Best Quadratic Unbiased Estimators of Integrated Variance in the Presence of Market Microstructure Noise[*]

Yixiao Sun
Department of Economics
University of California, San Diego

First version: February 2006
This Version: July 2006

Abstract

We consider the best quadratic unbiased estimators of the integrated variance in the presence of independent market microstructure noise. We establish the asymptotic normality of a feasible best quadratic unbiased estimator under the assumption of constant volatility and show that it is asymptotically efficient when the market microstructure noise is normal. Since the class of quadratic estimators includes all the existing estimators of the integrated variance as special cases, the best quadratic unbiased estimator outperforms the existing estimators in terms of root mean squared error, often by a large margin.

JEL Classification: C13; C22

Keywords: Best Quadratic Unbiased Estimator, Market Microstructure Noise, Quadratic Form Estimator, Realized Variance, Realized Volatility.

1 Introduction

The paper considers estimating the integrated variance using high frequency asset price data. A salient feature of high frequency data is the presence of market microstructure noise. This feature renders both the classical quadratic variation estimator (e.g. Anderson, Bollersllev, Diebold and Labys (2003), Bardorff-Nielsen and Shephard (2002)) and the kernel-based estimators (e.g. Zhou (1996), Hansen and Lunde (2006, hereafter HL)) inconsistent for the integrated variance. To tackle the inconsistency problem, Zhang, Mykland, Aït-Sahalia (2005, hereafter ZMA) propose a subsampling-based estimator, which is consistent and subsequently refined by Zhang (2006). In practice, sparse sampling has been recommended to reduce the market microstructure contamination. Optimal sampling schemes have been investigated by Bandi and Russell (2004a, 2004b, hereafter BR) and ZMA (2005).

In this paper, we consider a class of quadratic estimators that includes all the above existing estimators as special cases. Given a sequence of high frequency asset returns $r = (r_1, r_2, ..., r_m)$, a quadratic estimator is a linear function of the cross-products $\{r_i r_j\}$. In other words, it can be written as a quadratic form $\hat{V}_Q = r'Wr := \sum_{i=1}^{m} \sum_{j=1}^{m} W(i,j) r_i r_j$ where W is a symmetric and positive definite matrix with elements $W(i,j)$. Our objective is to choose W to minimize the variance of $\hat{V}_Q$ subject to some unbiasedness conditions. The resulting estimator is a best quadratic unbiased (BQU) estimator. We consider two sets of unbiasedness conditions. The first set uses prior information on model parameters while the second does not. Different unbiasedness conditions lead to different BQU estimators. In this paper, we establish explicit multi-window representations of the two BQU estimators and provide a geometric interpretation of both estimators. More specifically, let $h^{(k)} = (h_1^{(k)}, ..., h_j^{(k)}, ..., h_m^{(k)})$ where $h_j^{(k)} = \sqrt{2/(m+1)} \sin(\pi k j/(m+1))$ be an orthonormal basis in $\mathbb{R}^m$ and $\alpha^{(k)} h^{(k)}$ be the projection of r onto the subspace spanned by $h^{(k)}$. Then each BQU estimator is a weighted sum of $(\alpha^{(k)})^2$, the squared lengths of the projections. Interestingly, the classical quadratic variation estimator can be written as the simple unweighted sum of these squared lengths.

The two BQU estimators are infeasible, as they depend on the unknown parameter λ, the signal-to-noise ratio. Replacing the unknown λ by a consistent pilot estimate yields the feasible best quadratic unbiased estimators. Under the assumption of constant volatility, we establish the asymptotic normality of the two feasible BQU estimators and show that they converge to the true realized variance at the rate of $m^{-1/4}$, the best attainable

1

rate for nonparametric variance estimators. More importantly, when the market noise is normally distributed, one of the feasible BQU estimators is asymptotically as efficient as the maximum likelihood estimator but computationally much simpler and more robust to model mis-specifications.

Quadratic estimators have been employed in estimating variance components in the statistical literature. The monograph of Rao and Kleffe (1988) provides an extensive survey of this literature. In the time series literature, quadratic estimators have been used in estimating the variance of a sample mean; see Song and Schmeiser (1993). The multi-taper estimator of a spectral density (e.g. Percival and Walden (1993, Ch 7)) also belongs to the class of quadratic estimators. Some long run variance estimators in the econometrics literature can be written as quadratic estimators (see Sun (2004)). Therefore, the idea of best quadratic unbiased estimators has a long history but its usage in the present context is new. We provide a systematic analysis of the BQU estimators under the infill asymptotics, the type of asymptotics that is suitable for high frequency financial data. The use of the infill asymptotics has a great number of technical implications and makes the analysis far from trivial. We show that the class of quadratic estimators not only unifies the previous literature on the integrated variance estimation but also leads to a new estimator that dominates the existing ones.

The paper that is closest to this paper is Bardorff-Nielsen, Hansen, Lunde, and Shephard (2006, hereafter BNHLS) where a weighted sum of autocovariances (or modified autocovariances) is used as the estimator of the integrated variance. BNHLS consider choosing the weights optimally to minimize the asymptotic variance of their estimator, subject to some unbiasedness conditions. It is easy to see that the BNHLS method amounts to solving the BQU-type problem but restricting the weighting matrix to be a symmetric Toeplitz matrix. This restriction is not innocuous at least in finite samples because the optimal weighting matrix given in this paper is not a Toeplitz matrix. In particular, for one of the BQU estimators, the optimal weighting matrix is the sum of a Toeplitz matrix and a Hankel matrix. As a result, the BNHLS estimator is not optimal under the constant volatility assumption maintained in this paper.

In this paper, we compare the finite sample performances of the BQU estimators with those of the HL, ZMA, BNHLS estimators and the multi-scale estimator (hereafter MS estimator) of Zhang (2006). The HL and ZMA estimators are implemented using their respective optimal truncation lags given in BR (2005). We consider two different levels of microstructure noise contaminations and three different noise distributions: normal, χ^2_1

2

and t_5. We employ three sets of parameter values that are representative of the S&P 100 stocks. Both constant volatility models and stochastic volatility models are considered. Our simulation results show that, under the assumption of constant volatility, one of the BQU estimators has the smallest root mean squared error (RMSE) among all the estimators considered. This BQU estimator reduces the RMSEs of the HL and ZMA estimators by 30% to 40% and the RMSE of the BNHLS and MS estimators by 5% to 10%. What is perhaps surprising is that the same BQU estimator also performs very well in stochastic volatility models. It dominates the HL and ZMA estimators and is outperformed by the BNHLS and MS estimators only when the noise contamination is very large.

The rest of the paper is organized as follows. Section 2 outlines the basic assumptions and introduces the quadratic estimator. Section 3 proposes a BQU estimator and establishes its multi-window representation. The next section investigates the BQU estimator under alternative unbiasedness conditions. Section 5 compares the BQU estimators with the existing estimators via Monte-Carlo experiments. Section 6 concludes. Proofs are given in the appendix.

2 The Model and Estimator

Following BR (2004b, 2005), we assume that the log-price process is given by:

$$p_t = p_t^e + \eta_t, \ t \in [0, T], \tag{1}$$

where p_t^e is the efficient logarithmic-price process and η_t is the noise process. Denote a trading day by $h = [0, 1]$, which is divided into m subperiods $t_i - t_{i-1}$ with $i = 1, 2, ..., m$ so that $t_0 = 0$ and $t_m = 1$. Now define

$$\underbrace{p_{t_i} - p_{t_{i-1}}}_{r_i} = \underbrace{p_{t_i}^e - p_{t_{i-1}}^e}_{r_i^e} + \underbrace{\eta_{t_i} - \eta_{t_{i-1}}}_{\varepsilon_i} \tag{2}$$

where r_i is an observed continuously-compounded intra-daily return, r_i^e is an efficient continuously-compounded intra-daily return, ε_i is a market microstructure contamination. As in the previous literature, we maintain the following assumptions:

Assumption 1 The efficient logarithmic-price process p_t^e is a stochastic volatility local martingale, defined by

$$p_t^e = \int_0^t \sigma_s dB_s \tag{3}$$

3

where $\{B_t, t \geq 0\}$ is a standard Brownian motion that is independent of the càdlàg spot volatility process $\{\sigma_s, s \geq 0\}$.

Assumption 2 The logarithmic price contaminations η_t are $\mathrm{iid}(0, \sigma_\eta^2)$ with a bounded fourth moment and independent of p_t^e.

The empirical validity of Assumption 2 depends on, for example, the sampling frequency and the nature of price measurements. When transaction prices are sampled at more than moderate frequencies such as every 15 ticks, there is little evidence against this assumption. See Bandi and Russell (2005) for more detailed discussions.

The object of interest is the integrated price variance over the trading day, i.e.

$$V = \int_0^1 \sigma_s^2 ds. \tag{4}$$

The problem of estimating V is, in some ways, similar to the estimation of the long run variance in time series analysis. It is not surprising that kernel-based estimators have been suggested in the literature. For a given sequence of weights $w_0, w_1, ..., w_q$, kernel-based estimators are defined as

$$\hat{V}_K = w_0 \hat{\gamma}_0 + 2 \sum_{s=1}^{q} w_s \hat{\gamma}_s, \text{ where } \hat{\gamma}_s = \sum_{i=1}^{m-s} r_i r_{i+s}. \tag{5}$$

Examples of the kernel-based estimators include Zhou (1996), HL (2006) and BNHLS (2006). The HL estimator is based on the Bartlett-type kernel and is given by[1]

$$\hat{V}_{HL} = \left(\frac{m-1}{m} \frac{q-1}{q} \right) \hat{\gamma}_0 + 2 \sum_{s=1}^{q} \left(1 - \frac{s}{q} \right) \hat{\gamma}_s. \tag{6}$$

The subsampling-based estimator of ZMA (2005) can be regarded as a modified version of the above estimator (see BR (2005) and BNHLS (2006)). It is defined as

$$\hat{V}_{ZMA} = \left(1 - \frac{m-q+1}{mq} \right) \hat{\gamma}_0 + 2 \sum_{s=1}^{q} \left(1 - \frac{s}{q} \right) \hat{\gamma}_s - \frac{1}{q} \vartheta_q \tag{7}$$

---

[1]Hansen and Lunde propose several estimators of the integrated variance. The estimator that is usually associated with HL is defined by

$$\hat{V} = \hat{\gamma}_0 + 2 \sum_{s=1}^{q} \frac{m}{m-s} \hat{\gamma}_s.$$

See Hansen and Lunde (2004). In this paper, we call the Bartlett-kernel-based estimator the HL estimator as it is almost identical to the estimator applied by HL (2005).

where the modification term ϑ_q satisfies

$$\vartheta_1 = 0, \ \vartheta_q = \vartheta_{q-1} + (r_1 + ... + r_{q-1})^2 + (r_{m-q+2} + ... + r_m)^2, \text{ for } q \geq 2. \quad (8)$$

It is the modification term, which the subsampling approach entails by construction, that makes the ZMA estimator consistent. In a recent paper, BR (2005) consider choosing q to minimize the mean squared error of $\hat{V}_{HL}$ and $\hat{V}_{ZMA}$.

The BNHLS estimator is based on the class of flat top kernels where a unit weight is imposed on the first autocovariance. The estimator can be represented as

$$\hat{V}_{BNHLS} = \tilde{\gamma}_0 + \sum_{s=1}^{q} k \left(\frac{s-1}{q} \right) (\tilde{\gamma}_s + \tilde{\gamma}_{-s}) \text{ where } \tilde{\gamma}_s = \sum_{i=q+1}^{m-q} r_i r_{i+s} \quad (9)$$

and $k(\cdot)$ is a kernel function. BNHLS (2006) propose using the following modified Tukey-Hanning kernel and bandwidth selection rule[2]:

$$k_{MTH}(x) = \left(1 - \cos \pi (1-x)^2 \right)/2, \quad (10)$$

$$q = 5.74\sqrt{m}\sigma_\eta/V. \quad (11)$$

A closely related estimator is the multi-scale (MS) estimator by Zhang (2006). BNHLS (2006) show that the multi-scale estimator is asymptotically equivalent to the BNHLS estimator based on the cubic kernel

$$k_{MS}(x) = 1 - 3x^2 + 2x^3. \quad (12)$$

The corresponding bandwidth selection rule is

$$q = 3.68\sqrt{m}\sigma_\eta/V. \quad (13)$$

In the sequel, we use $\hat{V}_{MS}$ to denote the multi-scale estimator and use $\hat{V}_{BNHLS}$ to refer to the modified-Tukey-Hanning-kernel-based BNHLS estimator.

In this paper, we consider an estimator of the form:

$$\hat{V}_Q = r'Wr \text{ where } r = (r_1, r_2, ..., r_m)' \quad (14)$$

and W is a symmetric and positive definite matrix. This estimator includes the estimators $\hat{V}_{HL}$, $\hat{V}_{ZMA}$, $\hat{V}_{BNHLS}$ and $\hat{V}_{MS}$ as special cases. Some algebraic manipulations show that

---

[2]BNHLS consider various kernels in their paper. The modified Tukey-Hanning kernel delivers the smallest asymptotic variance among all the kernels considered for the simple Brownian motion plus noise model.

the respective weighting matrices for these four estimators are

$$W_{HL}(i,i) = \frac{m-1}{m}\frac{q-1}{q},$$

$$W_{HL}(i,j) = \left(1 - \frac{|i-j|}{q}\right)\{|i-j| \le q\}, i \ne j, \tag{15}$$

$$W_{ZMA}(i,i) = 1 - \frac{m-q+1}{mq} - [q-i]\{i \le q-1\} - [q-m-1+i]\{i \ge m-q+2\},$$

$$W_{ZMA}(i,j) = \left(1 - \frac{|i-j|}{q}\right)\{|i-j| \le q\} - \frac{1}{q}[q-\max(i,j)]\{i \le q-1\}\{j \le q-1\}$$

$$- \frac{1}{q}[q-m-1+\min(i,j)]\{i \ge m-q+2\}\{j \ge m-q+2\}, \; i \ne j,$$

$$\tag{16}$$

$$W_{BNHLS}(i,j) = 1, \tag{17}$$

$$W_{BNHLS}(i,j) = \{q+1 \le i \le m-q\}\{i-q \le j \le i+q\}k_{MTH}\left(\frac{|i-j|-1}{q}\right), \; i \ne j,$$

and

$$W_{MS}(i,j) = 1, \tag{18}$$

$$W_{MS}(i,j) = \{q+1 \le i \le m-q\}\{i-q \le j \le i+q\}k_{MS}\left(\frac{|i-j|-1}{q}\right), \; i \ne j.$$

In the preceding equations, $\{\cdot\}$ is the indicator function.

What is more interesting is that the kernel-based estimator with an optimal sampling scheme is also a special case of the quadratic estimator. As an example, consider $m = 6$, $r = (r_1, r_2, r_3, r_4, r_5, r_6)$ and

$$W = \begin{pmatrix} W(1,1) & W(1,2) & W(1,3) \\ W(2,1) & W(2,2) & W(2,3) \\ W(3,1) & W(3,2) & W(3,3) \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{19}$$

Then the quadratic estimator $\hat{V}_Q = r'Wr$ uses only the observations $\{r_1, r_3, r_5\}$. The underlying sampling scheme is to sample every other observations. Obviously, the basic idea applies to any sample size and sampling scheme.

The next theorem characterizes the finite sample bias and variance of the quadratic estimator[3].

---

[3]All population moments are made conditional on the stochastic volatility process $\sigma^2(s), s \in [0, 1]$. For notational convenience, we write $E(\cdot) := E\left(\cdot|\sigma^2(s), s \in [0, 1]\right)$. $\text{Var}(\cdot)$ and $\text{Cov}(\cdot)$ are similarly defined.

Theorem 1 *Assume that $\eta's$ are mean zero random variables with $E\eta^4/\sigma_\eta^4 = \kappa_4$, then*

$$E\hat{V}_Q = tr(W\Omega) \ and \ var\left(\hat{V}_Q\right) = 2tr\left(W\Omega W\Omega\right) + \sigma_\eta^4(\kappa_4 - 3)\sum_{i=1}^{m+1}\omega_{ii}^2 \qquad (20)$$

where $\Omega = Err'$, ω_{ij} is the (i,j)-th element of the matrix $D'WD$ and D is the $m \times (m+1)$ matrix:

$$D = \begin{pmatrix} 1 & -1 & 0 & ... & 0 \\ 0 & 1 & -1 & 0 & ... \\ ... & \ddots & \ddots & \ddots & 0 \\ 0 & ... & 0 & 1 & -1 \end{pmatrix}. \qquad (21)$$

If we assume that η is normal as in BR (2005), we have $\kappa_4 = 3$ and $var(\hat{V}_Q) = 2tr\left(W\Omega W\Omega\right)$. When $\kappa_4 > 3$,

$$\sigma_\eta^4\sum_{i=1}^{m+1}\omega_{ii}^2 \leq \sigma_\eta^4\sum_{i=1}^{m+1}\sum_{j=1}^{m+1}\omega_{ij}^2 = tr\left(\sigma_\eta^4 D'WDD'WD\right) \leq tr\left(W\Omega W\Omega\right) \qquad (22)$$

and thus

$$var(\hat{V}_Q) \leq \phi tr\left(W\Omega W\Omega\right) \qquad (23)$$

for some ϕ that depends on κ_4 but does not depend on the matrix W. In the next section, we assume that $\kappa_4 = 3$ and select the weighting matrix to minimize $tr\left(W\Omega W\Omega\right)$. When $\kappa_4 > 3$, our procedure amounts to minimizing an upper bound of the finite sample variance.

It can be shown that the formula $E\hat{V}_Q - V = tr(W\Omega) - V$ is the same as that given in BR (2005, Theorems 1 and 2) for the HL and ZMA estimators. However, the variance formula is different. The main difference is that $Q = \int_0^1 \sigma_s^4 ds$, the so-called quarticity introduced in Barndorff-Nielsen and Shephard (2002), appears in BR's formula while our formula does not involve the quarticity. The difference can be explained by noting that BR (2005) employ the additional approximation: $plim_{m\to\infty}(1/3)\,mE\sum_{i=1}^m\left(r_i^e\right)^4 = Q$ while we employ the exact relationship $(1/3)\,mE\sum_{i=1}^m\left(r_i^e\right)^4 = m\sum_{i=1}^m(E\left(r_i^e\right)^2)^2$. In general $Q \neq m\sum_{i=1}^m(E\left(r_i^e\right)^2)^2$. Strictly speaking, the bias and variance obtained by BR (2005) are not the exact finite sample ones as they use large-m approximations in their proofs. In contrast, Theorem 1 gives the exact finite sample mean and variance under the conditional normal assumption (3).

3   The Best Quadratic Unbiased (BQU) Estimator

In this section, we first find the optimal weighting matrix for the quadratic estimator under the first choice of the unbiasedness condition and then establish the multi-window representation of the BQU estimator.

3.1   Optimal Weighting Matrix

Given Theorem 1, we seek to minimize the variance of $\hat{V}_Q$ subject to an unbiasedness condition. More specifically, we assume $\kappa_4 = 3$ and solve the following optimization problem:

$$\min tr\left(W\Omega W\Omega\right) \text{ s.t. } tr(W\Omega) = V \text{ and } W > 0. \tag{24}$$

Here '$W > 0$' signifies the positive definiteness of W. Using the theorem by Rao (1973) given in the appendix, we find that the solution is

$$W_{BQU} = \frac{V}{m}\Omega^{-1} \tag{25}$$

and the minimum variance is

$$var(\hat{V}_Q) = \frac{2V^2}{m}. \tag{26}$$

We call the quadratic estimator with the optimal weight W_{BQU} a best quadratic unbiased (BQU) estimator and denote it as $\hat{V}_{BQU}$.

If we seek a quadratic estimator with minimum mean squared error, we first solve the problem:

$$\min tr\left(W\Omega W\Omega\right) \text{ s.t. } tr(W\Omega) = b \text{ and } W > 0, \tag{27}$$

for any b. The solution is $W = m^{-1}b\Omega^{-1}$. The bias of the resulting quadratic estimator is $b - V$ and the variance is $2b^2/m$. To minimize the MSE, we then choose b to minimize $(b - V)^2 + 2b^2/m$. The optimal b is $b^0 = mV/(m + 2)$. The optimal W and the minimum MSE are

$$W^0 = \frac{V}{m + 2}\Omega^{-1}, \text{ MSE}^0 = \frac{2V^2}{m + 2}, \tag{28}$$

respectively. Comparing the optimal MSE with the minimum variance in (26), we find that the difference is very small when m is large. Therefore, when m is large, it does not make much difference whether one uses the minimum MSE estimator or the BQU estimator. In the rest of the paper, we focus on the BQU estimator.

Assume that $\sigma_i^2 := Er_i^2 = V/m$ for all i, then

$$\Omega = \begin{pmatrix} \frac{1}{m}V & 0 & \cdots & 0 \\ 0 & \frac{1}{m}V & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & \cdots & \frac{1}{m}V \end{pmatrix} + \begin{pmatrix} 2\sigma_\eta^2 & -\sigma_\eta^2 & \cdots & 0 \\ -\sigma_\eta^2 & 2\sigma_\eta^2 & -\sigma_\eta^2 & \cdots \\ \cdots & \cdots & \ddots & \cdots \\ 0 & \cdots & -\sigma_\eta^2 & 2\sigma_\eta^2 \end{pmatrix}, \tag{29}$$

and thus

$$W_{BQU} = m^{-1}V\Omega^{-1} = -\lambda\left(A(\lambda)\right)^{-1} \tag{30}$$

where $\lambda = V/(m\sigma_\eta^2)$ is the signal-to-noise ratio and

$$A(\lambda) = \begin{pmatrix} -(\lambda+2) & 1 & \cdots & 0 \\ 1 & -(\lambda+2) & 1 & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & 1 \\ 0 & & 1 & -(\lambda+2) \end{pmatrix} \tag{31}$$

is a $m \times m$ matrix that depends only on λ. To emphasize the dependence of W_{BQU} and $\hat{V}_{BQU}$ on λ, we sometimes write them as $W_{BQU}(\lambda)$ and $\hat{V}_{BQU}(\lambda)$, respectively.

The next theorem gives an explicit expression for W_{BQU} under the assumption $\sigma_i^2 = V/m$ for $i = 1, 2, ..., m$. This assumption is realistic when volatility does not change much within the day or the sampling is conducted in business time (e.g. Oomen (2005)). We use this assumption to derive a closed form solution for the optimal weighting matrix. BR (2005) also make this assumption in their derivations.

Theorem 2 *Assume that $\eta's$ are mean zero random variables with $\kappa_4 = 3$ and $\sigma_i^2 = V/m$ for $i = 1, 2, ..., m$. The weighting matrix for the BQU estimator is given by*

$$W_{BQU}(i,j) = \frac{\lambda}{\sqrt{\lambda(\lambda+4)}} \frac{\cosh\{[m+1-|i-j|]\alpha\} - \cosh\{[m+1-i-j]\alpha\}}{\sinh((m+1)\alpha)}, \tag{32}$$

where

$$\alpha = \log\left(\frac{1}{2}\lambda + \frac{1}{2}\sqrt{4\lambda + \lambda^2} + 1\right). \tag{33}$$

To understand the BQU estimator, we write

$$W_{BQU} = \left(W_{BQU}^1 + W_{BQU}^2\right) - \left(W_{BQU}^3 + W_{BQU}^4\right),$$
$$\hat{V}_{BQU} = \left(\hat{V}_{BQU}^1 + \hat{V}_{BQU}^2\right) - \left(\hat{V}_{BQU}^3 + \hat{V}_{BQU}^4\right), \tag{34}$$

9

where $\hat{V}_{BQU}^k = r'W_{BQU}^k r$ and the elements of W_{BQU}^k are:

$$W_{BQU}^1(i,j) = \frac{\lambda}{\sqrt{\lambda\,(\lambda+4)}} \frac{1}{1 - \exp(-2\,(m+1)\,\alpha)} \exp\left(-\,|i-j|\,\alpha\right),$$

$$W_{BQU}^2(i,j) = \frac{\lambda}{\sqrt{\lambda\,(\lambda+4)}} \frac{\exp(-2\,(m+1)\,\alpha)}{1 - \exp(-2\,(m+1)\,\alpha)} \exp\left(|i-j|\,\alpha\right),$$

$$W_{BQU}^3(i,j) = \frac{\lambda}{\sqrt{\lambda\,(\lambda+4)}} \frac{1}{1 - \exp(-2\,(m+1)\,\alpha)} \exp(-(i+j)\alpha),$$

$$W_{BQU}^4(i,j) = \frac{\lambda}{\sqrt{\lambda\,(\lambda+4)}} \frac{\exp(-2\,(m+1)\,\alpha)}{1 - \exp(-2\,(m+1)\,\alpha)} \exp\left((i+j)\alpha\right). \tag{35}$$

It now follows that $W_{BQU}(i,j) = W_{BQU}(m+1-i, m+1-j)$, which ensures the invariance of the BQU estimator to data reversibility. This invariance property means that the BQU estimator has the same value no matter whether it is applied to the original sample $\{r_i\}$ or its reversed version $\{s_i\}$ with $s_i = r_{m+1-i}$. If r_i is covariance stationary, then invariance property is desirable because $cov(r_i, r_j) = cov(s_i, s_j)$ for all i and j.

The sum $W_{BQU}^1 + W_{BQU}^2$ is a Toeplitz matrix while the sum of $W_{BQU}^3 + W_{BQU}^4$ is a Hankel matrix. Note that a Toeplitz matrix is a matrix with constant values along negative-sloping diagonals while the Hankel matrix is a matrix with constant values along positive-sloping diagonals. A quadratic estimator with a Toeplitz matrix as the weight can be written as a kernel-based estimator. To see this, we use $\hat{V}_{BQU}^1$ as an example. Some algebraic manipulations show that

$$\hat{V}_{BQU}^1 = \sum_{s=-(m+1)}^{m-1} w(s)\hat{\gamma}_s \tag{36}$$

where

$$w(s) = \frac{\lambda}{\sqrt{\lambda\,(\lambda+4)}} \frac{\exp\left(-\alpha\,|s|\right)}{1 - \exp(-2\,(m+1)\,\alpha)}. \tag{37}$$

Therefore, $\hat{V}_{BQU}^1$ is a kernel-based estimator with an exponential type kernel, i.e. $k_\rho(x) = 1/(2\rho)\exp(-\rho\,|x|)$ for $\rho > 0$. The exponential kernel $k_\rho(x)$ is positive definite because $k_\rho(r-s)$ is the covariance kernel of the OU process $\{Z_t\}$ defined by $dZ_t = -\rho Z_t + dB_t$ where B_t is the standard Brownian motion. In other words, $k_\rho(r-s) = EZ_r Z_s$. Similarly, we can show that $\hat{V}_{BQU}^2$ is a kernel-based estimator with an exponential type kernel.

Exponential kernels of this type have not been used before in long run variance estimation and appear in spectral density estimation only in the Abel estimate (c.f. Hannan, 1970, p. 279). The long run variance estimator that is closest to $\hat{V}_{BQU}^1$ and $\hat{V}_{BQU}^2$ is the kernel-based estimator proposed by Phillips, Sun and Jin (PSJ, 2005a, 2005b). PSJ exponentiate

the conventional kernels and use the resulting kernels in the long run variance estimation without truncation. As the PSJ estimator, the kernel weight $w(s)$ used in $\hat{V}^1_{BQU}$ decays to zero at an exponential rate and there is no truncation lag. Note that $\alpha \to 0$ as $m \to \infty$. It is easy to show that $w(s)$ becomes more concentrated around zero as α decreases. In effect, the action of α shrinking to zero plays a role similar to that of a bandwidth parameter in that very high order autocovariances are progressively downweighted as $m \to \infty$.

Note that $W^3_{BQU} + W^4_{BQU}$ is a special Hankel matrix as its (i,j)-th entry can be expressed as $a_i a_j$ for some a_i. Given this, $\hat{V}^3_{BQU}$ and $\hat{V}^4_{BQU}$ can be written as

$$\hat{V}^3_{BQU} = \frac{\lambda}{\sqrt{\lambda(\lambda+4)}} \frac{1}{1 - \exp(-2(m+1)\alpha)} \left(\sum_{i=1}^{m} r_i \exp(-\alpha i) \right)^2,$$

$$\hat{V}^4_{BQU} = \frac{\lambda}{\sqrt{\lambda(\lambda+4)}} \frac{1}{1 - \exp(-2(m+1)\alpha)} \left(\sum_{i=1}^{m} r_i \exp(-(m+1-i)\alpha) \right)^2. \tag{38}$$

So both $\hat{V}^3_{BQU}$ and $\hat{V}^4_{BQU}$ are squares of a weighted sum of the observed returns. In $\hat{V}^3_{BQU}$ more weights are attached to the first few observations while in $\hat{V}^4_{BQU}$ more weights are attached to the last few observations. As a result, the sum $\hat{V}^3_{BQU} + \hat{V}^4_{BQU}$ effectively ignores the middle part of the observations and captures mainly the edge effect.

We proceed to relate the BQU estimator to the maximum likelihood estimator when the noise is normal. Under the normality assumption, the log-likelihood function (ignoring the constant term) is

$$\log L = -\frac{1}{2} \log |\Omega| - \left(\frac{1}{2} r' \Omega^{-1} r \right). \tag{39}$$

It can be rewritten as

$$\log L = -\frac{m}{2} \log \sigma_\eta^2 - \frac{1}{2} \log |-A(\lambda)| + \left(\frac{1}{2\sigma_\eta^2} r' \left(A(\lambda) \right)^{-1} r \right), \tag{40}$$

or

$$\log L = -\frac{m}{2} \log mV\lambda - \frac{1}{2} \log |-A(\lambda)| + \left(\frac{1}{2mV\lambda} r' \left(A(\lambda) \right)^{-1} r \right). \tag{41}$$

As a result, the MLE's of σ_η^2 and V satisfy:

$$\hat{\sigma}^2_{\eta,MLE} = -m^{-1} r' \left(A(\hat{\lambda}_{MLE}) \right)^{-1} r \tag{42}$$

$$\hat{V}_{MLE} = -\hat{\lambda}_{MLE} r' \left(A(\hat{\lambda}_{MLE}) \right)^{-1} r \tag{43}$$

where $\hat{\lambda}_{MLE} = \hat{V}_{MLE}/(m\hat{\sigma}^2_{\eta,MLE})$. The above equations are highly nonlinear and difficult to solve explicitly. However, they could be solved iteratively by using the following steps:

11

(i) Choose a starting value $\lambda^{(0)}$ for λ and plug it into (42) and (43) to get $V^{(1)}$ and $\left(\sigma_\eta^2\right)^{(1)}$.
(ii) Compute $\lambda^{(1)} = V^{(1)}/(m\left(\sigma_\eta^2\right)^{(1)})$ and plug $\lambda^{(1)}$ into (42) and (43) to get updated values $V^{(2)}$ and $\left(\sigma_\eta^2\right)^{(2)}$. ($iii$) Repeat (ii) until the sequence $(V^{(k)}, \left(\sigma_\eta^2\right)^{(k)})$ converges. It is now obvious that the BQU estimator is the first iterative step in solving the MLE problem when the true value of λ is used as the starting value. The feasible BQU estimator given at the end of this section is the first iterative step when a consistent estimate of λ is used.

3.2 Multi-Window Representation

Recall that $W_{BQU} = m^{-1}V\Omega^{-1}$. Since Ω^{-1} is a positive definite symmetric matrix, it has a spectral decomposition:

$$\Omega^{-1} = \sum_{k=1}^{m} \delta_k^{-1} h^{(k)}\left(h^{(k)}\right)' \tag{44}$$

where δ_k is the eigenvalue of Ω and $h^{(k)}$ is the corresponding eigenvector. It now follows that the BQU estimator $\hat{V}_{BQU}$ can be written as

$$\hat{V}_{BQU} = m^{-1}V\sum_{k=1}^{m} \delta_k^{-1} r' h^{(k)}\left(h^{(k)}\right)' r. \tag{45}$$

With analytical expressions for the eigenvalues and eigenvectors, we can obtain an alternative representation of the BQU estimator in the next theorem.

Theorem 3 *Assume that $\eta's$ are mean zero random variables with $\kappa_4 = 3$ and $\sigma_i^2 = V/m$ for $i = 1, 2, ..., m$. The BQU estimator $\hat{V}_{BQU}$ can be represented as*

$$\hat{V}_{BQU} = \sum_{k=1}^{m} w_{BQU,k}\left(\sum_{\ell=1}^{m} r_\ell h_\ell^{(k)}\right)^2 \tag{46}$$

where

$$w_{BQU,k} = \frac{\lambda}{\left(\lambda + 2 - 2\cos\frac{k\pi}{(m+1)}\right)} \tag{47}$$

and

$$h^{(k)} = (h_1^{(k)}, ..., h_m^{(k)})' = \left(\sqrt{\frac{2}{m+1}}\sin\frac{\pi k}{m+1}, ..., \sqrt{\frac{2}{m+1}}\sin\ell\frac{\pi k}{m+1}, ...\right)' \tag{48}$$

is an eigenvector of Ω with the corresponding eigenvalue

$$\delta_k = \sigma_\eta^2\left(\lambda + 2 - 2\cos\frac{k\pi}{(m+1)}\right). \tag{49}$$

In addition, $\left\{h^{(k)}\right\}_{k=1}^{m}$ forms a complete orthonormal system in $\mathbb{R}^m$.

Theorem 3 shows that the BQU estimator has a multi-taper or multi-window representation, a term we now clarify. For a given stationary and mean zero time series $x_1, x_2, ..., x_T$, a multi-window estimator of its spectral density at frequency θ_0 is defined to be

$$\hat{S}(\theta_0) = \sum_{k=1}^{K} \omega_k \left| \sum_{t=1}^{T} x_t v_t^{(k)} \exp(-\theta_0 \sqrt{-1}) \right|^2 \tag{50}$$

where ω_k is a constant, $v^{(k)} = (v_1^{(k)}, v_2^{(k)}, ..., v_T^{(k)})$ is a sequence of constants called a data window (or taper), and K is the number of data windows used. In the multi-window spectral analysis of Thomson (1982), K is normally chosen to be much less than the sample size. According to this definition, the BQU estimator $\hat{V}_{BQU}$ is a multi-window estimator at frequency $\theta_0 = 0$. The underlying data window is $h^{(k)}$ and the number of data windows is m. Since the window weight $w_{BQU,k}$ depends on the data generating process, $\hat{V}_{BQU}$ belongs to the class of multi-window estimators with adaptive weighting.

The geometrical interpretation of the multi-window representation is that the return series r is projected onto the subspace spanned by the windows $h^{(k)}$, $k = 1, 2, ..., m$. Since $\{h^{(k)}\}_{k=1}^{m}$ is a complete orthonormal system in $\mathbb{R}^m$, we may write

$$r = \sum_{k=1}^{m} \alpha^{(k)} h^{(k)}, \text{ where } \alpha^{(k)} = \left(r, h^{(k)} \right) \tag{51}$$

and $(\cdot, \cdot)$ is the usual inner product in $\mathbb{R}^m$. By Parseval's identity, we have $\sum_{i=1}^{m} r_i^2 = \sum_{k=1}^{m} \left(\alpha^{(k)} \right)^2$. Note that $\sum_{i=1}^{m} r_i^2$ is the classical quadratic variation estimator of the realized volatility. So the quadratic variation estimator is the simple unweighted summation of $\left(\alpha^{(k)} \right)^2$. In contrast, the BQU estimator is a weighted sum of $\left(\alpha^{(k)} \right)^2$, i.e.

$$\hat{V}_{BQU} = \sum_{k=1}^{m} w_{BQU,k} \left(\alpha^{(k)} \right)^2. \tag{52}$$

Some simple algebraic manipulations show that

$$E \left(\alpha^{(k)} \right)^2 = tr(h^{(k)\prime} \Omega h^{(k)}) = \delta_k tr \left(\left(h^{(k)} \right)' h^{(k)} \right) = \delta_k \tag{53}$$

and when $\kappa_4 = 3$,

$$var \left(\left(\alpha^{(k)} \right)^2 \right) = 2tr \left(h^{(k)} h^{(k)\prime} \Omega h^{(k)} h^{(k)\prime} \Omega \right) = 2 \left(tr(h^{(k)\prime} \Omega h^{(k)}) \right)^2 = 2\delta_k^2 \tag{54}$$

13

and

$$cov\left(\left(\alpha^{(k_1)}\right)^2,\ \left(\left(\alpha^{(k_2)}\right)^2\right)^2\right) \tag{55}$$

$$= Er'h^{(k_1)}h^{(k_1)'}rr'h^{(k_2)}h^{(k_2)'}r - \left(Er'h^{(k_1)}h^{(k_1)'}r\right)\left(Er'h^{(k_1)}h^{(k_1)'}r\right)$$

$$= 2tr(h^{(k_1)}h^{(k_1)'}\Omega h^{(k_2)}h^{(k_2)'}\Omega) + tr(h^{(k_1)}h^{(k_1)'}\Omega)tr(h^{(k_2)}h^{(k_2)'}\Omega) - \delta_{k_1}\delta_{k_2}$$

$$= 0,$$

where we have used: for any two $m \times m$ matrices W_1 and W_2,

$$E\left(r'W_1r\right)\left(r'W_2r\right) = 2tr(W_1\Omega W_2\Omega) + tr(W_1\Omega)tr(W_2\Omega). \tag{56}$$

Therefore, $\{\left(\alpha^{(k)}\right)^2\}$ are uncorrelated and each of them has mean δ_k and variance $2\delta_k^2$. So, for each k, $mw_{BQU,k}\left(\alpha^{(k)}\right)^2$ is a unbiased estimator of V but it is inconsistent because its variance does not die out as $m \to \infty$. By averaging over the m uncorrelated terms, the BQU estimator becomes consistent.

A great virtue of the multi-window formulation is that $\hat{V}_{BQU}$ may be computed in real time using fast Fourier transforms. It follows from Theorem 3 that

$$\hat{V}_{BQU} = \frac{1}{2(m+1)}\sum_{k=1}^{m} w_{BQU,k}\left(\sum_{\ell=1}^{2m+2}\tilde{r}_{\ell}^{(k)}\sin\ell\frac{2\pi k}{2(m+1)}\right)^2 \tag{57}$$

where

$$\tilde{r}^{(k)} = \begin{cases} (r_1, r_2, ..., r_m, 0, -r_1, -r_2, ..., -r_m, 0) & k \text{ is odd}, \\ (r_1, r_2, ..., r_m, 0,\ \ r_1,\ \ r_2, ...,\ \ r_m, 0) & k \text{ is even}. \end{cases} \tag{58}$$

Note that $\sum_{\ell=1}^{2m+2}\tilde{r}_{\ell}^{(k)}\sin\ell 2\pi k/(2(m+1))$ is the imaginary part of the discrete Fourier transform of the augmented sequence $\tilde{r}^{(k)}$. To compute $\hat{V}_{BQU}$, we can first obtain the discrete Fourier transforms of the two series given in (58) and then take the weighted sum of these discrete Fourier transforms.

The BQU estimator $\hat{V}_{BQU}$ depends on the unknown quantity λ, the signal-to-noise ratio. In practice, we can employ a consistent estimator $\tilde{\lambda}$ of λ to implement the BQU estimator. We denote the resulting feasible BQU estimator as $\hat{V}_{BQU}(\tilde{\lambda})$. Let $\tilde{V}$ and $\tilde{\sigma}_\eta^2$ be consistent estimators of V and σ_η^2, respectively, we can take $\tilde{\lambda}$ to be $\tilde{\lambda} = \tilde{V}/(m\tilde{\sigma}_\eta^2)$. When $\tilde{V} - V = O_p\left(m^{-1/4}\right)$ and $\tilde{\sigma}_\eta^2 - \sigma_\eta^2 = O_p\left(m^{-1/2}\right)$, we have $\tilde{\lambda} - \lambda = O\left(m^{-5/4}\right)$. Using the multi-window representation, we can establish the asymptotic normality of $\hat{V}_{BQU}\left(\lambda\right)$ and $\hat{V}_{BQU}(\tilde{\lambda})$ in the next theorem.

14

Theorem 4 *Assume that $\eta's$ are mean zero random variables with $\kappa_4 = 3$ and $\sigma_i^2 = V/m$ for $i = 1, 2, ..., m$. If $\tilde{\lambda} - \lambda = O\left(m^{-5/4}\right)$, then as $m \to \infty$*

    (i) $\sqrt{m}(\hat{V}_{BQU}(\lambda) - V) \to_d N(0, 2V^2)$;

    (ii) $\sqrt[4]{m}(\hat{V}_{BQU}(\tilde{\lambda}) - V) = \sqrt[4]{m}\left(\tilde{V} - V\right) + o_p(1)$.

Theorem 4(i) shows that the infeasible BQU estimator converges to V at the rate of $1/\sqrt{m}$. This rate is faster than the best nonparametric rate $1/\sqrt[4]{m}$. Unfortunately and not surprisingly, this rate can not be achieved in practice. Theorem 4(ii) shows that the feasible estimator $\hat{V}_{BQU}(\tilde{\lambda})$ converges to V at the rate of only $1/\sqrt[4]{m}$. In addition, $\sqrt[4]{m}(\hat{V}_{BQU}(\tilde{\lambda}) - V)$ and $\sqrt[4]{m}(\tilde{V}(\lambda) - V)$ are asymptotically equivalent in large samples. This suggests that there is no asymptotic gain in using the feasible BQU estimator $\hat{V}_{BQU}(\tilde{\lambda})$. However, in finite samples, the feasible BQU estimator $\hat{V}_{BQU}(\tilde{\lambda})$ may have a smaller variance than the initial estimator $\tilde{V}$. Simulation results not reported here show that this is the case when $\tilde{V}$ is the ZMA estimator.

4   The BQU Estimator Under Alternative Unbiasedness Conditions

The unbiasedness condition $tr(W\Omega) = V$ in the previous section relies crucially on prior information on σ_η^2 and V. When we employ some preliminary estimates $\tilde{\sigma}_\eta^2$ and $\tilde{V}$ to implement the BQU estimator $\hat{V}_{BQU}$, the unbiasedness condition becomes $tr(W\tilde{\Omega}) = \tilde{V}$. Due to the estimation uncertainty, the resulting BQU estimator is not unbiased any more. In this section, we take advantage of the structure of Ω and impose alternative unbiasedness conditions. These new conditions ensure the unbiasedness of the BQU estimator regardless of the values of σ_η^2 and V.

When $\sigma_i^2 = V/m$ for $i = 1, 2, ..., m$, the finite sample mean of the quadratic estimator $\hat{V}_Q$ can be written as

$$V\frac{1}{m}\sum_{i=1}^{m} W(i,i) - \left(\sum_{i=2}^{m} W(i,i-1) + \sum_{i=1}^{m-1} W(i,i+1) - 2\sum_{i=1}^{m} W(i,i)\right)\sigma_\eta^2$$

$$= 1/m tr(W) \cdot V - tr(W\Gamma) \cdot \sigma_\eta^2 \tag{59}$$

where $\Gamma = A(0)$ is a constant matrix and $A(\lambda)$ is defined in (31). To ensure unbiasedness without using any information on V and σ_η^2, we can let

$$tr(W) = m \text{ and } tr(W\Gamma) = 0. \tag{60}$$

We now minimize $tr(W\Omega W\Omega)$ over W subject to the positive definiteness of W and the above two conditions. Using the Theorem of Rao (1973) given in the appendix, we find the solution is

$$W^*_{BQU} = \Omega^{-1}\sigma^2_\eta \left(c_1 I + c_2 \Gamma\right) \sigma^2_\eta \Omega^{-1}, \tag{61}$$

where c_1 and c_2 satisfy

$$c_1 \sigma^4_\eta tr(\Omega^{-1}\Omega^{-1}) + c_2 \sigma^4_\eta tr(\Omega^{-1}\Gamma\Omega^{-1}) = m, \tag{62}$$

$$c_1 \sigma^4_\eta tr\left(\Omega^{-1}\Omega^{-1}\Gamma\right) + c_2 \sigma^4_\eta tr\left(\Omega^{-1}\Gamma\Omega^{-1}\Gamma\right) = 0.$$

We call the quadratic estimator with weight W^*_{BQU} the BQU* estimator and denote it as $\hat{V}^*_{BQU}$. The next theorem gives a representation of this estimator.

Theorem 5 *Assume that $\eta's$ are mean zero random variables with $\kappa_4 = 3$ and $\sigma^2_i = V/m$ for $i = 1, 2, ..., m$. The BQU* estimator can be represented as*

$$\hat{V}^*_{BQU} = \sum_{k=1}^{m} w^*_{BQU,k} \left(\sum_{\ell=1}^{m} r_\ell h^{(k)}_\ell\right)^2 \tag{63}$$

where

$$w^*_{BQU,k} = \left(\lambda + 2 - 2\cos\frac{k\pi}{(m+1)}\right)^{-2}\left[c_1 - c_2\left(2 - 2\cos\frac{k\pi}{(m+1)}\right)\right] \tag{64}$$

$$c_1 = \frac{m\beta_{2,2}}{\beta_{2,0}\beta_{2,2} - \beta^2_{2,1}}, c_2 = \frac{m\beta_{2,1}}{\left(\beta_{2,0}\beta_{2,2} - \beta^2_{2,1}\right)} \tag{65}$$

and

$$\beta_{2,j} = \sum_{k=1}^{m} \left(\lambda + 2 - 2\cos\frac{k\pi}{(m+1)}\right)^{-2}\left(2 - 2\cos\frac{k\pi}{(m+1)}\right)^j, \ j = 0, 1, 2. \tag{66}$$

Theorem 5 shows that the BQU* estimator $\hat{V}^*_{BQU}$ is also a multi-window quadratic estimator. The difference between $\hat{V}^*_{BQU}$ and $\hat{V}_{BQU}$ is that they impose different weights on the data windows. All the qualitative results for $\hat{V}_{BQU}$ in Section 3.2 remain valid for $\hat{V}^*_{BQU}$. In particular, $\hat{V}^*_{BQU}$ can be computed using fast Fourier transforms.

As the BQU estimator $\hat{V}_{BQU}$, the BQU* estimator $\hat{V}^*_{BQU}$ also depends on the unknown quantity λ. Plugging in a consistent estimator $\tilde{\lambda}$ of λ, we can obtain the feasible version of $\hat{V}^*_{BQU}$. We denote the feasible estimator as $\hat{V}^*_{BQU}(\tilde{\lambda})$. The next theorem establishes the asymptotic normality of both $\hat{V}^*_{BQU}(\lambda)$ and $\hat{V}^*_{BQU}(\tilde{\lambda})$.

Theorem 6 *Assume that $\eta's$ are mean zero random variables with $\kappa_4 = 3$ and $\sigma_i^2 = V/m$ for $i = 1, 2, ..., m$. If $\tilde{\lambda} - \lambda = O\left(m^{-5/4}\right)$, then, as $m \to \infty$*

(i) $m^{1/4}\left(\hat{V}_{BQU}^*(\lambda) - V\right) \to_d N(0, 8V^2\sqrt{V/\sigma_\eta^2})$;

(ii) $m^{1/4}\left(\hat{V}_{BQU}^*(\tilde{\lambda}) - V\right) \to_d N(0, 8V^2\sqrt{V/\sigma_\eta^2})$.

It follows from Theorem 6 that the feasible estimator $\hat{V}_{BQU}^*(\tilde{\lambda})$ converges to the true realized variance at the rate of $m^{-1/4}$, the best attainable rate in the present context. Furthermore, the theorem shows that the feasible and infeasible estimators are asymptotically equivalent. The estimation uncertainty in the pilot estimator $\tilde{\lambda}$ does not affect the asymptotic distribution of $\hat{V}_{BQU}^*(\tilde{\lambda})$. This result is analogous to that for a two-step estimator where the estimation uncertainty in the first step does not factor into the asymptotic variance of the second step estimator.

Under the normality assumption, it can be shown that the MLE satisfies

$$m^{1/4}\left(\hat{V}_{MLE} - V\right) \to_d N(0, 8V^2\sqrt{V/\sigma_\eta^2}). \tag{67}$$

See, for example, BNHLS (2006). Comparing this with Theorem 6(ii), we find that the feasible BQU estimator $\hat{V}_{BQU}^*(\tilde{\lambda})$ has the same asymptotic variance as the MLE. Therefore, the estimator $\hat{V}_{BQU}^*(\tilde{\lambda})$ is asymptotically as efficient as the MLE under the normality assumption. The advantage of $\hat{V}_{BQU}^*(\tilde{\lambda})$ is that it is computationally simpler and does not rely on the normality assumption.

5   Finite Sample Performance

In this section, we first compare the finite sample performances of the BQU and BQU* estimators with those of the HL, ZMA, BNHLS and MS estimators when the model is correctly specified and model parameters are assumed to be known. This comparison is used as a benchmark. We then compare the finite sample performances of these estimators when the model is possibly misspecified and model parameters have to be estimated. The second comparison allows us to evaluate the performances of different estimators in realistic situations.

5.1   Theoretical Comparison

For the theoretical comparison, we assume

$$p_t = \sqrt{V}B_t + \sigma_\eta\eta_t \tag{68}$$

where B_t is a standard Brownian motion and η_t is iid N(0,1) and independent of B_t. We calibrate the parameters V and σ_η^2 and the sample size based on the following three stocks: Goldman Sachs (GS), SBC communications (SBC) and EXXON Mobile Corporation (XOM). These three stocks are also considered in BR(2005) and are thought to be representative of the median and extreme features of the S&P100 stocks. Using the TAQ data set, BR (2004b) obtain parameter values for σ_η^2 and V and sample size m. They are reproduced in Table 1 and assumed to be the true parameter values in this subsection.

Table 1: Parameter Values for GS, SBC and XOM

	GS	SBC	XOM
σ_η^2	0.87×10^{-7}	1.89×10^{-7}	2.10×10^{-7}
V	0.00042	0.00041	0.00018
m	2247	2034	2630

We compare six estimators: HL, ZMA, BNHLS, MS, BQU and BQU* estimators. For the HL and ZMA estimators, we use Theorem 1 and choose q to minimize

$$q^* = \arg\min_q \left(tr(W\Omega) - V\right)^2 + 2tr\left(W\Omega W\Omega\right) \tag{69}$$

where W is given in (15) and (16), respectively. The minimum value of the objective function gives us the minimum MSE. For the BNHLS and MS estimators, we choose q according to equations (11) and (13), respectively. By construction, the biases of these two estimators are zero and their minimum MSE's are $2tr\left(W\Omega W\Omega\right)$. To obtain the BQU and BQU* estimators, we plug the values of λ and m into Theorems 3 and 5, respectively. The minimum MSE of $\hat{V}_{BQU}$ is $2V^2/m$ while that of $\hat{V}_{BQU}^*$ is $2mc_1\sigma_\eta^4$.

Table 2 reports the biases, standard deviations and root mean squared errors (RMSE) of the six different estimators. The table shows that the BQU estimator $\hat{V}_{BQU}$ has the smallest RMSE among the six estimators. The RMSE of $\hat{V}_{BQU}$ is less than 1/3 of those of the HL and ZMA estimators for all three stocks. In terms of RMSE, the second best estimator is the BQU* estimator $\hat{V}_{BQU}^*$, whose RMSE is smaller than that of the BNHLS estimator by 15.75%, 11.90% and 5.55%, respectively for GS, SBC and XOM. The BNHLS estimator outperforms the MS estimator by a small margin. Finally, the table shows that the BNHLS and MS estimators dominate the HL and ZMA estimators.

The RMSE's in Table 2 are close to those in BR (2005). The data-driven optimal truncation lags are also close to those given there. While the optimal q's for both the HL and ZMA estimators in BR(2005) are 13, 14 and 15 respectively, the optimal q's we obtain

for these two estimators are 15, 15 and 16, respectively. The small difference in q does not matter very much as the RMSE as a function of q is very flat in the neighborhood of $q = 14$. These observations suggest that the difference in the variance formulae is of little practical importance.

Table 2: Finite Sample Performances of Different Estimators ($\times 10^{-4}$)

	GS			SBC			XOM		
	Bias	STD	RMSE	Bias	STD	RMSE	Bias	Std	RMSE
HL	-0.2817	0.3962	0.4862	-0.2752	0.4093	0.4932	-0.1131	0.1679	0.2025
ZMA	-0.3044	0.3950	0.4987	-0.2997	0.4077	0.5060	-0.1221	0.1672	0.2071
BNHLS	0.0000	0.2946	0.2946	0.0000	0.3306	0.3306	0.0000	0.1483	0.1483
MS	0.0000	0.2996	0.2996	0.0000	0.3380	0.3380	0.0000	0.1497	0.1497
BQU	0.0000	0.1253	0.1253	0.0000	0.1286	0.1286	0.0000	0.0496	0.0496
BQU*	0.0000	0.2624	0.2624	0.0000	0.2978	0.2978	0.0000	0.1430	0.1430

It is important to point out that the RMSE comparison is not feasible as we do not know the true values of σ_η^2 and V. In practice, we have to obtain preliminary estimates of σ_η^2 and V before implementing each of the six estimators. The use of preliminary estimates introduces additional randomness into each estimator in finite samples.

To deepen our understanding of the quadratic estimators, we plot the weighting matrix $W(i, j)$ for each estimator against (i, j). We choose SBC as an example since it represents the median of the S&P 100 stocks in terms of the ratio σ_η^2/V. The optimal q for both the HL and ZMA estimators is $q^* = 15$ while the optimal q's for the BNHLS and MS estimators are $q^* = 5$ and 3, respectively. Since the surfaces are symmetric front to back, Figures 1(a)–3(b) plot the front of each surface. Four observations can be made from these figures. First, most of the diagonal elements of W are close to one for the estimators $\hat{V}_{HL}, \hat{V}_{ZMA}, \hat{V}_{BNHLS}, \hat{V}_{MS}$ and $\hat{V}_{BQU}^*$. This observation remains valid for other parameter configurations. In contrast, for the BQU estimator $\hat{V}_{BQU}$, the diagonal elements of W_{BQU} are less than 0.5. Figures not reported here reveal that these diagonal elements are very sensitive to the parameter configuration. This is not surprising as the BQU estimator $\hat{V}_{BQU}$ relies on precise information on λ to ensure its unbiasedness and consistency. Second, most of the volume of the surface lies in the ridge whose center is the straight line from $(1, 1)$ to (m, m). The surface of the BQU* estimator is more concentrated than the HL and ZMA estimators but less concentrated than the BNHLS and MS estimators. Define the effective truncation lag as q such that $W(i, j)$ is essentially zero for $|i - j| > q$, then the effective truncation lag for the five estimators $\hat{V}_{MS}, \hat{V}_{BNHLS}, \hat{V}_{BQU}^*, \hat{V}_{HL}$ and $\hat{V}_{ZMA}$ are 3, 5, 8,

15 and 15, respectively. Third, these figures illustrate that the kernel weight in the HL, ZMA, BNHLS and MS estimators decays polynomially and becomes zero after certain lag while the weight in the BQU* estimator decays exponentially and only approaches zero. The difference in the decaying rate and effective truncation lag implies that the HL, ZMA, BNHLS and MS estimators may be suboptimal. Finally, compared with the HL estimator which entails no edge effect, the ZMA estimator entails a large edge effect with the weights assigned to the first few cross products ($r_i r_j$ for $i, j \leq 15$) being substantially smaller. In contrast, the edge effect entailed by the BNHLS, MS, BQU and BQU* estimators is relatively small as it applies to fewer observations at the beginning and towards the end of the time series.

5.2 Simulation Evidence

To compare the finite performance of different estimators in more realistic situations, we consider the following stochastic volatility model:

$$dp_t^e = \mu dt + \sigma_t dB_{1t}, \ \sigma_t = \sqrt{V} \exp(\beta_0 + \beta_1 \tau_t),$$

$$d\tau_t = \alpha \tau_t dt + dB_{2t}, \ \text{cov}(dB_{1t}, dB_{2t}) = \rho,$$

$$p_t = p_t^e + \sigma_\eta \eta_t$$

where B_{1t} and B_{2t} are correlated standard Brownian motions. This model is also simulated in Goncalves and Meddahi (2004), Huang and Tauchen (2005), and BNHLS (2006).

In our simulation experiment, we consider two sets of parameter values for $\mu, \alpha, \beta_0, \beta$ and ρ :

$$\mu = 0, \beta_1 = 0, \alpha = -0.025, \ \beta_0 = 0, \ \rho = -0.3.$$

and

$$\mu = 0.03, \beta_1 = 0.125, \alpha = -0.025, \ \beta_0 = \beta_1^2/2\alpha, \ \rho = -0.3.$$

Under the first set of parameter values, the model reduces to a model with constant volatility. Under the second set of parameter values, which is taken from BNHLS (2006), the model is a genuine stochastic volatility model. The constant volatility model satisfies the assumptions of the two BQU estimators while the stochastic volatility model does not. The stochastic volatility model is used to check the robustness of the BQU estimators to the deviation from maintained assumptions.

As in the theoretical comparison, the values of V and σ_η^2 are chosen to match those in Table 1. It is easy to show that for the DGP considered, $E(\sigma_t^2) = V$. For each value of V

20

in Table 1, we also set the corresponding σ_η^2 to be 10 times those in Table 1 and obtain models with different levels of noise contaminations. The sampling frequency is chosen to match the m values in Table 1. More specifically, we normalize one second to be $1/23400$ and simulate over the unit interval $[0,1]$, which is thought to cover 6.5 hours. The efficient log-process p_t^e is generated using an Euler scheme based on 23400 intervals. Given the price process p_t, $t = 1, 2, ..., 23400$, we sample every ℓ points and obtain the sample observations $p_1, p_{1+\ell}, p_{1+2\ell}, ...$ where $l = [23400/m]$.

To check the robustness of the BQU and BQU* estimators to the noise distribution, we consider three distributions for η_t : $\eta_t \sim iidN(0,1)$, $\eta_t \sim iid\left(\chi_1^2 - 1\right)/\sqrt{2}$ where χ_1^2 is the χ^2 distribution with one degree of freedom and $\eta_t \sim iid\ t_5/\sqrt{5/3}$ where t_5 is the t distribution with five degrees of freedom. The latter two distributions are considered because they exhibit asymmetry and heavy tails, respectively.

To implement the HL and ZMA estimators, we first obtain pilot estimates of V and σ_η^2. The pilot estimate $\tilde{V}$ we use is the ZMA estimate with q set equal to 10. The pilot estimate $\tilde{\sigma}_\eta^2$ is obtained as follows:

$$\tilde{\sigma}_\eta^2 = \frac{1}{2m}\sum_{t=1}^m r_t^2 - \frac{\tilde{V}}{2m}. \tag{70}$$

The consistency of $\tilde{\sigma}_\eta^2$ for σ_η^2 follows immediately from the consistency $1/(2m)\sum r_t^2$ for σ_η^2 as $m \to \infty$. The second term in the above expression is a finite sample adjustment and vanishes as $m \to \infty$. Next, we plug the pilot estimates $\tilde{\sigma}_\eta^2$ and $\tilde{V}$ into (69) and solve the minimization problem to obtain $\tilde{q}^*$ for the HL and ZMA estimators. We compute the HL and ZMA estimators using their respective data-driven q's.

To implement the BNHLS and MS estimators, we use the formulae in equations (9)–(13). These formulae call for pilot estimates of V and σ_η^2. In the simulation study, we use the ZMA estimate $\hat{V}_{ZMA}$ as the pilot estimate of V and obtain the pilot estimate of σ_η^2 by plugging $\hat{V}_{ZMA}$ into (70). Finally, given the data-driven BNHLS estimate $\hat{V}_{BNHLS}$, we estimate the signal-to-noise ratio by

$$\hat{\lambda}_{BNHLS} = \frac{\hat{V}_{BNHLS}}{m \times 1/(2m)\left(\sum_{t=1}^m r_t^2 - \hat{V}_{BNHLS}\right)} = \frac{2\hat{V}_{BNHLS}}{\left(\sum_{t=1}^m r_t^2 - \hat{V}_{BNHLS}\right)}. \tag{71}$$

The feasible BQU and BQU* estimators are given by $\hat{V}_{BQU}(\hat{\lambda}_{BNHLS})$ and $\hat{V}_{BQU}^*(\hat{\lambda}_{BNHLS})$, respectively.

We use 10000 replications. Tables 3-4 report the finite sample bias, standard deviation and RMSE of each estimator. We report the cases when the noises follow normal and t

21

distributions. To save space, we omit the table for chi-squared noises as the performance ranking of different estimators is the same as that in Tables 3-4. The reported statistics are computed by binning the estimates according to the value of $\int_0^1 \sigma_s^2 ds$, calculating bias, standard deviation and RMSE within each bin, and then averaging across the bins, weighted by the number of elements in each bin.

We now discuss the simulation results. First, under the assumption of constant volatility, the BQU* estimator $\hat{V}_{BQU}^*(\hat{\lambda}_{BNHLS})$ has the smallest RMSE among all estimators and for all model parameters considered. Compared with the HL and ZMA estimators, the BQU* estimator achieves a RMSE reduction of 30% to 40% regardless of the level of noise contaminations and their distributions. Compared with the BNHLS, MS and BQU estimators, the BQU* estimator achieves a RMSE reduction of 5% to 10% under different model configurations. The superior RMSE performance arises because the BQU* estimator has a smaller absolute bias and variance than other estimators. Comparing with Table 1, we find that the RMSE of the BQU* estimator is close to that of the infeasible RMSE, reflecting the asymptotic result given in Theorem 6. Simulation results not reported here show that the BQU* estimator is not sensitive to the plug-in value of λ used. More specifically, let $\hat{\lambda}_{ZMA}$ be defined in the same way as $\hat{\lambda}_{BNHLS}$ but with $\hat{V}_{BNHLS}$ replaced by $\hat{V}_{ZMA}$ in equation (71), then the finite sample RMSEs of $\hat{V}_{BQU}^*(\hat{\lambda}_{ZMA})$ and $\hat{V}_{BQU}^*(\hat{\lambda}_{BNHLS})$ are almost indistinguishable.

Second, among the stochastic volatility models, the BQU* estimator outperforms other estimators in terms RMSE in 10 out of 12 cases reported in Tables 3-4. The two exceptions are stochastic volatility models with higher level of noise contaminations and with model parameters calibrated to XOM. In these two cases $\sigma_\eta = 1.4491 \times 10^{-3}$, so the standard deviation of the noise is 0.1% of the value of the asset price. This level of noise contaminations may be regarded as high indeed. Even in the two worst scenarios, the RMSE of the BQU* estimator is lower than those of the HL and ZMA estimators and is at most 15% higher than those of the BNHLS and MS estimators.

Third, for both constant volatility models and stochastic volatility models, the RMSE performances of the BNHLS, MS and BQU estimators are close to each other in almost all cases. Comparing the results for constant volatility models with Table 1, we find that the estimation uncertainty in $\hat{q}$ has a very small effect on the variances of the BNHLS and MS estimators. This result is consistent with the finding in BNHLS (2006). However, due to the estimation uncertainty in $\hat{\lambda}_{BNHLS}$, the variance of $\hat{V}_{BQU}(\hat{\lambda}_{BNHLS})$ is substantially inflated. This is consistent with Theorem 4.

6　Conclusions

In this paper we have investigated the best quadratic unbiased estimators of the integrated variance in the presence of market microstructure noise. It is shown that the feasible BQU estimators are asymptotically normal with convergence rate $m^{-1/4}$, the best attainable rate for nonparametric variance estimators. More importantly, we show that one of our feasible BQU estimators is asymptotically as efficient as the maximum likelihood estimator in the constant volatility plus normal noise model.

The present study can be extended in several ways, and we briefly discuss some possibilities as follows. First, our results are obtained under the assumption that the mean of the intra-daily return is zero. To obtain a BQU estimator that is invariant to the unknown but time invariant mean, we need to impose the conditions that each row of the weighting matrix sums up to zero. Using the Theorem of Rao (1973), we can easily find the optimal weighting matrix for this case. Second, throughout the paper, we have assumed that the market microstructure noise is independent across time. Our theoretical framework can be extended to allow for serial dependence but an analytical expression for the optimal weighting matrix and an explicit multi-window representation of the resulting BQU estimator may not be readily available. Alternatively, if the noise is assumed to be k-dependent, we can first sample every k data points, construct the BQU estimator based on each subsample, and then take an average of the subsampled BQU estimators. Finally, we have assumed constant volatility throughout the paper. Although the assumption may be realistic for some stocks when the sampling is conducted in calendar time and for all stocks when the sampling is conducted in business time, it is desirable to relax this assumption.
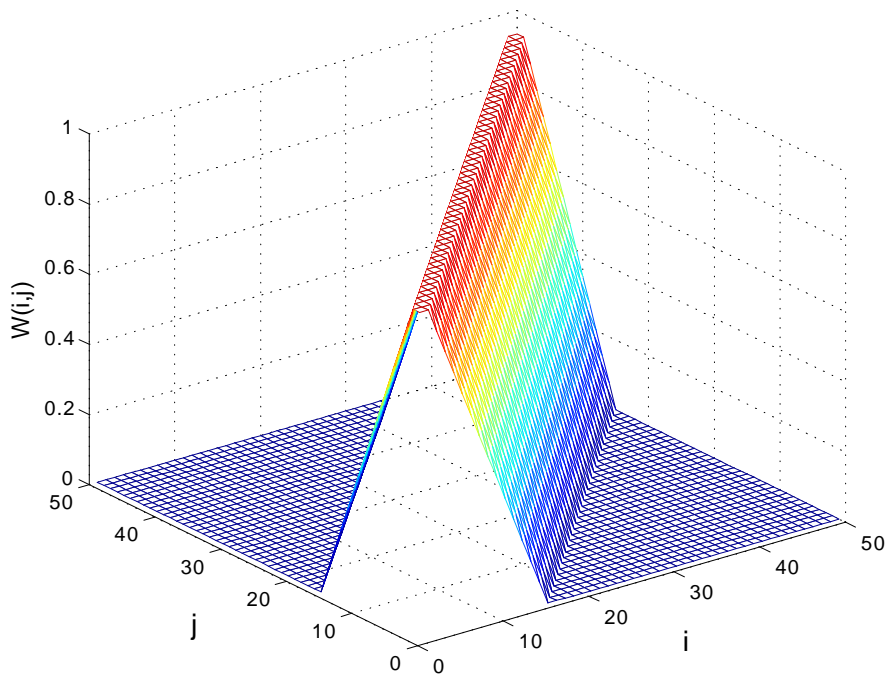
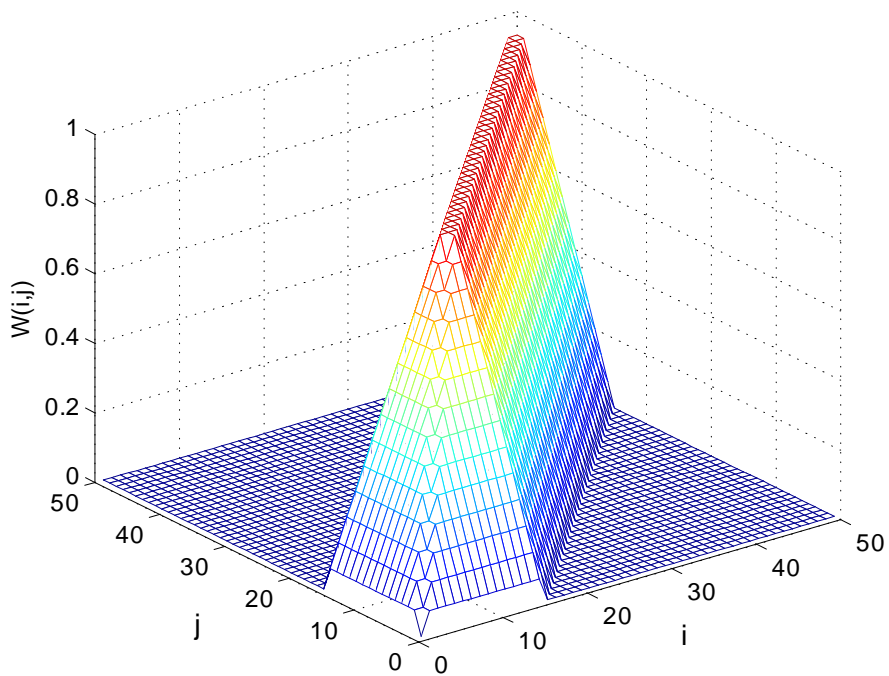Figure 1(a): Graph of $W_{HL}(i, j)$ against (i, j)

Figure 1(b): Graph of $W_{ZMA}(i, j)$ against (i, j)
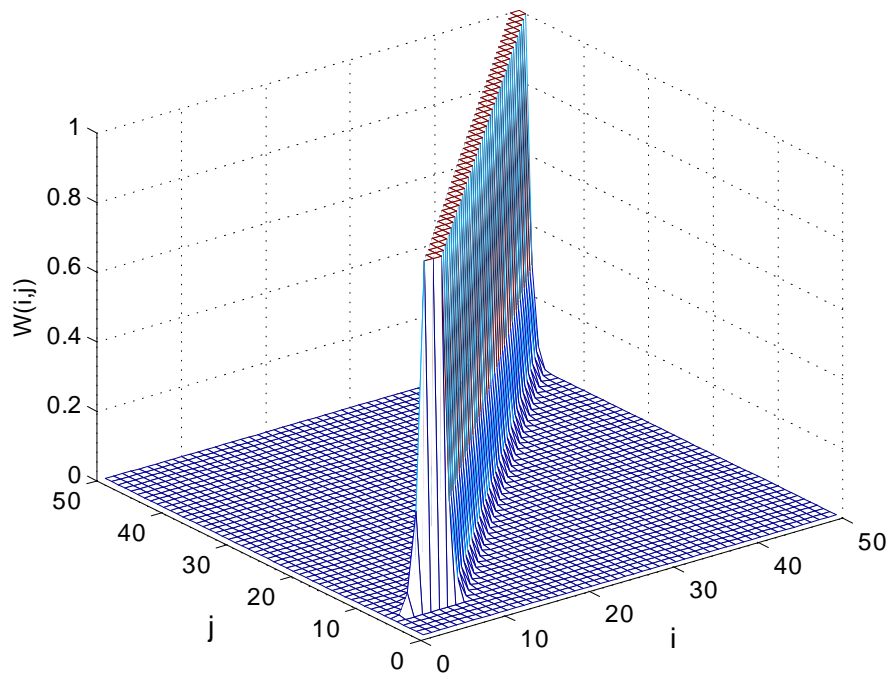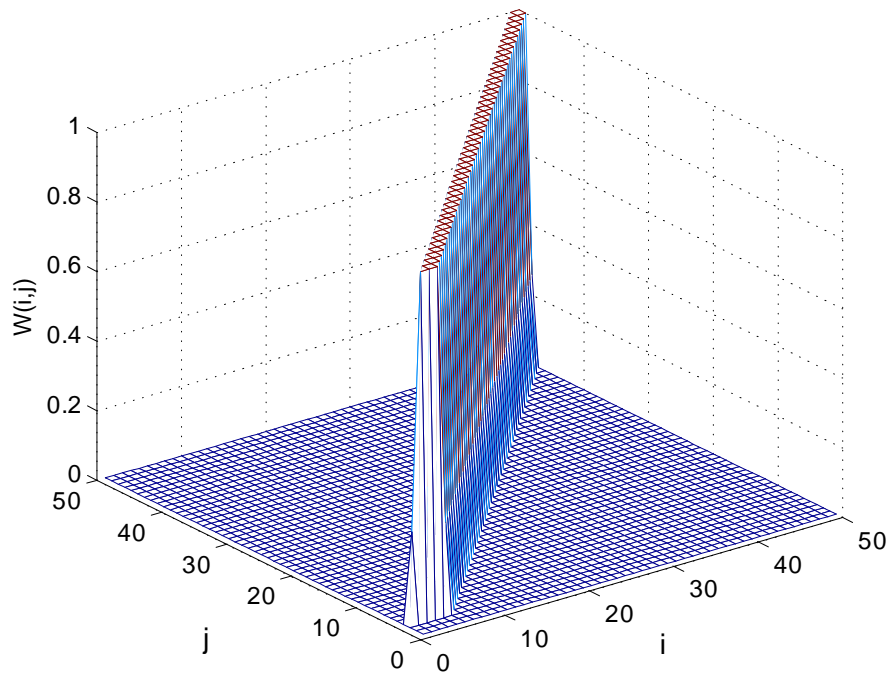
Figure 2(a): Graph of $W_{BNHLS}(i,j)$ against (i,j)
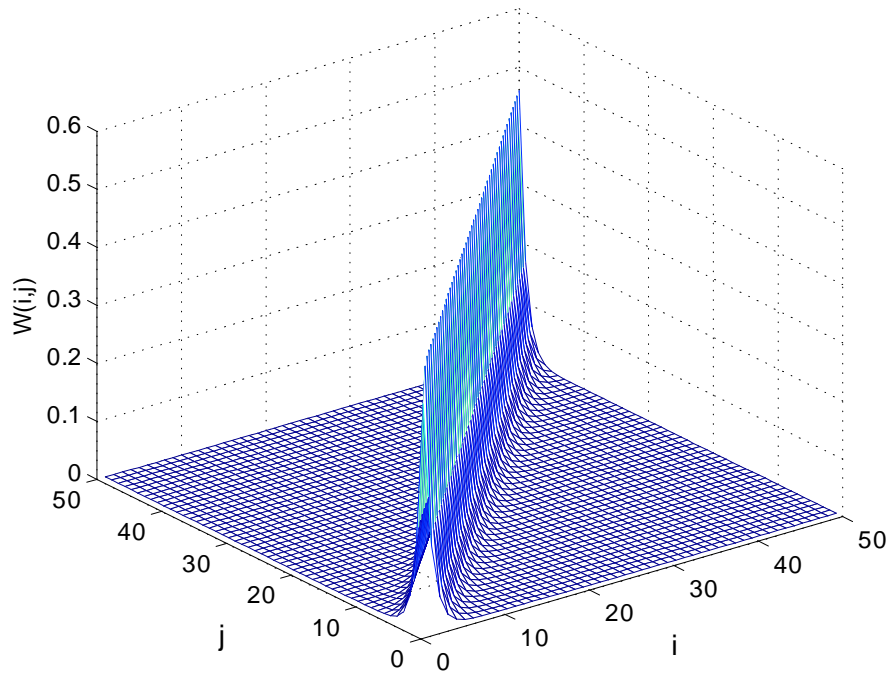

Figure 2(b): Graph of $W_{MS}(i,j)$ against (i,j)
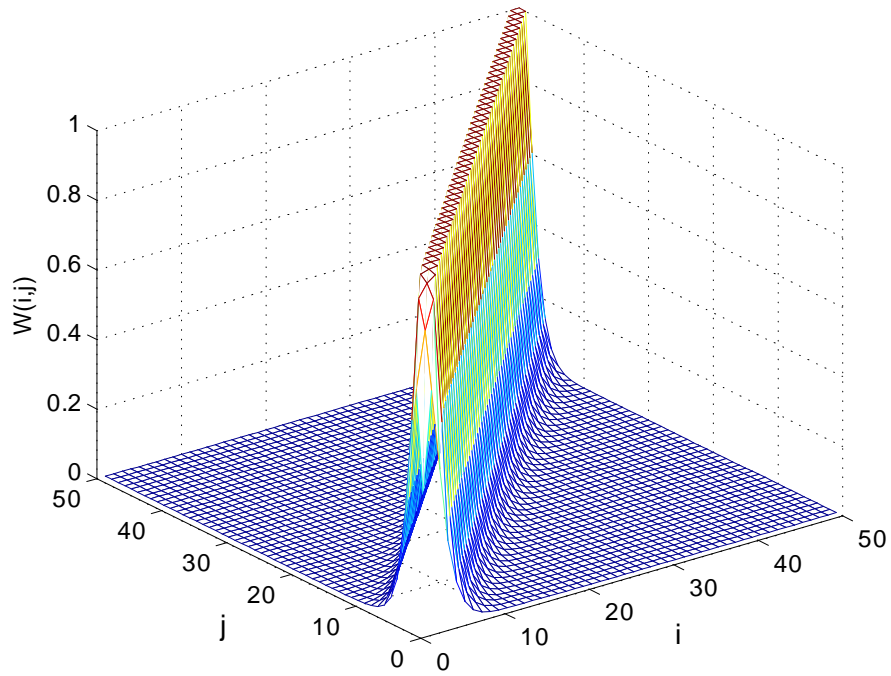
Figure 3(a): Graph of $W_{BQU}(i,j)$ against (i,j)

Figure 3(b): Graph of $W^*_{BQU}(i,j)$ against (i,j)

Table 3: Finite Sample Performances of Different Estimators with Normal Noise (10^{-4})

	GS			SBC			XOM		
	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
	Constant Volatility, Lower Contamination								
HL	-0.2514	0.4052	0.4768	-0.2626	0.4011	0.4794	-0.0876	0.1756	0.1962
ZMA	-0.2336	0.4115	0.4731	-0.2376	0.4124	0.4759	-0.0884	0.1715	0.1929
BNHLS	-0.0168	0.2824	0.2829	-0.0292	0.3145	0.3158	-0.0158	0.1458	0.1466
MS	-0.0112	0.2949	0.2950	-0.0214	0.3227	0.3234	-0.0110	0.1456	0.1460
BQU	-0.0138	0.2792	0.2795	-0.0240	0.3129	0.3138	-0.0132	0.1478	0.1484
BQU*	0.0027	0.2577	0.2577	0.0033	0.2888	0.2888	0.0009	0.1388	0.1388
	Constant Volatility, Higher Contamination								
HL	-0.2352	0.4379	0.4970	-0.3354	0.4861	0.5905	-0.0640	0.6032	0.6065
ZMA	-0.2225	0.4375	0.4908	-0.1935	0.4988	0.5350	-0.0450	0.2650	0.2688
BNHLS	-0.0596	0.4046	0.4089	-0.0874	0.4704	0.4784	-0.0472	0.2399	0.2445
MS	-0.0429	0.4094	0.4116	-0.0649	0.4775	0.4818	-0.0339	0.2445	0.2468
BQU	-0.0509	0.4108	0.4139	-0.0776	0.4777	0.4839	-0.0431	0.2433	0.2470
BQU*	-0.0037	0.3882	0.3882	-0.0022	0.4580	0.4580	0.0007	0.2346	0.2346
	Stochastic Volatility, Lower Contamination								
HL	-0.2030	1.0015	1.0226	-0.2110	1.0123	1.0351	-0.0300	0.4258	0.4265
ZMA	-0.1732	1.0295	1.0440	-0.1737	1.0347	1.0492	-0.0300	0.4210	0.4217
BNHLS	-0.0018	0.4961	0.4965	-0.0072	0.5404	0.5405	0.0116	0.2364	0.2367
MS	0.0025	0.4993	0.4995	0.0006	0.5420	0.5420	0.0159	0.2390	0.2396
BQU	-0.0336	0.6012	0.6040	-0.0038	0.5386	0.5387	0.0131	0.2357	0.2360
BQU*	0.0138	0.4784	0.4792	0.0256	0.5288	0.5294	0.0341	0.2276	0.2302
	Stochastic Volatility, Higher Contamination								
HL	-0.1652	1.0377	1.0508	-0.3043	1.3822	1.4176	-0.1217	1.1695	1.1757
ZMA	-0.1520	1.0463	1.0570	-0.1105	1.0785	1.0833	0.0456	0.5106	0.5128
BNHLS	-0.0103	0.6540	0.6541	-0.0154	0.7332	0.7329	0.0422	0.3523	0.3547
MS	0.0029	0.6528	0.6528	0.0051	0.7455	0.7458	0.0561	0.3538	0.3580
BQU	-0.0041	0.6540	0.6540	-0.0085	0.7348	0.7344	0.0441	0.3566	0.3592
BQU*	0.0477	0.6326	0.6342	0.0731	0.7037	0.7072	0.1025	0.3957	0.4089

Table 4: Finite Sample Performances of Different Estimators with t-Noise (10^{-4})

	GS			SBC			XOM		
	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
	Constant Volatility, Lower Contamination								
HL	-0.2504	0.4015	0.4731	-0.2611	0.3981	0.4760	-0.0881	0.1752	0.1961
ZMA	-0.2343	0.4105	0.4726	-0.2388	0.4111	0.4754	-0.0890	0.1723	0.1939
BNHLS	-0.0204	0.2782	0.2789	-0.0305	0.3218	0.3232	-0.0166	0.1468	0.1477
MS	-0.0149	0.2900	0.2903	-0.0219	0.3282	0.3289	-0.0120	0.1464	0.1468
BQU	-0.0172	0.2753	0.2758	-0.0254	0.3201	0.3211	-0.0139	0.1487	0.1494
BQU*	0.0008	0.2592	0.2592	0.0016	0.2973	0.2973	-0.0002	0.1403	0.1403
	Constant Volatility, Higher Contamination								
HL	-0.2314	0.4358	0.4934	-0.3571	0.5189	0.6299	-0.0488	0.6163	0.6182
ZMA	-0.2214	0.4379	0.4906	-0.1922	0.5035	0.5389	-0.0483	0.2696	0.2739
BNHLS	-0.0568	0.4047	0.4087	-0.0862	0.4845	0.4921	-0.0496	0.2404	0.2455
MS	-0.0433	0.4063	0.4085	-0.0625	0.4937	0.4976	-0.0363	0.2461	0.2488
BQU	-0.0482	0.4106	0.4134	-0.0763	0.4914	0.4972	-0.0457	0.2434	0.2477
BQU*	0.0003	0.3911	0.3910	-0.0023	0.4711	0.4710	-0.0032	0.2333	0.2333
	Stochastic Volatility, Lower Contamination								
HL	-0.2055	0.8528	0.8771	-0.2144	0.8491	0.8759	-0.0324	0.3510	0.3525
ZMA	-0.1805	0.8745	0.8926	-0.1797	0.8599	0.8782	-0.0330	0.3491	0.3506
BNHLS	0.0039	0.4303	0.4304	-0.0010	0.4674	0.4675	0.0167	0.2174	0.2183
MS	0.0085	0.4315	0.4318	0.0054	0.4698	0.4700	0.0202	0.2192	0.2203
BQU	-0.0031	0.5473	0.5476	0.0024	0.4648	0.4650	0.0182	0.2163	0.2173
BQU*	0.0242	0.4149	0.4157	0.0310	0.4394	0.4407	0.0386	0.1984	0.2023
	Stochastic Volatility, Higher Contamination								
HL	-0.1666	0.8832	0.8988	-0.3082	1.1115	1.1536	-0.0917	0.9461	0.9512
ZMA	-0.1566	0.8961	0.9095	-0.1099	0.9388	0.9453	0.0411	0.4263	0.4289
BNHLS	-0.0043	0.6018	0.6021	-0.0093	0.6838	0.6842	0.0491	0.3239	0.3281
MS	0.0088	0.6125	0.6128	0.0105	0.6779	0.6782	0.0610	0.3248	0.3310
BQU	0.0019	0.5993	0.5996	-0.0022	0.6863	0.6867	0.0516	0.3284	0.3329
BQU*	0.0546	0.5386	0.5418	0.0770	0.6368	0.6418	0.1073	0.3510	0.3676

7 Appendix of Proofs

Theorem A.1 (Rao (1973)) *Let W, Ω, and $\{U_i, i = 1, 2, .., k\}$ be positive definite and symmetric matrices. Let B be any arbitrary matrix and $\{p_i, i = 1, ..., k\}$ be constants. The solution to the following minimization problem $\min_{W \geq 0} tr W \Omega W \Omega$ subject to*

$$WB = 0,$$
$$Tr(WU_i) = p_i, \ \ i = 1, 2, ..., k, \tag{A.1}$$

is

$$W = \sum_{i=1}^{k} \lambda_i C' \Omega^{-1} U_i \Omega^{-1} C \tag{A.2}$$

where $\lambda_1, ..., \lambda_k$ are roots of

$$\sum_{i=1}^{k} \lambda_i tr \left(C' \Omega^{-1} U_i \Omega^{-1} C U_j \right) = p_j, j = 1, 2, ..., k \tag{A.3}$$

and

$$C = I - B \left(B' \Omega^{-1} B \right)^{-1} B' \Omega^{-1}. \tag{A.4}$$

Proof of Theorem 1. It is easy to see that the mean of $\hat{V}_Q$ is

$$E \hat{V}_Q = Etr(r'Wr) = tr(EWrr') = tr(W\Omega). \tag{A.5}$$

Let $\eta = (\eta_{t_0}, \eta_{t_1}, ..., \eta_{t_m})'$, $r^e = (r_1^e, r_2^e, ..., r_m^e)'$ and $\varepsilon = (\varepsilon_1, ..., \varepsilon_m)'$. Then

$$\varepsilon = D\eta, \ r = r^e + \varepsilon \text{ and } \Omega = \Omega_e + \Omega_\varepsilon. \tag{A.6}$$

where $\Omega_e = E r^e r^{e\prime}$ and $\Omega_\varepsilon = E \varepsilon \varepsilon' = \sigma_\eta^2 DD'$. To prove the variance formula, we first compute

$$
\begin{aligned}
E \left(\varepsilon' W \varepsilon \right) \left(\varepsilon' W \varepsilon \right) &= E \left(\eta' D' W D \eta \right) \left(\eta' D' W D \eta \right) \\
&= var(\eta' D' W D \eta) + \left(E \left(\eta' D' W D \eta \right) \right)^2 \\
&= \sigma_\eta^4 (\kappa_4 - 3) \sum_{i=1}^{m+1} \omega_{ii}^2 + 2\sigma_\eta^4 tr \left(D' W D D' W D \right) + \left(tr \left(\sigma_\eta^2 D' W D \right) \right)^2 \\
&= \sigma_\eta^4 (\kappa_4 - 3) \sum_{i=1}^{m+1} \omega_{ii}^2 + 2 tr \left(W \Omega_\varepsilon W \Omega_\varepsilon \right) + \left(tr \left(W \Omega_\varepsilon \right) \right)^2,
\end{aligned}
\tag{A.7}
$$

where we have used the following result: for any $(m+1) \times (m+1)$ matrix $G = (g_{ij})$, it holds that

$$var(\eta G' \eta) = \sigma_\eta^4 (\kappa_4 - 3) \sum_{i=1}^{m+1} g_{ii}^2 + 2\sigma_\eta^4 tr(GG'). \qquad (A.8)$$

To prove this result, we note that

$$var(\eta G' \eta) = cov\left(\sum_{i=1}^{m+1} \sum_{j=1}^{m+1} g_{ij} \eta_i \eta_j, \sum_{k=1}^{m+1} \sum_{l=1}^{m+1} g_{kl} \eta_k \eta_l \right)$$

$$= \sum_{i=1}^{m+1} \sum_{j=1}^{m+1} \sum_{k=1}^{m+1} \sum_{l=1}^{m+1} g_{ij} g_{kl} cov(\eta_i \eta_j, \eta_k \eta_l)$$

$$= \sum_{i=1}^{m+1} g_{ii}^2 cov(\eta_i^2, \eta_i^2) + 2 \sum_{i=1}^{m+1} \sum_{j \neq i}^{m+1} g_{ij}^2 cov(\eta_i \eta_j, \eta_i \eta_j)$$

$$= \sum_{i=1}^{m+1} g_{ii}^2 \left(E\eta_i^4 - \left(E\eta_i^2 \right)^2 \right) + 2 \left(E\eta_i^2 \right)^2 \sum_{i=1}^{m+1} \sum_{j \neq i}^{m+1} g_{ij}^2$$

$$= \sum_{i=1}^{m+1} g_{ii}^2 \left(E\eta_i^4 - 3 \left(E\eta_i^2 \right)^2 \right) + 2 \left(E\eta_i^2 \right)^2 \sum_{i=1}^{m+1} \sum_{j=1}^{m+1} g_{ij}^2$$

$$= \sigma_\eta^4 (\kappa_4 - 3) \sum_{i=1}^{m+1} g_{ii}^2 + 2\sigma_\eta^4 tr(GG') \qquad (A.9)$$

as desired. Now,

$$E\left(r'Wr \right) \left(r'Wr \right)$$

$$= E\left(r^{e\prime} W r^e + \varepsilon' W \varepsilon + 2\varepsilon' W r^e \right) \left(r^{e\prime} W r^e + \varepsilon' W \varepsilon + 2\varepsilon' W r^e \right)$$

$$= E\left(r^{e\prime} W r^e \right) \left(r^{e\prime} W r^e \right) + 2 \left(E r^{e\prime} W r^e \right) \left(E \varepsilon' W \varepsilon \right)$$

$$+ 4E \varepsilon' W r^e \varepsilon' W r^e + E\left(\varepsilon' W \varepsilon \right) \left(\varepsilon' W \varepsilon \right)$$

$$= 2tr\left(W\Omega_e W\Omega_e \right) + \left(tr(W\Omega_e) \right)^2 + 2tr(W\Omega_e) tr(W\Omega_\varepsilon) + 4tr\left(W\Omega_e W\Omega_\varepsilon \right)$$

$$+ \sigma_\eta^4 (\kappa_4 - 3) \sum_{i=1}^{m+1} \omega_{ii}^2 + 2tr\left(W\Omega_\varepsilon W\Omega_\varepsilon \right) + \left(tr\left(W\Omega_\varepsilon \right) \right)^2$$

$$= 2tr\left(W\Omega W\Omega \right) + \left(tr(W\Omega) \right)^2 + \sigma_\eta^4 (\kappa_4 - 3) \sum_{i=1}^{m+1} \omega_{ii}^2. \qquad (A.10)$$

As a result

$$var\left(r'Wr \right) = E\left(r'Wr \right) \left(r'Wr \right) - \left(tr(W\Omega) \right)^2$$

$$= 2tr\left(W\Omega W\Omega \right) + \sigma_\eta^4 (\kappa_4 - 3) \sum_{i=1}^{m+1} \omega_{ii}^2 \qquad (A.11)$$

which completes the proof of the theorem. ∎

Proof of Theorem 2. We want to find the analytic expression for the inverse of the $m \times m$ matrix $A(\lambda)$. To reflect the dimension of the matrix, we write $A(\lambda) := A_m(\lambda) := A_m$. In view of the formula $A_m^{-1} = A_m^* / \det(A_m)$, where A_m^* is the adjoint matrix of A_m, it suffices to find the cofactor of each element and the determinant of A_m.

First, it is easy to see that

$$\det(A_m) = a \det(A_{m-1}) - \det(A_{m-2}) \tag{A.12}$$

and $\det(A_1) = a$, $\det(A_2) = a^2 - 1$ where $a = -(\lambda + 2)$. To solve the difference equation (A.12), we write it as

$$[\det(A_m) - k_1 \det(A_{m-1})] = k_2 [\det(A_{m-1}) - k_1 \det(A_{m-2})] \tag{A.13}$$

where

$$k_1 + k_2 = a \text{ and } k_1 k_2 = 1. \tag{A.14}$$

In other words,

$$k_1 = \frac{1}{2}a + \frac{1}{2}\sqrt{a^2 - 4}, \ k_2 = \frac{1}{2}a - \frac{1}{2}\sqrt{a^2 - 4}. \tag{A.15}$$

Solving (A.13) recursively, we get

$$[\det(A_m) - k_1 \det(A_{m-1})] = [\det(A_2) - k_1 \det(A_1)] k_2^{m-2}. \tag{A.16}$$

But

$$\det(A_2) - k_1 \det(A_1) = (k_1 + k_2)^2 - 1 - k_1(k_1 + k_2) = k_2^2 \tag{A.17}$$

and thus

$$[\det(A_m) - k_1 \det(A_{m-1})] = k_2^m. \tag{A.18}$$

It now follows from (A.18) that

$$\det(A_m) = k_1^{m-1} a + \sum_{i=0}^{m-2} k_2^{m-i} k_1^i = k_1^{m-1} a + k_2^m \sum_{i=0}^{m-2} k_2^{-2i}$$

$$= k_2^{-m+1}(k_2^{-1} + k_2) + k_2^m \frac{1 - \left(k_2^{-2}\right)^{m-1}}{1 - k_2^{-2}}$$

$$= \frac{k_2^{m+1} - k_2^{-(m+1)}}{k_2 - k_2^{-1}}. \tag{A.19}$$

31

Second, we compute the cofactors of matrix A_m. Direct calculations show that, when $i \leq j$, the (i,j)-th cofactor satisfies

$$A_m^*(i,j) = (-1)^{i+j} \det(A_{i-1}) \det(A_{m-j}). \tag{A.20}$$

By symmetry, the (i,j)-th cofactor for $i \geq j$ is

$$A_m^*(i,j) = (-1)^{i+j} \det(A_{j-1}) \det(A_{m-i}). \tag{A.21}$$

Finally, let $B_m = A_m^{-1}$, then when $i \leq j$, the (i,j)-th element of B_m is

$$\begin{aligned}
B_m(i,j) &= (-1)^{i+j} \det(A_{i-1}) \det(A_{m-j}) / \det(A_m) \\
&= (-1)^{i+j} \frac{\left(k_2^i - k_2^{-i}\right)\left(k_2^{m-j+1} - k_2^{-(m-j+1)}\right)}{\left(k_2 - k_2^{-1}\right)\left(k_2^{m+1} - k_2^{-(m+1)}\right)} \\
&= (-1)^{i+j} \frac{\left(k_2^{m+i-j+1} + k_2^{-(m+i-j+1)}\right) - \left(k_2^{m-i-j+1} + k_2^{-(m-i-j+1)}\right)}{\left(k_2 - k_2^{-1}\right)\left(k_2^{m+1} - k_2^{-(m+1)}\right)}.
\end{aligned} \tag{A.22}$$

Similarly, we can find the (i,j)-th element of B_m when $i \geq j$. A unified expression for $B_m(i,j)$ for any i and j is then given by

$$\begin{aligned}
B_m(i,j) &= (-1)^{i+j} \frac{\left(k_2^{m-|i-j|+1} + k_2^{-(m-|i-j|+1)}\right) - \left(k_2^{m-i-j+1} + k_2^{-(m-i-j+1)}\right)}{\left(k_2 - k_2^{-1}\right)\left(k_2^{m+1} - k_2^{-(m+1)}\right)} \\
&= -\frac{\left(k^{m-|i-j|+1} + k^{-(m-|i-j|+1)}\right) - \left(k^{m-i-j+1} + k^{-(m-i-j+1)}\right)}{\left(k - k^{-1}\right)\left(k^{m+1} - k^{-(m+1)}\right)}
\end{aligned} \tag{A.23}$$

where

$$k = -k_2 = \frac{1}{2}(\lambda + 2) + \frac{1}{2}\sqrt{\lambda^2 + 4\lambda} \geq 0. \tag{A.24}$$

In view of $k = \exp(\alpha)$, we get

$$B_m(i,j) = -\frac{1}{\sqrt{\lambda(\lambda+4)}} \frac{\cosh\{[m+1-|i-j|]\alpha\} - \cosh\{[m+1-i-j]\alpha\}}{\sinh((m+1)\alpha)}, \tag{A.25}$$

from which we obtain

$$\begin{aligned}
W_{BQU}(i,j) &= -\lambda B_m(i,j) \tag{A.26} \\
&= \frac{\lambda}{\sqrt{\lambda(\lambda+4)}} \frac{\cosh\{[m+1-|i-j|]\alpha\} - \cosh\{[m+1-i-j]\alpha\}}{\sinh((m+1)\alpha)}
\end{aligned}$$

as desired. ∎

Proof of Theorem 3. We start by finding analytic expressions for the eigenvalues and eigenvectors of matrix $A = A_m(\lambda)$. Let μ be an eigenvalue and $h = (h_1, h_2, ..., h_m)$ be the corresponding eigenvector, then $A_m h = \mu h$ and $\det(A_m - \mu) = 0$.

First, we compute the eigenvalues of A. Following the same procedure as in the proof of Theorem 2, we have

$$\det(A_m - \mu) = \frac{\rho^{m+1} - \rho^{-(m+1)}}{\rho - \rho^{-1}}, \tag{A.27}$$

where

$$\rho = \frac{1}{2}(a - \mu) - \frac{1}{2}\sqrt{(a - \mu)^2 - 4}. \tag{A.28}$$

Now $\det(A_m - \mu) = 0$ implies that $\rho^{2(m+1)} = 1$ but $\rho^2 \neq 1$. As a result

$$\rho = \exp\frac{j\pi}{(m+1)}i, \ j = 1, 2, ..., m, m + 2, ..., 2m + 1, \tag{A.29}$$

where $i = \sqrt{-1}$. Let

$$\frac{1}{2}(a - \mu) + \frac{1}{2}\sqrt{(a - \mu)^2 - 4} = \exp\frac{j\pi}{(m+1)}i, \tag{A.30}$$

for $j = m + 2, ..., 2m + 1$, we find that

$$\mu = a + 2\cos\frac{j\pi}{(m+1)}, \ j = 1, 2, ..., m. \tag{A.31}$$

Next, we compute the eigenvectors of A_m. It follows from $A_m h = \mu h$ that

$$ah_1 + h_2 = \left(a + 2\cos\frac{j\pi}{(m+1)}\right)h_1,$$

$$h_{k-1} + ah_k + h_{k+1} = \left(a + 2\cos\frac{j\pi}{(m+1)}\right)h_k, k = 2, 3, ..., m - 1 \tag{A.32}$$

$$h_{m-1} + ah_m = \left(a + 2\cos\frac{j\pi}{(m+1)}\right)h_m.$$

Equation (A.32) can be rewritten as

$$h_{k+1} - 2\cos\frac{j\pi}{(m+1)}h_k + h_{k-1} = 0, \ k \geq 2. \tag{A.33}$$

Let

$$\rho_1 = \exp\left(-\frac{j\pi}{(m+1)}i\right) \text{ and } \rho_2 = \exp\left(\frac{j\pi}{(m+1)}i\right) \tag{A.34}$$

then

$$[h_{k+1} - \rho_1 h_k] = \rho_2[h_k - \rho_1 h_{k-1}] = [h_2 - \rho_1 h_1]\rho_2^{k-1}. \tag{A.35}$$

But

$$h_2 - \rho_1 h_1 = \left(2\cos\frac{j\pi}{(m+1)}\right)h_1 - \rho_1 h_1 = \rho_2 h_1,$$

33

and thus

$$h_{k+1} - \rho_1 h_k = \rho_2^k h_1. \tag{A.36}$$

It now follows from (A.36) that

$$h_{k+1} = \rho_1^k h_1 + \left(\sum_{i=0}^{k-1} \rho_1^i \rho_2^{k-i} \right) h_1 = \frac{\rho_1^k \rho_2 - \rho_1^k \rho_2^{-1} + \rho_2^{k+1} - \rho_2^{-k+1}}{\rho_2 - \rho_2^{-1}} h_1$$

$$= \frac{\rho_2^{k+1} - \rho_2^{-(k+1)}}{\rho_2 - \rho_2^{-1}} h_1 = \sin\left(\frac{j\pi(k+1)}{(m+1)} \right) \sin^{-1} \frac{j\pi}{(m+1)} h_1 \tag{A.37}$$

Let $h_1 = \sin \frac{j\pi}{(m+1)}$, then

$$h_k = \sin \frac{j\pi k}{(m+1)} \quad \text{for } k = 1, ..., m-1. \tag{A.38}$$

It is easy to see that $h_m = \sin \frac{j\pi m}{(m+1)}$ satisfies $h_{m-1} + a h_m = \left(a + 2\cos \frac{j\pi}{(m+1)} \right) h_m$. As a result, the eigenvalues of A are

$$\mu_j = a + 2\cos \frac{j\pi}{(m+1)}, \quad j = 1, 2, ..., m \tag{A.39}$$

and the corresponding eigenvectors are

$$h^j = \left(\sin \frac{j\pi}{(m+1)}, \sin \frac{2j\pi}{(m+1)}, ..., \sin \frac{mj\pi}{(m+1)} \right). \tag{A.40}$$

Since

$$\sum_{k=1}^{m} \sin^2 \frac{j\pi k}{(m+1)} = \frac{m+1}{2}, \quad j = 1, 2, ..., m, \tag{A.41}$$

the orthonormal eigenvectors are

$$h^j = \frac{2}{\sqrt{m+1}} \left(\sin \frac{j\pi}{(m+1)}, \sin \frac{2j\pi}{(m+1)}, ..., \sin \frac{mj\pi}{(m+1)} \right). \tag{A.42}$$

Note that (45) can be rewritten as

$$\hat{V}_{BQU} = -\lambda \sum_{j=1}^{m} \mu_j^{-1} r' h^j \left(h^j \right)' r = -\lambda \sum_{j=1}^{m} \mu_j^{-1} \left(r' h_j \right)^2 \tag{A.43}$$

Plugging μ_j and h^j into the above expression completes the proof of the theorem. ∎

Proof of Theorem 4. (i) Since $\hat{V}_{BQU}(\lambda)$ is a weighted summation of $\left(\alpha^{(k)} \right)^2$, $\left(\alpha^{(k)} \right)^2$ is not correlated with $\left(\alpha^{(j)} \right)^2$ for any $k \neq j$, and each $\alpha^{(k)}$ is asymptotically normal, we obtain the asymptotic normality of $\hat{V}_{BQU}(\lambda)$ immediately. It remains to compute the asymptotic variance of $\hat{V}_{BQU}(\lambda)$. But the variance of $\hat{V}_{BQU}(\lambda)$ is

$$\sum_{k=1}^{m} 2\delta_k^2 w_{BQU,k} = 2\lambda^2 \sigma_\eta^2 = 2V^2/m, \tag{A.44}$$

34

which implies

$$\sqrt{m}\left(\hat{V}_{BQU}(\lambda) - V\right) \to N(0, 2V^2). \tag{A.45}$$

(ii) The feasible BQU estimator $\hat{V}_{BQU}(\tilde{\lambda})$ can be written as

$$\hat{V}_{BQU}(\tilde{\lambda}) = \frac{\tilde{V}}{m} \sum_{k=1}^{m} \tilde{\delta}_k^{-1} \left(\sum_{\ell=1}^{m} r_\ell h_\ell^{(k)} \right)^2 \tag{A.46}$$

where $\tilde{\delta}_k = \tilde{\sigma}_\eta^2 \left(\tilde{\lambda} + 2 - 2\cos\frac{k\pi}{(m+1)} \right)$. Using the consistency of $\tilde{\sigma}_\eta^2$ and $\tilde{V}$, we have

$$\begin{aligned}
\frac{\delta_k}{\tilde{\delta}_k} &= \frac{\sigma_\eta^2 \left(\lambda + 2 - 2\cos\frac{k\pi}{(m+1)} \right)}{\tilde{\sigma}_\eta^2 \left(\tilde{\lambda} + 2 - 2\cos\frac{k\pi}{(m+1)} \right)} = \frac{\lambda + 2 - 2\cos\frac{k\pi}{(m+1)}}{\lambda + 2 - 2\cos\frac{k\pi}{(m+1)} + O_p\left(m^{-5/4}\right)} (1 + o_p(1)) \\
&= \left[1 + O_p\left(m^{-5/4} \left(\lambda + 2 - 2\cos\frac{k\pi}{(m+1)} \right)^{-1} \right) \right] (1 + o_p(1)) \\
&= 1 + o_p(1) \tag{A.47}
\end{aligned}$$

uniformly over $k = 1, 2, ..., m$. Now consider

$$\begin{aligned}
\sqrt{m}\left(\frac{\hat{V}_{BQU}(\tilde{\lambda})}{\tilde{V}} - 1 \right) &= \frac{1}{\sqrt{m}} \sum_{k=1}^{m} \left[\left(\delta_k \tilde{\delta}_k^{-1} \right) \delta_k^{-1} \left(\sum_{\ell=1}^{m} r_\ell h_\ell^{(k)} \right)^2 - 1 \right] \\
&= \frac{1}{\sqrt{m}} \sum_{k=1}^{m} \left[(1 + o_p(1)) \delta_k^{-1} \left(\sum_{\ell=1}^{m} r_\ell h_\ell^{(k)} \right)^2 - 1 \right] \\
&= \frac{1}{\sqrt{m}} \sum_{k=1}^{m} \left[\delta_k^{-1} \left(\sum_{\ell=1}^{m} r_\ell h_\ell^{(k)} \right)^2 - 1 \right] (1 + o_p(1)) \\
&\to_d N(0, 2). \tag{A.48}
\end{aligned}$$

Consequently, $\sqrt{m}\left(\hat{V}_{BQU}(\tilde{\lambda}) - \tilde{V}\right) \to_d N(0, 2V^2)$ and

$$\begin{aligned}
m^{1/4}\left(\hat{V}_{BQU}(\tilde{\lambda}) - V\right) &= m^{1/4}\left(\tilde{V} - V\right) + m^{1/4}\left(\hat{V}_{BQU}(\tilde{\lambda}) - \tilde{V}\right) \\
&= m^{1/4}\left(\tilde{V} - V\right) + o_p\left(m^{-1/4}\right)
\end{aligned}$$

as stated. ∎

Proof of Theorem 5. Let $H = (h^{(1)}, h^{(2)}, ..., h^{(m)})$ and

$$\begin{aligned}
\Delta_\Omega &= diag(\lambda + 2 - 2\cos(j\pi/(m+1))), \\
\Delta_\Gamma &= diag(-2 + 2\cos(j\pi/(m+1))). \tag{A.49}
\end{aligned}$$

then

$$\Omega\sigma_\eta^{-2} = -A(\lambda) = H\Delta_\Omega H' \text{ and } \Gamma = A(0) = H\Delta_\Gamma H'. \tag{A.50}$$

As a result

$$
\begin{aligned}
W_{BQU}^* &= \Omega^{-1}\sigma_\eta^2\left(c_1 I + c_2\Gamma\right)\sigma_\eta^2\Omega^{-1} \\
&= H\Delta_\Omega^{-1}H'(c_1 I + c_2 H\Delta_\Gamma H')H\Delta_\Omega^{-1}H' \\
&= H\Delta_\Omega^{-1}(c_1 I + c_2\Delta_\Gamma)\Delta_\Omega^{-1}H' \\
&= H\left[diag(w_{BQU,1}^*, ..., w_{BQU,m}^*)\right]^{-1}H'.
\end{aligned} \tag{A.51}
$$

Hence

$$\hat{V}_{BQU}^* = r'W_{BQU}^* r = \sum_{k=1}^m w_{BQU,k}^* \left(\sum_{\ell=1}^m r_\ell h_\ell^{(k)}\right)^2$$

as desired.

In view of (A.50) and $H'H = I$, it is easy to see that equations in (62) is equivalent to

$$
\begin{aligned}
c_1\beta_{20} - c_2\beta_{21} &= m \\
-c_1\beta_{21} + c_2\beta_{22} &= 0.
\end{aligned} \tag{A.52}
$$

Solving the above equation system leads to the stated formula. ■

Proof of Theorem 6. It is easy to show that

$$E\hat{V}_{BQU}^*(\lambda) = V \text{ and } Var(\hat{V}_{BQU}^*(\lambda)) = 2mc_1\sigma_\eta^4.$$

To evaluate the order of magnitude of $Var(\hat{V}_{BQU}^*(\lambda))$, we start by establishing the orders of magnitude of $\beta_{2,j}$, for $j = 0, 1, 2$. First, approximating the sum by integral, we get

$$
\begin{aligned}
\beta_{2,0} &= \sum_{k=1}^m \frac{1}{\left(\lambda + 2\left(1 - \cos\frac{k\pi}{m+1}\right)\right)^2} \\
&= m\int_{1/m}^{1-1/m} \frac{1}{(\lambda + 2(1 - \cos\pi x))^2}dx(1 + o(1)) \\
&= m\int_{1/m}^{1-1/m} \frac{1}{(\lambda + \pi^2 x^2)^2}dx(1 + o(1))
\end{aligned} \tag{A.53}
$$

where the last line holds because $2(1 - \cos\pi x) = \pi^2 x^2(1 + o(1)$ as $x \to 0$. Now we compute

36

the above integral explicitly:

$$\int_{1/m}^{1-1/m} \frac{1}{(\lambda + \pi^2 x^2)^2} dx$$

$$= -\frac{1}{4\pi^3 x^2 \lambda + 4\pi\lambda^2} \left(2\pi x - (\lambda^2 + \pi^2 x^2 \lambda) \times 2\sqrt{\frac{1}{\lambda^3}} \arctan\left(\frac{\sqrt{\lambda}}{\pi x}\right) \right)_{x=1/m} (1 + o(1))$$

$$= -\frac{1}{4\pi\lambda^2} \left(\frac{2\pi}{m} - 2\sqrt{\lambda} \arctan \frac{m\sqrt{\lambda}}{\pi} \right) (1 + o(1))$$

$$= -\frac{1}{4\pi\lambda^2} \left(\frac{2\pi}{m} - 2\sqrt{\lambda} \left(\frac{\pi}{2} - \frac{\pi}{m\sqrt{\lambda}} \right) \right) (1 + o(1))$$

$$= \frac{1}{4\lambda^2} \sqrt{\lambda}(1 + o(1)) = \frac{1}{4} \left(\frac{V}{\sigma_\eta^2} \right)^{-3/2} m^{3/2}(1 + o(1)). \tag{A.54}$$

Combining (A.53) and (A.54) yields

$$\beta_{2,0} = \frac{1}{4} \left(\frac{V}{\sigma_\eta^2} \right)^{-3/2} m^{5/2}(1 + o(1)). \tag{A.55}$$

Second, using a similar approach, we get

$$\beta_{2,1} = m \int_{1/m}^{1-1/m} \frac{\pi^2 x^2}{(\lambda + \pi^2 x^2)^2} dx (1 + o(1))$$

$$= \frac{m(1 + o(1))}{4\pi(\pi^2 x^2 + \lambda)} \left(2\pi x + 2(\pi^2 x^2 + \lambda) \sqrt{\frac{1}{\lambda}} \arctan\left(\frac{\sqrt{\lambda}}{\pi x}\right) \right)_{x=1/m}$$

$$= \frac{m}{4\pi\lambda} \left(\frac{2\pi}{m} + 2\sqrt{\lambda} \arctan \frac{m\sqrt{\lambda}}{\pi} \right) (1 + o(1))$$

$$= \frac{m}{4\pi\lambda} \left(\frac{2\pi}{m} + 2\sqrt{\lambda} \left(\frac{\pi}{2} - \frac{\pi}{m\sqrt{\lambda}} \right) \right) (1 + o(1))$$

$$= \frac{m}{4\sqrt{\lambda}}(1 + o(1)) = \frac{1}{4} \left(\frac{\sigma_\eta^2}{V} \right)^{1/2} m^{3/2}(1 + o(1)). \tag{A.56}$$

Finally, it is easy to see that

$$\beta_{2,2} = m \int_{1/m}^{1-1/m} \frac{(1 - \cos \pi x)^2}{(\lambda + 2(1 - \cos \pi x))^2} dx (1 + o(1))$$

$$= m(1 + o(1)). \tag{A.57}$$

Now

$$c_1 = \frac{m\beta_2}{\beta_0 \beta_2 - \beta_1^2} = \frac{m \times m(1 + o(1))}{(\sigma_\eta^2/V)^{3/2} m^{5/2} m/4 - (\sigma_\eta^2/V) m^3/16}$$

$$= 4\lambda^{3/2}(1 + o(1)) = 4m^{-\frac{3}{2}} (V/\sigma_\eta^2)^{\frac{3}{2}} (1 + o(1)), \tag{A.58}$$

37

and

$$c_2 = c_1\beta_1/\beta_2 = \lambda\left(1 + o\left(1\right)\right) = m^{-1}\left(V/\sigma_\eta^2\right)(1 + o(1)). \tag{A.59}$$

It follows from (A.58) that

$$\lim_{m\to\infty} var\left(m^{1/4}\left(\hat{V}_{BQU}^* - V\right)\right) = 8V^2\left(\sigma_\eta^2/V\right)^{\frac{1}{2}}. \tag{A.60}$$

The asymptotic normality of $m^{1/4}\left(\hat{V}_{BQU}^* - V\right)$ follows from that of $\sum_{\ell=1}^{m} r_\ell h_\ell^{(k)}$ and the asymptotic independence between $\sum_{\ell=1}^{m} r_\ell h_\ell^{(k_1)}$ and $\sum_{\ell=1}^{m} r_\ell h_\ell^{(k_2)}$ for any $k_1 \neq k_2$.

(ii) A Taylor expansion gives

$$m^{1/4}\left(\hat{V}_{BQU}^*(\tilde{\lambda}) - V\right)$$
$$= m^{1/4}\left(\hat{V}_{BQU}^*(\lambda) - V\right) + \frac{1}{m}\left(\frac{\partial}{\partial\lambda}\hat{V}_{BQU}^*(\lambda)\right)m^{5/4}\left(\tilde{\lambda} - \lambda\right) + o_p\left(1\right) \tag{A.61}$$

where λ^* is between λ and $\tilde{\lambda}$.

We now compute the probability limit of $\frac{1}{m}\left(\frac{\partial}{\partial\lambda}\hat{V}_{BQU}^*(\lambda)\right)$. The mean of $\frac{1}{m}\frac{\partial}{\partial\lambda}\hat{V}_{BQU}^*(\lambda)$ is

$$\frac{1}{m}E\left(\frac{\partial}{\partial\lambda}\hat{V}_{BQU}^*(\lambda)\right) = \frac{1}{m}Er'\frac{\partial W_{BQU}^*(\lambda)}{\partial\lambda}r$$
$$= \frac{1}{m^2}tr\left(\frac{\partial W_{BQU}^*(\lambda)}{\partial\lambda}\right)V - \frac{1}{m}tr\left(\frac{\partial W_{BQU}^*(\lambda)}{\partial\lambda}\Gamma\right)\sigma_\eta^2$$
$$= \frac{1}{m^2}\left[\frac{\partial}{\partial\lambda}tr\left(W_{BQU}^*(\lambda)\right)\right]V - \frac{1}{m}\left[\frac{\partial}{\partial\lambda}tr\left(W_{BQU}^*(\lambda)\Gamma\right)\right]\sigma_\eta^2$$
$$= 0, \tag{A.62}$$

where the last equality follows because by definition $tr(W_{BQU}^*(\lambda)) = m$ and $tr(W_{BQU}^*(\lambda)\Gamma) = 0$. The variance of $\frac{1}{m}\frac{\partial}{\partial\lambda}\hat{V}_{BQU}^*(\lambda)$ is

$$2m^{-2}\sum_{k=1}^{m}\left(\frac{\partial w_{BQU,k}^*}{\partial\lambda}\delta_k\right)^2$$
$$= \frac{2}{m^2}\sum_{k=1}^{m}4\left(\lambda + 2 - 2\cos\frac{k\pi}{(m+1)}\right)^{-4}\left[c_1 - c_2\left(2 - 2\cos\frac{k\pi}{(m+1)}\right)\right]^2$$
$$+ \frac{2}{m^2}\sum_{k=1}^{m}\left(\lambda + 2 - 2\cos\frac{k\pi}{(m+1)}\right)^{-2}\left[\frac{\partial c_1}{\partial\lambda} - \frac{\partial c_2}{\partial\lambda}\left(2 - 2\cos\frac{k\pi}{(m+1)}\right)\right]^2$$
$$- \frac{2}{m^2}\sum_{k=1}^{m}4\left(\lambda + 2 - 2\cos\frac{k\pi}{(m+1)}\right)^{-3}\left[c_1 - c_2\left(2 - 2\cos\frac{k\pi}{(m+1)}\right)\right]$$
$$\times\left[\frac{\partial c_1}{\partial\lambda} - \frac{\partial c_2}{\partial\lambda}\left(2 - 2\cos\frac{k\pi}{(m+1)}\right)\right] \tag{A.63}$$

$$= \frac{2}{m^2} \left\{ 4c_1^2 \beta_{4,0} + 4c_2^2 \beta_{4,2} - 8c_1 c_2 \beta_{4,1} + \dot{c}_1^2 \beta_{2,0} + \dot{c}_2^2 \beta_{2,2} \right.$$
$$\left. -2\dot{c}_1 \dot{c}_2 \beta_{2,1} - 4\beta_{3,0} c_1 \dot{c}_1 - 4\beta_{3,2} c_2 \dot{c}_2 + 4\beta_{3,1} c_1 \dot{c}_2 + 4\beta_{3,1} \dot{c}_1 c_2 \right\} \tag{A.64}$$

where

$$\beta_{i,j} = \sum_{k=1}^{m} \left(\lambda + 2 - 2\cos\frac{k\pi}{(m+1)} \right)^{-i} \left[c_1 - c_2 \left(2 - 2\cos\frac{k\pi}{(m+1)} \right) \right]^j$$

for $i = 2, 3, 4, j = 0, 1, 2$ and for a variable, say c, we denote $\dot{c} = \partial c(\lambda)/\partial \lambda$.

To evaluate the variance of $\frac{1}{m} \frac{\partial}{\partial \lambda} \hat{V}_{BQU}^*(\lambda)$, we first establish the asymptotic approximation for $\beta'_{i,j} s$. Approximating the sum by integral, we get

$$\beta_{4,0} = \sum_{k=1}^{m} \left(\lambda + 2 - 2\cos\frac{k\pi}{(m+1)} \right)^{-4}$$

$$= m \int_{1/m}^{1-1/m} \frac{1}{(\lambda + \pi^2 x^2)^4} dx \, (1 + o(1))$$

$$= \frac{m(1 + o(1))}{96\pi\lambda^3 (\pi^2 x^2 + \lambda)^3} \left\{ 30\pi^5 x^5 + 80\pi^3 x^3 \lambda + 66\pi x \lambda^2 \right.$$

$$\left. + 2\left(15\lambda^6 + 45\pi^2 x^2 \lambda^5 + 45\pi^4 x^4 \lambda^4 + 15\pi^6 x^6 \lambda^3 \right) \sqrt{\frac{1}{\lambda^7}} \arctan\left(\frac{\sqrt{\lambda}}{\pi x} \right) \right\}_{x=1/m} \tag{A.65}$$

$$= m \frac{1}{96\pi\lambda^6} \left(66\pi x \lambda^2 + 2\left(15\lambda^6 \right) \sqrt{\frac{1}{\lambda^7}} \left(\frac{\pi}{2} - \frac{\pi}{m\sqrt{\lambda}} \right) \right) (1 + o(1))$$

$$= \frac{5}{32} m \sqrt{\frac{1}{\lambda^7}} (1 + o(1));$$

$$\beta_{4,1} = \sum_{k=1}^{m} \left(\lambda + 2 - 2\cos\frac{k\pi}{(m+1)} \right)^{-4} \left(2 - 2\cos\frac{k\pi}{(m+1)} \right)$$

$$= m \int_{1/m}^{1-1/m} \frac{\pi^2 x^2}{(\lambda + \pi^2 x^2)^4} dx \, (1 + o(1))$$

$$= \frac{-m(1 + o(1))}{96\pi\lambda^2 (\pi^2 x^2 + \lambda)^3} \left\{ 6\pi^5 x^5 + 16\pi^3 x^3 \lambda - 6\pi x \lambda^2 \right. \tag{A.66}$$

$$\left. -2\left(3\lambda^5 + 9\pi^2 x^2 \lambda^4 + 9\pi^4 x^4 \lambda^3 + 3\pi^6 x^6 \lambda^2 \right) \sqrt{\frac{1}{\lambda^5}} \arctan\left(\frac{\sqrt{\lambda}}{\pi x} \right) \right\}_{x=1/m}$$

$$= \frac{-m}{96\pi\lambda^5} \left(-2\left(3\lambda^5 \right) \sqrt{\frac{1}{\lambda^5}} \left(\frac{\pi}{2} - \frac{\pi}{m\sqrt{\lambda}} \right) \right) (1 + o(1))$$

$$= \frac{1}{32} m \sqrt{\frac{1}{\lambda^5}} (1 + o(1));$$

$$\beta_{4,2} = \sum_{k=1}^{m} \left(\lambda + 2 - 2\cos\frac{k\pi}{(m+1)}\right)^{-4} \left(2 - 2\cos\frac{k\pi}{(m+1)}\right)^2$$

$$= m \int_{1/m}^{1-1/m} \frac{\pi^4 x^4}{(\lambda + \pi^2 x^2)^4} dx \, (1 + o(1)) \tag{A.67}$$

$$= -\frac{m}{96\pi\lambda \left(\pi^2 x^2 + \lambda\right)^3} \left\{16\pi^3 x^3\lambda - 6\pi^5 x^5 + 6\pi x\lambda^2\right.$$

$$+ 2\left(3\lambda^4 + 3\pi^6 x^6\lambda + 9\pi^2 x^2\lambda^3 + 9\pi^4 x^4\lambda^2\right)\sqrt{\frac{1}{\lambda^3}}\arctan\left(\frac{\sqrt{\lambda}}{\pi x}\right)\right\}(1 + o(1))$$

$$= \frac{m}{96\lambda^4}\left(6\lambda^4\sqrt{\frac{1}{\lambda^3}}\frac{1}{2}\right)(1 + o(1))$$

$$= \frac{m}{32}\sqrt{\frac{1}{\lambda^3}}\,(1 + o(1)).$$

Similarly,

$$\beta_{3,0} = \frac{3}{16}m\sqrt{\frac{1}{\lambda^5}}\,(1 + o(1)),$$

$$\beta_{3,1} = \frac{m}{16}\sqrt{\frac{1}{\lambda^3}}\,(1 + o(1)), \tag{A.68}$$

$$\beta_{3,2} = \frac{m}{16}\sqrt{\frac{1}{\lambda^3}}\,(1 + o(1)).$$

To summarize the asymptotic approximations of the $\beta'_{i,j}s$, we get, up to a multiplicative factor $(1 + o(1))$:

$$\beta_{4,0} = \frac{5}{32}m\sqrt{\frac{1}{\lambda^7}}, \beta_{3,0} = \frac{3}{16}m\sqrt{\frac{1}{\lambda^5}}, \beta_{2,0} = \frac{1}{4}m\sqrt{\frac{1}{\lambda^3}},$$

$$\beta_{4,1} = \frac{1}{32}m\sqrt{\frac{1}{\lambda^5}}, \beta_{3,1} = \frac{1}{16}m\sqrt{\frac{1}{\lambda^3}}, \beta_{2,1} = \frac{1}{4}m\sqrt{\frac{1}{\lambda}}, \tag{A.69}$$

$$\beta_{4,2} = \frac{1}{32}m\sqrt{\frac{1}{\lambda^3}}, \beta_{3,2} = \frac{3}{16}m\sqrt{\frac{1}{\lambda}}, \beta_{2,2} = m.$$

Next, we establish the asymptotic approximations for $\dot{c}_1$ and $\dot{c}_2$. Note that

$$\dot{c}_1 = \frac{m\dot{\beta}_{2,2}}{\beta_{2,0}\beta_{2,2} - \beta_{2,1}^2} - \frac{m\beta_{2,2}\left(\dot{\beta}_{2,0}\beta_{2,2} + \beta_{2,0}\dot{\beta}_{2,2} - 2\beta_{2,1}\dot{\beta}_{2,1}\right)}{\left(\beta_{2,0}\beta_{2,2} - \beta_{2,1}^2\right)^2}$$

$$= -\frac{2m\beta_{3,2}}{\beta_{2,0}\beta_{2,2} - \beta_{2,1}^2} + \frac{2m\beta_{2,2}\left(\beta_{3,0}\beta_{2,2} + \beta_{2,0}\beta_{3,2} - 2\beta_{2,1}\beta_{3,1}\right)}{\left(\beta_{2,0}\beta_{2,2} - \beta_{2,1}^2\right)^2}. \tag{A.70}$$

Using results in (A.69), we have

$$-\frac{2m\beta_{3,2}}{\beta_{2,0}\beta_{2,2} - \beta_{2,1}^2} = -\frac{2m\left(\frac{3}{16}m\sqrt{\frac{1}{\lambda}}\right)}{\left(\frac{1}{4}m\sqrt{\frac{1}{\lambda^3}}\right)m}(1 + o(1)) = -\frac{3}{2}\lambda\left((1 + o(1))\right) \tag{A.71}$$

40

and

$$\frac{2m\beta_{2,2}\left(\beta_{3,0}\beta_{2,2}+\beta_{2,0}\beta_{3,2}-2\beta_{2,1}\beta_{3,1}\right)}{\left(\beta_{2,0}\beta_{2,2}-\beta_{2,1}^2\right)^2}$$

$$=2\times\frac{m^2\left(\left(\frac{3}{16}m\sqrt{\frac{1}{\lambda^5}}\right)m+\left(\frac{1}{4}m\sqrt{\frac{1}{\lambda^3}}\right)\left(\frac{3}{16}m\sqrt{\frac{1}{\lambda}}\right)-2\left(\frac{1}{4}m\sqrt{\frac{1}{\lambda}}\right)\left(\frac{1}{16}m\sqrt{\frac{1}{\lambda^3}}\right)\right)}{\left(\left(\frac{1}{4}m\sqrt{\frac{1}{\lambda^3}}\right)m\right)^2}(1+o(1))$$

$$=\left(6\sqrt{\lambda}+\frac{1}{2}\lambda\right)(1+o(1)). \tag{A.72}$$

As a result,

$$\dot{c}_1=\left(6\sqrt{\lambda}-\lambda\right)(1+o(1))=6\sqrt{\lambda}(1+o(1)). \tag{A.73}$$

Similarly, we can show that

$$\dot{c}_2=1+o(1). \tag{A.74}$$

Combining (A.69) with (A.73) and (A.74), we get:

$$var\left(\frac{1}{m}\frac{\partial}{\partial\lambda}\hat{V}_{BQU}^*(\lambda)\right)$$

$$=\frac{2}{m^2}\left\{4\left(4\lambda^{3/2}\right)^2\left(\frac{5}{32}m\sqrt{\frac{1}{\lambda^7}}\right)+4\lambda^2\left(\frac{1}{32}m\sqrt{\frac{1}{\lambda^3}}\right)-8\left(4\lambda^{3/2}\right)(\lambda)\left(\frac{1}{32}m\sqrt{\frac{1}{\lambda^5}}\right)\right.$$

$$+\left(6\sqrt{\lambda}\right)^2\left(\frac{1}{4}m\sqrt{\frac{1}{\lambda^3}}\right)+(1)(m)-2\left(6\sqrt{\lambda}\right)(1)\left(\frac{1}{4}m\sqrt{\frac{1}{\lambda}}\right)$$

$$-4\left(\frac{3}{16}m\sqrt{\frac{1}{\lambda^5}}\right)\left(4\lambda^{3/2}\right)\left(6\sqrt{\lambda}\right)-4\left(\frac{3}{16}m\sqrt{\frac{1}{\lambda}}\right)(\lambda)(1)$$

$$\left.+4\left(\frac{1}{16}m\sqrt{\frac{1}{\lambda^3}}\right)\left(4\lambda^{3/2}\right)+4\left(\frac{1}{16}m\sqrt{\frac{1}{\lambda^3}}\right)\left(6\sqrt{\lambda}\right)(\lambda)\right\}(1+o(1))$$

$$=\frac{1}{4}\frac{\lambda^5}{m}\left(\frac{8}{\sqrt{\lambda^{11}}}+\frac{28}{\lambda^5}-\frac{5}{\sqrt{\lambda^9}}-32\right)(1+o(1))$$

$$=\frac{1}{4}\frac{\lambda^5}{m}\frac{8}{\sqrt{\lambda^{11}}}(1+o(1))=\frac{2}{m\sqrt{\lambda}}(1+o(1))=o(1). \tag{A.75}$$

Therefore

$$\frac{1}{m}\left(\frac{\partial}{\partial\lambda}\hat{V}_{BQU}^*(\lambda)\right)=o_p(1). \tag{A.76}$$

In view of the above result and $m^{5/4}\left(\tilde{\lambda}-\lambda\right)=O_p(1)$, we get

$$m^{1/4}\left(\hat{V}_{BQU}^*(\tilde{\lambda})-V\right)=m^{1/4}\left(\hat{V}_{BQU}^*(\lambda)-V\right)+o_p(1) \tag{A.77}$$

as stated. ∎

References

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

[2] Bandi, F.M., and J.R. Russell (2004a). Realized variance, microstructure noise, and optimal sampling. Working paper.

[3] Bandi, F.M., and J.R. Russell (2004b). Separating market microstructure noise from volatility. *Journal of Financial Economics*, forthcoming.

[4] Bandi, F.M., and J.R. Russell (2005). On the finite sample properties of kernel-based integrated variance estimators. Working paper.

[5] 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.

[6] Barndorff-Nielsen, O.E., P. Hansen, A. Lunde and N. Shephard (2006). Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise. Working paper.

[7] Goncalves, S. and N. Meddahi (2004). Bootstrapping realized volatility. Working paper, CIRANO, Montreal.

[8] Hannan, E. J. (1970). *Multiple Time Series*, New York: John Wiley.

[9] Hansen, P.R., and A. Lunde (2004). An unbiased measure of realized variance. Working paper. Available at http://ssrn.com/paper=524602.

[10] Hansen, P. R. and A. Lunde (2005). A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics,* 3, 525-554.

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

[12] Huang, X. and G. Tauchen (2005). The relative contribution of jumps to total price variation. *Journal of Financial Econometrics* 3, 456-499.

[13] Oomen, R.C.A. (2005). Properties of bias-corrected realized variance in calendar time and business time. *Journal of Financial Econometrics*, 3, 555-577

[14] Percival, D.B. and A.T. Walden (1993). *Spectral Analysis for Physical Applications*, Cambridge University Press.

[15] Phillips, P. C. B., Y. Sun and S. Jin (2005a). Long run variance estimation and robust regression testing using sharp origin kernels with no truncation, forthcoming, *Journal of Statistical Planning and Inference*.

[16] Phillips, P. C. B., Y. Sun and S. Jin (2005b). Spectral density estimation and robust hypothesis testing using steep origin kernels without truncation. forthcoming, *International Economic Review*.

[17] Rao, C.R. (1973). *Linear Statistical Inference and its Applications,* New York: John Wiley

[18] Rao, C.R. and J. Kleffe (1988). *Estimation of Variance Components and Applications*, Elsevier Science Publication Company.

[19] Song, W. T. and B. Schmeiser (1993). Variance of the sample mean: properties and graphs of quadratic-forms estimators. *Operation Research,* 41: 501-517.

[20] Sun, Y. (2004). Estimation of the long-run average relationship in nonstationary panel time series. *Econometric Theory,* 20, 1227-1260

[21] Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *IEEE Proceedings.*, 70, 1055-1096.

[22] 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

[23] Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. forthcoming, *Bernoulli.*

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