eta^2 V\right) + \frac{1}{m^3}(-4V^2),$$

where the first equality derives from Lemma A.4 of Barndorff-Nielsen et al. (2005). Next, we consider the terms:

$$2\frac{1}{q^2}\sum_{s=1}^{q}\sum_{k=1}^{q-1}(q-s)(q-k)cov\left(\widehat{\gamma}_s, z_k\right) = 2\frac{1}{q^2}\sum_{s=1}^{q-1}\sum_{k=1}^{q-1}(q-s)(q-k)cov\left(\widehat{\gamma}_s, z_k\right)$$

since $cov(\widehat{\gamma}_s, z_{s-i}) = 0$ for all s and all $i \geq 1$. Write

$$\underbrace{2\frac{1}{q^2}\sum_{s=1}^{q-1}(q-s)^2 cov\left(\widehat{\gamma}_s, z_s\right)}_{\alpha'} + \underbrace{2\,\frac{1}{q^2}\sum_{s=1}^{q-1}\sum_{k>s}(q-s)(q-k)cov\left(\widehat{\gamma}_s, z_k\right)}_{\beta'}.$$

Start with $\alpha^{'}$. Using again Lemma A.4 of Barndorff-Nielsen et al. (2005),

$$\begin{aligned}
\alpha^{'} &= 2\frac{1}{q^2}\sum_{s=1}^{q-1}(q-s)^2 cov\left(\widehat{\gamma}_s, z_s\right) \\
&= 2\frac{1}{q^2}(q-1)^2[-4\sigma_\eta^4 - 2\sigma_\eta^2\sigma_1^2] + 2\frac{1}{q^2}\sum_{s=2}^{q-1}(q-s)^2[-2\sigma_\eta^4 - 2\sigma_\eta^2\sigma_1^2] \\
&= -2\sigma_\eta^4 - \frac{4\sigma_\eta^4}{q^2} + \frac{22}{3}\frac{\sigma_\eta^4}{q} + 2\sigma_\eta^2\frac{V}{m} - \frac{2}{3}\sigma_\eta^2 V\frac{1}{qm} - \frac{4}{3}\sigma_\eta^4 q - \frac{4}{3}\sigma_\eta^2 V\frac{q}{m}.
\end{aligned}$$

Now turn to $\beta^{'}$. Write

$$\begin{aligned}
&\beta^{'} = 2\frac{1}{q^2}\sum_{s=1}^{q-1}\sum_{k>s}(q-s)(q-k)cov\left(\widehat{\gamma}_s, z_k\right) \\
= \;& \frac{2}{q^2}\sum_{s=1}^{q-2}(q-(s+1))(q-s)\left[4\sigma_\eta^4 + 4\sigma_\eta^2\sigma_1^2 + 2\sigma_\eta^2(\sigma_{s+1}^2 - \sigma_2^2) + 2\sigma_1^2\sigma_{s+1}^2\right] \\
&+\frac{2}{q^2}\sum_{s=1}^{q-3}(q-(s+2))(q-s)\left[-2\sigma_\eta^4 + 2\sigma_\eta^2\left((\sigma_2^2-\sigma_1^2)-(\sigma_3^2-\sigma_2^2)\right) + 2\sigma_2^2\sigma_{s+2}^2\right] \\
&+\frac{2}{q^2}\sum_{s=1}^{q-4}\sum_{k>s+2}(q-k)(q-s)\left[2\sigma_\eta^2\left((\sigma_s^2-\sigma_{s-1}^2)-(\sigma_{s+1}^2-\sigma_s^2)\right) + 2\sigma_s^2\sigma_k^2\right] \\
= \;& \frac{2}{q^2}\sum_{s=1}^{q-2}(q-(s+1))(q-s)\left[4\sigma_\eta^4 + 4\sigma_\eta^2\frac{V}{m} + 2\frac{V^2}{m^2}\right] \\
&+\frac{2}{q^2}\sum_{j=1}^{q-3}(q-(s+2))(q-s)\left[-2\sigma_\eta^4 + 2\frac{V^2}{m^2}\right] + \frac{2}{q^2}\sum_{s=1}^{q-4}\sum_{k>s+2}(q-k)(q-s)\left[2\frac{V^2}{m^2}\right].
\end{aligned}$$

Simple algebra leads to

$$\begin{aligned}
\beta^{'} \;=\;& -2\sigma_\eta^4 + \frac{4}{3}\sigma_\eta^4 q + \frac{1}{q}\frac{2}{3}\sigma_\eta^4 - 4\sigma_\eta^4\frac{1}{q^2} - 8\sigma_\eta^2\frac{V}{m} - \frac{17}{4}\frac{V^2}{m^2} \\
&+\frac{16}{3}\sigma_\eta^2\frac{V}{qm} + \frac{7}{2}\frac{V^2}{qm^2} - \frac{5}{4}\frac{V^2}{m^2}q + \frac{1}{4}\frac{V^2}{m^2}q^2 + \frac{8}{3}\sigma_\eta^2\frac{V}{m}q + 2\frac{1}{q^2}\frac{V^2}{m^2}.
\end{aligned}$$

When computing the covariance between α and β we can treat the terms involving $\widetilde{z}$ similarly as the terms involving z, in that

$$\begin{aligned}
cov(\widehat{\gamma}_0, \widetilde{z}_1) &= 10\sigma_\eta^4 + 8\sigma_\eta^2\sigma_m^2 + 2\sigma_m^4, \\
cov(\widehat{\gamma}_0, \widetilde{z}_2) &= -4\sigma_\eta^4 + 4\sigma_\eta^2\left(\sigma_{m-1}^2 - \sigma_m^2\right) + 2\sigma_{m-1}^4, \\
cov(\widehat{\gamma}_0, \widetilde{z}_j) &= 4\sigma_\eta^2\left(\sigma_{m-j+1}^2 - \sigma_{m-j+2}^2\right) + 2\sigma_{m-j+1}^4 \qquad j \geq 3,
\end{aligned}$$

and

$$
\begin{aligned}
cov(\widehat{\gamma}_1, \widetilde{z}_1) &= -4\sigma_\eta^4 - 2\sigma_\eta^2\sigma_m^2, \\
cov(\widehat{\gamma}_j, \widetilde{z}_j) &= -2\sigma_\eta^4 - 2\sigma_\eta^2\sigma_m^2 \qquad j \geq 2, \\
cov(\widehat{\gamma}_j, \widetilde{z}_{j+1}) &= 4\sigma_\eta^4 + 2\sigma_\eta^2\sigma_m^2 + 2\sigma_\eta^2(\sigma_{m-j}^2 - \sigma_{m-1}^2) + 2\sigma_m^2\sigma_{m-j}^2, \\
cov(\widehat{\gamma}_j, \widetilde{z}_{j+2}) &= -2\sigma_\eta^4 + 2\sigma_\eta^2\left[(\sigma_{m-1}^2 - \sigma_m^2) - (\sigma_{m-2}^2 - \sigma_{m-1}^2)\right] + 2\sigma_{m-1}^2\sigma_{m-j-1}^2, \\
cov(\widehat{\gamma}_j, \widetilde{z}_{j+i}) &= 2\sigma_\eta^2\left[(\sigma_{m-i+1}^2 - \sigma_{m-i+2}^2) - (\sigma_{m-i}^2 - \sigma_{m-i+1}^2)\right] + 2\sigma_{m-i+1}^2\sigma_{m-j-i+1}^2, \qquad i \geq 3.
\end{aligned}
$$

Putting all the elements together, we obtain

$$
\begin{aligned}
&Var\left(\widehat{V}^{ZMA}\right) = 4\sigma_\eta^4 + \frac{20}{m}\sigma_\eta^4 + \left(-\frac{4}{m^2}\sigma_\eta^4 + \frac{24}{m^2}\sigma_\eta^2 V + \frac{13}{3}\frac{1}{m^2}Q\right) + \frac{2}{m^3}Q \\
&+8\frac{m}{q^2}\sigma_\eta^4 + \frac{2}{m^3q^2}Q - \frac{1}{3}\frac{q^2}{m^2}Q - \frac{4}{m^3q}Q + \left(-12\sigma_\eta^4 + 8\sigma_\eta^2 V\right)\frac{1}{q} + \left(-8\sigma_\eta^4 - 8\sigma_\eta^2 V\right)\frac{1}{q^2} \\
&+\left(\frac{8}{3}\sigma_\eta^2 V + \frac{4}{3}Q\right)\frac{q}{m} + \left(4\sigma_\eta^4 + 16\sigma_\eta^2 V + 2Q\right)\frac{1}{mq^2} + \left(8\sigma_\eta^4 - 48\sigma_\eta^2 V - 8Q\right)\frac{1}{m^2q} \\
&+\left(-4\sigma_\eta^4 + 24\sigma_\eta^2 V + 4Q\right)\frac{1}{m^2q^2} + \left(-\frac{56}{3}\sigma_\eta^2 V - \frac{10}{3}Q - 24\sigma_\eta^4\right)\frac{1}{mq} + 4\sigma_\eta^4 + \frac{1}{q}\left(4\sigma_\eta^4\right) \\
&+\frac{1}{q^2}(-8\sigma_\eta^4) + \frac{1}{m^2}(\frac{4}{3}V^2) + \frac{1}{mq}\left(-\frac{8}{3}V\sigma_\eta^2\right) + \frac{q}{m}\left(\frac{8}{3}\sigma_\eta^2 V\right) + \frac{q^2}{m^2}\left(\frac{2}{3}V^2\right) + \frac{1}{qm^2}(-\frac{2}{3}V^2) \\
&-\frac{4}{3}\frac{q}{m^2}V^2 - 24\sigma_\eta^4 + \frac{32}{q}\sigma_\eta^4 - \frac{8}{q^2}\sigma_\eta^4 - \frac{1}{m}(24\sigma_\eta^4 + 32\sigma_\eta^2 V) - \frac{1}{m^2}(32\sigma_\eta^2 V - 8V^2) \\
&-4\frac{V^2}{m^3}q - 4\frac{V^2}{m^2}q - \frac{1}{qm}\left(-32\sigma_\eta^4 - 64\sigma_\eta^2 V\right) - \frac{1}{qm^2}\left(4V^2 - 64\sigma_\eta^2 V\right) - \frac{1}{qm^3}\left(4V^2\right) - \frac{1}{q^2m^2}\left(32\sigma_\eta^2 V\right) \\
&-\frac{1}{mq^2}\left(8\sigma_\eta^4 + 32\sigma_\eta^2 V\right) - \frac{1}{m^3}(-8V^2) + 8\sigma_\eta^4 + \frac{16\sigma_\eta^4}{q^2} - \frac{88}{3}\frac{\sigma_\eta^4}{q} - 8\sigma_\eta^2\frac{V}{m} + \frac{8}{3}\sigma_\eta^2 V\frac{1}{qm} \\
&+\frac{16}{3}\sigma_\eta^4 q + \frac{16}{3}\sigma_\eta^2 V\frac{q}{m} + 8\sigma_\eta^4 - \frac{16}{3}\sigma_\eta^4 q - \frac{1}{q}\frac{8}{3}\sigma_\eta^4 + 16\sigma_\eta^4\frac{1}{q^2} + 32\sigma_\eta^2\frac{V}{m} + 17\frac{V^2}{m^2} \\
&-\frac{64}{3}\sigma_\eta^2\frac{V}{qm} - 14\frac{V^2}{qm^2} + 5\frac{V^2}{m^2}q - \frac{V^2}{m^2}q^2 - \frac{32}{3}\sigma_\eta^2\frac{V}{m}q - 8\frac{1}{q^2}\frac{V^2}{m^2}.
\end{aligned}
$$

Now rewrite, in terms of ϕ, as

$$
\begin{aligned}
&Var\left(\widehat{V}^{ZMA}\right) = (-4\sigma_\eta^4 - 8V\sigma_\eta^2)\frac{1}{m} + \left(-4\sigma_\eta^4 - 8\sigma_\eta^2 V + \frac{13}{3}Q + \frac{79}{3}V^2\right)\frac{1}{m^2} \\
&+\frac{1}{m^3}\left(2Q + 8V^2\right) + 8\frac{1}{\phi^2 m}\sigma_\eta^4 - \frac{1}{3}(Q + V^2)\phi^2 - \frac{1}{3}V^2\frac{\phi}{m} - 4V^2\frac{\phi}{m^2} + \frac{2}{m^5\phi^2}Q - \frac{4}{m^4\phi}\left(Q + V^2\right) \\
&+\left(-8\sigma_\eta^4 + 8\sigma_\eta^2 V\right)\frac{1}{\phi m} + \left(8\sigma_\eta^4 - 8\sigma_\eta^2 V\right)\frac{1}{\phi^2 m^2} + \left(\frac{4}{3}Q\right)\phi + \left(-4\sigma_\eta^4 - 16\sigma_\eta^2 V + 2Q\right)\frac{1}{m^3\phi^2} \\
&+\left(8\sigma_\eta^4 + 16\sigma_\eta^2 V - 8Q - \frac{56}{3}V^2\right)\frac{1}{m^3\phi} + \left(-4\sigma_\eta^4 - 8\sigma_\eta^2 V + 4Q - 8V^2\right)\frac{1}{m^4\phi^2} \\
&+\left(24\sigma_\eta^2 V - \frac{10}{3}Q + 8\sigma_\eta^4\right)\frac{1}{m^2\phi}.
\end{aligned}
$$

Then,

$$Var\left(\widehat{V}^{ZMA}\right) = K - \frac{1}{3}(Q+V^2)\phi^2 + \left(-\frac{1}{3}V^2\frac{1}{m} - 4V^2\frac{1}{m^2} + \frac{4}{3}Q\right)\phi$$
$$+\left[-\frac{4}{m^4}\left(Q+V^2\right) + \left(\frac{8\sigma_\eta^4 + 16\sigma_\eta^2 V - 8Q - \frac{56}{3}V^2}{m^3}\right) + \left(\frac{24\sigma_\eta^2 V - \frac{10}{3}Q + 8\sigma_\eta^4}{m^2}\right) + \left(\frac{-8\sigma_\eta^4 + 8\sigma_\eta^2 V}{m}\right)\right]\frac{1}{\phi}$$
$$+\left[\frac{2}{m^5}Q + \left(\frac{-4\sigma_\eta^4 - 8\sigma_\eta^2 V + 4Q - 8V^2}{m^4}\right) + \left(\frac{-4\sigma_\eta^4 - 16\sigma_\eta^2 V + 2Q}{m^3}\right) + \left(\frac{8\sigma_\eta^4 - 8\sigma_\eta^2 V}{m^2}\right) + \frac{8}{m}\sigma_\eta^4\right]\frac{1}{\phi^2},$$

where

$$K = (-4\sigma_\eta^4 - 8V\sigma_\eta^2)\frac{1}{m} + \left(-4\sigma_\eta^4 - 8\sigma_\eta^2 V + \frac{13}{3}Q + \frac{79}{3}V^2\right)\frac{1}{m^2} + \frac{1}{m^3}\left(2Q + 8V^2\right).$$

As for the bias term, write

$$\begin{aligned}
\frac{1}{q}E(\vartheta_q) &= \frac{1}{q}\sum_{s=1}^{q-1}(q-s)E(z_s) + \frac{1}{q}\sum_{s=1}^{q-1}(q-s)E(\widetilde{z}_s) \\
&= \frac{1}{q}(q-1)2\sigma_\eta^2 + \frac{1}{q}\sum_{s=1}^{q-1}(q-s)\sigma_s^2 + \frac{1}{q}(q-1)2\sigma_\eta^2 + \frac{1}{q}\sum_{s=1}^{q-1}(q-s)\sigma_{m+1-s}^2 \\
&= \left(4 - \frac{4}{q}\right)\sigma_\eta^2 + \frac{q}{m}V - \frac{V}{m}.
\end{aligned}$$

Hence,

$$\begin{aligned}
bias\left(\widehat{V}^{ZMA}\right) &= \left(\frac{mq - m + q - 1}{mq}\right)\left(V + 2m\sigma_\eta^2\right) - V - 2\left(\frac{q-1}{q}\right)\sigma_\eta^2\,(m-1) - \left(4 - \frac{4}{q}\right)\sigma_\eta^2 - \frac{q}{m}V + \frac{V}{m} \\
&= -\frac{V}{q} + \frac{2V}{m} - \frac{q}{m}V - \frac{V}{qm}.
\end{aligned}$$

Therefore, in terms of ϕ,

$$\begin{aligned}
\left(bias\left(\widehat{V}^{ZMA}\right)\right)^2 &= \frac{V^2}{\phi^2 m^2} + \frac{4V^2}{m^2} + \phi^2 V^2 + \frac{V^2}{\phi^2 m^4} - \frac{4V^2}{\phi m^2} + \frac{2V^2}{m} + \frac{2V^2}{\phi^2 m^3} - 4\frac{\phi}{m}V^2 - 4\frac{V^2}{\phi m^3} + \frac{2V^2}{m^2} \\
&= \left(\frac{6V^2}{m^2} + \frac{2V^2}{m}\right) + V^2\phi^2 - \frac{4V^2}{m}\phi + \left(-4\frac{V^2}{m^2} - 4\frac{V^2}{m^3}\right)\frac{1}{\phi} + \left(\frac{V^2}{m^2} + \frac{2V^2}{m^3} + \frac{V^2}{m^4}\right)\frac{1}{\phi^2}.
\end{aligned}$$

■

Proof of Theorem 3. Write

$$\begin{aligned}
\widehat{V}^{BNHLS} &= \widehat{\gamma}_0 + \sum_{s=1}^{q} k\left(\frac{s-1}{q}\right)\left(\widehat{\gamma}_s + \widehat{\gamma}_{-s}\right) \\
&= \widehat{\gamma}_0^{r^e,r^e} + \widehat{\gamma}_0^{\varepsilon,\varepsilon} + \widehat{\gamma}_0^{r^e,\varepsilon} + \widehat{\gamma}_0^{\varepsilon,r^e} + \\
&\quad + \sum_{s=1}^{q} k\left(\frac{s-1}{q}\right)\left(\widehat{\gamma}_s^{r^e,r^e} + \widehat{\gamma}_s^{\varepsilon,\varepsilon} + \widehat{\gamma}_s^{r^e,\varepsilon} + \widehat{\gamma}_s^{\varepsilon,r^e} + \widehat{\gamma}_{-s}^{r^e,r^e} + \widehat{\gamma}_{-s}^{\varepsilon,\varepsilon} + \widehat{\gamma}_{-s}^{r^e,\varepsilon} + \widehat{\gamma}_{-s}^{\varepsilon,r^e}\right) \\
&= w^\perp\widetilde{\gamma}^{r^e,r^e} + w^\perp\widetilde{\gamma}^{\varepsilon,\varepsilon} + w^\perp\widetilde{\gamma}^{r^e,\varepsilon} + w^\perp\widetilde{\gamma}^{\varepsilon,r^e},
\end{aligned}$$

where $\widehat{\gamma}_s^{y,x} = \sum_{i=1}^{m} y_i x_{i-s}$ with $s = -q, ..., q$,

$$\widetilde{\gamma}^{y,x}=\left(\widehat{\gamma}_0^{y,x},\widehat{\gamma}_1^{y,x}+\widehat{\gamma}_{-1}^{y,x},...,\widehat{\gamma}_q^{y,x}+\widehat{\gamma}_{-q}^{y,x}\right),$$

and

$$w=\left(1,1,k\left(\frac{1}{q}\right),...,k\left(\frac{q-1}{q}\right)\right)^\intercal.$$

Now, notice that

$$\begin{aligned}E\left(\widehat{V}^{BNHLS}\right) &= w^\intercal E\left(\widetilde{\gamma}^{r^e,r^e}\right)+w^\intercal E(\widetilde{\gamma}^{\varepsilon,\varepsilon})+w^\intercal E\left(\widetilde{\gamma}^{r^e,\varepsilon}\right)+w^\intercal E\left(\widetilde{\gamma}^{\varepsilon,r^e}\right)\\ &= V+2m\sigma_\eta^2-m\sigma_\eta^2-m\sigma_\eta^2=V.\end{aligned}$$

Furthermore,

$$Var\left(\widehat{V}^{BNHLS}\right)=w^\intercal Var\left(\widetilde{\gamma}^{r^e,r^e}\right)w+w^\intercal Var(\widetilde{\gamma}^{\varepsilon,\varepsilon})w+w^\intercal Var\left(\widetilde{\gamma}^{r^e,\varepsilon}+\widetilde{\gamma}^{\varepsilon,r^e}\right)w,$$

since

$$\begin{aligned}Cov\left(w^\intercal\widetilde{\gamma}^{r^e,r^e},w^\intercal\widetilde{\gamma}^{\varepsilon,\varepsilon}\right) &= 2mV\sigma_\eta^2-2mV\sigma_\eta^2=0,\\ Cov\left(w^\intercal\widetilde{\gamma}^{r^e},w^\intercal\widetilde{\gamma}^{r^e,\varepsilon}+w^\intercal\widetilde{\gamma}^{\varepsilon,r^e}\right) &= 0,\end{aligned}$$

and $Cov\left(w^\intercal\widetilde{\gamma}^{\varepsilon},w^\intercal\widetilde{\gamma}^{r^e,\varepsilon}+w^\intercal\widetilde{\gamma}^{\varepsilon,r^e}\right)=0$. The final result can be easily derived by suitably re-expressing the quadratic forms in Theorem 1 of Barndorff-Nielsen et al. (2005, 2008). ■

References

[1] Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica* **71**, 579-625.

[2] Andrews, D.W.K. (1991). An heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* **59**, 817-854.

[3] Andrews, D.W.K., and J.C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* **60**, 953-966.

[4] Awartani, B, V. Corradi, and W. Distaso (2009). Assessing market microstructure effects via realized volatility measure with an application to the Dow Jones Industrial Average stocks. *Journal of Business and Economic Statistics* **27**, 251-265.

[5] Bandi, F.M., and J.R. Russell (2003). Microstructure noise, realized volatility, and optimal sampling. *Working paper.*

[6] Bandi, F.M., and J.R. Russell (2006a). Separating market microstructure noise from volatility. *Journal of Financial Economics* **79**, 655-692.

[7] Bandi, F.M., and J.R. Russell (2006b). Comment on "Realized variance and market microstructure noise" by Hansen and Lunde. *Journal of Business and Economic Statistics* **24**, 167-173.

[8] Bandi, F.M., and J.R. Russell (2008). Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies* **75**, 339-369.

[9] Bandi, F.M., J.R. Russell, and C. Yang (2008). Realized variance forecasting and option pricing. *Journal of Econometrics* **147**, 34-46.

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

[11] Barndorff-Nielsen, O.E., P. Hansen, A. Lunde and N. Shephard (2005). Regular and modified kernel-based estimators of integrated variance: the case with independent noise. *Working paper.*

[12] Barndorff-Nielsen, O.E., P. Hansen, A. Lunde and N. Shephard (2008). Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* **76**, 1481-1536.

[13] Bartlett, M.S. (1950). Periodogram analysis and continuous spectra. *Biometrica* **37**, 1-16.

[14] Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* **7**, 174-196.

[15] Ghysels, E., Santa-Clara P., and R. Valkanov (2006). Predicting volatility: Getting the most out of return data sampled at different frequencies. *Journal of Econometrics* **131**, 59-95.

[16] Hansen, P.R., and A. Lunde (2005). A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics* **4**, 525-554.

[17] Hansen, P.R., and A. Lunde (2006). Realized variance and market microstructure noise (with discussion). *Journal of Business and Economic Statistics* **24**, 127-218.

[18] Kiefer, N.M., and T. J. Vogelsang (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory* **21**, 1130-1164.

[19] Neave, H.R. (1970). An improved formula for the asymptotic variance of the spectrum estimates. *Annals of Mathematical Statistics* **41**, 70-77.

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

[21] Oomen, R.C.A. (2005). Properties of bias-corrected realized variance under alternative sampling schemes. *Journal of Financial Econometrics* **3**, 555-577.

[22] Oomen, R.C.A. (2006). Properties of realized variance under alternative sampling schemes. *Journal of Business and Economic Statistics* **24**, 219-233.

[23] Renault, E. and B. Werker (2008). Causality effects in return volatility measures with random times. *Working paper*.

[24] Sun, Y. (2006). Best quadratic unbiased estimators of integrated variance. *Working paper.*

[25] Zhang, L., P. Mykland, and Y. Aït-Sahalia (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* **100**, 1394-1411.

[26] Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. *Bernoulli* **12**, 1019-1043.

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

Table 1. Features of the noise and price process for alternative representative stocks: Goldman Sachs (GS), SBC Communications (SBC), and EXXON Mobile Corporation (XOM)

		GS	**SBC**	**XOM**
Second moment of the noise - σ_{η}^{2}		0.87e-07	1.89e-07	2.1e-07
Integrated variance – V		0.00042	0.00041	0.00018
Integrated quarticity – Q		2.31e-07	2.1e-07	4.1e-08
Average number of intra-daily obs – m		2,247	2,034	2,630

Table 2. The finite-sample MSE-optimal number of auto-covariances for alternative integrated variance estimators.

		GS	**SBC**	**XOM**
(1) Bartlett - q_{Bar}^{*}		13	14	15
(2) Two-scale - q_{ZMA}^{*}		13	14	15
(3) Corrected Bartlett - $q_{Bar\_adj}^{*}$		3	4	8
(4) Corrected two-scale - $q_{ZMA\_adj}^{*}$		3	4	8
(5) Corrected two-scale 2 - $q_{ZMA\_adj2}^{*}$		3	5	8
(6) Flat-top Bartlett - $q_{BNHLS\_Bartlett}^{*}$		2	3	6
(7) Flat-top cubic - $q_{BNHLS\_Cubic}^{*}$		2	3	5
(8) Flat-top Tukey - $q_{BNHLS\_TH}^{*}$		3	4	8

(1) is the Bartlett kernel estimator (q is derived in Theorem 1), (2) is the two-scale estimator (q is derived in Theorem 2), (3) is the bias-corrected (as in Corollary 1) Bartlett kernel estimator (q is derived in Corollary 1), (4) is the bias-corrected (as in Corollary 2) two-scale estimator (q is derived in Corollary 2), (5) is the two-scale estimator bias-corrected as in Zhang et al. (2005) (q is derived in Corollary 3), (6) is the flat-top Bartlett estimator (q is derived in Theorem 3), (7) is the flat-top cubic estimator (q is derived in Theorem 3), (8) is the flat-top Tukey-Hanning estimator (q is derived in Theorem 3).

Table 3. The finite sample MSEs of alternative integrated variance estimators with different choices of the number of autocovariances (*q*) and sampling frequency (*1/m)*.

		GS	SBC	XOM
(1) Realized variance - $1/m^{*}_{\hat{\gamma}_0}$		5e-09	7.5e-09	2.5e-09
(2) Bartlett - q^{*}_{Bar}		2.82e-09	2.82e-09	4.78e-10
(3) Two-scale - q^{*}_{ZMA}		2.95e-09	2.95e-09	4.98e-10
(4) Corrected Bartlett - $q^{*}_{Bar\_adj}$		9.13e-10	1.12e-09	2.71e-10
(5) Corrected two-scale - $q^{*}_{ZMA\_adj}$		9.07e-10	1.11e-09	2.67e-10
(6) Corrected two-scale 2 - $q^{*}_{ZMA\_adj2}$		9.13e-10	1.12e-09	2.71e-10
(7) Two-scale - $\tilde{q}_{ZMA}$		1.5e-07	4.1e-08	1.36e-09
(8) Flat-top Bartlett - $q^{*}_{BNHLS\_Bartlett}$		8.99e-10	1.12e-09	2.65e-10
(9) Flat-top cubic - $q^{*}_{BNHLS\_Cubic}$		8.99e-10	1.16e-09	2.79e-10
(10) Flat-top Tukey - $q^{*}_{BNHLS\_TH}$		8.65e-10	1.09e-09	2.60e-10

(1) is realized variance optimized using the finite sample methods in Bandi and Russell (2006, 2008), (2) is the Bartlett kernel estimator with a finite-sample optimal number of auto-covariances as derived in Theorem 1, (3) is the two-scale estimator with a finite-sample optimal number of sub-samples as derived in Theorem 2, (4) is the bias-corrected Bartlett kernel with a finite-sample optimal number of auto-covariances as derived in Corollary 1, (5) is the bias-corrected two-scale estimator with a finite-sample optimal number of sub-samples as derived in Corollary 2, (6) is the two-scale estimator bias-corrected as in Zhang et al. (2005) with a finite-sample optimal number of sub-samples as derived in Corollary 3, (7) is the two-scale estimator with an asymptotically-optimal number of sub-samples as derived in Zhang et al. (2005), (8) is the flat-top Bartlett kernel estimator with a finite-sample optimal number of auto-covariances as derived in Theorem 3, (9) is the flat-top cubic kernel estimator with a finite-sample optimal number of auto-covariances as derived in Theorem 3, (10) is the flat-top Tukey-Hanning kernel estimator with a finite-sample optimal number of auto-covariances as derived in Theorem 3**.**

Table 4. The asymptotic MSE-optimal number of auto-covariances for alternative integrated variance estimators.

		GS	SBC	XOM
(1) Two-scale - $\tilde{q}_{ZMA}$		1	2	4
(2) Corrected two-scale - $\tilde{q}_{ZMA}$		1	2	4
(3) Flat-top Bartlett - $\tilde{q}_{BNHLS\_Bartlett}$		1	2	4
(4) Flat-top cubic - $\tilde{q}_{BNHLS\_Cubic}$		2	3	6

(1) is the two-scale estimator (*q* is derived in Zhang et al., 2005), (2) is the bias-corrected (as in Zhang et al., 2005) two-scale estimator (*q* is derived in Zhang et al., 2005), (3) is the flat-top Bartlett kernel estimator (*q* is derived in Barndorff-Nielsen et al., 2008), (4) is the flat-top cubic kernel estimator (*q* is derived in Barndorff-Nielsen et al., 2008).

Table 5. The asymptotically-optimal MSEs of alternative integrated variance estimators with different choices of the number of auto-covariances (q).

		GS	SBC	XOM
(1) Two-scale - $\tilde{q}_{ZMA}$		2.58e-10	4.2e-10	1.39e-10
(2) Corrected two-scale - $\tilde{q}_{ZMA}$		2.58e-10	4.2e-10	1.39e-10
(3) Flat-top Bartlett - $\tilde{q}_{BNHLS\_Bartlett}$		2.58e-10	4.2e-10	1.39e-10
(4) Flat-top cubic - $\tilde{q}_{BNHLS\_Cubic}$		5.7e-10	8.3e-10	2.25e-10

(1) is the two-scale estimator with an asymptotically-optimal (as in Zhang et al., 2005) number of auto-covariances, (2) is the bias-corrected (as in Zhang et al., 2005) two-scale estimator with an asymptotically-optimal (as in Zhang et al., 2005) number of auto-covariances, (3) is the flat-top Bartlett kernel estimator with an asymptotically-optimal (as in Barndorff-Nielsen et al., 2008) number of auto-covariances, (4) is the flat-top cubic kernel estimator with an asymptotically-optimal (as in Barndorff-Nielsen et al., 2008) number of auto-covariances.

Table 6. Average utilities and standard errors associated with alternative integrated variance estimators

		Average utilities	Standard errors
(1) Flat-top cubic - $q^*_{BNHLS\_Cubic}$		1.433e-04	2.21e-06
(2) Flat-top Bartlett - $q^*_{BNHLS\_Bartlett}$		1.433e-04	2.19e-06
(3) Flat-top Tukey - $q^*_{BNHLS\_TH}$		1.429e-04	2.23e-06
(4) Flat-top cubic - $\tilde{q}_{BNHLS\_Cubic}$		1.429e-04	2.29e-06
(5) Two-scale - q^*_{ZMA}		1.427e-04	2.46e-06
(6) Bartlett - q^*_{Bar}		1.427e-04	2.46e-06
(7) VIX		1.426e-04	1.47e-06
(8) Flat-top Tukey - $\tilde{q}_{BNHLS\_TH}$		1.425e-04	2.29e-06
(9) Flat-top Bartlett - $\tilde{q}_{BNHLS\_Bartlett}$		1.425e-04	2.29e-06
(10) Corrected two-scale - $q^*_{ZMA\_adj}$		1.420e-04	2.37e-06
(11) Corrected two-scale - $\tilde{q}_{ZMA}$		1.412e-04	2.55e-06
(12) Two-scale - $\tilde{q}_{ZMA}$		1.402e-04	2.66e-06

In the table, the symbols q^* and $\tilde{q}$ stand for number of auto-covariances chosen using finite sample MSE-based methods, as in this paper, and asymptotic methods, as in the original papers, respectively.

Table 7. Pair-wise t-statistics for the null of equal mean utilities for estimators optimized using finite sample methods and asymptotic methods. * represents significance at the 5% level.

Flat-top Bartlett	Flat-top cubic	Flat-top Tukey	Corrected two-scale	Two-scale
2.12*	1.15	1.62	2.61*	2.35*