Edgeworth expansions in small noise asymptotics*

Lan Zhang, Per A. Mykland, and Yacine Aït-Sahalia

First Draft: November 2004.

Abstract

The paper considers Edgeworth expansions for estimators of volatility. Unlike the usual exapsions, we have found that in order to obtain meaningful terms, one needs to let the size of the noise to go zero asymptotically. This is reflected in our expansions. The results have application to Cornish-Fisher inversion and bootstrapping.
KEY WORDS: Bootstrapping; Edgeworth expasion; Measurement error; Subsampling; Market Microstructure; Martingale; Bias-correction; Realized volatility.

*Lan Zhang is Assistant Professor, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, and Assistant Professor, Department of Finance, University of Illinois at Chicago, Chicago, IL 60607. E-mail: lzhang@stat.cmu.edu. Per A. Mykland is Professor, Department of Statistics, The University of Chicago, Chicago, IL 60637. E-mail: mykland@galton.uchicago.edu. Yacine Aït-Sahalia is Professor, Department of Economics and Bendheim Center for Finance, Princeton University and NBER, Princeton, NJ 08544-1021. E-mail: yacine@princeton.edu.
We gratefully acknowledge the support of the National Science Foundation under grants DMS-0204639 (Zhang and Mykland) and SBR-0111140 (Aït-Sahalia).

1 Introduction

In this paper, we consider the Edgeworth expansion on the volatility estimator when the price process is noisy.

Let $\{Y_{t_i}\}$, $0 = t_0 \leq t_1 \leq \cdots t_n = T$, be the observed (log) price of a security at time $t_i \in [0, T]$. Suppose that these observed prices can be decomposed into an underlying (log) price process X (the signal) and a noise term. That is, at each observation time t_i, one can write

$$Y_{t_i} = X_{t_i} + \epsilon_{t_i}. \tag{1}$$

Let the signal (latent) process X follows an Itô process

$$dX_t = \mu_t dt + \sigma_t dB_t \tag{2}$$

where B_t is a standard Brownian motion. Typically, μ_t, the drift coefficient, and σ_t^2, the instantaneous variance of the returns process X_t, will be (continuous) stochastic processes.

Let the noise ϵ_{t_i} in (1) satisfy the following assumption,

$$\epsilon_{t_i} \ i.i.d. \text{ with } E\epsilon_{t_i} = 0, \text{ and } \ Var(\epsilon_{t_i}) = E\epsilon^2. \text{ Also } \epsilon \perp\!\!\!\perp X \text{ process} \tag{3}$$

where $\perp\!\!\!\perp$ denotes independence between two random quantities. Note that our interest in the noise is only at the observation times t_i's, so, model (1) does not require that ϵ_t exists for every t.

In Zhang, Mykland, and Aït-Sahalia (2003), our focus is to construct a statistically sound estimator for integrated volatility $\int_0^T \sigma_t^2 dt$ of the true process, assuming model (1) and that Y_{t_i}'s can be observed highly frequently. In search for a final estimator, we have touched a sequence of RV estimators, which are from the statistically least desiarble to the most desirable: the *all* estimator $[Y,Y]^{(all)}$, the *sparse* estimator $[Y,Y]^{(sparse)}$, the *optimal, sparse* estimator $[Y,Y]^{(sparse,opt)}$, the *averaging* estimator $[Y,Y]^{(avg)}$, the *optimal, averaging* estimator $[Y,Y]^{(avg,opt)}$, and the final *two scale* estimator (TSRV) $\widehat{\langle X, X\rangle}$. While the TSRV is consistent, the first four estimators are biased, typically in proportion to the sampling frequency. When one looks at the stochastic terms in all five estimators, they should be asymptotically normal. However, simulation results show that the distribution of the stochastic term in the sparse estimators and the averaging estimator is far from normality. We argue that the lack of normality is caused by the coexistence of small effective sample size and small noise. In the current paper, we provide Edgeworth expansions to the sparse estimators and the averaging estimator.

What makes the situation unusual is that the errors ϵ are very small, and if they are taken to be of order $O_p(1)$, their impact on the Edgeworth expansion may be exaggerated. Consequently, the coefficients in the expansion may not accurately reflect which terms are important.

To deal with this, we here find expansions under the hypothesis that the size of $|\epsilon|$ goes to zero, as stated precisely at the beginning of Section 4.

We provide the coefficients, both conditional and unconditional, in the expansions for $[Y, Y]^{(sparse)}$ and $[Y, Y]^{(avg)}$. In particular, we shall see that not only does the latter have substantially less bias than the former, but it is also much closer to a normal distribution, cf. the end of Section 4.

With the help of Cornish-Fisher expansions, our Edgeworth expansions can be used for the purpose of setting intervals that are more accurate than the ones based on the normal distribution, see, for example, Hall (1992). Since our expansions also hold in a triangular array setting, they can also be used to analyse the behavior of bootstrapping distributions (for earlier theory on bootstrapping in this setting, see Goncalves and Meddahi (2005)).

2 Estimators

Our estimators have the following forms.

First, $[Y, Y]_T^{(all)}$ uses all the observations,

$$[Y, Y]_T^{(all)} = \sum_{t_i \in \mathcal{G}} (Y_{t_{i+1}} - Y_{t_i})^2, \tag{4}$$

where $\mathcal{G}$ contains all the observation times t_i's in $[0, T]$, $0 = t_0 \leq t_1, \ldots, \leq t_n = T$.

The sparse estimator uses a subsample of the data,

$$[Y, Y]_T^{(sparse)} = \sum_{t_j, t_{j,+} \in \mathcal{H}} (Y_{t_{j,+}} - Y_{t_j})^2, \tag{5}$$

where $\mathcal{H}$ is a strict subset of $\mathcal{G}$, with sample size n_{sparse}, $n_{sparse} < n$. And, if $t_i \in \mathcal{H}$, then $t_{i,+}$ denotes the following elements in $\mathcal{H}$.

The optimal estimator $[Y, Y]^{(sparse,opt)}$ has the same form as in (5) except replacing n_{sparse} with $n^*{}_{sparse}$, where $n^*{}_{sparse}$ is determined by minimizing MSE of the estimator.

The averaging estimator maintains a slow sampling scheme while using all the data,

$$[Y, Y]_T^{(avg)} = \frac{1}{K} \sum_{k=1}^{K} \underbrace{\sum_{t_j, t_{j,+} \in \mathcal{G}^{(k)}} (Y_{t_{j,+}} - Y_{t_j})^2}_{[Y,Y]_T^{(k)}}, \tag{6}$$

where $\mathcal{G}^{(k)}$'s are disjoint with union $\mathcal{G}$. Let n_k be the number of time points in $\mathcal{G}_k$, $\bar{n} = \sum_k n_k$ would then be the average sample size across different grids $\mathcal{G}_k$, $k = 1, \ldots, K$.

One can also consider the optimal, averaging estimator $[Y, Y]^{(avg,opt)}$, by substituting $\bar{n}$ by $\bar{n}^*$ where the latter is selected to balance the bias-variance trade-off in the error of averaging estimator.

A special case of (6) when the sampling points are regularly allocated is of the form,

$$[Y,Y]_T^{(avg)} = \frac{1}{K}\sum_{t_j,t_{j+K}\in\mathcal{G}} (Y_{t_{j+K}} - Y_{t_j})^2,$$

where the sum-squared returns are computed only from subsampling every K-th observation times, and then averaged with equal weights.

The TSRV has the form of

$$\widehat{\langle X,X\rangle}_T = [Y,Y]_T^{(avg)} - \frac{\bar{n}}{n}[Y,Y]_T^{(all)} \tag{7}$$

that is, the volatility estimator $\widehat{\langle X,X\rangle}_T$ combines the sum squared estimators from two different time scales, $[Y,Y]_T^{(avg)}$ from the returns on a slow time scale whereas $[Y,Y]_T^{(all)}$ from the returns on a fast time scale. $\bar{n}$ in (7) is the average sample size across different grids.

From model (1), the distributions of various estimators can be studied by decomposing the sum-squared returns $[Y,Y]$,

$$[Y,Y]_T = [X,X]_T + 2[X,\epsilon]_T + [\epsilon,\epsilon]_T. \tag{8}$$

The above decomposion applies to all the estimators in this section, with the samples suitably selected.

3 Why Do We Need the Edgeworth Expansions?

3.1 Asymptotical Normality in theory: sparse estimator and averaging estimator

3.1.1 sparse estimator

For the sparse estimator, we have shown in Zhang, Mykland, and Aït-Sahalia (2003) that

$$\begin{aligned}&[Y,Y]_T^{(sparse)} \\ \overset{\mathcal{L}}{\approx}\ & \langle X,X\rangle_T + \underbrace{2n_{\text{sparse}}E\epsilon^2}_{\text{bias due to noise}} + \underbrace{[\underbrace{Var([\epsilon,\epsilon]_T^{(sparse)}) + 8[X,X]_T^{(sparse)}E\epsilon^2}_{\text{due to noise}} + \underbrace{\frac{2T}{n_{\text{sparse}}}\int_0^T \sigma_t^4 dt}_{\text{due to discretization}}}_{\text{total variance}}]^{1/2} Z_{\text{total}},\end{aligned} \tag{9}$$

where $Var([\epsilon,\epsilon]_T^{(sparse)}) = 4n_{\text{sparse}}E\epsilon^4 - 2Var(\epsilon^2)$, and Z_{total} is standard normal.

If the sample size n_{sparse} is large relative to the noise, the variance due to noise in (9) would be dominated by $Var([\epsilon,\epsilon]_T^{(sparse)})$ which is of order $n_{sparse}E\epsilon^4$. However, at the co-presence of small

n_{sparse} and small noise (say, $E\epsilon^2$), $8[X,X]_T^{sparse}E\epsilon^2$ is not necessarily smaller than $Var([\epsilon,\epsilon]_T^{(sparse)})$. One then needs to add $8[X,X]_T^{sparse}E\epsilon^2$ into the approximation. We call this correction *small-sample, small-error adjustment*. This type of adjustment is often useful, since the magnitude of the microstructure noise is typically smallish as documented in the empirical literature (cite).

Of course, n_{sparse} is selected either arbitrarily or in some ad hoc manner. In contrast, the sampling frequency in the optimal-sparse estimator $[Y,Y]^{(sparse,opt)}$ can be determined by minimizing the MSE of the estimator analytically. Distribution-wise, the optimal-sparse estimator has the same form as in (9), but, one replaces n_{sparse} by the optimal sampling frequency n^*_{sparse}, where for equidistant observations,

$$n^*_{\text{sparse}} = \left(E\epsilon^2\right)^{-2/3}\left(\frac{T}{4}\int_0^T \sigma_t^4 dt\right)^{1/3}. \tag{10}$$

n^*_{sparse} is optimal in the sense of minimizing the mean square error of the sparse estimator.

No matter whether n_{sparse} is selected optimally or not, one can see from (9) that the de-biased sparse estimator would be asymptotically normal.

3.1.2 averaging estimator

The optimal-sparse estimator only uses a fraction n^*_{sparse}/n of the data, and it faces the arbitrarity of picking the beginning point of the sample. The averaging estimator overcomes both shortcomings.

Based on the decomposition (8), analysis in Zhang, Mykland, and Aït-Sahalia (2003) leads to

$$[Y,Y]_T^{(avg)} \tag{11}$$

$$\stackrel{\mathcal{L}}{\approx} \langle X,X\rangle_T + \underbrace{2\bar{n}E\epsilon^2}_{\text{bias due to noise}} + \underbrace{[\underbrace{Var([\epsilon,\epsilon]_T^{(avg)}) + \frac{8}{K}[X,X]_T^{(avg)}E\epsilon^2}_{\text{due to noise}} + \underbrace{\frac{4T}{3\bar{n}}\int_0^T \sigma_t^4 dt}_{\text{due to discretization}}}_{\text{total variance}}]^{1/2} Z_{\text{total}},$$

where

$$Var([\epsilon,\epsilon]_T^{(avg)}) = 4\frac{\bar{n}}{K}E\epsilon^4 - \frac{2}{K}Var(\epsilon^2),$$

and Z_{total} is a standard normal term.

For the optimal-averaging estimator $[Y,Y]^{(avg,opt)}$, its distribution has the same form as in (11) but substituting $\bar{n}$ with the optimal sub-sampling size $\bar{n}^*$. To find $\bar{n}^*$, one determines K^* from the bias-variance trade-off in (11) and then set $K^* \approx n/\bar{n}^*$. In the equidistantly sampled case,

$$\bar{n}^* = \left(\frac{T}{6(E\epsilon^2)^2}\int_0^T \sigma_t^4 dt\right)^{1/3}. \tag{12}$$

If one removes the bias in $[Y,Y]_T^{(avg)}$ or in $[Y,Y]_T^{(avg,opt)}$, the next term would follow asymptotically normal.

However, the distributions of the de-biased sparse estimators and the de-biased averaging estimators are not exactly normal from the simulation results.

4 Edgeworth Expansions for the Distribution of the Estimators

An Edgeworth expansion up to second order can be found separately for each of the components in (8) by first considering expansions for $n^{-1/2}([\epsilon,\epsilon]^{(all)} - 2nE\epsilon^2)$ and $n^{-1/2}K([\epsilon,\epsilon]_T^{(avg)} - 2\bar{n}E\epsilon^2)$. Each of these can be represented exactly as a triangular array of martingales. Results deriving such an expansion can be found in Mykland (1993, 1995b,a).

It is easily seen that in the current case, the expansion takes on the usual Edgeworth form, see for example Section 5.3 of McCullagh (1987). Note that with the exception of the term of type $n^{-1/2}K([\epsilon,\epsilon]_T^{(avg)} - 2\bar{n}E\epsilon^2)$, the expansion can also be found from Bickel, Götze, and van Zwet (1986).

We assume that the "size" of the law of ϵ goes to zero, formally that $E|\epsilon|^p \to 0$ for all $p \in (0, 8]$. In particlular, say, $O_p(E|\epsilon|^5) = o_p(E|\epsilon|^4)$.

4.1 Conditional Cumulants

We start with the conditional cumulants for $[Y,Y]$ and $[Y,Y]^{(avg)}$, given the latent process X. All the expressions about $[Y,Y]$ hold for both $[Y,Y]^{(all)}$ and $[Y,Y]^{(sparse)}$, in the former case, n remains to be the total sample size in $\mathcal{G}$ while in the latter n is replaced by n_{sparse}. Similar notations apply for $[\epsilon,\epsilon]$ and for $[X,X]$.

4.1.1 third-order conditional cumulants

Denote

$$c_3(n) \triangleq cum_3([\epsilon,\epsilon] - 2nE\epsilon^2), \tag{13}$$

where $[\epsilon,\epsilon] = \sum_{i=0}^{n-1} (\epsilon_{t_{i+1}} - \epsilon_{t_i})^2$.

From Lemma 1 in the Appendix,

$$\begin{aligned} c_3(n) &= 8\left[(n-\frac{3}{4})\ cum_3(\epsilon^2) - 7(n-\frac{6}{7})\ cum_3(\epsilon)^2 + 6(n-\frac{1}{2})\ var(\epsilon)var(\epsilon^2)\right] \\ &= O_p(nE[\epsilon^6]) \end{aligned} \tag{14}$$

and also because the ϵ's from the different grids are independent,

$$cum_3\left(K([\epsilon,\epsilon]^{(avg)} - 2\bar{n}E\epsilon^2)\right) = \sum_{k=1}^{K} cum_3([\epsilon,\epsilon]^{(k)} - 2n_k E\epsilon^2) = Kc_3(\bar{n}). \tag{15}$$

For the conditional third cumulant of $[Y,Y]$:

$$\begin{aligned} cum_3([Y,Y]_T|X) &= cum_3([\epsilon,\epsilon]_T + 2[X,\epsilon]|X) \\ &= cum_3([\epsilon,\epsilon]_T) + 6cum([\epsilon,\epsilon]_T,[\epsilon,\epsilon]_T,[X,\epsilon]_T|X) \\ &+ 12cum([\epsilon,\epsilon]_T,[X,\epsilon]_T,[X,\epsilon]_T|X) + 8cum_3([X,\epsilon]_T|X) \end{aligned}$$

To proceed, define

$$a_i = \begin{cases} 1 & \text{if } 1 \le i \le n-1 \\ \frac{1}{2} & \text{if } i = 0, n \end{cases} \tag{16}$$

and

$$b_i = \begin{cases} \Delta X_{t_{i-1}} - \Delta X_{t_i} & \text{if } 1 \le i \le n-1 \\ \Delta X_{t_{n-1}} & \text{if } i = n \\ -\Delta X_{t_0} & \text{if } i = 0 \end{cases} \tag{17}$$

Note that $[X,\epsilon]_T = \sum_{i=0}^{n} b_i \epsilon_{t_i}$.

Then it follows that

$$cum([\epsilon,\epsilon]_T,[\epsilon,\epsilon]_T,[X,\epsilon]_T|X) = \sum_{i=0}^{n} b_i cum([\epsilon,\epsilon]_T,[\epsilon,\epsilon]_T,\epsilon_{t_i}) = (b_0+b_n)[2E\epsilon^2E\epsilon^3 - 3E\epsilon^5] = O_p(n^{-1/2}E[|\epsilon|^5])$$

because $cum([\epsilon,\epsilon]_T,[\epsilon,\epsilon]_T,\epsilon_{t_i}) = cum([\epsilon,\epsilon]_T,[\epsilon,\epsilon]_T,\epsilon_{t_1})$, for $i = 1,\cdots,n-1$.

Also

$$\begin{aligned} &cum([\epsilon,\epsilon]_T,[X,\epsilon]_T,[X,\epsilon]_T|X) \\ &= cum(2\sum_{i=0}^{n} a_i\epsilon_{t_i}^2, \sum_{j=0}^{n} b_j\epsilon_{t_j}, \sum_{k=0}^{n} b_k\epsilon_{t_k}|X) - cum(2\sum_{i=0}^{n-1} \epsilon_{t_i}\epsilon_{t_{i+1}}, \sum_{j=0}^{n} b_j\epsilon_{t_j}, \sum_{k=0}^{n} b_k\epsilon_{t_k}|X) \\ &= 2\sum_{i=0}^{n} a_i b_i^2 Var(\epsilon^2) - 4\sum_{i=0}^{n-1} b_i b_{i+1}(Var(\epsilon))^2 \\ &= 4[X,X]_T E\epsilon^4 + O_p(n^{-1/2}E[\epsilon^4]) \end{aligned}$$

Finally,

$$\begin{aligned} cum_3([X,\epsilon]_T|X) &= \sum_{i=0}^{n} b_i^3 cum_3(\epsilon) \\ &= E(\epsilon^3)[-3\sum_{i=1}^{n-1}(\Delta X_{t_{i-1}})^2(\Delta X_{t_i}) + 3\sum_{i=1}^{n-1}(\Delta X_{t_{i-1}})(\Delta X_{t_i})^2] \\ &= O_p(n^{-1/2}E[|\epsilon|^3]) \end{aligned}$$

Summing up,

$$cum_3([Y,Y]_T|X) = cum_3([\epsilon,\epsilon]_T) + 48[X,X]E\epsilon^4 + O_p(n^{-1/2}E[|\epsilon|^3]), \tag{18}$$

where $cum_3([\epsilon,\epsilon]_T)$ is given in (14).

For $[Y,Y]_T^{(avg)}$, it is now obvious that

$$cum_3(K[Y,Y]_T^{(avg)}|X) = cum_3(K[\epsilon,\epsilon]_T^{(avg)}) + 48K[X,X]_T^{(avg)}E\epsilon^4 + O_p(K\bar{n}^{-1/2}E[|\epsilon|^3]) \tag{19}$$

4.1.2 fourth-order conditional cumulants

For the fourth-order cumulant, denote

$$c_4(n) \quad \triangleq \quad cum_4([\epsilon,\epsilon]^{(all)} - 2nE\epsilon^2). \tag{20}$$

It follows from Lemma 2 in the Appendix that

$$\begin{aligned} c_4(n) \quad = \quad & 16\{(n-\frac{7}{8})cum_4(\epsilon^2) + n(E\epsilon^4)^2 - 3n(E\epsilon^2)^4 + 12(n-1)var(\epsilon^2)E\epsilon^4 \\ & -32(n-\frac{17}{16})E\epsilon^3 cov(\epsilon^2,\epsilon^3) + 24(n-\frac{7}{4})E\epsilon^2(E\epsilon^3)^2 + 12(n-\frac{3}{4})cum_3(\epsilon^2)E\epsilon^2\} \end{aligned}$$

Also here,

$$cum_4\left(K([\epsilon,\epsilon]^{(avg)} - 2\bar{n}E\epsilon^2)\right) = \sum_{k=1}^{K} cum_4([\epsilon,\epsilon]^{(k)} - 2n_k E\epsilon^2) = Kc_4(\bar{n}). \tag{21}$$

For the conditional fourth-order cumulant, we know that

$$\begin{aligned} cum_4([Y,Y]|X) = {} & cum_4([\epsilon,\epsilon]_T) + 24cum([\epsilon,\epsilon]_T,[\epsilon,\epsilon]_T,[X,\epsilon]_T,[X,\epsilon]_T|X) \\ & + 8cum([\epsilon,\epsilon]_T,[\epsilon,\epsilon]_T,[\epsilon,\epsilon]_T,[X,\epsilon]_T|X) \\ & + 32cum([\epsilon,\epsilon]_T,[X,\epsilon]_T,[X,\epsilon]_T,[X,\epsilon]_T|X) \\ & + 16cum_4([X,\epsilon]|X). \end{aligned} \tag{22}$$

Similar arguement as in deriving the third cumulant shows that the latter three terms in the right hand side of (22) are of order $O_p(n^{-1/2}E[|\epsilon|^5])$.

For the second term in equation (22),

$$cum([\epsilon,\epsilon]_T,[\epsilon,\epsilon]_T,[X,\epsilon]_T,[X,\epsilon]_T|X) \tag{23}$$

$$= \sum_{i,j} b_i b_j cum([\epsilon,\epsilon]_T,[\epsilon,\epsilon]_T,\epsilon_{t_i},\epsilon_{t_j})$$

$$= \sum_i b_i^2 cum([\epsilon,\epsilon]_T,[\epsilon,\epsilon]_T,\epsilon_{t_i},\epsilon_{t_i}) + 2\sum_{i=0}^{n-1} b_i b_{i+1} cum([\epsilon,\epsilon]_T,[\epsilon,\epsilon]_T,\epsilon_{t_i},\epsilon_{t_{i+1}}) \tag{24}$$

Note that $cum([\epsilon,\epsilon]_T, [\epsilon,\epsilon]_T, \epsilon_{t_i}, \epsilon_{t_i})$ and $cum([\epsilon,\epsilon]_T, [\epsilon,\epsilon]_T, \epsilon_{t_i}, \epsilon_{t_{i+1}})$ are independent of i, except close to the edges. One can take α and β to be

$$\alpha = n^{-1}\sum_i cum([\epsilon,\epsilon]_T, [\epsilon,\epsilon]_T, \epsilon_{t_i}, \epsilon_{t_i})$$
$$\beta = n^{-1}\sum_i cum([\epsilon,\epsilon]_T, [\epsilon,\epsilon]_T, \epsilon_{t_i}, \epsilon_{t_{i+1}}).$$

Now following the two identities:

$$cum([\epsilon,\epsilon]_T, [\epsilon,\epsilon]_T, \epsilon_i, \epsilon_i) = cum_3([\epsilon,\epsilon]_T, [\epsilon,\epsilon]_T, \epsilon_i^2) - 2(Cov([\epsilon,\epsilon]_T, \epsilon_i))^2 - Var([\epsilon,\epsilon]_T)E\epsilon^2$$
$$cum([\epsilon,\epsilon]_T, [\epsilon,\epsilon]_T, \epsilon_i, \epsilon_{i+1}) = cum_3([\epsilon,\epsilon]_T, [\epsilon,\epsilon]_T, \epsilon_i\epsilon_{i+1}) - 2Cov([\epsilon,\epsilon]_T, \epsilon_i)Cov([\epsilon,\epsilon]_T, \epsilon_{i+1}),$$

also observing that that $Cov([\epsilon,\epsilon]_T, \epsilon_i) = Cov([\epsilon,\epsilon]_T, \epsilon_{i+1})$, except at the edges,

$$2(\alpha - \beta) = n^{-1}cum_3([\epsilon,\epsilon]_T) - 2Var([\epsilon,\epsilon]_T)E\epsilon^2 + O_p(n^{-1/2}E[|\epsilon|^6])$$

Hence, (24) becomes

$$\begin{aligned}
&cum_4([\epsilon,\epsilon]_T, [\epsilon,\epsilon]_T, [X,\epsilon]_T, [X,\epsilon]_T|X) \\
&= \sum_{i=0}^n b_i^2\alpha + 2\sum_{i=0}^n b_i b_{i+1}\beta + O_p(n^{-1/2}E[|\epsilon|^6]) \\
&= n^{-1}[X,X]_T cum_3([\epsilon,\epsilon]_T) - 2[X,X]_T Var([\epsilon,\epsilon]_T)E\epsilon^2 + O_p(n^{-1/2}E[|\epsilon|^6])
\end{aligned}$$

where the last line is because

$$\sum_{i=0}^n b_i^2 = 2[X,X]_T + O_p(n^{-1/2}), \quad \sum_{i=0}^n b_i b_{i+1} = -[X,X]_T + O_p(n^{-1/2}).$$

Therefore in the final analysis,

$$\begin{aligned}
cum_4([Y,Y]|X) &= cum_4([\epsilon,\epsilon]_T) + 24[X,X]_T n^{-1}cum_3([\epsilon,\epsilon]_T) - 48[X,X]_T Var([\epsilon,\epsilon]_T)E\epsilon^2 \\
&\quad + O_p(n^{-1/2}E[|\epsilon|^5])
\end{aligned} \tag{25}$$

For the average estimator,

$$\begin{aligned}
&cum_4(K[Y,Y]^{(avg)}|X) \\
&= cum_4(K[\epsilon,\epsilon]_T^{(avg)}) + 24K[X,X]_T^{(avg)}\frac{c_3(\bar{n})}{\bar{n}} - 48E\epsilon^2\sum_k [X,X]_T^{(k)}Var([\epsilon,\epsilon]_T^{(k)}) + O_p(K\bar{n}^{-1/2}E[|\epsilon|^5])
\end{aligned}$$

4.2 Unconditional Cumulants

To pass to the unconditional third cumulant, we use the general formulas (Brillinger (1969), Speed (1983), see also Chapter 2 in McCullagh (1987)):

$$\begin{aligned} cum_3(A) &= E[cum_3(A|\mathcal{F})] + 3Cov[Var(A|\mathcal{F}), E(A|\mathcal{F})] + cum_3[E(A|\mathcal{F})] \\ cum_4(A) &= E[cum_4(A|\mathcal{F})] + 4Cov[cum_3(A|\mathcal{F}), E(A|\mathcal{F})] + 3Var[Var(A|\mathcal{F})] \\ &\quad + 6cum_3(Var(A|\mathcal{F}), E(A|\mathcal{F}), E(A|\mathcal{F})) + cum_4(E(A|\mathcal{F})) \end{aligned}$$

4.2.1 unconditional cumulants for sparse estimators

Recall that, in Zhang, Mykland, and Aït-Sahalia (2003), we have derived that

$$E([Y, Y]_T \mid X \text{ process}) = [X, X]_T + 2nE\epsilon^2 \tag{26}$$

also

$$Var([Y, Y]_T|X) = \underbrace{4nE\epsilon^4 - 2Var(\epsilon^2)}_{Var([\epsilon,\epsilon]_T)} + 8[X, X]_T E\epsilon^2 + O_p(E|\epsilon|^2 n^{-1/2}), \tag{27}$$

Put together the third and fourth conditional cumulants with the results from (26)-(27), and we obtain the unconditional cumulants as in the following:

$$\begin{aligned} cum_3([Y, Y]_T - \langle X, X\rangle_T) = c_3(n) &+ 48E(\epsilon^4)E[X, X] \\ &+ 24Var(\epsilon)Cov([X, X]_T, [X, X]_T - \langle X, X\rangle_T) \\ &+ cum_3([X, X]_T - \langle X, X\rangle_T) + O(n^{-1/2}E[|\epsilon|^3]) \end{aligned} \tag{28}$$

and

$$\begin{aligned} cum_4([Y, Y]_T - \langle X, X\rangle_T) = c_4(n) &+ 24\frac{1}{n}c_3(n)E[X, X]_T - 48E[X, X]_T E\epsilon^2 Var([\epsilon, \epsilon]_T) \\ &+ 192E\epsilon^4 Cov([X, X]_T, [X, X]_T - \langle X, X\rangle_T) \\ &+ 192(Var(\epsilon))^2 Var([X, X]_T) \\ &+ 48Var(\epsilon)cum_3([X, X]_T, [X, X]_T - \langle X, X\rangle_T, [X, X]_T - \langle X, X\rangle_T) \\ &+ cum_4([X, X]_T - \langle X, X\rangle_T) + O(n^{-1/2}E[|\epsilon|^5]) \end{aligned} \tag{29}$$

Example 1 (constant σ and equidistant case)**.** *Suppose $\Delta t_i = \Delta t = T/n$ (the equidistant case), and when $\sigma_t = \sigma$ is a constant, $\langle X, X\rangle_T$ is a constant which does not contribute to either of the above cumulants. Also, $[X, X]_T$ has distribution $\sigma^2\Delta t\chi^2_n$ (χ^2 with n degrees of freedom), so that $cum_p([X, X]_T) = \sigma^{2p}(\Delta t)^p n \times cum_p(\chi^2_1) = n^{-(p-1)}(\sigma^2 T)^p cum_p(\chi^2_1)$; recall that*

	$p=1$	$p=2$	$p=3$	$p=4$
$cum_p(\chi^2_1)$	1	2	8	54

It follows that in this case

$$
\begin{aligned}
cum_3([Y,Y]_T - \langle X,X\rangle_T) &= c_3(n) + 48E(\epsilon^4)(\sigma^2 T) \\
&+ 48Var(\epsilon)n^{-1}(\sigma^2 T)^2 \\
&+ 8n^{-2}(\sigma^2 T)^3 + O(n^{-1/2}E[|\epsilon|^3])
\end{aligned}
\tag{30}
$$

Similarly for the fourth cumulant

$$
\begin{aligned}
cum_4([Y,Y]_T - \langle X,X\rangle_T) &= c_4(n) + 24\frac{1}{n}c_3(n)(\sigma^2 T) - 48(\sigma^2 T)E\epsilon^2 Var([\epsilon,\epsilon]_T) \\
&+ 384(E\epsilon^4 + Var(\epsilon)^2)n^{-1}(\sigma^2 T)^2 \\
&+ 384Var(\epsilon)n^{-2}(\sigma^2 T)^3 \\
&+ 54n^{-3}(\sigma^2 T)^4 + O(n^{-1/2}E[|\epsilon|^5])
\end{aligned}
\tag{31}
$$

∎

It is obvious that one needs $\epsilon_n = o_p(n^{-1/2})$ to keep all the terms in (30) and (31) non-neglegible.

In the case of optimal-sparse estimator, (10) lends to $\epsilon = O_p(n^{-3/4})$, in particular $\epsilon = o_p(n^{-1/2})$. Hence, the expression works in this case, and also for many suboptimal choices of n.

For the special case of constant σ and equidistant sampling times, the optimal sampling size is

$$
n^*_{\text{sparse}} = \left(\frac{T}{2}\frac{\sigma^2}{E\epsilon^2}\right)^{2/3}. \tag{32}
$$

Plug (32) into (30) and (31) for the choice of n, then

$$
\begin{aligned}
&cum_3([Y,Y]_T^{(sparse,opt)} - \langle X,X\rangle_T) \\
&\qquad = 48(\sigma^2 T)^{\frac{4}{3}} 2^{\frac{2}{3}} (E\epsilon^2)^{\frac{5}{3}} + 8(\sigma^2 T)^{\frac{5}{3}}(2E\epsilon^2)^{\frac{4}{3}} + O(E|\epsilon|^{\frac{11}{3}})
\end{aligned}
\tag{33}
$$

and

$$
\begin{aligned}
&cum_4([Y,Y]_T^{(sparse,opt)} - \langle X,X\rangle_T) \\
&\qquad = -192(\sigma^2 T)^{\frac{5}{3}} E\epsilon^4 \left(\frac{E\epsilon^2}{4}\right)^{\frac{1}{3}} + 384(E\epsilon^4 + Var(\epsilon)^2)(\sigma^2 T)^{\frac{4}{3}}(2E\epsilon^2)^{\frac{2}{3}} \\
&\qquad + 384(\sigma^2 T)^{\frac{5}{3}} 2^{\frac{4}{3}} (E\epsilon^2)^{\frac{7}{3}} + 54(\sigma^2 T)^2 (2E\epsilon^2)^2 + O(E|\epsilon|^{\frac{17}{3}})
\end{aligned}
\tag{34}
$$

respectively.

Note that under the optimal sampling,

$$
\begin{aligned}
Var([Y,Y]_T^{(sparse,opt)}) &= E\left(Var([Y,Y]_T^{(sparse,opt)} \mid X)\right) + Var\left(E([Y,Y]_T^{(sparse,opt)} \mid X)\right) \\
&= 8 < X,X >_T E\epsilon^2 + \frac{2}{n}(\sigma^2 T)^2 + o_p(E\epsilon^2) \\
&= 2(\sigma^2 T)^{\frac{4}{3}}(2E\epsilon^2)^{\frac{2}{3}} + O_p(E\epsilon^2),
\end{aligned}
$$

hence,

$$\begin{aligned} cum_3\left((E\epsilon^2)^{-1/3}([Y,Y]_T^{(sparse,opt)} - \langle X,X\rangle_T)\right) &= O((E|\epsilon|)^{2/3}) \\ cum_4\left((E\epsilon^2)^{-1/3}([Y,Y]_T^{(sparse,opt)} - \langle X,X\rangle_T)\right) &= O((E|\epsilon|)^{4/3}) \end{aligned}$$

in other words, the third-order and the fourth-order cumulants indeed vanish as $n \to \infty$ and $E\epsilon^2 \to 0$.

4.2.2 unconditional cumulants for averaging estimator

Similarly, for the averaging estimators,

$$E([Y,Y]_T^{(avg)} \mid X \text{ process}) = [X,X]_T^{(avg)} + 2\bar{n}E\epsilon^2, \tag{35}$$

$$Var([Y,Y]_T^{(avg)}|X) = Var([\epsilon,\epsilon]_T^{(avg)}) + \frac{8}{K}[X,X]_T^{(avg)}E\epsilon^2 + O_p(E[|\epsilon|^2(nK)^{-1/2}]), \tag{36}$$

with

$$Var([\epsilon,\epsilon]_T^{(avg)}) = 4\frac{\bar{n}}{K}E\epsilon^4 - \frac{2}{K}Var(\epsilon^2).$$

Invoking the relations between the conditional and the unconditional cumulants, one gets the unconditional cumulants for the average estimator:

$$\begin{aligned} cum_3([Y,Y]_T^{(avg)} - \langle X,X\rangle_T) = {} & \frac{1}{K^2}c_3(\bar{n}) + 48\frac{1}{K^2}E(\epsilon^4)E[X,X]_T^{(avg)} \\ & + 24\frac{1}{K}Var(\epsilon)Cov([X,X]_T^{(avg)},[X,X]_T^{(avg)} - \langle X,X\rangle_T) \\ & + cum_3([X,X]_T^{(avg)} - \langle X,X\rangle_T) + O(K^{-2}\bar{n}^{-1/2}E[|\epsilon|^3]) \end{aligned} \tag{37}$$

and

$$\begin{aligned} cum_4([Y,Y]_T - \langle X,X\rangle_T) = {} & \frac{1}{K^3}c_4(\bar{n}) + 24\frac{1}{K^3}\frac{c_3(\bar{n})}{\bar{n}}E[X,X]_T^{(avg)} - 48\frac{1}{K^4}E[X,X]_T^{(avg)}E\epsilon^2 Var([\epsilon,\epsilon]_T^{(all)}) \\ & + 192\frac{1}{K^2}E\epsilon^4 Cov([X,X]_T^{(avg)},[X,X]_T^{(avg)} - \langle X,X\rangle_T) \\ & + 192\frac{1}{K^2}(Var(\epsilon))^2 Var([X,X]_T^{(avg)}) \\ & + 48\frac{1}{K}Var(\epsilon)cum_3([X,X]_T^{(avg)},[X,X]_T^{(avg)} - \langle X,X\rangle_T,[X,X]_T^{(avg)} - \langle X,X\rangle_T) \\ & + cum_4([X,X]_T^{(avg)} - \langle X,X\rangle_T) + O(K^{-3}\bar{n}^{-1/2}E[|\epsilon|^5]) \end{aligned} \tag{38}$$

In the special case where σ_t is constant, and $\Delta t_i = \Delta$, $[X,X]_T^{(avg)}$ has the same distribution as $[X,X]_T^{(all)}$, namely $\sigma^2\Delta t\chi_n^2$.

Thus,

$$cum_3([Y,Y]_T^{(avg)} - \langle X,X\rangle_T) = 24\frac{1}{K}Var(\epsilon)2(\bar{n}K)^{-1}(\sigma^2T)^2 + 8(\bar{n}K)^{-2}(\sigma^2T)^3 + O(K^{-2}\bar{n}^{-1/2}E[|\epsilon|^3]) \tag{39}$$

and

$$\begin{aligned} cum_4([Y,Y]_T^{(avg)} - \langle X,X\rangle_T) = & -48\frac{1}{K^4}(\sigma^2T)E\epsilon^2(4E\epsilon^4\bar{n}K) \\ & + 192\frac{1}{K^2}(E\epsilon^4 + Var(\epsilon))^2)2(\bar{n}K)^{-1}(\sigma^2T)^2 \\ & + 48\frac{1}{K}Var(\epsilon)8(\bar{n}K)^{-2}(\sigma^2T)^3 \\ & + 54(\bar{n}K)^{-3}(\sigma^2T)^4 + O(K^{-3}\bar{n}^{-1/2}E[|\epsilon|^5]) \end{aligned} \tag{40}$$

Also, the optimal average subsampling size for the constant σ is,

$$\bar{n}^* = \left(\frac{\sigma^4T^2}{6(E\epsilon^2)^2}\right)^{1/3}.$$

Also, the unconditional variance of the averaging estimator, under the optimal sampling,

$$\begin{aligned} Var([Y,Y]_T^{(avg,opt)}) &= \underbrace{\frac{8}{K}E\epsilon^2(\sigma^2T) + o(E\epsilon^2K^{-1})}_{E\left(Var([Y,Y]_T^{(avg,opt)} \mid X)\right)} + \underbrace{(\sigma^2T)^2 2(\bar{n}^*K)^{-1}}_{Var\left(E([Y,Y]_T^{(avg,opt)} \mid X)\right)} \\ &= \frac{2}{K}6^{\frac{1}{3}}(E\epsilon^2)^{\frac{2}{3}}(\sigma^2T)^{\frac{4}{3}} + O(E\epsilon^2K^{-1}) \end{aligned}$$

hence,

$$\begin{aligned} cum_3\left((E\epsilon^2)^{-1/3}K^{1/2}([Y,Y]_T^{(avg,opt)} - \langle X,X\rangle_T)\right) &= O((E|\epsilon|)^{2/3}K^{-1/2}) \to 0, \\ cum_4\left((E\epsilon^2)^{-1/3}K^{1/2}([Y,Y]_T^{(avg,opt)} - \langle X,X\rangle_T)\right) &= O((E|\epsilon|)^{4/3}K^{-1}) \to 0, \end{aligned}$$

as $n \to \infty$ and $E\epsilon^2 \to 0$.

By comparing to the expression for the sparse case, it is clear that the average volatility is substantially closer to normal that the sparsely sampled volatility.

5 Appendix: Proofs of $c_3(n)$ and $c_4(n)$

Lemma 1.

$$c_3(n) = 8\left[(n-\frac{3}{4})\ cum_3(\epsilon^2) - 7(n-\frac{6}{7})\ cum_3(\epsilon)^2 + 6(n-\frac{1}{2})\ var(\epsilon)var(\epsilon^2)\right]$$

□

Proof of Lemma 1:

Let a_i be defined as in (16).

One can then write

$$\begin{aligned} c_3(n) &= cum_3(2\sum_{i=0}^{n} a_i(\epsilon_{t_i}^2 - E\epsilon^2) - 2\sum_{i=0}^{n-1}\epsilon_{t_i}\epsilon_{t_{i+1}}), \\ &= 8[cum_3(\sum_{i=0}^{n} a_i\epsilon_{t_i}^2) - cum_3(\sum_{i=0}^{n-1}\epsilon_{t_i}\epsilon_{t_{i+1}}) - 3cum(\sum_{i=0}^{n} a_i\epsilon_{t_i}^2, \sum_{j=0}^{n} a_j\epsilon_{t_j}^2, \sum_{k=0}^{n-1}\epsilon_{t_k}\epsilon_{t_{k+1}}) \\ &\quad +3cum(\sum_{i=0}^{n} a_i\epsilon_{t_i}^2, \sum_{j=0}^{n-1}\epsilon_{t_j}\epsilon_{t_{j+1}}, \sum_{k=0}^{n-1}\epsilon_{t_k}\epsilon_{t_{k+1}})] \end{aligned} \tag{41}$$

where

$$cum(\sum_{i=0}^{n} a_i\epsilon_{t_i}^2, \sum_{j=0}^{n} a_j\epsilon_{t_j}^2, \sum_{k=0}^{n-1}\epsilon_{t_k}\epsilon_{t_{k+1}}) = 2\sum_{k=0}^{n-1} a_k a_{k+1} cum(\epsilon_{t_k}^2, \epsilon_{t_{k+1}}^2, \epsilon_{t_k}\epsilon_{t_{k+1}}) = 2(n-1)(E\epsilon^3)^2 \tag{42}$$

since $\sum_{k=0}^{n-1} a_k a_{k+1} = n-1$, and the summation is non-zero only when $(i=k, j=k+1)$ or $(i=k+1, j=k)$.

Also,

$$cum(\sum_{i=0}^{n} a_i\epsilon_{t_i}^2, \sum_{j=0}^{n-1}\epsilon_{t_j}\epsilon_{t_{j+1}}, \sum_{k=0}^{n-1}\epsilon_{t_k}\epsilon_{t_{k+1}}) = 2\sum_{j=0}^{n-1} a_j cum(\epsilon_{t_j}^2, \epsilon_{t_j}\epsilon_{t_{j+1}}, \epsilon_{t_j}\epsilon_{t_{j+1}}) = 2(n-\frac{1}{2})(E\epsilon^2)Var(\epsilon^2) \tag{43}$$

since $\sum_{j=0}^{n-1} a_j = n - \frac{1}{2}$, and the summation is non-zero only when $j = k = (i$, or $i-1)$.

And finally,

$$cum(\sum_{i=0}^{n-1}\epsilon_{t_i}\epsilon_{t_{i+1}}, \sum_{j=0}^{n-1}\epsilon_{t_j}\epsilon_{t_{j+1}}, \sum_{k=0}^{n-1}\epsilon_{t_k}\epsilon_{t_{k+1}}) = \sum_{i=0}^{n-1} cum_3(\epsilon_{t_i}\epsilon_{t_{i+1}}) = n(E\epsilon^3)^2, \tag{44}$$

$$cum(\sum_{i=0}^{n} a_i\epsilon_{t_i}^2, \sum_{j=0}^{n} a_j\epsilon_{t_j}^2, \sum_{k=0}^{n} a_k\epsilon_{t_k}^2) = \sum_{i=0}^{n} a_i^3 cum_3(\epsilon_{t_i}^2) = (n-\frac{3}{4})cum_3(\epsilon^2), \tag{45}$$

with $\sum_{i=0}^{n} a_i^3 = n - \frac{3}{4}$.

Inserting (42)-(45) in (41) yields (14). ∎

Lemma 2.

$$\begin{aligned} c_4(n) &= 16\{(n-\frac{7}{8})cum_4(\epsilon^2) + n(E\epsilon^4)^2 - 3n(E\epsilon^2)^4 + 12(n-1)var(\epsilon^2)E\epsilon^4 \\ &\quad -32(n-\frac{17}{16})E\epsilon^3 cov(\epsilon^2, \epsilon^3) + 24(n-\frac{7}{4})E\epsilon^2(E\epsilon^3)^2 + 12(n-\frac{3}{4})cum_3(\epsilon^2)E\epsilon^2\} \end{aligned}$$

□

Proof of Lemma 2:

$$
\begin{aligned}
& cum(\sum_{i=0}^{n} a_i\epsilon_{t_i}^2, \sum_{j=0}^{n-1}\epsilon_{t_j}\epsilon_{t_{j+1}}, \sum_{k=0}^{n-1}\epsilon_{t_k}\epsilon_{t_{k+1}}, \sum_{l=0}^{n-1}\epsilon_{t_l}\epsilon_{t_{l+1}}) \\
= & \sum_{i=0}^{n}\sum_{j=k=0}^{n-1} a_i[1_{\{l=j\}}1_{\{i=j,\text{or } j+1\}} + \binom{3}{2}(1_{\{l=j+1,i=j+2\}} + 1_{\{l=i=j-1\}})]cum(\epsilon_{t_i}^2, \epsilon_{t_j}\epsilon_{t_{j+1}}, \epsilon_{t_k}\epsilon_{t_{k+1}}, \epsilon_{t_l}\epsilon_{t_{l+1}}) \\
= & 2(n-\frac{1}{2})E\epsilon^3 cov(\epsilon^2,\epsilon^3) + 6(n-\frac{3}{2})(E\epsilon^3)^2 E\epsilon^2
\end{aligned} \tag{46}
$$

$$
\begin{aligned}
& cum(\sum_{i=0}^{n} a_i\epsilon_{t_i}^2, \sum_{j=0}^{n} a_j\epsilon_{t_j}^2, \sum_{k=0}^{n-1}\epsilon_{t_k}\epsilon_{t_{k+1}}, \sum_{l=0}^{n-1}\epsilon_{t_l}\epsilon_{t_{l+1}}) \\
= & \sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n-1}\sum_{l=0}^{n-1} a_i a_j[1_{\{i=j,k=l,i=(k+1 \text{ or } k)\}} + (1_{\{l=k-1,(i,j)=(k+1,k-1)[2]\}} + 1_{\{l=k+1,(i,j)=(k,k+2)[2]\}}) \\
& +1_{\{k=l,(i,j)=(k,k+1)[2]\}} \\
= & 2(n-\frac{3}{4})cum_3(\epsilon^2)E\epsilon^2 + 4(n-2)(E\epsilon^3)^2 E\epsilon^2 + 2(n-1)(Var(\epsilon^2))^2
\end{aligned} \tag{47}
$$

where the notation $(i,j) = (k+1,k-1)[2]$ means that $(i = k+1, j = k-1)$, or $(j = k+1, i = k-1)$. The last equation above holds because $\sum_{i=1}^{n} a_i^2 = n - \frac{3}{4}$, $\sum_{i=1}^{n-1} a_{i-1}a_{i+1} = n-2$, and $\sum_{i=0}^{n-1} a_i a_{i+1} = n-1$.

$$
\begin{aligned}
& cum(\sum_{i=0}^{n} a_i\epsilon_{t_i}^2, \sum_{j=0}^{n} a_j\epsilon_{t_j}^2, \sum_{k=0}^{n} a_k\epsilon_{t_k}^2, \sum_{l=0}^{n-1}\epsilon_{t_l}\epsilon_{t_{l+1}}) \\
= & \sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n-1}\sum_{l=0}^{n-1} a_i a_j a_k \binom{3}{2}[1_{\{i=j=l,k=l+1\}} + 1_{\{i=j=l+1,k=l\}}]cum(\epsilon_{t_i}^2, \epsilon_{t_j}^2, \epsilon_{t_k}^2, \epsilon_{t_l}\epsilon_{t_{l+1}}) \\
= & 6\sum_{i=0}^{n-1} a_i^2 a_{i+1} cum(\epsilon^2,\epsilon^2,\epsilon)E\epsilon^3 \\
= & 6(n-\frac{5}{4})cum(\epsilon^2,\epsilon^2,\epsilon)E\epsilon^3,
\end{aligned} \tag{48}
$$

since $\sum_{i=0}^{n-1} a_i^2 a_{i+1} = n - \frac{5}{4}$.

$$
\begin{aligned}
& cum_4(\sum_{i=0}^{n-1} \epsilon_{t_i}\epsilon_{t_{i+1}}) \\
= \; & \sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\sum_{k=0}^{n-1}\sum_{l=0}^{n-1}[1_{\{i=j=k=l\}} + \binom{4}{2} 1_{\{i=j,k=l,i=(k+1,k-1)\}}]cum(\epsilon_{t_i}\epsilon_{t_{i+1}}, \epsilon_{t_j}\epsilon_{t_{j+1}}, \epsilon_{t_k}\epsilon_{t_{k+1}}, \epsilon_{t_l}\epsilon_{t_{l+1}}) \\
= \; & n((E\epsilon^4)^2 - 3(E\epsilon^2)^4) + 12(n-1)(E\epsilon^2)^2 Var(\epsilon^2) \qquad (49)
\end{aligned}
$$

$$
cum_4(\sum_{i=0}^{n} a_i \epsilon_{t_i}^2) = \sum_{i=0}^{n} a_i^4 cum_4(\epsilon^2) = (n - \frac{7}{8}) cum_4(\epsilon^2) \qquad (50)
$$

Putting together (46)-(50):

$$
\begin{aligned}
c_4(n) \; = \; & cum_4(2\sum_{i=0}^{n} a_i\epsilon_{t_i}^2 - 2\sum_{i=0}^{n-1}\epsilon_{t_i}\epsilon_{t_{i+1}}), \\
= \; & 16[cum_4(\sum_{i=0}^{n} a_i\epsilon_{t_i}^2) + cum_4(\sum_{i=0}^{n-1}\epsilon_{t_i}\epsilon_{t_{i+1}}) - \binom{4}{1} cum(\sum_{i=0}^{n} a_i\epsilon_{t_i}^2, \sum_{j=0}^{n} a_j\epsilon_{t_j}^2, \sum_{k=0}^{n} a_k\epsilon_{t_k}^2, \sum_{l=0}^{n-1}\epsilon_{t_l}\epsilon_{t_{l+1}}) \\
& -\binom{4}{3} cum(\sum_{i=0}^{n} a_i\epsilon_{t_i}^2, \sum_{j=0}^{n-1}\epsilon_{t_j}\epsilon_{t_{j+1}}, \sum_{k=0}^{n-1}\epsilon_{t_k}\epsilon_{t_{k+1}}, \sum_{l=0}^{n-1}\epsilon_{t_l}\epsilon_{t_{l+1}}) \\
& +\binom{4}{2} cum(\sum_{i=0}^{n} a_i\epsilon_{t_i}^2, \sum_{j=0}^{n} a_j\epsilon_{t_j}^2, \sum_{k=0}^{n-1}\epsilon_{t_k}\epsilon_{t_{k+1}}, \sum_{l=0}^{n-1}\epsilon_{t_l}\epsilon_{t_{l+1}})] \\
\overset{(46)-(50)}{=} \; & 16\{(n - \frac{7}{8})cum_4(\epsilon^2) + n(E\epsilon^4)^2 - 3n(E\epsilon^2)^4 + 12(n-1)var(\epsilon^2)E\epsilon^4 \\
& -32(n - \frac{17}{16})E\epsilon^3 cov(\epsilon^2, \epsilon^3) + 24(n - \frac{7}{4})E\epsilon^2(E\epsilon^3)^2 + 12(n - \frac{3}{4})cum_3(\epsilon^2)E\epsilon^2\} \qquad (51)
\end{aligned}
$$

since $cov(\epsilon^2, \epsilon^3) = E\epsilon^5 - E\epsilon^2 E\epsilon^3$ and $cum(\epsilon^2, \epsilon^2, \epsilon) = E\epsilon^5 - 2E\epsilon^2 E\epsilon^3$.

∎

REFERENCES

Bickel, P. J., Götze, F., and van Zwet, W. R. (1986), "The Edgeworth Expansion for U-statistics of Degree Two," *The Annals of Statistics*, 14, 1463–1484.

Brillinger, D. R. (1969), "The Calculation of Cumulants via Conditioning," *Annals of the Institute of Statistical Mathematics*, 21, 215–218.

Goncalves, S. and Meddahi, N. (2005), "Bootstrapping realized volatility," Tech. rep., Université de Montréal.

Hall, P. (1992), *The bootstrap and Edgeworth expansion*, New York: Springer.

McCullagh, P. (1987), *Tensor Methods in Statistics*, London, U.K.: Chapman and Hall.

Mykland, P. A. (1993), "Asymptotic Expansions for Martingales," *Annals of Probability*, 21, 800–818.

— (1995a), "Embedding and Asymptotic Expansions for Martingales," *Probability Theory and Related Fields*, 103, 475–492.

— (1995b), "Martingale Expansions and Second Order Inference," *Annals of Statistics*, 23, 707–731.

Speed, T. P. (1983), "Cumulants and Partition Lattices," *The Australian Journal of Statistics*, 25, 378–388.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2003), "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," Tech. rep., Carnegie-Mellon University.