Proofs to

"Least Squares Inference on Integrated Volatility and the Relationship between Efficient Prices and Noise"

1 Preliminaries

Under Assumptions 1, 2 and 3, we have that

$$\mathrm{E}\left[RV^{h,s}(N_{h,s})\right] = IV + 2N_{h,s}\omega^2,$$

and hence we have the regression

$$y_{h,s} = c + \beta_0 N_{h,s} + \varepsilon_{h,s}, \quad s = 1, \ldots S,\ h = 1, \ldots, s,$$

where $y_{h,s} = RV^{h,s}(N_{h,s})$ and the total number of observations in the regression is $N_{tot} = S(S+1)/2$. Set $N_{h,s} = N_s$ as $N_{h,s} \approx \frac{N}{s}$, $s = 1, \ldots, S$ up to a rounding error. The above regression can be written in a matrix form as

$$Y = X\theta + \varepsilon,$$

where $\theta = (c, \beta_0)'$. From now on, we condition on the trading times t_j, $j = 1, \ldots, N$, which is equivalent to conditioning on the regressor matrix X.

Set $\mathrm{Var}[\varepsilon] = \Xi = \Xi(N, S)$ (we will usually suppress the dependence on N and S). Hansen & Lunde (2006) (Equation 2) show that[1]

$$\mathrm{Var}[y_{h,s}] = \mathrm{Var}[\varepsilon_{h,s}] = 12\kappa\omega^4 N_s + 8\omega^2 \int_0^1 \sigma_s^2 ds - (6\kappa - 2)\omega^4 + \frac{2}{N_s}\int_0^1 \sigma_s^4 H'(s) ds + o\left(\frac{1}{N_s}\right), \quad (1)$$

[1] In their case with regular sampling $H'(t) = 1$, here we combine their result with a result contained in Equation (25) of Zhang, Mykland & Aït-Sahalia (2005). Alternatively, our expression follows directly from Equation (27) in Zhang et al. (2005) after accounting for the difference in notation.

which is a diagonal element of Ξ. Denoting the OLS estimator $\hat{\theta} = (\hat{c}, \hat{\beta}_0)'$, we have that

$$\mathrm{Var}[\hat{\theta}] = (X'X)^{-1} X' \Xi X (X'X)^{-1}.$$

Denote by X_1 the first row of $(X'X)^{-1}X'$. Then

$$\mathrm{Var}[\hat{c}] = X_1 \Xi X_1'.$$

2 Auxiliary Lemma

Lemma 1. *Under Assumptions 1, 2 and 3, it holds that for any s, s', h, h' for which $s = s'$ and $h = h'$ are not simultaneously fulfilled,*[2]

$$\mathrm{Cov}[RV^{h,s}(N_{h,s}), RV^{h',s'}(N_{h',s'})] = \begin{cases} \frac{2IQ\min(s,s')}{N}, & \text{if } (\star) \\ \frac{2IQ\min(s,s')}{N} + 4\omega^2 \int_{\mathcal{O}} \sigma_s^2 ds + \frac{N\omega^4(12\kappa-4)}{\mathrm{lcm}(s,s')}, & \text{otherwise} \end{cases},$$

where $(\star) : \{t_{js+h}\}_{j=1,\dots,N_{h,s}} \bigcap \{t_{is'+h'}\}_{i=1,\dots,N_{h',s'}} = \varnothing$ *and the set* $\mathcal{O}$ *is defined in the following proof.* $\mathrm{lcm}(s, s')$ *stands for the least common multiplier of* s *and* s'.[3]

Proof. Write the covariance $\mathrm{Cov}[RV^{h,s}(N_{h,s}), RV^{h',s'}(N_{h',s'})]$ explicitly as

$$\mathrm{Cov}[RV^{h,s}(N_{h,s}), RV^{h',s'}(N_{h',s'})] = \mathrm{Cov}\left[\sum_{j=1}^{N_{h,s}} r_{js+h}^2, \sum_{i=1}^{N_{h',s'}} r_{is'+h'}^2\right].$$

[2]The case $s = s'$ and $h = h'$ corresponds to the variance of $RV^{h,s}(N_{h,s})$, which is given in Equation (1).
[3]It holds that $\max(s, s') \leq \mathrm{lcm}(s, s') \leq ss'$. For coprime s and s', $\mathrm{lcm}(s, s') = ss'$.

This expression can be decomposed as

$$\begin{aligned}
&\mathrm{Cov}\left[\sum_{j=1}^{N_{h,s}} r_{js+h}^2, \sum_{i=1}^{N_{h',s'}} r_{is'+h'}^2\right] \\
&= \mathrm{Cov}\left[\sum_{j=1}^{N_{h,s}} r_{js+h}^{*2}, \sum_{i=1}^{N_{h',s'}} r_{is'+h'}^{*2}\right] + 2\,\mathrm{Cov}\left[\sum_{j=1}^{N_{h,s}} r_{js+h}^{*2}, \sum_{i=1}^{N_{h',s'}} r_{is'+h'}^{*} e_{is'+h'}\right] + \mathrm{Cov}\left[\sum_{j=1}^{N_{h,s}} r_{js+h}^{*2}, \sum_{i=1}^{N_{h',s'}} e_{is'+h'}^2\right] \\
&+ 2\,\mathrm{Cov}\left[\sum_{j=1}^{N_{h,s}} r_{js+h}^{*} e_{js+h}, \sum_{i=1}^{N_{h',s'}} r_{is'+h'}^{*2}\right] + 4\,\mathrm{Cov}\left[\sum_{j=1}^{N_{h,s}} r_{js+h}^{*} e_{js+h}, \sum_{i=1}^{N_{h',s'}} r_{is'+h'}^{*} e_{is'+h'}\right] + 2\,\mathrm{Cov}\left[\sum_{j=1}^{N_{h,s}} r_{js+h}^{*} e_{js+h}, \sum_{i=1}^{N_{h',s'}} e_{is'+h'}^2\right] \\
&+ \mathrm{Cov}\left[\sum_{j=1}^{N_{h,s}} e_{js+h}^2, \sum_{i=1}^{N_{h',s'}} r_{is'+h'}^{*2}\right] + 2\,\mathrm{Cov}\left[\sum_{j=1}^{N_{h,s}} e_{js+h}^2, \sum_{i=1}^{N_{h',s'}} r_{is'+h'}^{*} e_{is'+h'}\right] + \mathrm{Cov}\left[\sum_{j=1}^{N_{h,s}} e_{js+h}^2, \sum_{i=1}^{N_{h',s'}} e_{is'+h'}^2\right] \\
&= \mathrm{Cov}\left[\sum_{j=1}^{N_{h,s}} r_{js+h}^{*2}, \sum_{i=1}^{N_{h',s'}} r_{is'+h'}^{*2}\right] + 4\,\mathrm{Cov}\left[\sum_{j=1}^{N_{h,s}} r_{js+h}^{*} e_{js+h}, \sum_{i=1}^{N_{h',s'}} r_{is'+h'}^{*} e_{is'+h'}\right] + \mathrm{Cov}\left[\sum_{j=1}^{N_{h,s}} e_{js+h}^2, \sum_{i=1}^{N_{h',s'}} e_{is'+h'}^2\right],
\end{aligned}$$

where the last equation follows because all other terms are zero. The first term is a covariance between two estimators for IV in the absence of noise. Consider the case when the s'-mesh is a proper subgrid of the s-mesh (or vice versa), e.g., the $s = 2, h = 1$ (comprising the ticks $t_1, t_3, t_5, \ldots$) and $s' = 4, h' = 1$ (comprising the ticks $t_1, t_5, t_9, \ldots$) combination. In these cases, using Lemma 2.1 in Hausman (1978), it follows that the covariance between them is equal to the variance of the more efficient one, i.e.,

$$\begin{aligned}
\mathrm{Cov}\left[\sum_{j=1}^{N_{h,s}} r_{js+h}^{*2}, \sum_{i=1}^{N_{h',s'}} r_{is'+h'}^{*2}\right] &= \begin{cases} \mathrm{Var}\left[\sum_{j=1}^{N_{h,s}} r_{js+h}^{*2}\right] & \text{, if } N_{h,s} \geq N_{h',s'} \\ \mathrm{Var}\left[\sum_{i=1}^{N_{h',s'}} r_{is'+h'}^{*2}\right] & \text{, otherwise.} \end{cases} \\
&= \frac{2}{\max(N_{h,s}, N_{h',s'})} \int_0^1 \sigma_s^4 H'(s) ds + o(1/N) = \frac{c^* \min(s, s')}{N} + o(1/N),
\end{aligned}$$

where the expression for the variance of a sparse realized variance under irregular sampling follows from Equation (25) in Zhang et al. (2005). In the remaining cases, when one of the subgrids does not represent a proper subset of the other one, the above result can be seen as a first order approximation. In these cases, the covariance is in fact smaller than the variance of the more efficient one, but the difference between the two terms is of a smaller order than the variance.

The second term vanishes if $\{t_{js+h}\}_{j=1,\ldots,N_{h,s}} \bigcap \{t_{is'+h'}\}_{i=1,\ldots,N_{h',s'}} = \varnothing$ since then the summands are uncorrelated. In the remaining cases we have $\{t_{js+h}\}_{j=1,\ldots,N_{h,s}} \bigcap \{t_{is'+h'}\}_{i=1,\ldots,N_{h',s'}} = \mathcal{A}$, which is a set with $\frac{N}{\mathrm{lcm}(s,r)}$ elements. For $\{\{\{t_{js+h}\} \in \mathcal{A}\} \bigcup \{\{t_{is'+h'}\} \in \mathcal{A}\}\}$, denote $t^* = \max(t_{(j-1)s+h}, t_{(i-1)s'+h'})$ and $t_* = \min(t_{js+h}, t_{is'+h'})$, where the dependence on i, j, s, h, s', h' is deliberately suppressed. Since $e_{js+h} = u_{(j-1)s+h} - u_{js+h}$, we have that for each individual summand in the second term, there are 3 possibilities:

$$\text{Cov}\left[r^*_{js+h}e_{js+h}, r^*_{is'+h'}e_{is'+h'}\right] = \begin{cases} 0, & \text{if } t_{js+h} \neq t_{is'+h'} \text{ and } t_{(j-1)s+h} \neq t_{(i-1)s'+h'} \\ \omega^2 \int_{t^*}^{t_{js+h}} \sigma_s^2 ds, & \text{if } t_{js+h} = t_{is'+h'} \\ \omega^2 \int_{t_{(j-1)s+h}}^{t_*} \sigma_s^2 ds, & \text{if } t_{(j-1)s+h} = t_{(i-1)s'+h'} \end{cases}.$$

It follows that

$$\begin{aligned} 4\,\text{Cov}\left[\sum_{j=1}^{N_{h,s}} r^*_{js+h}e_{js+h}, \sum_{i=1}^{N_{h',s'}} r^*_{is'+h'}e_{is'+h'}\right] &= 4\omega^2 \sum_{\mathcal{A}} \left(\int_{t^*}^{t_{js+h}} \sigma_s^2 ds + \int_{t_{(j-1)s+h}}^{t_*} \sigma_s^2 ds \right) \\ &= 4\omega^2 \sum_{\mathcal{A}} \left(\int_{t^*}^{t_*} \sigma_s^2 ds \right) = 4\omega^2 \int_{\mathcal{O}} \sigma_s^2 ds, \end{aligned}$$

where $\mathcal{O} = \bigcup_{t_{js+h}\in\mathcal{A}, t_{is'+h'}\in\mathcal{A}} [t_*, t^*]$. Since the set $\mathcal{A}$ has $\frac{N}{\text{lcm}(s,s')}$ elements and each of the integrals $\int_{t^*}^{t_*} \sigma_s^2 ds$ is of order $O\left(\frac{1}{\max(N_{h,s},N_{h',s'})}\right)$ and $\frac{1}{\max(N_{h,s},N_{h',s'})} = \frac{\min(s',s)}{N}$, it follows that $\int_{\mathcal{O}} \sigma_s^2 ds$ is of order $O\left(\frac{\min(s,s')}{\text{lcm}(s,s')}\right)$.

The third term is also zero whenever $\{t_{js+h}\}_{j=1,\ldots,N_{h,s}} \bigcap \{t_{is'+h'}\}_{i=1,\ldots,N_{h',s'}} = \varnothing$. In the remaining cases we have that for each $j, i : t_{js+h} \in \mathcal{A}, t_{is'+h'} \in \mathcal{A}$ there are four correlated pairs of noise terms, e.g., if $t_{js+h} = t_{is'+h'}$, then the following four pairs are correlated: $e^2_{js+h}, e^2_{is'+h'}$; $e^2_{(j-1)s+h}, e^2_{is'+h'}$; $e^2_{js+h}, e^2_{(i-1)s'+h'}$ and $e^2_{(j-1)s+h}, e^2_{(i-1)s'+h'}$. Take, for example, the first pair and consider its covariance:

$$\begin{aligned} \text{Cov}\left[e^2_{js+h}, e^2_{is'+h'}\right] &= \text{E}\left[e^2_{js+h}e^2_{is'+h'}\right] - \text{E}\left[e^2_{js+h}\right]\text{E}\left[e^2_{is'+h'}\right] \\ &= \text{E}\left[u^2_{js+h}u^2_{is'+h'}\right] + \text{E}\left[u^2_{(j-1)s+h}u^2_{is'+h'}\right] + \text{E}\left[u^2_{js+h}u^2_{(i-1)s'+h'}\right] \\ &\quad + \text{E}\left[u^2_{(j-1)s+h}u^2_{(i-1)s'+h'}\right] - \text{E}\left[e^2_{js+h}\right]\text{E}\left[e^2_{is'+h'}\right] \\ &= \mu_4 + 3\omega^4 - 4\omega^4 = \mu_4 - \omega^4 = (3\kappa - 1)\omega^4. \end{aligned}$$

The remaining three pairs can be similarly shown to have the same covariance. Thus, it follows

$$\text{Cov}\left[\sum_{j=1}^{N_{h,s}} e^2_{js+h}, \sum_{i=1}^{N_{h',s'}} e^2_{is'+h'}\right] = \frac{N\omega^4(12\kappa - 4)}{\text{lcm}(s,s')}.$$

□

3 Proof of Theorem 1

Calculating X_1

We have

$$X'X = \begin{pmatrix} N_{tot} & \sum\limits_{s,h} N_s \\ \sum\limits_{s,h} N_s & \sum\limits_{s,h} N_s^2 \end{pmatrix},$$

and in the following we suppress the double summation indices s, h when unambiguous. Then

$$\det(X'X) = N_{tot} \sum N_s^2 - \left(\sum N_s\right)^2,$$

and

$$(X'X)^{-1} = \frac{1}{\det(X'X)} \begin{pmatrix} \sum N_s^2 & -\sum N_s \\ -\sum N_s & N_{tot} \end{pmatrix}.$$

The first row of $(X'X)^{-1}$ is

$$\left(\frac{\sum N_s^2}{N_{tot} \sum N_s^2 - (\sum N_s)^2} \quad -\frac{\sum N_s}{N_{tot} \sum N_s^2 - (\sum N_s)^2} \right).$$

Set

$$A = \frac{1}{N_{tot} B - C^2}, \text{ with } B = \sum N_s^2 \text{ and } C = \sum N_s.$$

We then have:

$$X_1 = \begin{pmatrix} AB - ACN_1 \\ \\ \left.\begin{matrix} AB - ACN_2 \\ AB - ACN_2 \end{matrix}\right\} 2 \text{ times} \\ \vdots \\ \left.\begin{matrix} AB - ACN_S \\ \vdots \\ AB - ACN_S \end{matrix}\right\} S \text{ times} \end{pmatrix}'.$$

Calculating $\mathrm{Var}[\hat{c}]$

Given the block structure of X_1 and Ξ, we can write

$$X_1 \Xi X_1' = \sum_{s=1}^{S} \sum_{r=1}^{S} \sum_{i=1}^{s} \sum_{j=1}^{r} X_1^{(s)} X_1^{(r)} \xi_{ij}^{(s,r)}.$$

where $\xi_{ij}^{(s,r)}$ is the ij element in the (s,r)-block of Ξ. Let us look at the terms A, B and C. For B we have

$$\lim_{S\to\infty} B = \lim_{S\to\infty} \sum N_s^2 \;\; = \;\; \lim_{S\to\infty} \sum_{s=1}^{S} \sum_{h=1}^{s} N_s^2 = \lim_{S\to\infty} \sum_{s=1}^{S} s N_s^2 = N^2 \lim_{S\to\infty} \sum_{s=1}^{S} \frac{1}{s} = N^2 \lim_{S\to\infty} (\ln(S) + \gamma_0)$$

with γ_0 the Euler-Mascheroni constant. Similarly, we can derive $C = NS$. It follows that

$$\begin{aligned} \lim_{S\to\infty} A &= \lim_{S\to\infty} \frac{1}{\frac{S(S+1)}{2} N^2(\ln(S) + \gamma_0) - N^2 S^2} \\ &= \lim_{S\to\infty} \frac{2}{N^2(S^2 \ln(S) + S^2(\gamma_0 - 2) + S\ln(S) + S\gamma_0)} \end{aligned}$$

The expression

$$\mathrm{Var}[\hat{c}] = X_1 \Xi X_1' = \sum_{s=1}^{S} \sum_{r=1}^{S} \sum_{i=1}^{s} \sum_{j=1}^{r} X_1^{(s)} X_1^{(r)} \xi_{ij}^{(s,r)}$$

can be decomposed as

$$\begin{aligned}
&\sum_{s=1}^{S}\sum_{r=1}^{S}\sum_{i=1}^{s}\sum_{j=1}^{r} X_1^{(s)} X_1^{(r)} \xi_{ij}^{(s,r)} \\
&\quad = \sum_{s=1}^{S}\sum_{i=1}^{s}\sum_{j=1}^{s} \left(X_1^{(s)}\right)^2 \xi_{ij}^{(s,s)} + \sum_{s=1}^{S}\sum_{r\neq s}^{S}\sum_{i=1}^{s}\sum_{j=1}^{r} X_1^{(s)} X_1^{(r)} \xi_{ij}^{(s,r)} \\
&\quad = \sum_{s=1}^{S}\sum_{i=1}^{s} \left(X_1^{(s)}\right)^2 \xi_{ii}^{(s,s)} + \sum_{s=1}^{S}\sum_{i=1}^{s}\sum_{j\neq i}^{s} \left(X_1^{(s)}\right)^2 \xi_{ij}^{(s,s)} + \sum_{s=1}^{S}\sum_{r\neq s}^{S}\sum_{i=1}^{s}\sum_{j=1}^{r} X_1^{(s)} X_1^{(r)} \xi_{ij}^{(s,r)} \quad (2)
\end{aligned}$$

The term $\sum_{s=1}^{S}\sum_{i=1}^{s}\left(X_1^{(s)}\right)^2 \xi_{ii}^{(s,s)}$

Since $X_1^{(s)}$ does not depend on i, we have

$$\sum_{s=1}^{S}\sum_{i=1}^{s}\left(X_1^{(s)}\right)^2 \xi_{ii}^{(s,s)} = \sum_{s=1}^{S}\left(X_1^{(s)}\right)^2 \sum_{i=1}^{s} \xi_{ii}^{(s,s)}$$

We have that (ignoring the $o\left(\frac{1}{N_s}\right)$ term)

$$\xi_{ii}^{(s,s)} = \underbrace{aN_s + b}_{\text{noise error}} + \underbrace{\frac{c}{N_s}}_{\text{discretization error}},$$

where, by comparing to Equation (1), we see that

$$a = 12\kappa\omega^4, \quad b = 8\omega^2\int_0^1 \sigma_s^2 ds - (6\kappa - 2)\omega^4, \quad c = 2\int_0^1 \sigma_s^4 H'(s) ds$$

The inner sum is

$$\sum_{i=1}^{s} \xi_{ii}^{(s,s)} = \sum_{i=1}^{s}\left(a\frac{N}{s} + b + \frac{cs}{N}\right) = aN + bs + \frac{cs^2}{N}.$$

Further

$$\left(X_1^{(s)}\right)^2 = \left(AB - AC\frac{N}{s}\right)^2 = A^2B^2 - 2A^2BC\frac{N}{s} + A^2C^2\frac{N^2}{s^2}$$

Finally, we have

$$\sum_{s=1}^{S}\sum_{i=1}^{s}\left(X_1^{(s)}\right)^2\xi_{ii}^{(s,s)} = \sum_{s=1}^{S}\left(aN+bs+\frac{cs^2}{N}\right)\left(A^2B^2-2A^2BC\frac{N}{s}+A^2C^2\frac{N^2}{s^2}\right)$$

Since

$$\sum_{s=1}^{S}s^2=\frac{1}{6}(2S^3+3S^2+S), \qquad \sum_{s=1}^{S}s=\frac{1}{2}(S^2+S),$$
$$\lim_{S\to\infty}\left(\sum_{s=1}^{S}\frac{1}{s}-\ln(S)\right)=\gamma_0, \qquad \lim_{S\to\infty}\left(\sum_{s=1}^{S}\frac{1}{s^2}-\frac{\pi^2}{6}\right)=0,$$
$$A^2B^2\in O\left(\frac{1}{S^4}\right), \qquad A^2BC\in O\left(\frac{1}{NS^3\ln(S)}\right),$$
$$A^2C^2\in O\left(\frac{1}{N^2S^2(\ln(S))^2}\right)$$

we obtain that as $S\to\infty$ and $N\to\infty$, $\sum_{s=1}^{S}\sum_{i=1}^{s}\left(X_1^{(s)}\right)^2\xi_{ii}^{(s,s)}$ is dominated by $\frac{2\pi^2aN}{3(S(\ln(S)+\gamma_0)+(\ln(S)+\gamma_0)-2S)^2}$ which is of order $O\left(\frac{N}{S^2(\ln(S))^2}\right)$.

The term $\sum_{s=1}^{S}\sum_{i=1}^{s}\sum_{j\neq i}^{s}\left(X_1^{(s)}\right)^2\xi_{ij}^{(s,s)}$

For this term we need the covariance between two realized variances computed at the same sampling frequency (within an (s,s)-block) but with non-overlapping grids. As we are working under an iid noise framework, this covariance is not affected by the noise. Using the same arguments as Barndorff-Nielsen & Shephard (2002), it follows that this covariance is equal to

$$\xi_{ij}^{(s,s)}=\mathrm{Cov}\left[RV^{h,s}(N_{h,s}),RV^{h',s}(N_{h',s})\right]=\frac{2}{N_s}\int_0^1\sigma_s^4H'(s)ds+o\left(\frac{1}{N_s}\right)=\frac{c}{N_s}+o\left(\frac{1}{N_s}\right).$$

Then we have

$$\sum_{s=1}^{S}\sum_{i=1}^{s}\sum_{j\neq i}^{s}\left(X_1^{(s)}\right)^2\xi_{ij}^{(s,s)}=\sum_{s=1}^{S}\left(X_1^{(s)}\right)^2\sum_{i=1}^{s}\sum_{j\neq i}^{s}\frac{cs}{N}=\sum_{s=1}^{S}\left(X_1^{(s)}\right)^2\frac{s^2(s-1)c}{N}$$

Substituting in $\left(X_1^{(s)}\right)^2$ yields

$$\begin{aligned}\sum_{s=1}^{S}\left(X_1^{(s)}\right)^2\frac{s^2(s-1)c}{N} &= \sum_{s=1}^{S}\left(A^2B^2-2A^2BC\frac{N}{s}+A^2C^2\frac{N^2}{s^2}\right)\frac{s^2(s-1)c}{N}\\ &= \sum_{s=1}^{S} cA^2B^2\frac{s^2(s-1)}{N}-2cA^2BCs(s-1)+cA^2C^2N(s-1).\end{aligned}$$

This sum is of order $O\left(\frac{1}{N}\right)$ and thus negligible.

The term $\sum_{s=1}^{S}\sum_{r\neq s}^{S}\sum_{i=1}^{s}\sum_{j=1}^{r}X_1^{(s)}X_1^{(r)}\xi_{ij}^{(s,r)}$

For this term we use Lemma 1. The covariance $\xi_{ij}^{(s,r)}$ is affected by whether the numbers s and r are coprime or not. Consider first the case (I) when s and r are coprime. This implies that the number of common observations in an s-subgrid and r-subgrid is $\frac{N}{sr}$ for all s-subgrids and r-subgrids. From Lemma 1, it follows that in this case the covariance $\xi_{ij}^{(s,r)}$ can be written as

$$\xi_{ij}^{(s,r)}=a^*\frac{N}{sr}+b^*\int_{\mathcal{O}}\sigma_s^2ds+\frac{c^*\min(s,r)}{N},$$

where

$$a^*=12\kappa\omega^4-4\omega^4,\quad b^*=4\omega^2,\quad c^*=2\int_0^1\sigma_s^4H'(s)ds.$$

In the second case (II) s and r are not coprime. In such an (s,r)-block there are two possibilities: (II.1) in $\mathrm{lcm}(s,r)$ out of the sr elements in the block, the number of common points on both subgrids is $\frac{N}{\mathrm{lcm}(s,r)}$, (II.2) in the remaining $sr-\mathrm{lcm}(s,r)$ cases the subgrids do not share observations. In case (II.1) we have

$$\xi_{ij}^{(s,r)}=a^*\frac{N}{\mathrm{lcm}(s,r)}+b^*\int_{\mathcal{O}}\sigma_s^2ds+\frac{c^*\min(s,r)}{N},$$

while in case (II.2) it holds that

$$\xi_{ij}^{(s,r)} = \frac{c^* \min(s,r)}{N}.$$

As in all cases (I, II.1 and II.2), $\xi_{ij}^{(s,r)}$ does not depend on i and j, and because for coprime s and r, $\operatorname{lcm}(s,r) = sr$, we can write in general that

$$\begin{aligned}
\sum_{i=1}^{s}\sum_{j=1}^{r}\xi_{ij}^{(s,r)} &= \left(a^* \frac{N}{\operatorname{lcm}(s,r)} + b^* \int\limits_{\mathcal{O}} \sigma_s^2 ds + \frac{c^* \min(s,r)}{N}\right)\operatorname{lcm}(s,r) + \frac{c^* \min(s,r)}{N}(sr - \operatorname{lcm}(s,r)) \\
&= a^* N + b^* \int\limits_{\mathcal{O}} \sigma_s^2 ds \operatorname{lcm}(s,r) + \frac{c^* sr \min(s,r)}{N} \\
&\approx a^* N + b^* \min(s,r) + \frac{c^* sr \min(s,r)}{N}
\end{aligned}$$

where the last approximation is employed for operational reasons in the sense that $\int_{\mathcal{O}} \sigma_s^2 ds$ term is of order $O\left(\frac{\min(s,r)}{\operatorname{lcm}(s,r)}\right)$ (and as we show in the sequel, terms involving b^* are asymptotically negligible). As the matrix Ξ is symmetric, we express

$$\sum_{s=1}^{S}\sum_{r\neq s}^{S} X_1^{(s)} X_1^{(r)} \sum_{i=1}^{s}\sum_{j=1}^{r}\xi_{ij}^{(s,r)} = 2\sum_{s=1}^{S}\sum_{r>s}^{S} X_1^{(s)} X_1^{(r)} \sum_{i=1}^{s}\sum_{j=1}^{r}\xi_{ij}^{(s,r)}.$$

Substituting in the above derived equation for $\sum_{i=1}^{s}\sum_{j=1}^{r}\xi_{ij}^{(s,r)}$, $X_1^{(s)}$ and $X_1^{(r)}$ results in

$$\begin{aligned}
2\sum_{s=1}^{S}\sum_{r>s}^{S} X_1^{(s)} X_1^{(r)} \sum_{i=1}^{s}\sum_{j=1}^{r}\xi_{ij}^{(s,r)} &= 2\sum_{s=1}^{S}\sum_{r>s}^{S} A^2 \left(B - C\frac{N}{s}\right)\left(B - C\frac{N}{r}\right)\left(a^* N + b^* s + \frac{c^* s^2 r}{N}\right) \\
= 2\Bigg(&\frac{S(S-1)}{2} a^* A^2 B^2 N + b^* A^2 B^2 \sum_{s=1}^{S}\sum_{r>s}^{S} s + \frac{c^* A^2 B^2}{N}\sum_{s=1}^{S}\sum_{r>s}^{S} s^2 r \\
&- a^* A^2 BCN^2 \sum_{s=1}^{S}\sum_{r>s}^{S}\left(\frac{1}{r} + \frac{1}{s}\right) - b^* A^2 BCN \sum_{s=1}^{S}\sum_{r>s}^{S}\left(1 + \frac{s}{r}\right) - c^* A^2 BC \sum_{s=1}^{S}\sum_{r>s}^{S}\left(s^2 + sr\right) \\
&+ a^* A^2 C^2 N^3 \sum_{s=1}^{S}\sum_{r>s}^{S}\frac{1}{rs} + b^* A^2 C^2 N^2 \sum_{s=1}^{S}\sum_{r>s}^{S}\frac{1}{r} + c^* A^2 C^2 N \sum_{s=1}^{S}\sum_{r>s}^{S} s\Bigg).
\end{aligned}$$

We first show that the terms involving b^* are asymptotically negligible. This can be confirmed by considering that $\sum_{s=1}^{S}\sum_{r>s}^{S} s \in O\left(S^3\right)$, $\sum_{s=1}^{S}\sum_{r>s}^{S}\left(1 + \frac{s}{r}\right) \in O\left(S^2\right)$ and $\sum_{s=1}^{S}\sum_{r>s}^{S}\frac{1}{r} \in$

$O(S)$. The term $b^* A^2 B^2 \sum_{s=1}^{S} \sum_{r>s}^{S} s$ is dominant and of order $O\left(\frac{1}{S}\right)$ and hence asymptotically negligible. Next, we look at limits of terms involving a^*. To this end consider the sums

$$\sum_{s=1}^{S}\sum_{r>s}^{S}\left(\frac{1}{r}+\frac{1}{s}\right) = \sum_{s=1}^{S}\left(\sum_{r=1}^{S}\frac{1}{r}-\sum_{r=1}^{s}\frac{1}{r}\right)+\sum_{s=1}^{S}\frac{1}{s}(S-s)$$
$$= 2\sum_{s=1}^{S}(\ln(S)+\gamma_0) - \sum_{s=1}^{S}(\ln(s)+\gamma_0) - S = S(\ln(S)+\gamma_0) - 0.5\ln(S) - 0.5\ln(2\pi) + o(1).$$
$$\sum_{s=1}^{S}\sum_{r>s}^{S}\frac{1}{rs} = \sum_{s=1}^{S}\frac{1}{s}\left(\sum_{r=1}^{S}\frac{1}{r}-\sum_{r=1}^{s}\frac{1}{r}\right) = \sum_{s=1}^{S}\frac{1}{s}(\ln(S)+\gamma_0) - \sum_{s=1}^{S}\frac{1}{s}(\ln(s)+\gamma_0)$$
$$= (\ln(S)+\gamma_0)^2 - 0.5\,(\ln(S))^2 - \gamma_1 - \gamma_0(\ln(S)+\gamma_0) = 0.5\,(\ln(S))^2 + \gamma_0\ln(S) - \gamma_1 + o(1).$$

where we have used that $\lim_{S\to\infty}\left(\sum_{s=1}^{S}\ln(s) - \ln\left(\sqrt{2\pi S}\left(\frac{S}{e}\right)^S\right)\right) = 0$ by Sterling's approximation and $\lim_{S\to\infty}\left(\sum_{s=1}^{S}\frac{\ln(s)}{s} - 0.5\,(\ln(S))^2\right) = \gamma_1$, where γ_1 is the first Stieltjes constant equal to approximately -0.0728 (see, e.g., Havil (2003)). Thus, we obtain

$$\begin{aligned}\frac{S(S-1)}{2}A^2B^2N &= \frac{1}{2}\frac{4N(S^2-S)(\ln(S)+\gamma_0)^2}{\left(S^2(\ln(S)+\gamma_0)+S(\ln(S)+\gamma_0)-2S^2\right)^2}\\ &= \frac{2N(S^2-S)}{\left(S^2+S-\frac{2S^2}{\ln(S)+\gamma_0}\right)^2} = \frac{2N}{\left(S+1-\frac{2S}{\ln(S)+\gamma_0}\right)^2} + O\left(\frac{N}{S^3}\right).\end{aligned}$$

$$\begin{aligned}-A^2BCN^2\sum_{s=1}^{S}\sum_{r>s}^{S}\left(\frac{1}{r}+\frac{1}{s}\right) &= -\frac{4NS(\ln(S)+\gamma_0)\left(S(\ln(S)+\gamma_0)-0.5\ln(S)-0.5\ln(2\pi)\right)}{\left(S^2(\ln(S)+\gamma_0)+S(\ln(S)+\gamma_0)-2S^2\right)^2}\\ &= -\frac{4N}{\left(S+1-\frac{2S}{\ln(S)+\gamma_0}\right)^2} + O\left(\frac{N}{S^3}\right).\end{aligned}$$

$$\begin{aligned}A^2C^2N^3\sum_{s=1}^{S}\sum_{r>s}^{S}\frac{1}{rs} &= \frac{4NS^2\left(0.5\,(\ln(S))^2+\gamma_0\ln(S)-\gamma_1\right)}{\left(S^2(\ln(S)+\gamma_0)+S(\ln(S)+\gamma_0)-2S^2\right)^2}\\ &= \frac{NS^2\left(2\,(\ln(S)+\gamma_0)^2-2\gamma_0^2-4\gamma_1\right)}{\left(S^2(\ln(S)+\gamma_0)+S(\ln(S)+\gamma_0)-2S^2\right)^2}\\ &= \frac{2N}{\left(S+1-\frac{2S}{\ln(S)+\gamma_0}\right)^2} - \frac{N(2\gamma_0^2+4\gamma_1)}{\left(S(\ln(S)+\gamma_0)+(\ln(S)+\gamma_0)-2S\right)^2}.\end{aligned}$$

Summing up the three terms, we obtain $-\frac{(2\gamma_0^2+4\gamma_1)N}{(S(\ln(S)+\gamma_0)+(\ln(S)+\gamma_0)-2S)^2}+O\left(\frac{N}{S^3}\right)$. It remains to calculate the terms with c^*. We have $\sum_{s=1}^{S}\sum_{r>s}^{S} s^2 r = 1/15S^5+1/24S^4-1/12S^3-1/24S^2+1/60S$, $\sum_{s=1}^{S}\sum_{r>s}^{S}(s^2+sr) = 5/24S^4+1/12S^3-5/24S^2-1/12S$, and $\sum_{s=1}^{S}\sum_{r>s}^{S} s = 1/6S^3-1/6S$. Considering the order of the terms A^2B^2, A^2BC and A^2C^2, the leading term turns out to be

$$\frac{c^*A^2B^2}{N}\sum_{s=1}^{S}\sum_{r>s}^{S} s^2 r = \frac{4Sc^*}{15N\left(1-\frac{2}{\ln(S)+\gamma_0}+\frac{1}{S}\right)^2}+O\left(\frac{1}{N}\right).$$

Final Result

Let $N\to\infty$ and $S=\alpha N^\beta$ for $\alpha>0$ and $\beta\in[0.5,1)$. Summing everything up together results in

$$\mathrm{Var}[\hat{c}] = \frac{2(\pi^2 a-6(\gamma_0^2+2\gamma_1)a^*)N}{3\left(S(\ln(S)+\gamma_0)+(\ln(S)+\gamma_0)-2S\right)^2}+\frac{8Sc^*}{15N\left(1-\frac{2}{\ln(S)+\gamma_0}+\frac{1}{S}\right)^2}+O\left(N^{-1/2}\right).$$

Let $\eta=\frac{2}{3}(\pi^2 a-6(\gamma_0^2+2\gamma_1)a^*)$ and $\delta=\frac{8c^*}{15}$. Substituting $S=\alpha N^\beta$ we can rewrite

$$\mathrm{Var}[\hat{c}] = \frac{\eta}{\beta^2\alpha^2}N^{1-2\beta}(\ln(N))^{-2}+\delta\alpha N^{\beta-1}+o\left(N^{1-2\beta}(\ln(N))^{-2}\right)+o\left(N^{\beta-1}\right).$$

4 Proof of Corollary 1

The choice of β determines the speed of convergence and the dominating terms. The highest speed of convergence of the estimator is achieved when $N^{1-2\beta}(\ln(N))^{-2}=N^{\beta-1}$ which holds for $\beta_N=\frac{2}{3}\left(1-\frac{\ln(\ln(N))}{\ln(N)}\right)$ converging to $\beta=2/3$ from below. For $\beta_N=\frac{2}{3}\left(1-\frac{\ln(\ln(N))}{\ln(N)}\right)$, we have that $N^{1-2\beta_N}(\ln(N))^{-2}=N^{\beta_N-1}=N^{-1/3}(\ln(N))^{-2/3}$. Thus, with $S=\alpha N^{\beta_N}$, the asymptotic variance can be expressed as

$$\lim_{N\to\infty}\mathrm{Var}\left[N^{1/6}(\ln(N))^{1/3}\,\hat{c}\right] = \lim_{N\to\infty}\frac{\eta}{\beta_N^2\alpha^2}+\delta\alpha = \frac{9\eta}{4\alpha^2}+\delta\alpha.$$

Minimizing this expression with respect to α gives

$$\alpha^* = \sqrt[3]{\frac{9\eta}{2\delta}}$$

for which $\mathrm{Var}\left[N^{1/6}(\ln(N))^{1/3}\,\hat{c}\right] = 2.48\sqrt[3]{\delta^2\eta}$.

Recalling that $a = 12\kappa\omega^4$, $a^* = 12\kappa\omega^4 - 4\omega^4$, $c^* = 2\int_0^1 \sigma_s^4 H'(s)ds$, denoting $IQ = \int_0^1 \sigma_s^4 H'(s)ds$ and setting $\kappa = 1$ (normal noise) we can write $\eta = 8\omega^4(\pi^2 - 4(\gamma_0^2 + 2\gamma_1))$ and $\delta = \frac{16IQ}{15}$ and thus

$$\alpha^* = \sqrt[3]{\frac{33.75\omega^4(\pi^2 - 4(\gamma_0^2 + 2\gamma_1))}{IQ}}.$$

5 Proof of Theorem 2

This proof follows closely the proof of Theorem 1. Denote by X_2 the second row of $(X'X)^{-1}X'$. Then

$$\mathrm{Var}[\hat{\beta}_0] = X_2 \Xi X_2'.$$

Since $\hat{\omega}^2 = \hat{\beta}_0/2$ it follows that

$$\mathrm{Var}[\hat{\omega}^2] = \frac{1}{4} X_2 \Xi X_2'.$$

Using notation from above, X_2 is given by:

$$X_2 = \begin{pmatrix} -AC + N_{tot}AN_1 \\ \\ \left.\begin{matrix} -AC + N_{tot}AN_2 \\ -AC + N_{tot}AN_2 \end{matrix}\right\} \text{2 times} \\ \vdots \\ \left.\begin{matrix} -AC + N_{tot}AN_S \\ \vdots \\ -AC + N_{tot}AN_S \end{matrix}\right\} S \text{ times} \end{pmatrix}'.$$

Calculating $\mathrm{Var}[\hat{\beta}_0]$

Given the block structure of X_1 and Ξ, we can write

$$X_2 \Xi X_2' = \sum_{s=1}^{S} \sum_{r=1}^{S} \sum_{i=1}^{s} \sum_{j=1}^{r} X_2^{(s)} X_2^{(r)} \xi_{ij}^{(s,r)}.$$

Using the decomposition in Eq. (2), with $X_2^{(\cdot)}$ in the place of $X_1^{(\cdot)}$, we examine each term separately.

The term $\sum_{s=1}^{S} \sum_{i=1}^{s} \left(X_2^{(s)}\right)^2 \xi_{ii}^{(s,s)}$

Since $X_2^{(s)}$ does not depend on i, we have

$$\sum_{s=1}^{S} \sum_{i=1}^{s} \left(X_2^{(s)}\right)^2 \xi_{ii}^{(s,s)} = \sum_{s=1}^{S} \left(X_2^{(s)}\right)^2 \sum_{i=1}^{s} \xi_{ii}^{(s,s)}$$

where

$$\left(X_2^{(s)}\right)^2 = \left(-AC + N_{tot}A\frac{N}{s}\right)^2 = A^2C^2 - 2A^2CN_{tot}\frac{N}{s} + A^2N_{tot}^2\frac{N^2}{s^2}.$$

Thus, we obtain

$$\sum_{s=1}^{S}\sum_{i=1}^{s}\left(X_2^{(s)}\right)^2\xi_{ii}^{(s,s)} \quad = \quad \sum_{s=1}^{S}\left(aN+bs+\frac{cs^2}{N}\right)\left(A^2C^2-2A^2CN_{tot}\frac{N}{s}+A^2N_{tot}^2\frac{N^2}{s^2}\right)$$

Using results from above, it follows that as $S\to\infty$ and $N\to\infty$, the leading term in the expression is given by $\frac{\pi^2 a}{6N(\ln(S))^2}$.

The term $\sum_{s=1}^{S}\sum_{i=1}^{s}\sum_{j\neq i}^{s}\left(X_2^{(s)}\right)^2\xi_{ij}^{(s,s)}$

As in Theorem 1, this term is of smaller order than the previous term and thus asymptotically negligible.

The term $\sum_{s=1}^{S}\sum_{r\neq s}^{S}\sum_{i=1}^{s}\sum_{j=1}^{r}X_2^{(s)}X_2^{(r)}\xi_{ij}^{(s,r)}$

As above we examine the cases I: s and r coprime, II.1: s and r not coprime with number of common points on both subgrids $\frac{N}{\text{lcm}(s,r)}$ ($\text{lcm}(s,r)$ elements), and II.2: s and r not coprime with no common points on the subgrids ($sr-\text{lcm}(s,r)$ elements). Proceeding as in the proof of Theorem 1, it can be shown that terms involving b^* are asymptotically negligible. From the terms involving a^*, the dominating term can be shown to be $\frac{a^*}{N}$.

The three terms involving c^* are of the same order, so it becomes important to consider them in more detail. The first one, $2\frac{c^*A^2C^2}{N}\sum_{s=1}^{S}\sum_{r\neq s}^{S}s^2r$ has a leading term given by $\frac{8c^*S^3}{15N^3(\ln(S))^2}$, the second one $-2c^*A^2CN_{tot}\sum_{s=1}^{S}\sum_{r\neq s}^{S}(s^2+sr)$ has a leading term given by $-\frac{5c^*S^3}{6N^3(\ln(S))^2}$, and the third one $2C^*A^2N_{tot}^2N\sum_{s=1}^{S}\sum_{r\neq s}^{S}s$ has a leading term given by $\frac{c^*S^3}{3N^3(\ln(S))^2}$. Summing up the three terms, we obtain $\frac{c^*S^3}{30N^3(\ln(S))^2}$.

Final Result

Let $N\to\infty$ and $S=\alpha N^\beta$ for $\alpha>0$ and $\beta\in[0.5,1)$. Summing everything up together results in

$$\text{Var}[\hat{\omega}^2]=\frac{1}{4}\text{Var}[\hat{\beta}_0]=\frac{1}{4}\left(\frac{a^*}{N}+\frac{c^*S^3}{30N^3(\ln(S))^2}\right)+o\left(N^{-1}\right).$$

Substituting $S = \alpha N^{\beta}$ in the above equation results in

$$\mathrm{Var}[\hat{\omega}^2] = \frac{1}{4}\left(\frac{a^*}{N} + \frac{c^*\alpha^3 N^{3\beta}}{30N^3(\ln(\alpha)+\beta\ln(N))^2}\right) + o\left(N^{-1}\right).$$

The two terms in the brackets are of the same order, $O\left(N^{-1}\right)$, if $\beta = \frac{2}{3}\left(\frac{\ln(\ln(N))}{\ln(N)} + 1\right)$. It follows that for $\beta < \frac{2}{3}\left(\frac{\ln(\ln(N))}{\ln(N)} + 1\right)$

$$\lim_{N\to\infty} \mathrm{Var}[\hat{\omega}^2] = \frac{a^*}{4N} + o\left(N^{-1}\right).$$

References

Barndorff-Nielsen, O. E. & Shephard, N. (2002), 'Econometric analysis of realized volatility and its use in estimating stochastic volatility models', *Journal of the Royal Statistical Society Series B* **64**(2), 253–280.

Hansen, P. R. & Lunde, A. (2006), 'Realized variance and market microstructure noise', *Journal of Business and Economic Statistics* **24**, 127–161.

Hausman, J. A. (1978), 'Specification test in econometrics', *Econometrica* **46**, 1251–1272.

Havil, J. (2003), *Exploring Euler's Constant*, Princeton University Press, New Jersey.

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