

Edgeworth Expansions for Realized Volatility and Related Estimators*

Lan Zhang
University of Illinois at Chicago

Per A. Mykland
The University of Chicago

Yacine Aït-Sahalia
Princeton University and NBER

This Version: January 12, 2009 .

Abstract

This paper shows that the asymptotic normal approximation is often insufficiently accurate for volatility estimators based on high frequency data. To remedy this, we derive Edgeworth expansions for such estimators. The expansions are developed in the framework of small-noise asymptotics. The results have application to Cornish-Fisher inversion and help setting intervals more accurately than those relying on normal distribution.

KEYWORDS: Bias-correction; Edgeworth expansion; Market microstructure; Martingale; Realized volatility; Two Scales Realized Volatility.

JEL CODES: C13; C14; C15; C22

*Financial support from the NSF under grants SBR-0350772 (Aït-Sahalia) and DMS-0204639, DMS 06-04758, and SES 06-31605 (Mykland and Zhang) is gratefully acknowledged.

1. Introduction

Volatility estimation from high frequency data has received substantial attention in the recent literature.¹ A phenomenon which has been gradually recognized, however, is that the standard estimator, realized volatility or realized variance (RV, hereafter), can be unreliable if the microstructure noise in the data is not explicitly taken into account. Market microstructure effects are surprisingly prevalent in high frequency financial data. As the sampling frequency increases, the noise becomes progressively more dominant, and in the limit swamps the signal. Empirically, sampling a typical stock price every few seconds can lead to volatility estimates that overestimate the true volatility by a factor of two or more. As a result, the usual prescription in the literature is to sample sparsely, with the recommendations ranging from five to thirty minutes, even if the data are available at much higher frequencies.

More recently various RV-type estimators have been proposed to take into account of the market microstructure impact. For example, in the parametric setting, Aït-Sahalia et al. (2005a) proposed likelihood corrections for volatility estimation; in the nonparametric context, Zhang et al. (2005) proposed five different RV-like estimation strategies, culminating with a consistent estimator based on combining two time scales, which we called TSRV (two scale realized volatility). A natural generalization of TSRV, based on multiple time scales, can improve the estimator's efficiency (Zhang (2006)). It is also possible to deal with more complex noise structures (see Aït-Sahalia et al. (2005b), Barndorff-Nielsen et al. (2008)).²

One thing in common among various RV-type estimators is that the limit theory predicts that these estimators should be asymptotically mixed normal. Without noise, the asymptotic normality of RV estimates dates back to at least Jacod (1994) and Jacod and Protter (1998). When microstructure noise is present, the asymptotic normality of the standard RV estimator (as well as that of the subsequent refinements that are robust to the presence of microstructure noise, such as TSRV) was established in Zhang et al. (2005).

However, simulation and/or empirical results do not always agree well with what the asymptotic theory predicts. As we shall see in Section 5, the distributions of various RV type estimators are not

¹See, e.g., Dacorogna et al. (2001), Andersen et al. (2001b), Zhang (2001), Barndorff-Nielsen and Shephard (2002), Meddahi (2002), Oomen (2002) and Mykland and Zhang (2006)

²Other strategies have been proposed to deal with the microstructure issue in RV estimation. Zhou (1996) proposed adjusting the usual RV estimator by adding one or more lagged correction terms; this approach was further investigated by Hansen and Lunde (2004); Bandi and Russell (2003) advocated using an optimally sampled sparse data set, as in one of the intermediary estimators (“third best”) in Zhang et al. (2005). None of these estimators are consistent, however.

normal, in particular, they are skewed and heavy-tailed. This phenomenon of non-normal (unconditional) distribution of RV estimators is somewhat different from those documented in the literature. For example, using a simple RV estimator, *i.e.*, the sum of squared high frequency returns, Zumbach et al. (1999) and Andersen et al. (2001b) found that the finite-sample (unconditional) distributions of foreign exchange RV are approximately lognormal. A non-normal distribution for the simple RV was also reported when using the 30 Dow Jones Industrial Average stocks data (Andersen et al. (2001a)) and in simulation experiments (Barndorff-Nielsen and Shephard (2005)). In these early studies, the distribution of simple RV approaches normality as sample size increases. In comparison, the current paper shows that this lack of normality phenomenon extends to other RV-type estimators, including those having accounted for microstructure noise. Furthermore, it persists even with fairly large sample sizes!

We argue that the lack of normality can be caused by the coexistence of a small effective sample size and small noise. As a first-order remedy, we derive Edgeworth expansions for the RV-type estimators when the observations of the price process are noisy. What makes the situation unusual is that the errors (noises) 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 develop expansions under the hypothesis that the size of $|\epsilon|$ goes to zero, as stated precisely at the beginning of Section 4. We will document that this approach predicts the small sample behavior of the estimators better than the approach where $|\epsilon|$ is of fixed size. In this sense, we are dealing with an unusual type of Edgeworth expansion.

One can argue that it is counterfactual to let the size of ϵ go to zero as the number of observations go to infinity. We should emphasize that we do not literally mean that the noise goes down the more observations one gets. The form of asymptotics is merely a device to generate appropriate accurate distributions. Another problem where this type of device is used is for ARMA processes with nearly unit root (see, e.g. Chan and Wei (1987)), or the local-to-unity paradigm. Nobody actually thinks that the root gets closer to unity as the sample size grows, but the device nonetheless produces useful asymptotic distributions. In our setting, the assumption that the size of ϵ goes down has produced useful results in (Sections 2 and 3) of Zhang et al. (2005). For the problem discussed there, shrinking ϵ is the only known way of discussing bias-variance tradeoff rigorously in the presence of a leverage effect. Note that a similar use of triangular array asymptotics has been used by Delattre and Jacod (1997) in the context of rounding, and by Gloter and Jacod (2000) in the context of additive error. Another interpretation is that of small-sigma asymptotics, cf. the discussion in Section 4.1 below.

It is worth mentioning that jumps are not the likely causes leading to the non-normality in RV distributions in Section 5, as we model both the underlying returns and the volatility as continuous processes. Also, it is important to note that our analysis focuses on normalized RV-type estimators, rather than studentized RV which has more immediate implementation in practice. In other words, our Edgeworth expansion has the limitation of conditioning on volatility processes, while hopefully it sheds some light on how an Edgeworth correction can be done for RV-type estimators while allowing for the presence of microstructure noise. For an Edgeworth expansion applicable to the studentized (basic) RV estimator when there is no noise, one can consult Gonçalves and Meddahi (2005). Their expansion is used for assessing the accuracy of the bootstrap in comparison to the first order asymptotic approach. See also Gonçalves and Meddahi (2008). Edgeworth expansions for realized volatility are also developed by Lieberman and Phillips (2006) for inference on long memory parameters.

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. Since our expansions hold in a triangular array setting, they can also be used to analyze the behavior of bootstrapping distributions. A nice side result in our development, which may be of use in other contexts, shows how to calculate the third and fourth cumulants of integrals of Gaussian processes with respect to Brownian motion. This can be found in Proposition 4.

The paper is organized as follows. In Section 2, we briefly recall the estimators under consideration. Section 3 gives their first order asymptotic properties, and reports initial simulation results which show that the normal asymptotic distribution can be unsatisfactory. So, in Section 4, we develop Edgeworth expansions. In Section 5, we examine the behavior of our small-sample Edgeworth corrections in simulations. Section 6 concludes. Proofs are in the Appendix.

2. Data Structure and Estimators

Let $\{Y_{t_i}\}$, $0 = t_0 \leq t_1 \leq \dots \leq t_n = T$, be the observed (log) price of a security at time $t_i \in [0, T]$. The basic modelling assumption we make is that these observed prices can be decomposed into an underlying (log) price process X (the signal) and a noise term ϵ , which captures a variety of phenomena collectively known as market microstructure noise. That is, at each observation time t_i , we have

$$Y_{t_i} = X_{t_i} + \epsilon_{t_i}. \tag{2.1}$$

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

$$dX_t = \mu_t dt + \sigma_t dB_t, \quad (2.2)$$

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. We do not, in general, assume that the volatility process, when stochastic, is orthogonal to the Brownian motion driving the price process. We will, however, make this assumption in Section 4.3.

Let the noise ϵ_{t_i} in (2.1) satisfy the following assumption,

$$\epsilon_{t_i} \text{ i.i.d. with } E(\epsilon_{t_i}) = 0, \text{ and } \text{Var}(\epsilon_{t_i}) = E\epsilon^2. \text{ Also } \epsilon \perp\!\!\!\perp X \text{ process,} \quad (2.3)$$

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 (2.1) does not require that ϵ_t exists for every t . We are interested in estimating

$$\langle X, X \rangle_T = \int_0^T \sigma_t^2 dt, \quad (2.4)$$

the integrated volatility or quadratic variation of the true price process X , assuming model (2.1), and assuming that Y_{t_i} 's can be observed at high frequency. In particular, we focus on estimators that are nonparametric in nature, and as we will see, are extensions of the realized volatility or realized variance (RV) estimators. These estimators require nothing more than summing up squared returns sampled at some frequency within a fixed time period, and taking different linear combinations of such sums.

Following Zhang et al. (2005), we consider five RV-type estimators. Ranked from the statistically least desirable to the most desirable, we start with (1) the “all” estimator $[Y, Y]^{(\text{all})}$, where RV is based on the entire sample and consecutive returns are used; (2) the sparse estimator $[Y, Y]^{(\text{sparse})}$, where the RV is based on a sparsely sampled returns series. Its sampling frequency is often arbitrary or selected in an ad hoc fashion; (3) the optimal, sparse estimator $[Y, Y]^{(\text{sparse, opt})}$, which is similar to $[Y, Y]^{(\text{sparse})}$ except that the sampling frequency is pre-determined to be optimal in the sense of minimizing root mean squared error (MSE); (4) the averaging estimator $[Y, Y]^{(\text{avg})}$, which is constructed by averaging the sparse estimators and thus also utilizes the entire sample, and finally (5) two scales estimator (TSRV) $\widehat{\langle X, X \rangle}$, which combines the RV estimators from two time scales, $[Y, Y]^{(\text{avg})}$ and $[Y, Y]^{(\text{all})}$, using the latter as a means to bias-correct the former. We showed that the combination of two time scales results in a consistent estimator. TSRV is the first estimator proposed in the literature to have this property. The first four estimators are biased; the magnitude of their bias is typically proportional to the sampling frequency.

Specifically, our estimators have the following form. First, $[Y, Y]_T^{(\text{all})}$ uses all the observations

$$[Y, Y]_T^{(\text{all})} = \sum_{t_i \in \mathcal{G}} (Y_{t_{i+1}} - Y_{t_i})^2, \quad (2.5)$$

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

The sparse estimator uses a subsample of the data,

$$[Y, Y]_T^{(\text{sparse})} = \sum_{t_j, t_{j,+} \in \mathcal{H}} (Y_{t_{j,+}} - Y_{t_j})^2, \quad (2.6)$$

where \mathcal{H} is a strict subset of \mathcal{G} , with sample size n_{sparse} , $n_{\text{sparse}} < n$. And, if $t_i \in \mathcal{H}$, then $t_{i,+}$ denotes the following elements in \mathcal{H} . The optimal sparse estimator $[Y, Y]^{(\text{sparse, opt})}$ has the same form as in (2.6) except replacing n_{sparse} with n_{sparse}^* , where n_{sparse}^* is determined by minimizing MSE of the estimator (an explicit formula for doing so is given in Zhang et al. (2005)).

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

$$[Y, Y]_T^{(\text{avg})} = \frac{1}{K} \sum_{k=1}^K \underbrace{\sum_{t_j, t_{j,+} \in \mathcal{G}^{(k)}} (Y_{t_{j,+}} - Y_{t_j})^2}_{=[Y, Y]_T^{(k)}}, \quad (2.7)$$

where $\mathcal{G}^{(k)}$'s are disjoint subsets of the full set of observation times with union \mathcal{G} . Let n_k be the number of time points in \mathcal{G}_k and $\bar{n} = K^{-1} \sum_{k=1}^K n_k$ the average sample size across different grids \mathcal{G}_k , $k = 1, \dots, K$. One can also consider the optimal, averaging estimator $[Y, Y]^{(\text{avg, opt})}$, 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 (see again Zhang et al. (2005) for an explicit formula.) A special case of (2.7) arises when the sampling points are regularly allocated:

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

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

The TSRV estimator has the form of

$$\widehat{\langle X, X \rangle}_T = (1 - \frac{\bar{n}}{n})^{-1} \left([Y, Y]_T^{(\text{avg})} - \frac{\bar{n}}{n} [Y, Y]_T^{(\text{all})} \right) \quad (2.8)$$

that is, the volatility estimator $\widehat{\langle X, X \rangle}_T$ combines the sum of squares estimators from two different time scales, $[Y, Y]_T^{(\text{avg})}$ from the returns on a slow time scale whereas $[Y, Y]_T^{(\text{all})}$ is computed from

the returns on a fast time scale. \bar{n} in (2.8) is the average sample size across different grids. Note that this is what is called the “adjusted” TSRV in Zhang et al. (2005).

In the model (2.1), the distributions of various estimators can be studied by decomposing the sum-of-squared returns $[Y, Y]$,

$$[Y, Y]_T = [X, X]_T + 2[X, \epsilon]_T + [\epsilon, \epsilon]_T. \quad (2.9)$$

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

3. Small Sample Accuracy of the Normal Asymptotic Distribution

We now briefly recall the distributional theory for each of these five estimators which we developed in Zhang et al. (2005); all five have asymptotically (mixed) normal distributions. As we will see, however, this asymptotic distribution is not particularly accurate in small samples.

3.1. Asymptotic Normality for the Sparse Estimators

For the sparse estimator, we have shown that

$$\begin{aligned}
 [Y, Y]_T^{(\text{sparse})} &\stackrel{\mathcal{L}}{\approx} \langle X, X \rangle_T + \underbrace{2n_{\text{sparse}}E\epsilon^2}_{\text{bias due to noise}} \quad (3.1) \\
 &+ \underbrace{\left[\underbrace{\text{Var}([\epsilon, \epsilon]_T^{(\text{sparse})}) + 8[X, X]_T^{(\text{sparse})}E\epsilon^2}_{\text{due to noise}} + \underbrace{\frac{2T}{n_{\text{sparse}}} \int_0^T \sigma_t^4 dt}_{\text{due to discretization}} \right]^{1/2} Z_{\text{total}}}_{\text{total variance}},
 \end{aligned}$$

where $\text{Var}([\epsilon, \epsilon]_T^{(\text{sparse})}) = 4n_{\text{sparse}}E\epsilon^4 - 2\text{Var}(\epsilon^2)$, and Z_{total} is standard normal term. The symbol “ $\stackrel{\mathcal{L}}{\approx}$ ” means that when suitably standardized, the two sides have the same limit in law.

If the sample size n_{sparse} is large relative to the noise, the variance due to noise in (3.1) would be dominated by $\text{Var}([\epsilon, \epsilon]_T^{(\text{sparse})})$ which is of order $n_{\text{sparse}}E\epsilon^4$. However, with the dual presence of small n_{sparse} and small noise (say, $E\epsilon^2$), $8[X, X]_T^{(\text{sparse})}E\epsilon^2$ is not necessarily smaller than $\text{Var}([\epsilon, \epsilon]_T^{(\text{sparse})})$. One then needs to add $8[X, X]_T^{(\text{sparse})}E\epsilon^2$ 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, cf. the discussion in the introduction to Zhang et al. (2005).

Of course, n_{sparse} is selected either arbitrarily or in some ad hoc manner. By contrast, the sampling frequency in the optimal-sparse estimator $[Y, Y]^{(\text{sparse, opt})}$ can be determined by minimizing the MSE of the estimator analytically. Distribution-wise, the optimal-sparse estimator has the same form as in (3.1), but, one replaces n_{sparse} by the optimal sampling frequency n_{sparse}^* given below in (4.11). No matter whether n_{sparse} is selected optimally or not, one can see from (3.1) that after suitably adjusted for the bias term, the sparse estimators are asymptotically normal.

3.2. Asymptotic Normality for the Averaging Estimator

The optimal-sparse estimator only uses a fraction n_{sparse}^*/n of the data; one also has to pick the beginning (or ending) point of the sample. The averaging estimator overcomes both shortcomings. Based on the decomposition (2.9), we have

$$\begin{aligned}
 [Y, Y]_T^{(\text{avg})} &\stackrel{\mathcal{L}}{\approx} \langle X, X \rangle_T + \underbrace{2\bar{n}E\epsilon^2}_{\text{bias due to noise}} \\
 &+ \underbrace{\left[\text{Var}([\epsilon, \epsilon]_T^{(\text{avg})}) + \frac{8}{K} [X, X]_T^{(\text{avg})} E\epsilon^2 + \frac{4T}{3\bar{n}} \int_0^T \sigma_t^4 dt \right]}_{\text{total variance}}^{1/2} Z_{\text{total}},
 \end{aligned} \tag{3.2}$$

where

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

and Z_{total} is a standard normal term.

The distribution of the optimal averaging estimator $[Y, Y]^{(\text{avg, opt})}$ has the same form as in (3.2) except that we substitute \bar{n} with the optimal sub-sampling average size \bar{n}^* . To find \bar{n}^* , one determines K^* from the bias-variance trade-off in (3.2) and then set $\bar{n}^* \approx n/K^*$. If one removes the bias in either $[Y, Y]_T^{(\text{avg})}$ or $[Y, Y]_T^{(\text{avg, opt})}$, it follows from (3.2) that the next term is, again, asymptotically normal.

3.3. The Failure of Asymptotic Normality

In practice, things are, unfortunately, somewhat more complicated than the story that emerges from equations (3.1) and (3.2). The distributions of the sparse estimators and the averaging estimator can be, in fact, quite far from normal. We provide an illustration of this using simulations. The simulation design is described in Section 5.1 below, but here we give a preview to motivate our following theoretical development of small sample corrections to these asymptotic distributions.

Figure 1 reports the QQ plots of the standardized distribution of the five estimators before any Edgeworth correction is applied, as well as the histograms of the estimates. It is clear that the sparse, the sparse-optimal and the averaging estimators are not normally distributed, in particular, they are positively skewed and show some degree of leptokurtosis. On the other hand, the “all” estimator and the TSRV estimator appear to be normally distributed. The apparent normality of the “all” estimator is mainly due to the large sample size (one second sampling over 6.5 hours); it is thus fairly irrelevant to talk about its small-sample behavior.

Overall, we conclude from these QQ plots that the small-sample distribution of the TSRV estimator is close to normality, while the small-sample distribution of the other estimators departs from normality. As mentioned in Section 5.1, n is very large in this simulation.

It should be emphasized that bias is not the cause of the non-normality. Apart from TSRV, all the estimators have substantial bias. This bias, however, does not change the shape of the distribution of the estimator, it only changes where the distribution is centered.

4. Edgeworth Expansions for the Distribution of the Estimators

4.1. The Form of the Edgeworth Expansion in Terms of Cumulants

In situations where the normal approximation is only moderately accurate, improved accuracy can be obtained by appealing to Edgeworth expansions, as follows. Let θ be a quantity to be estimated, such as $\theta = \int_0^T \sigma_t^2 dt$, and let $\hat{\theta}_n$ be an estimator, say the sparse or average realized volatility, and suppose that α_n is a normalizing constant so that $T_n = \alpha_n(\hat{\theta}_n - \theta)$ is asymptotically normal. A better approximation to the density f_n of T_n can then be obtained through the Edgeworth expansion. Typically, second order expansions are sufficient to capture skewness and kurtosis, as follows:

$$f_n(x) = \frac{\phi(z)}{\text{Var}(T_n)^{1/2}} \left[1 + \frac{1}{6} \frac{\text{Cum}_3(T_n)}{\text{Var}(T_n)^{3/2}} h_3(z) + \frac{1}{24} \frac{\text{Cum}_4(T_n)}{\text{Var}(T_n)^2} h_4(z) + \frac{1}{72} \frac{\text{Cum}_3(T_n)^2}{\text{Var}(T_n)^3} h_6(z) + \dots \right] \quad (4.1)$$

where $\text{Cum}_i(T_n)$ is the i -th order cumulant of T_n , $z = (x - E(T_n))/\text{Var}(T_n)^{1/2}$, and where the Hermite polynomials h_i are given by $h_3(z) = z^3 - 3z$, $h_4(z) = z^4 - 6z^2 + 3$, $h_6(z) = z^6 - 15z^4 + 45z^2 - 15$. The neglected terms are typically of smaller order in n than the explicit terms. We shall refer to the explicit terms in (4.1) as the usual Edgeworth form. For broad discussions of Edgeworth expansions, and definitions of cumulants, see e.g., Chapter XVI of Feller (1971) and Chapter 5.3 of McCullagh (1987).

In some cases, Edgeworth expansions can only be found for distribution functions, in which case

the form is obtained by integrating equation (4.1) term by term. In either situation, the Edgeworth approximations can be turned into expansions for p -values, and to Cornish-Fisher expansions for critical values; see formula (5.2) below. For more detail, we refer the reader to, e.g., Hall (1992).

Let us now apply this to the problem at hand here. An Edgeworth expansion of the usual form, up to second order, can be found separately for each of the components in (2.9) 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 then be represented exactly as a triangular array of martingales. The remaining terms are also, to relevant order, martingales. Results deriving expansions for martingales can be found in Mykland (1993), Mykland (1995b) and Mykland (1995a). See also Bickel et al. (1986) for $n^{-1/2}([\epsilon, \epsilon]^{(all)} - 2nE\epsilon^2)$.

To implement the expansions, however, one needs the form of the first four cumulants of T_n . We assume that the “size” of the law of ϵ goes to zero, formally that ϵ for sample size n is of the form $\tau_n\zeta$, *i.e.*, $P_n(\epsilon/\tau_n \leq x) = P(\zeta \leq x)$, where the left hand probability is for sample size n , and the right hand probability is independent of n . Here, $E\zeta^8 < \infty$ and does not depend on sample size, and τ_n is nonrandom and goes to zero as $n \rightarrow \infty$. Note that under our assumption, $\text{Var}(\epsilon) = O(\tau_n^2)$, so the assumption is similar to the small-sigma asymptotics which goes back to Kadane (1971). Finally, while in our case this is a way of setting up asymptotics, there is empirical work on whether the noise decreases with n ; see, in particular, Awartani et al. (2006).

No matter what assumptions are made on the noise (on τ_n), one should not expect cumulants in (4.1) to have standard convergence rates. The typical situation for an asymptotically normal statistic T_n is that the p 'th cumulant, $p \geq 2$, is of order $O(n^{-(p-2)/2})$, see, for example, Chapters 2.3-2.4 of Hall (1992), along with Wallace (1958), Bhattacharya and Ghosh (1978), and the discussion in Mykland (2001) and the references therein. While the typical situation does remain in effect for realized volatility in the no-noise and no-leverage case (which is, after all, a matter simply of observations that are independent but non-identically distributed), the picture changes for more complex statistics. To see that non-standard rates can occur even in the absence of microstructure noise, consult (4.28)-(4.29) in Section 4.3.2 below.

An important question which arises in connection with Edgeworth expansions is the comparison of Cornish-Fisher inversion with bootstrapping. The latter has been developed in the no-noise case by Gonçalves and Meddahi (2005). A comparison of this type is beyond the scope of this paper, but is clearly called for.

4.2. Conditional Cumulants

We start by deriving explicit expressions for the conditional cumulants for $[Y, Y]$ and $[Y, Y]^{(avg)}$, given the latent process X . All the expressions we give below 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} . We use a similar notation for $[\epsilon, \epsilon]$ and for $[X, X]$.

4.2.1 Third-Order Conditional Cumulants

Denote

$$c_3(n) \triangleq \text{Cum}_3([\epsilon, \epsilon] - 2nE\epsilon^2), \quad (4.2)$$

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

Lemma 1.

$$c_3(n) = 8 \left[\left(n - \frac{3}{4}\right) \text{Cum}_3(\epsilon^2) - 7\left(n - \frac{6}{7}\right) \text{Cum}_3(\epsilon)^2 + 6\left(n - \frac{1}{2}\right) \text{Var}(\epsilon) \text{Var}(\epsilon^2) \right]$$

From that Lemma, it follows that

$$c_3(n) = O(nE\epsilon^6) \quad (4.3)$$

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

$$\text{Cum}_3\left(K([\epsilon, \epsilon]^{(avg)} - 2\bar{n}E\epsilon^2)\right) = \sum_{k=1}^K \text{Cum}_3([\epsilon, \epsilon]^{(k)} - 2n_k E\epsilon^2) = Kc_3(\bar{n}).$$

For the conditional third cumulant of $[Y, Y]$, we have

$$\begin{aligned} \text{Cum}_3([Y, Y]_T | X) &= \text{Cum}_3([\epsilon, \epsilon]_T + 2[X, \epsilon] | X) \\ &= \text{Cum}_3([\epsilon, \epsilon]_T) + 6 \text{Cum}([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, [X, \epsilon]_T | X) \\ &\quad + 12 \text{Cum}([\epsilon, \epsilon]_T, [X, \epsilon]_T, [X, \epsilon]_T | X) + 8 \text{Cum}_3([X, \epsilon]_T | X). \end{aligned} \quad (4.4)$$

From this, we obtain:

Proposition 1.

$$\text{Cum}_3([Y, Y]_T | X) = \text{Cum}_3([\epsilon, \epsilon]_T) + 48[X, X]E\epsilon^4 + O_p(n^{-1/2}E[|\epsilon|^3]),$$

where $\text{Cum}_3([\epsilon, \epsilon]_T)$ is given in (4.3). Also

$$\text{Cum}_3(K[Y, Y]_T^{(avg)} | X) = \text{Cum}_3(K[\epsilon, \epsilon]_T^{(avg)}) + 48K[X, X]_T^{(avg)}E\epsilon^4 + O_p(K\bar{n}^{-1/2}E[|\epsilon|^3]).$$

4.2.2 Fourth-Order Conditional Cumulants

For the fourth-order cumulant, denote

$$c_4(n) \triangleq \text{Cum}_4([\epsilon, \epsilon]^{(all)} - 2nE\epsilon^2). \quad (4.5)$$

We have that:

Lemma 2.

$$\begin{aligned} c_4(n) = & 16\left\{ \left(n - \frac{7}{8}\right) \text{Cum}_4(\epsilon^2) + n(E\epsilon^4)^2 - 3n(E\epsilon^2)^4 + 12(n-1)\text{Var}(\epsilon^2)E\epsilon^4 \right. \\ & \left. - 32\left(n - \frac{17}{16}\right)E\epsilon^3 \text{Cov}(\epsilon^2, \epsilon^3) + 24\left(n - \frac{7}{4}\right)E\epsilon^2(E\epsilon^3)^2 + 12\left(n - \frac{3}{4}\right) \text{Cum}_3(\epsilon^2)E\epsilon^2 \right\} \end{aligned}$$

Also here,

$$\text{Cum}_4\left(K([\epsilon, \epsilon]^{(avg)} - 2\bar{n}E\epsilon^2)\right) = \sum_{k=1}^K \text{Cum}_4([\epsilon, \epsilon]^{(k)} - 2n_k E\epsilon^2) = Kc_4(\bar{n}).$$

For the conditional fourth-order cumulant, we know that

$$\begin{aligned} \text{Cum}_4([Y, Y]|X) = & \text{Cum}_4([\epsilon, \epsilon]_T) + 24 \text{Cum}([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, [X, \epsilon]_T, [X, \epsilon]_T|X) \\ & + 8 \text{Cum}([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, [X, \epsilon]_T|X) \\ & + 32 \text{Cum}([\epsilon, \epsilon]_T, [X, \epsilon]_T, [X, \epsilon]_T, [X, \epsilon]_T|X) + 16 \text{Cum}_4([X, \epsilon]|X). \end{aligned} \quad (4.6)$$

Similar argument as in deriving the third cumulant shows that the latter three terms in the right hand side of (4.6) are of order $O_p(n^{-1/2}E[|\epsilon|^5])$. Gathering terms of the appropriate order, we obtain:

Proposition 2.

$$\text{Cum}_4([Y, Y]|X) = \text{Cum}_4([\epsilon, \epsilon]_T) + 24[X, X]_T n^{-1} \text{Cum}_3([\epsilon, \epsilon]_T) + O_p(n^{-1/2}E[|\epsilon|^5])$$

Also, for the average estimator,

$$\text{Cum}_4(K[Y, Y]^{(avg)}|X) = \text{Cum}_4(K[\epsilon, \epsilon]_T^{(avg)}) + 24K[X, X]_T^{(avg)} \frac{c_3(\bar{n})}{\bar{n}} + O_p(K\bar{n}^{-1/2}E[|\epsilon|^5])$$

4.3. Unconditional Cumulants

To get the Edgeworth expansion form as in (4.1), we need unconditional cumulants for the estimator. To pass from conditional to unconditional third cumulants, we will use general formulae for this

purpose (see Brillinger (1969), Speed (1983), and also Chapter 2 in McCullagh (1987)):

$$\begin{aligned}\text{Cum}_3(A) &= E[\text{Cum}_3(A|\mathcal{F})] + 3 \text{Cov}[\text{Var}(A|\mathcal{F}), E(A|\mathcal{F})] + \text{Cum}_3[E(A|\mathcal{F})] \\ \text{Cum}_4(A) &= E[\text{Cum}_4(A|\mathcal{F})] + 4 \text{Cov}[\text{Cum}_3(A|\mathcal{F}), E(A|\mathcal{F})] + 3 \text{Var}[\text{Var}(A|\mathcal{F})] \\ &\quad + 6 \text{Cum}_3(\text{Var}(A|\mathcal{F}), E(A|\mathcal{F}), E(A|\mathcal{F})) + \text{Cum}_4(E(A|\mathcal{F})).\end{aligned}$$

In what follows, we apply these formulae to derive the unconditional cumulants for our estimators. The main development (after formula 4.8) will be for the case where there is no leverage effect. It should be noted that there are (other) cases, such as Bollerslev and Zhou (2002) and Corradi and Distaso (2006), involving leverage effect where (non-mixed) asymptotic normality holds. In such case, unconditional Edgeworth expansions may also be applicable.

4.3.1 Unconditional Cumulants for Sparse Estimators

In Zhang et al. (2005), we showed that

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

and also that

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

This allows us to obtain the unconditional cumulants as:

$$\begin{aligned}\text{Cum}_3([Y, Y]_T - \langle X, X \rangle_T) &= c_3(n) + 48E(\epsilon^4)E[X, X] \\ &\quad + 24 \text{Var}(\epsilon) \text{Cov}([X, X]_T, [X, X]_T - \langle X, X \rangle_T) \\ &\quad + \text{Cum}_3([X, X]_T - \langle X, X \rangle_T) + O(n^{-1/2}E[|\epsilon|^3])\end{aligned}\tag{4.7}$$

and

$$\begin{aligned}\text{Cum}_4([Y, Y]_T - \langle X, X \rangle_T) &= c_4(n) + 24\frac{1}{n}c_3(n)E[X, X]_T \\ &\quad + 192E\epsilon^4 \text{Cov}([X, X]_T, [X, X]_T - \langle X, X \rangle_T) \\ &\quad + 192(\text{Var}(\epsilon))^2 \text{Var}([X, X]_T) \\ &\quad + 48 \text{Var}(\epsilon) \text{Cum}_3([X, X]_T, [X, X]_T - \langle X, X \rangle_T, [X, X]_T - \langle X, X \rangle_T) \\ &\quad + \text{Cum}_4([X, X]_T - \langle X, X \rangle_T) + O(n^{-1/2}E[|\epsilon|^5])\end{aligned}\tag{4.8}$$

To calculate cumulants of $[X, X]_T - \langle X, X \rangle_T$, consider now the case where there is no leverage effect. For example, one can take σ_t to be conditionally nonrandom. Then

$$[X, X]_T = \sum_{i=1}^n \chi_{1,i}^2 \int_{t_{i-1}}^{t_i} \sigma_t^2 dt,$$

where the $\chi_{1,i}^2$ are i.i.d. χ_1^2 random variables. Hence, with implicit conditioning,

$$\text{Cum}_p([X, X]_T) = \text{Cum}_p(\chi_1^2) \sum_{i=1}^n \left(\int_{t_{i-1}}^{t_i} \sigma_t^2 dt \right)^p$$

The cumulants of the χ_1^2 distribution are as follows:

	$p = 1$	$p = 2$	$p = 3$	$p = 4$
$\text{Cum}_p(\chi_1^2)$	1	2	8	54

When the sampling points are equidistant, one then obtains the approximation

$$\text{Cum}_p([X, X]_T) = \text{Cum}_p(\chi_1^2) \left(\frac{T}{n} \right)^{p-1} \int_0^T \sigma_t^{2p} dt + O(n^{\frac{1}{2}-p})$$

under the assumption that σ_t^2 is an Itô process (often called a Brownian semimartingale). Hence, we have:

Proposition 3. *In the case where there is no leverage effect, conditionally on the path of σ_t^2 ,*

$$\begin{aligned} \text{Cum}_3([Y, Y]_T - \langle X, X \rangle_T) &= c_3(n) + 48E(\epsilon^4) \int_0^T \sigma_t^2 dt \\ &+ 48 \text{Var}(\epsilon) n^{-1} T \int_0^T \sigma_t^4 dt + 8n^{-2} T^2 \int_0^T \sigma_t^6 dt \\ &+ O(n^{-3/2} E[\epsilon^2]) + O(n^{-1/2} E[|\epsilon|^3]) + O(n^{-5/2}) \end{aligned} \quad (4.9)$$

Similarly for the fourth cumulant

$$\begin{aligned} \text{Cum}_4([Y, Y]_T - \langle X, X \rangle_T) &= c_4(n) + 24n^{-1} c_3(n) \int_0^T \sigma_t^2 dt \\ &+ 384(E\epsilon^4 + \text{Var}(\epsilon)^2) n^{-1} T \int_0^T \sigma_t^4 dt \\ &+ 384 \text{Var}(\epsilon) n^{-2} T^2 \int_0^T \sigma_t^6 dt + 54n^{-3} T^3 \int_0^T \sigma_t^8 dt \\ &+ O(n^{-1/2} E[|\epsilon|^5]) + O(n^{-3/2} E[\epsilon^4]) + O(n^{-5/2} E[\epsilon^2]) + O(n^{-7/2}) \end{aligned} \quad (4.10)$$

If one chooses $\epsilon = o_p(n^{-1/2})$ (i.e., $\tau_n = o(n^{-1/2})$), then all the explicit terms in (4.9) and (4.10) are non-negligible. In this case, the error term in equation (4.9) is of order $O(n^{-1/2}E[|\epsilon|^3]) + O(n^{-5/2})$, while that in equation (4.10) is of order $O(n^{-1/2}E[|\epsilon|^5]) + O(n^{-7/2})$. In the case of optimal-sparse estimator, it is shown in Zhang et al. (2005) (Section 2.3) that the optimal sampling frequency leads to $\epsilon = O_p(n^{-3/4})$, in particular $\epsilon = o_p(n^{-1/2})$.

For the special case of equidistant sampling times, the optimal sampling size is (*ibid*, equation (31), p. 1399)

$$n_{\text{sparse}}^* = \left(\frac{T}{4(E\epsilon^2)^2} \int_0^T \sigma_t^4 dt \right)^{1/3}. \quad (4.11)$$

Also, in this case, it is easy to see that the error terms in equations (4.9) and (4.10) are, respectively, $O(n^{-1/2}E[|\epsilon|^3])$ and $O(n^{-1/2}E[|\epsilon|^5])$. Plug (4.11) into (4.9) and (4.10) for the choice of n , and it follows that

$$\begin{aligned} \text{Cum}_3([Y, Y]_T^{(\text{sparse, opt})} - \langle X, X \rangle_T) &= 48 \left(T \int_0^T \sigma_t^4 dt \right)^{2/3} 2^{2/3} (E\epsilon^2)^{5/3} \\ &+ 8 \left(T \int_0^T \sigma_t^4 dt \right)^{-2/3} \left(T^2 \int_0^T \sigma_t^6 dt \right) (2E\epsilon^2)^{4/3} + O(E|\epsilon|^{11/3}) \end{aligned} \quad (4.12)$$

and

$$\begin{aligned} \text{Cum}_4([Y, Y]_T^{(\text{sparse, opt})} - \langle X, X \rangle_T) &= 384(E\epsilon^4 + \text{Var}(\epsilon)^2) \left(T \int_0^T \sigma_t^4 dt \right)^{2/3} (2E\epsilon^2)^{2/3} \\ &+ 384 \left(T \int_0^T \sigma_t^4 dt \right)^{-2/3} \left(T^2 \int_0^T \sigma_t^6 dt \right) 2^{4/3} (E\epsilon^2)^{7/3} \\ &+ 54 \left(T \int_0^T \sigma_t^4 dt \right)^{-1} \left(T^3 \int_0^T \sigma_t^8 dt \right) (2E\epsilon^2)^2 + O(E|\epsilon|^{17/3}) \end{aligned} \quad (4.13)$$

respectively.

But under optimal sampling, we have

$$\begin{aligned} \text{Var}([Y, Y]_T^{(\text{sparse, opt})}) &= E \left(\text{Var}([Y, Y]_T^{(\text{sparse, opt})} \mid X) \right) + \text{Var} \left(E([Y, Y]_T^{(\text{sparse, opt})} \mid X) \right) \\ &= 8 \langle X, X \rangle_T E\epsilon^2 + \frac{2}{n_{\text{sparse}}^*} \left(T \int_0^T \sigma_t^4 dt \right) + 4n_{\text{sparse}}^* E\epsilon^4 - 2 \text{Var}(\epsilon^2) \\ &= 2 \left(T \int_0^T \sigma_t^4 dt \right)^{\frac{2}{3}} (2E\epsilon^2)^{\frac{2}{3}} + O(E\epsilon^2), \end{aligned} \quad (4.14)$$

hence, if $s = \text{Var}([Y, Y]_T^{(\text{sparse, opt})})^{1/2}$,

$$\begin{cases} \text{Cum}_3 \left(s^{-1}([Y, Y]_T^{(\text{sparse, opt})} - \langle X, X \rangle_T) \right) = (E\epsilon^2)^{1/3}(\sigma^2 T)^{-1/3} 2^{5/6} + O((E|\epsilon|)^{4/3}) \\ \text{Cum}_4 \left(s^{-1}([Y, Y]_T^{(\text{sparse, opt})} - \langle X, X \rangle_T) \right) = (E\epsilon^2)^{2/3}(\sigma^2 T)^{-2/3} (27 \times 2^{1/3}) + O((E|\epsilon|)^2). \end{cases} \quad (4.15)$$

In other words, the third-order and the fourth-order cumulants indeed vanish as $n \rightarrow \infty$ and $E\epsilon^2 \rightarrow 0$, at rate $O((E\epsilon^2)^{1/3})$ and $O((E\epsilon^2)^{2/3})$, respectively.

4.3.2 Unconditional Cumulants for the Averaging Estimator

Similarly, for the averaging estimators,

$$E([Y, Y]_T^{(\text{avg})} | X \text{ process}) = [X, X]_T^{(\text{avg})} + 2\bar{n}E\epsilon^2, \quad (4.16)$$

$$\text{Var}([Y, Y]_T^{(\text{avg})} | X) = \text{Var}([\epsilon, \epsilon]_T^{(\text{avg})}) + \frac{8}{K} [X, X]_T^{(\text{avg})} E\epsilon^2 + O_p(E[|\epsilon|^2 (nK)^{-1/2}]), \quad (4.17)$$

with

$$\text{Var}([\epsilon, \epsilon]_T^{(\text{avg})}) = 4 \frac{\bar{n}}{K} E\epsilon^4 - \frac{2}{K} \text{Var}(\epsilon^2). \quad (4.18)$$

Also, from Zhang et al. (2005), for nonrandom σ_t , we have that

$$\text{Var}([X, X]_T^{(\text{avg})}) = \frac{K}{n} \frac{4}{3} T \int_0^T \sigma_t^4 dt + o\left(\frac{K}{n}\right). \quad (4.19)$$

Invoking the general relations between the conditional and the unconditional cumulants given above, we get the unconditional cumulants for the average estimator:

$$\begin{aligned} \text{Cum}_3([Y, Y]_T^{(\text{avg})} - \langle X, X \rangle_T) &= \frac{1}{K^2} c_3(\bar{n}) + 48 \frac{1}{K^2} E(\epsilon^4) E[X, X]_T^{(\text{avg})} \\ &+ 24 \frac{1}{K} \text{Var}(\epsilon) \text{Cov}([X, X]_T^{(\text{avg})}, [X, X]_T^{(\text{avg})} - \langle X, X \rangle_T) \\ &+ \text{Cum}_3([X, X]_T^{(\text{avg})} - \langle X, X \rangle_T) + O(K^{-2} \bar{n}^{-1/2} E[|\epsilon|^3]) \end{aligned} \quad (4.20)$$

and

$$\begin{aligned} \text{Cum}_4([Y, Y]_T^{(\text{avg})} - \langle X, X \rangle_T) &= \frac{1}{K^3} c_4(\bar{n}) + 24 \frac{1}{K^3} \frac{c_3(\bar{n})}{\bar{n}} E[X, X]_T^{(\text{avg})} \\ &+ 192 \frac{1}{K^2} E\epsilon^4 \text{Cov}([X, X]_T^{(\text{avg})}, [X, X]_T^{(\text{avg})} - \langle X, X \rangle_T) \\ &+ 192 \frac{1}{K^2} (\text{Var}(\epsilon))^2 \text{Var}([X, X]_T^{(\text{avg})}) \\ &+ 48 \frac{1}{K} \text{Var}(\epsilon) \text{Cum}_3([X, X]_T^{(\text{avg})}, [X, X]_T^{(\text{avg})} - \langle X, X \rangle_T, [X, X]_T^{(\text{avg})} - \langle X, X \rangle_T) \\ &+ \text{Cum}_4([X, X]_T^{(\text{avg})} - \langle X, X \rangle_T) + O(K^{-3} \bar{n}^{-1/2} E[|\epsilon|^5]) \end{aligned} \quad (4.21)$$

To calculate cumulants of $[X, X]_T^{(\text{avg})} - \langle X, X \rangle_T$ for the case where there is no leverage effect, we shall use the following proposition, which has some independent interest. We suppose that D_t is a process, $D_t = \int_0^t Z_s dW_s$. We also assume that (1) Z_s has mean zero, (2) is adapted to the filtration generated by W_t , and also (3) jointly Gaussian with W_t . The first two of these assumptions imply, by the martingale representation theorem, that one can write

$$Z_s = \int_0^s f(s, u) dW_u, \quad (4.22)$$

the third assumption yields that this $f(s, u)$ is nonrandom, with representation $\text{Cov}(Z_s, W_t) = \int_0^t f(s, u) du$ for $0 \leq t \leq s \leq T$.

Obviously, $\text{Var}(D_T) = \int_0^T E(Z_s^2) ds = \int_0^T \int_0^s f(s, u)^2 du ds$. The following result provides the third and fourth cumulants of D_T . Note that for $u \leq s$

$$\text{Cov}(Z_s, Z_u) = \int_0^u f(s, t) f(u, t) dt. \quad (4.23)$$

Proposition 4. *Under the assumptions above,*

$$\text{Cum}_3(D_T) = 6 \int_0^T ds \int_0^s \text{Cov}(Z_s, Z_u) f(s, u) du = 6 \int_0^T ds \int_0^s du \int_0^u f(s, u) f(s, t) f(u, t) dt \quad (4.24)$$

$$\begin{aligned} \text{Cum}_4(D_T) &= -12 \int_0^T ds \int_0^s dt \left(\int_0^t f(s, u) f(t, u) du \right)^2 \\ &\quad + 24 \int_0^T ds \int_0^s dx \int_0^x du \int_0^u dt (f(x, u) f(x, t) f(s, u) f(s, t) \\ &\quad + f(x, u) f(u, t) f(s, x) f(s, t) + f(x, t) f(u, t) f(s, x) f(s, u)) \end{aligned} \quad (4.25)$$

The proof is in the appendix. Note that it is possible to derive similar results in the multivariate case. See, for example, equation (E.3) in the appendix. For the application to our case, note that when σ_t is (conditionally or unconditionally) nonrandom, $D_T = [X, X]_T^{(\text{avg})} - \langle X, X \rangle_T$ is on the form discussed above, with

$$f(s, u) = \sigma_s \sigma_u \frac{2}{K} (K - \#t_j \text{ between } u \text{ and } s)^+. \quad (4.26)$$

This provides a general form of the low order cumulants of $[X, X]_T^{(\text{avg})}$. In the equidistant case, one can, in the equations above, to first order make the approximation

$$f(s, u) \approx 2\sigma_s \sigma_u \left(1 - \frac{s-u}{K\Delta t} \right)^+. \quad (4.27)$$

This yields, from Proposition 4,

$$\begin{aligned}
\text{Cum}_3([X, X]_T^{(\text{avg})}) &= 48 \left(\frac{K}{n}\right)^2 T^2 \int_0^T \sigma_t^6 dt \int_0^1 dy \int_0^1 dx (1-y)(1-x)(1-(x+y))^+ \\
&= 48 \left(\frac{K}{n}\right)^2 T^2 \int_0^T \sigma_t^6 dt \int_0^1 dz \int_{1-z}^1 dv zv(z+v-1) + o\left(\left(\frac{K}{n}\right)^2\right) \\
&= \frac{44}{10} \left(\frac{K}{n}\right)^2 T^2 \int_0^T \sigma_t^6 dt + o\left(\left(\frac{K}{n}\right)^2\right)
\end{aligned} \tag{4.28}$$

and

$$\begin{aligned}
\text{Cum}_4([X, X]_T^{(\text{avg})}) &= \left(\frac{K}{n}\right)^3 T^3 \int_0^T \sigma_t^8 dt \left\{ -192 \int_0^1 dy \left(\int_0^1 (1-(x+y))^+(1-x) dx \right)^2 \right. \\
&\quad + 384 \int_0^1 dz \int_0^1 dy \int_0^1 dw [(1-y)^+(1-(y+w))^+(1-(y+z))^+(1-(w+y+z))^+ \\
&\quad + (1-y)^+(1-w)^+(1-z)^+(1-(w+y+z))^+ \\
&\quad \left. + (1-(w+y))^+(1-w)^+(1-z)^+(1-(y+z))^+ \right\} + o\left(\left(\frac{K}{n}\right)^3\right) \\
&= \frac{1888}{105} \left(\frac{K}{n}\right)^3 T^3 \int_0^T \sigma_t^8 dt + o\left(\left(\frac{K}{n}\right)^3\right)
\end{aligned} \tag{4.29}$$

Thus, (4.20) and (4.21) lead to the following results:

Proposition 5. *In the case where there is no leverage effect, conditionally on the path of the σ_t^2 ,*

$$\begin{aligned}
\text{Cum}_3([Y, Y]_T^{(\text{avg})} - \langle X, X \rangle_T) &= \frac{1}{K^2} 8 \left\{ \left(\bar{n} - \frac{3}{4}\right) \text{Cum}_3(\epsilon^2) - 7\left(\bar{n} - \frac{6}{7}\right) \text{Cum}_3(\epsilon)^2 \right. \\
&\quad \left. + 6\left(\bar{n} - \frac{1}{2}\right) \text{Var}(\epsilon) \text{Var}(\epsilon^2) \right\} + 48 \frac{1}{K^2} E(\epsilon^4) \int_0^T \sigma_t^2 dt + \frac{96}{3} \frac{1}{n} E(\epsilon^2) T \int_0^T \sigma_t^4 dt \\
&\quad + \frac{44}{10} \left(\frac{K}{n}\right)^2 T^2 \int_0^T \sigma_t^6 dt + \text{smaller terms}
\end{aligned} \tag{4.30}$$

$$\begin{aligned}
\text{Cum}_4([Y, Y]_T^{(avg)} - \langle X, X \rangle_T) &= 16 \frac{\bar{n}}{K^3} \{ \text{Cum}_4(\epsilon^2) + (E\epsilon^4)^2 - 3(E\epsilon^2)^4 + 12\text{Var}(\epsilon^2)E\epsilon^4 \\
&\quad - 32E\epsilon^3 \text{Cov}(\epsilon^2, \epsilon^3) + 24E\epsilon^2(E\epsilon^3)^2 + 12 \text{Cum}_3(\epsilon^2)E\epsilon^2 \} + O\left(\frac{1}{K^3}E|\epsilon|^8\right) \\
&\quad + 192 \frac{1}{K^3} \{ \text{Cum}_3(\epsilon^2) - 7 \text{Cum}_3(\epsilon)^2 + 6 \text{Var}(\epsilon) \text{Var}(\epsilon^2) \} \int_0^T \sigma_t^2 dt + O\left(\frac{1}{nK^2}E|\epsilon|^6\right) \\
&\quad + 256 \frac{1}{nK} (E\epsilon^4 + (\text{Var}(\epsilon))^2) T \int_0^T \sigma_t^4 dt + o\left(\frac{1}{nK}E|\epsilon|^4\right) \\
&\quad + \frac{2112}{10} \frac{K}{n^2} \text{Var}(\epsilon) T^2 \int_0^T \sigma_t^6 dt + o\left(\frac{K}{n^2}E|\epsilon|^2\right) \\
&\quad + \frac{1888}{105} \left(\frac{K}{n}\right)^3 T^3 \int_0^T \sigma_t^8 dt + o\left(\left(\frac{K}{n}\right)^3\right) + \text{smaller terms}
\end{aligned} \tag{4.31}$$

Also, the optimal average subsampling size is,

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

The unconditional cumulants of the averaging estimator under the optimal sampling are

$$\begin{cases} \text{Cum}_3([Y, Y]_T^{(avg, opt)} - \langle X, X \rangle_T) = \frac{22}{5} \left(\frac{K}{n}\right)^2 T^2 \int_0^T \sigma_t^6 dt + o\left(\left(\frac{K}{n}\right)^2\right), \\ \text{Cum}_4([Y, Y]_T^{(avg, opt)} - \langle X, X \rangle_T) = \frac{1888}{105} \left(\frac{K}{n}\right)^3 T^3 \int_0^T \sigma_t^8 dt + o\left(\left(\frac{K}{n}\right)^3\right). \end{cases}$$

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

$$\begin{aligned}
\text{Var}([Y, Y]_T^{(avg, opt)}) &= \underbrace{\frac{8}{K} E\epsilon^2 \int_0^T \sigma_t^2 dt + 4 \frac{\bar{n}^*}{K} E\epsilon^4 - \frac{2}{K} \text{Var}(\epsilon^2)}_{=E(\text{Var}([Y, Y]_T^{(avg, opt)} | X))} \\
&\quad + \underbrace{\frac{K}{\bar{n}^*} \frac{4}{3} T \int_0^T \sigma_t^4 dt + o\left(\frac{K}{\bar{n}^*}\right)}_{=\text{Var}(E([Y, Y]_T^{(avg, opt)} | X))} \\
&= \frac{4}{3} 6^{\frac{1}{3}} (E\epsilon^2)^{\frac{2}{3}} \left(T \int_0^T \sigma_t^4 dt \right)^{\frac{2}{3}} + o(E|\epsilon|^{4/3})
\end{aligned} \tag{4.32}$$

hence, if we write $s = \text{Var}([Y, Y]_T^{(avg, opt)})^{1/2}$, we have that

$$\begin{cases} \text{Cum}_3 \left(s^{-1} ([Y, Y]_T^{(avg, opt)} - \langle X, X \rangle_T) \right) = (E\epsilon^2)^{1/3} (\sigma^2 T)^{-1/3} \left(\frac{11}{5} \times 2^{-11/6} \times 3^{5/3} \right) + O((E\epsilon^2)^{2/3}), \\ \text{Cum}_4 \left(s^{-1} ([Y, Y]_T^{(avg, opt)} - \langle X, X \rangle_T) \right) = (E\epsilon^2)^{2/3} (\sigma^2 T)^{-2/3} 6^{1/3} \frac{354}{35} + O((E\epsilon^2)) \end{cases} \tag{4.33}$$

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

It is interesting to note that the averaging estimator is no closer to normal than the sparse estimator. In fact, by comparing the expression for the third cumulants in (4.15) and (4.33), we find an increase in skewness of

$$\begin{aligned} \frac{\text{Cum}_3\left(s^{-1}([Y, Y]_T^{(\text{avg, opt})}) - \langle X, X \rangle_T\right)}{\text{Cum}_3\left(s^{-1}([Y, Y]_T^{(\text{sparse, opt})}) - \langle X, X \rangle_T\right)} &= \frac{(E\epsilon^2)^{1/3}(\sigma^2 T)^{-1/3}\left(\frac{11}{5} \times 2^{-11/6} \times 3^{5/3}\right) + O((E|\epsilon|)^{4/3})}{(E\epsilon^2)^{1/3}(\sigma^2 T)^{-1/3}2^{5/6} + O((E|\epsilon|)^{4/3})} \\ &= \frac{11}{5} \times 2^{-8/3} \times 3^{5/3} + O((E|\epsilon|)^{2/3}) \approx 216\%. \end{aligned} \quad (4.34)$$

This number does not fully reflect the change in skewness, since it is only a first order term and the higher order terms also matter, cf. our simulations in the next section. (The simulations use the most precise formulas above, see Table 1 for details.)

4.4. Cumulants for the TSRV Estimator

The same methods can be used to find cumulants for the two scales realized volatility (TSRV) estimator, $\widehat{\langle X, X \rangle}_T$. Since the distribution of TSRV is well approximated by its asymptotic normal distribution, we only sketch the results. When ϵ goes to zero sufficiently fast, the dominating term in the third and fourth unconditional cumulants for TSRV are, symbolically, the same as for the average volatility, namely

$$\begin{cases} \text{Cum}_3(\widehat{\langle X, X \rangle}_T - \langle X, X \rangle_T) = \frac{22}{5} \left(\frac{K}{n}\right)^2 T^2 \int_0^T \sigma_t^6 dt + o\left(\left(\frac{K}{n}\right)^2\right), \\ \text{Cum}_4(\widehat{\langle X, X \rangle}_T - \langle X, X \rangle_T) = \frac{1888}{105} \left(\frac{K}{n}\right)^3 T^3 \int_0^T \sigma_t^8 dt + o\left(\left(\frac{K}{n}\right)^3\right). \end{cases} \quad (4.35)$$

However, the value of K is quite different for TSRV than for the averaging volatility estimator. It is shown in Section 4 of Zhang et al. (2005) that for TSRV, the optimal choice of K is given by

$$K = \left(\frac{T}{12(E\epsilon^2)^2} \int_0^T \sigma_t^4 dt \right)^{-1/3} n^{2/3}. \quad (4.36)$$

As is seen from Table 1, this choice of K gives radically different distributional properties than those for the average volatility. This is consistent with the behavior in simulation. Thus, as predicted, the normal approximation works well in this case.

5. Simulation Results Incorporating the Edgeworth Correction

In this paper, we have discussed five estimators to deal with the microstructure noise in realized volatility. The five estimators, including $[Y, Y]_T^{(\text{all})}$, $[Y, Y]_T^{(\text{sparse})}$, $[Y, Y]_T^{(\text{sparse, opt})}$, $[Y, Y]_T^{(\text{avg})}$,

$\widehat{\langle X, X \rangle}_T$, are defined in Section 2. In this section, we focus on the case where the sampling points are regularly allocated. We first examine the empirical distributions of the five approaches in simulation. We then apply the the Edgeworth corrections as developed in Section 4, and compare the sample performance to those predicted by the asymptotic theory.

We simulate $M = 50,000$ sample paths from the standard Heston stochastic model

$$\begin{aligned} dX_t &= (\mu - \sigma_t^2/2) dt + \sigma_t dB_t \\ d\sigma_t^2 &= \kappa(\alpha - \sigma_t^2)dt + \gamma\sigma_t dW_t \end{aligned}$$

at a time interval $\Delta t = 1$ second, with parameter values $\mu = 0.05$, $\alpha = 0.04$, $\kappa = 5$, $\gamma = 0.05$ and $\rho = d\langle B, W \rangle_t/dt = -0.5$. As for the market microstructure noise ϵ , we assume that it is Gaussian with mean zero and standard deviation $(E\epsilon^2)^{1/2} = 0.0005$ (i.e., only 0.05% of the value of the asset price). On each simulated sample path, we estimate $\langle X, X \rangle_T$ over $T = 1$ day (i.e., $T = 1/252$ using annualized values) using the five estimation strategies described above: $[Y, Y]_T^{(\text{all})}$, $[Y, Y]_T^{(\text{sparse})}$, $[Y, Y]_T^{(\text{sparse, opt})}$, $[Y, Y]_T^{(\text{avg})}$ and the TSRV estimator, $\widehat{\langle X, X \rangle}_T$. We assume that a day consists of 6.5 hours of open trading, as is the case on the NYSE and NASDAQ. For $[Y, Y]_T^{(\text{sparse})}$, we use sparse sampling at a frequency of once every 5 minutes.

We shall see that even in this model – which includes leverage effect – the distributional approximation from our Edgeworth expansions is highly accurate.

For each estimator, we report the values of the standardized quantities

$$R = \frac{\text{estimator} - \langle X, X \rangle_T}{[\text{Var}(\text{estimator})]^{1/2}}. \quad (5.1)$$

For example, the variances of $[Y, Y]_T^{(\text{all})}$, $[Y, Y]_T^{(\text{sparse})}$ and $[Y, Y]_T^{(\text{sparse, opt})}$ are based on equation (4.14) with the sample size n , n_{sparse} and n_{sparse}^* respectively. And the variance of $[Y, Y]_T^{(\text{avg})}$ corresponds to (4.32) where the optimal subsampling size \bar{n}^* is adopted. The final estimator TSRV has variance

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

We now inspect how the simulation behavior of the five estimations compares to the second order Edgeworth expansion developed in the previous Section. The results are in Figure 1, and in Tables 1 and 2.

Table 1 reports the simulation results for the five estimation strategies. In each estimation strategy, “sample” represents the sample statistic from the M simulated paths; “Asymptotic (Normal)”

refers to the value predicted by the Normal asymptotic distribution (that is, without Edgeworth correction); “Asymptotic (Edgeworth)” refers to the value predicted by our theory (the asymptotic cumulants are given up to the approximation in the previous section; the relevant formula number is also given in Table 1). An inspection of Table 1 suggests that asymptotic normal theory (without higher order correction) is not adequate to capture the positive skewness and the leptokurtosis in each of the five (standardized) estimators, on the other hand, our expansion theory provides a good approximation to all four moments of the small sample distribution in each estimation scheme.

In Table 2, we report coverage probabilities computed as follows: for asymptotically standard normal T_n , let z_α be the upper $1 - \alpha$ quantile (so $\Phi(z_\alpha) = 1 - \alpha$) and set

$$w_{\alpha,n} = z_\alpha + \frac{1}{6} \text{Cum}_3(T_n)(z_\alpha^2 - 1) + \frac{1}{24} \text{Cum}_4(T_n)(z_\alpha^3 - 3z_\alpha) + \frac{1}{72} \text{Cum}_3(T_n)^2(-4z_\alpha^3 + 10z_\alpha). \quad (5.2)$$

The second order Cornish-Fisher corrected interval has actual coverage probability $P(T_n \leq w_{\alpha,n})$ (which should be close to $1 - \alpha$, but not exactly equal to it). The normal approximation gives a coverage probability $P(T_n \leq z_\alpha)$. We report these values for $\alpha = 0.10, 0.05$ and 0.01 . The results show that the Edgeworth-based coverage probabilities provide very accurate approximations to the sample ones, compared to the Normal-based coverage probabilities.

Figure 1 confirms that the sample distributions of all five estimators conform to our Edgeworth expansion. The nonlinearity in the QQ plots (left panels) reminds us that normal asymptotic theory without Edgeworth expansion fails to describe the sample behaviors of $[Y, Y]_T^{(\text{sparse})}$, $[Y, Y]_T^{(\text{sparse,opt})}$ and $[Y, Y]_T^{(\text{avg})}$. The histograms in the right panels display the standardized distribution of the five estimators obtained from simulation results, and the superimposed solid curve corresponds to the asymptotic distribution predicted by our Edgeworth expansion. The dashed curve represents the uncorrected $N(0, 1)$ distribution. By comparing the deviation between the dashed and solid curves, we can see how Edgeworth correction helps to capture the right skewness and leptokurtosis in the sample distribution of the (standardized) estimators.

6. Conclusions

We have developed and given formulas for Edgeworth expansions of several type of realized volatility estimators. Apart from the practical interest of having access to such expansions, there is an important conceptual finding. That is, a better expansion is obtained by using an asymptotics where the noise level goes to zero when the number of observations goes to infinity. Another lesson is that the asymptotic normal distribution is a more accurate approximation for the two scales

realized volatility (TSRV) than for the subsampled estimators, whose distributions definitely need to be Edgeworth-corrected in small samples.

In the process of developing the expansions, we also developed a general device for computing cumulants of the integrals of Gaussian processes with respect to Brownian motion (Proposition 4), and this result should have applications to other situations. The proposition is only stated for the third and fourth cumulant, but the same technology can potentially be used for higher order cumulants.

References

- Aït-Sahalia, Y., Mykland, P. A., Zhang, L., 2005a. How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies* 18, 351–416.
- Aït-Sahalia, Y., Mykland, P. A., Zhang, L., 2005b. Ultra high frequency volatility estimation with dependent microstructure noise. Tech. rep., Princeton University.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., Ebens, H., 2001a. The distribution of realized stock return volatility. *Journal of Financial Economics* 61, 43–76.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2001b. The distribution of exchange rate realized volatility. *Journal of the American Statistical Association* 96, 42–55.
- Awartani, B., Corradi, V., Distaso, W., 2006. Testing and modelling microstructure effects with an application to the dow jones industrial average, working Paper, University of Warwick.
- Bandi, F. M., Russell, J. R., 2003. Microstructure noise, realized volatility and optimal sampling. Tech. rep., University of Chicago Graduate School of Business.
- Barndorff-Nielsen, O., Shephard, N., 2005. How accurate is the asymptotic approximation to the distribution of realized variance? In: Andrews, D. W., Stock, J. J. (Eds.), *Identification and Inference for Econometric Models*, forthcoming. Cambridge University Press, Cambridge, UK.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N., 2008. Designing realized kernels to measure ex-post variation of equity prices in the presence of noise. *econometrica*, forthcoming .
- 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, B* 64, 253–280.
- Bhattacharya, R. N., Ghosh, J., 1978. On the validity of the formal edgeworth expansion. *Annals of Statistics* 6, 434–451.
- Bickel, P. J., Götze, F., van Zwet, W. R., 1986. The Edgeworth expansion for u -statistics of degree two. *The Annals of Statistics* 14, 1463–1484.

- Bollerslev, T., Zhou, H., 2002. Estimating stochastic volatility diffusions using conditional moments of integrated volatility. *Journal of Econometrics* 109, 33–65.
- Brillinger, D. R., 1969. The calculation of cumulants via conditioning. *Annals of the Institute of Statistical Mathematics* 21, 215–218.
- Chan, N. H., Wei, C. Z., 1987. Asymptotic inference for nearly nonstationary ar(1) processes. *Annals of Statistics* 15, 1050–1063.
- Corradi, V., Distaso, W., 2006. Semiparametric comparison of stochastic volatility models via realized measures. *Review of Economic Studies* 73, 635–667.
- Dacorogna, M. M., Gençay, R., Müller, U., Olsen, R. B., Pictet, O. V., 2001. *An Introduction to High-Frequency Finance*. Academic Press, San Diego.
- Delattre, S., Jacod, J., 1997. A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round-off errors. *Bernoulli* 3, 1–28.
- Feller, W., 1971. *An Introduction to Probability Theory and Its Applications, Volume 2*. John Wiley and Sons, New York.
- Gloter, A., Jacod, J., 2000. Diffusions with measurement errors: I - local asymptotic normality and ii - optimal estimators. Tech. rep., Université de Paris VI.
- Gonçalves, S., Meddahi, N., 2005. Bootstrapping realized volatility, forthcoming in *Econometrica*.
- Gonçalves, S., Meddahi, N., 2008. Edgeworth corrections for realized volatility. *Econometric Reviews* 27, 139–162.
- Hall, P., 1992. *The bootstrap and Edgeworth expansion*. Springer, New York.
- Hansen, P. R., Lunde, A., 2004. Realized variance and iid market microstructure noise. Tech. rep., Stanford University, Department of Economics.
- Jacod, J., 1994. Limit of random measures associated with the increments of a Brownian semimartingale. Tech. rep., Université de Paris VI.
- Jacod, J., Protter, P., 1998. Asymptotic error distributions for the euler method for stochastic differential equations. *Annals of Probability* 26, 267–307.
- Kadane, J. B., 1971. Comparison of k-class estimators when the disturbances are small. *Econometrica* 39, 723–737.
- Lieberman, O., Phillips, P. C., 2006. Refined inference on long memory in realized volatility, cowles Foundation Discussion Paper No. 1549.
- McCullagh, P., 1987. *Tensor Methods in Statistics*. Chapman and Hall, London, U.K.

- Meddahi, N., 2002. A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics* 17, 479–508.
- Mykland, P. A., 1993. Asymptotic expansions for martingales. *Annals of Probability* 21, 800–818.
- Mykland, P. A., 1994. Bartlett type identities for martingales. *Annals of Statistics* 22, 21–38.
- Mykland, P. A., 1995a. Embedding and asymptotic expansions for martingales. *Probability Theory and Related Fields* 103, 475–492.
- Mykland, P. A., 1995b. Martingale expansions and second order inference. *Annals of Statistics* 23, 707–731.
- Mykland, P. A., 2001. Likelihood computations without Bartlett identities. *Bernoulli* 7, 473–485.
- Mykland, P. A., Zhang, L., 2006. ANOVA for diffusions and itô processes. *Annals of Statistics* 34, 1931–1963.
- Oomen, R., 2002. Modelling realized variance when returns are serially correlated. Tech. rep., The University of Warwick, Warwick Business School.
- Speed, T. P., 1983. Cumulants and partition lattices. *The Australian Journal of Statistics* 25, 378–388.
- Wallace, D. L., 1958. Asymptotic approximations to distributions. *Annals of Mathematical Statistics* 29, 635–654.
- Zhang, L., 2001. From martingales to ANOVA: Implied and realized volatility. Ph.D. thesis, The University of Chicago, Department of Statistics.
- Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli* 12, 1019–1043.
- 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.
- Zhou, B., 1996. High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics* 14, 45–52.
- Zumbach, G., Dacorogna, M., Olsen, J., Olsen, R., 1999. Introducing a scale of market shocks. Tech. rep., Olsen & Associates.

Appendix: Proofs

A. Proof of Lemma 1

Let a_i be defined by

$$a_i = \begin{cases} 1 & \text{if } 1 \leq i \leq n-1 \\ \frac{1}{2} & \text{if } i = 0, n \end{cases} \quad (\text{A.1})$$

We can then write

$$\begin{aligned} c_3(n) &= \text{Cum}_3\left(2 \sum_{i=0}^n a_i(\epsilon_{t_i}^2 - E\epsilon^2) - 2 \sum_{i=0}^{n-1} \epsilon_{t_i} \epsilon_{t_{i+1}}\right), \\ &= 8[\text{Cum}_3\left(\sum_{i=0}^n a_i \epsilon_{t_i}^2\right) - \text{Cum}_3\left(\sum_{i=0}^{n-1} \epsilon_{t_i} \epsilon_{t_{i+1}}\right) \\ &\quad - 3 \text{Cum}\left(\sum_{i=0}^n a_i \epsilon_{t_i}^2, \sum_{j=0}^n a_j \epsilon_{t_j}^2, \sum_{k=0}^{n-1} \epsilon_{t_k} \epsilon_{t_{k+1}}\right) \\ &\quad + 3 \text{Cum}\left(\sum_{i=0}^n a_i \epsilon_{t_i}^2, \sum_{j=0}^{n-1} \epsilon_{t_j} \epsilon_{t_{j+1}}, \sum_{k=0}^{n-1} \epsilon_{t_k} \epsilon_{t_{k+1}}\right)] \end{aligned} \quad (\text{A.2})$$

where

$$\begin{aligned} \text{Cum}\left(\sum_{i=0}^n a_i \epsilon_{t_i}^2, \sum_{j=0}^n a_j \epsilon_{t_j}^2, \sum_{k=0}^{n-1} \epsilon_{t_k} \epsilon_{t_{k+1}}\right) &= 2 \sum_{k=0}^{n-1} a_k a_{k+1} \text{Cum}(\epsilon_{t_k}^2, \epsilon_{t_{k+1}}^2, \epsilon_{t_k} \epsilon_{t_{k+1}}) \\ &= 2(n-1)(E\epsilon^3)^2 \end{aligned} \quad (\text{A.3})$$

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,

$$\begin{aligned} \text{Cum}\left(\sum_{i=0}^n a_i \epsilon_{t_i}^2, \sum_{j=0}^{n-1} \epsilon_{t_j} \epsilon_{t_{j+1}}, \sum_{k=0}^{n-1} \epsilon_{t_k} \epsilon_{t_{k+1}}\right) &= 2 \sum_{j=0}^{n-1} a_j \text{Cum}(\epsilon_{t_j}^2, \epsilon_{t_j} \epsilon_{t_{j+1}}, \epsilon_{t_j} \epsilon_{t_{j+1}}) \\ &= 2\left(n - \frac{1}{2}\right)(E\epsilon^2) \text{Var}(\epsilon^2) \end{aligned} \quad (\text{A.4})$$

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

$$\text{Cum}\left(\sum_{i=0}^{n-1} \epsilon_{t_i} \epsilon_{t_{i+1}}, \sum_{j=0}^{n-1} \epsilon_{t_j} \epsilon_{t_{j+1}}, \sum_{k=0}^{n-1} \epsilon_{t_k} \epsilon_{t_{k+1}}\right) = \sum_{i=0}^{n-1} \text{Cum}_3(\epsilon_{t_i} \epsilon_{t_{i+1}}) = n(E\epsilon^3)^2, \quad (\text{A.5})$$

$$\text{Cum}\left(\sum_{i=0}^n a_i \epsilon_{t_i}^2, \sum_{j=0}^n a_j \epsilon_{t_j}^2, \sum_{k=0}^n a_k \epsilon_{t_k}^2\right) = \sum_{i=0}^n a_i^3 \text{Cum}_3(\epsilon_{t_i}^2) = \left(n - \frac{3}{4}\right) \text{Cum}_3(\epsilon^2), \quad (\text{A.6})$$

with $\sum_{i=0}^n a_i^3 = n - \frac{3}{4}$. Inserting (A.3)-(A.6) in (A.2) yields (4.3).

B. Proof of Proposition 1

To proceed, define

$$b_i = \begin{cases} \Delta X_{t_{i-1}} - \Delta X_{t_i} & \text{if } 1 \leq i \leq n-1 \\ \Delta X_{t_{n-1}} & \text{if } i = n \\ -\Delta X_{t_0} & \text{if } i = 0 \end{cases} \quad (\text{B.1})$$

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

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

because $\text{Cum}([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, \epsilon_{t_i}) = \text{Cum}([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, \epsilon_{t_1})$, for $i = 1, \dots, n-1$. Also, recalling the definition of a_i in (A.1)

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

Finally,

$$\begin{aligned} \text{Cum}_3([X, \epsilon]_T | X) &= \sum_{i=0}^n b_i^3 \text{Cum}_3(\epsilon) \\ &= E\epsilon^3 [-3 \sum_{i=1}^{n-1} (\Delta X_{t_{i-1}})^2 (\Delta X_{t_i}) + 3 \sum_{i=1}^{n-1} (\Delta X_{t_{i-1}}) (\Delta X_{t_i})^2] \\ &= O_p(n^{-1/2}E[|\epsilon|^3]) \end{aligned} \tag{B.3}$$

Gathering the terms above together, one now obtains the first part of Proposition 1. The second part of the result is then obvious.

C. Proof of Lemma 2

We have that:

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

$$\begin{aligned} &\text{Cum}(\sum_{i=0}^n a_i \epsilon_{t_i}^2, \sum_{j=0}^n a_j \epsilon_{t_j}^2, \sum_{k=0}^{n-1} \epsilon_{t_k} \epsilon_{t_{k+1}}, \sum_{l=0}^{n-1} \epsilon_{t_l} \epsilon_{t_{l+1}}) \\ &= \sum_{i=0}^n \sum_{j=0}^n \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} a_i a_j [1_{\{i=j, k=l, i=(k+1 \text{ or } k)\}} + 1_{\{l=k-1, (i,j)=(k+1, k-1)[2]\}} \\ &\quad + 1_{\{l=k+1, (i,j)=(k, k+2)[2]\}} + 1_{\{k=l, (i,j)=(k, k+1)[2]\}}] \\ &= 2(n - \frac{3}{4}) \text{Cum}_3(\epsilon^2) E\epsilon^2 + 4(n-2)(E\epsilon^3)^2 E\epsilon^2 + 2(n-1)(\text{Var}(\epsilon^2))^2 \end{aligned} \tag{C.2}$$

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 $\sum_{i=1}^n a_i^2 = n - 3/4$, $\sum_{i=1}^{n-1} a_{i-1}a_{i+1} = n - 2$, and $\sum_{i=0}^{n-1} a_i a_{i+1} = n - 1$. Next:

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

since $\sum_{i=0}^{n-1} a_i^2 a_{i+1} = n - 5/4$, and:

$$\begin{aligned}
\text{Cum}_4\left(\sum_{i=0}^{n-1} \epsilon_{t_i} \epsilon_{t_{i+1}}\right) &= \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} [1_{\{i=j=k=l\}} + \binom{4}{2} 1_{\{i=j, k=l, i=(k+1, k-1)\}}] \\
&\quad \times \text{Cum}(\epsilon_{t_i} \epsilon_{t_{i+1}}, \epsilon_{t_j} \epsilon_{t_{j+1}}, \epsilon_{t_k} \epsilon_{t_{k+1}}, \epsilon_{t_l} \epsilon_{t_{l+1}}) \\
&= n((E\epsilon^4)^2 - 3(E\epsilon^2)^4) + 12(n-1)(E\epsilon^2)^2 \text{Var}(\epsilon^2)
\end{aligned} \tag{C.4}$$

$$\text{Cum}_4\left(\sum_{i=0}^n a_i \epsilon_{t_i}^2\right) = \sum_{i=0}^n a_i^4 \text{Cum}_4(\epsilon^2) = (n - \frac{7}{8}) \text{Cum}_4(\epsilon^2). \tag{C.5}$$

Putting together (C.1)-(C.5):

$$\begin{aligned}
c_4(n) &= \text{Cum}_4\left(2 \sum_{i=0}^n a_i \epsilon_{t_i}^2 - 2 \sum_{i=0}^{n-1} \epsilon_{t_i} \epsilon_{t_{i+1}}\right), \\
&= 16[\text{Cum}_4\left(\sum_{i=0}^n a_i \epsilon_{t_i}^2\right) + \text{Cum}_4\left(\sum_{i=0}^{n-1} \epsilon_{t_i} \epsilon_{t_{i+1}}\right) \\
&\quad - \binom{4}{1} \text{Cum}\left(\sum_{i=0}^n a_i \epsilon_{t_i}^2, \sum_{j=0}^n a_j \epsilon_{t_j}^2, \sum_{k=0}^n a_k \epsilon_{t_k}^2, \sum_{l=0}^{n-1} \epsilon_{t_l} \epsilon_{t_{l+1}}\right) \\
&\quad - \binom{4}{3} \text{Cum}\left(\sum_{i=0}^n a_i \epsilon_{t_i}^2, \sum_{j=0}^{n-1} \epsilon_{t_j} \epsilon_{t_{j+1}}, \sum_{k=0}^{n-1} \epsilon_{t_k} \epsilon_{t_{k+1}}, \sum_{l=0}^{n-1} \epsilon_{t_l} \epsilon_{t_{l+1}}\right) \\
&\quad + \binom{4}{2} \text{Cum}\left(\sum_{i=0}^n a_i \epsilon_{t_i}^2, \sum_{j=0}^n a_j \epsilon_{t_j}^2, \sum_{k=0}^{n-1} \epsilon_{t_k} \epsilon_{t_{k+1}}, \sum_{l=0}^{n-1} \epsilon_{t_l} \epsilon_{t_{l+1}}\right)] \\
&= 16\left\{\left(n - \frac{7}{8}\right) \text{Cum}_4(\epsilon^2) + n(E\epsilon^4)^2 - 3n(E\epsilon^2)^4 + 12(n-1)\text{Var}(\epsilon^2)E\epsilon^4 \right. \\
&\quad \left. - 32\left(n - \frac{17}{16}\right)E\epsilon^3 \text{Cov}(\epsilon^2, \epsilon^3) + 24\left(n - \frac{7}{4}\right)E\epsilon^2(E\epsilon^3)^2 + 12\left(n - \frac{3}{4}\right) \text{Cum}_3(\epsilon^2)E\epsilon^2\right\}
\end{aligned} \tag{C.6}$$

since $\text{Cov}(\epsilon^2, \epsilon^3) = E\epsilon^5 - E\epsilon^2 E\epsilon^3$ and $\text{Cum}(\epsilon^2, \epsilon^2, \epsilon) = E\epsilon^5 - 2E\epsilon^2 E\epsilon^3$.

D. Proof of Proposition 2

It remains to deal with the second term in equation (4.6),

$$\begin{aligned} \text{Cum}([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, [X, \epsilon]_T, [X, \epsilon]_T | X) &= \sum_{i,j} b_i b_j \text{Cum}([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, \epsilon_{t_i}, \epsilon_{t_j}) \\ &= \sum_i b_i^2 \text{Cum}([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, \epsilon_{t_i}, \epsilon_{t_i}) + 2 \sum_{i=0}^{n-1} b_i b_{i+1} \text{Cum}([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, \epsilon_{t_i}, \epsilon_{t_{i+1}}) \end{aligned} \quad (\text{D.1})$$

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

$$\begin{aligned} \alpha &= n^{-1} \sum_i \text{Cum}([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, \epsilon_{t_i}, \epsilon_{t_i}) \\ \beta &= n^{-1} \sum_i \text{Cum}([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, \epsilon_{t_i}, \epsilon_{t_{i+1}}). \end{aligned}$$

Now following the two identities:

$$\begin{aligned} \text{Cum}([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, \epsilon_i, \epsilon_i) &= \text{Cum}_3([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, \epsilon_i^2) - 2(\text{Cov}([\epsilon, \epsilon]_T, \epsilon_i))^2 \\ \text{Cum}([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, \epsilon_i, \epsilon_{i+1}) &= \text{Cum}_3([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, \epsilon_i \epsilon_{i+1}) \\ &\quad - 2 \text{Cov}([\epsilon, \epsilon]_T, \epsilon_i) \text{Cov}([\epsilon, \epsilon]_T, \epsilon_{i+1}), \end{aligned}$$

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

$$2(\alpha - \beta) = n^{-1} \text{Cum}_3([\epsilon, \epsilon]_T) + O_p(n^{-1/2} E[|\epsilon|^6])$$

Hence, (D.1) becomes

$$\begin{aligned} \text{Cum}_4([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, [X, \epsilon]_T, [X, \epsilon]_T | X) \\ &= \sum_{i=0}^n b_i^2 \alpha + 2 \sum_{i=0}^n b_i b_{i+1} \beta + O_p(n^{-1/2} E[|\epsilon|^6]) \\ &= n^{-1} [X, X]_T \text{Cum}_3([\epsilon, \epsilon]_T) + O_p(n^{-1/2} E[|\epsilon|^6]) \end{aligned}$$

where the last line is because

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

The proposition now follows.

E. Proof of Proposition 4

The Bartlett identities for martingales, of this we use the cumulant version, with ‘‘cumulant variations’’, can be found in Mykland (1994). Set $Z_t^{(s)} = \int_0^{s \wedge t} f(s, u) dW_u$, which is taken to be a process in t for fixed s . For the third cumulant, by the third Bartlett identity

$$\text{Cum}_3(D_T) = 3 \text{Cov}(D_T, \langle D, D \rangle_T) = 3 \text{Cov}(D_T, \int_0^T Z_s^2 ds) = 3 \int_0^T \text{Cov}(D_T, Z_s^2) ds. \quad (\text{E.1})$$

To compute the integrand,

$$\begin{aligned}
\text{Cov}(D_T, Z_s^2) &= \text{Cov}(D_s, Z_s^2) \text{ since } D_t \text{ is a martingale} \\
&= \text{Cum}_3(D_s, Z_s, Z_s) \text{ since } ED_s = EZ_s = 0 \\
&= \text{Cum}_3(D_s, Z_s^{(s)}, Z_s^{(s)}) \\
&= 2 \text{Cov}(Z_s^{(s)}, \langle D, Z^{(s)} \rangle_s) + \text{Cov}(D_s, \langle Z^{(s)}, Z^{(s)} \rangle_s) \text{ by the third Bartlett identity} \\
&= 2 \text{Cov}(Z_s, \int_0^s Z_u f(s, u) du) \text{ by (4.22) and since } \langle Z^{(s)}, Z^{(s)} \rangle_s \text{ is nonrandom} \\
&= 2 \int_0^s \text{Cov}(Z_s, Z_u) f(s, u) du = 2 \int_0^s du \int_0^u f(s, u) f(s, t) f(u, t) dt \tag{E.2}
\end{aligned}$$

Combining the two last lines of (E.2) with equation (E.1) yields the result (4.24) in the Proposition.

Note that, more generally than (4.24), in the case of three different processes $D_T^{(i)}$, $i = 1, 2, 3$, one has

$$\text{Cum}_3(D_T^{(1)}, D_T^{(2)}, D_T^{(3)}) = 2 \int_0^T ds \int_0^s \text{Cov}(Z_s^{(1)}, Z_u^{(2)}) f^{(3)}(s, u) du \tag{E.3}$$

where the symbol “[3]” is used as in McCullagh (1987). We shall use this below.

For the fourth cumulant,

$$\text{Cum}_4(D_T) = -3 \text{Cov}(\langle D, D \rangle_T, \langle D, D \rangle_T) + 6 \text{Cum}_3(D_T, D_T, \langle D, D \rangle_T), \tag{E.4}$$

by the fourth Bartlett identity. For the first term

$$\begin{aligned}
\text{Cov}(\langle D, D \rangle_T, \langle D, D \rangle_T) &= \text{Cov}\left(\int_0^T Z_s^2 ds, \int_0^T Z_s^2 ds\right) \\
&= \int_0^T \int_0^T ds dt \text{Cov}(Z_s^2, Z_t^2) = 2 \int_0^T \int_0^T ds dt \text{Cov}(Z_s, Z_t)^2 \\
&= 4 \int_0^T ds \int_0^s dt \text{Cov}(Z_s, Z_t)^2 = 4 \int_0^T ds \int_0^s dt \left(\int_0^t f(s, u) f(t, u) du\right)^2 \tag{E.5}
\end{aligned}$$

For the other term in (E.4)

$$\text{Cum}_3(D_T, D_T, \langle D, D \rangle_T) = \int_0^T \text{Cum}_3(D_T, D_T, Z_s^2) ds$$

To calculate this, fix s , and set $D_t^{(1)} = D_t^{(2)} = D_t$, and $D_t^{(3)} = (Z_t^{(s)})^2 - \langle Z^{(s)}, Z^{(s)} \rangle_t$. Since $D_t^{(3)} = \int_0^t (2Z_u^{(s)} f(s, u)) dW_u$ for $t \leq s$, $D_t^{(3)}$ is on the form covered by the third cumulant equation (E.3), with Z (for $D^{(3)}$) $_u = 2Z_u^{(s)} f(s, u)$ and f (for $D^{(3)}$) $(a, t) = 2f(s, a)f(s, t)$ (for $t \leq a \leq s$). Then:

$$\begin{aligned}
\text{Cum}_3(D_T, D_T, \langle D, D \rangle_T) &= 4 \int_0^T ds \int_0^s dx \int_0^x du \int_0^u dt (f(x, u) f(x, t) f(s, u) f(s, t) \\
&\quad + f(x, u) f(u, t) f(s, x) f(s, t) + f(x, t) f(u, t) f(s, x) f(s, u)) \tag{E.6}
\end{aligned}$$

Combining equations (E.4), (E.5) and (E.6) yields the result (4.25) in the Proposition.

	ALL $[Y, Y]_T^{(\text{all})}$	SPARSE $[Y, Y]_T^{(\text{sparse})}$	SPARSE OPT $[Y, Y]_T^{(\text{sparse, opt})}$	AVG $[Y, Y]_T^{(\text{avg})}$	TSRV $\widehat{\langle X, X \rangle}_T$
Sample Bias ($\times 10^{-5}$)	1,171	3.89	2.23	1.918	0.00001
Asymptotic Bias ($\times 10^{-5}$)	1,170	3.90	2.19	1.923	0
Sample Mean	0.001	0.002	0.01	0.002	0.002
Asymptotic Mean	0	0	0	0	0
Sample Stdev	0.9997	1.006	1.006	0.996	1.01
Asymptotic Stdev	1	1	1	1	1
Sample Skewness	0.023	0.341	0.493	0.509	0.049
Asymp. Skewness (Normal)	0	0	0	0	0
Asymp. Skewness (Edgeworth)	0.025	0.340	0.490	0.511	0.043
Formula for cumulant	(4.9)	(4.9)	(4.12)	(4.30)	(4.35)
Sample Kurtosis	3.002	3.16	3.42	3.44	3.004
Asymp. Kurtosis (Normal)	3	3	3	3	3
Asymp. Kurtosis (Edgeworth)	3.001	3.17	3.41	3.37	3.005
Formula for cumulant	(4.10)	(4.10)	(4.13)	(4.31)	(4.35)

Table 1. Monte-Carlo simulations: This table reports the sample and asymptotic moments for the five estimators. The bias of the five estimators is computed relative to the true quadratic variation. Our theory predicts that the first four estimators are biased, with only TSRV being correctly centered. The mean, standard deviation, skewness and kurtosis are computed for the standardized distributions of the five estimators. As seen in the table, incorporating the Edgeworth correction provides a clear improvement in the fit of the asymptotic distribution, compared to the asymptotics based on the Normal distribution.

	ALL $[Y, Y]_T^{(\text{all})}$	SPARSE $[Y, Y]_T^{(\text{sparse})}$	SPARSE OPT $[Y, Y]_T^{(\text{sparse,opt})}$	AVG $[Y, Y]_T^{(\text{avg})}$	TSRV $\widehat{\langle X, X \rangle}_T$
		Theoretical Coverage Probability = 90%			
Normal-Based Coverage	89.9%	89.3%	89.0%	89.5%	89.6%
Edgeworth-Based Coverage	90.0%	89.8%	89.7%	89.8%	89.7%
		Theoretical Coverage Probability = 95%			
Normal-Based Coverage	94.9%	94.0%	93.6%	93.9%	94.5%
Edgeworth-Based Coverage	95.0%	94.9%	94.8%	94.5%	94.6%
		Theoretical Coverage Probability = 99%			
Normal-Based Coverage	98.9%	98.0%	98.0%	98.0%	98.8%
Edgeworth-Based Coverage	99.0%	99.0%	99.0%	98.6%	98.9%

Table 2. Monte-Carlo simulations: This table reports the coverage probability before and after the Edgeworth correction.

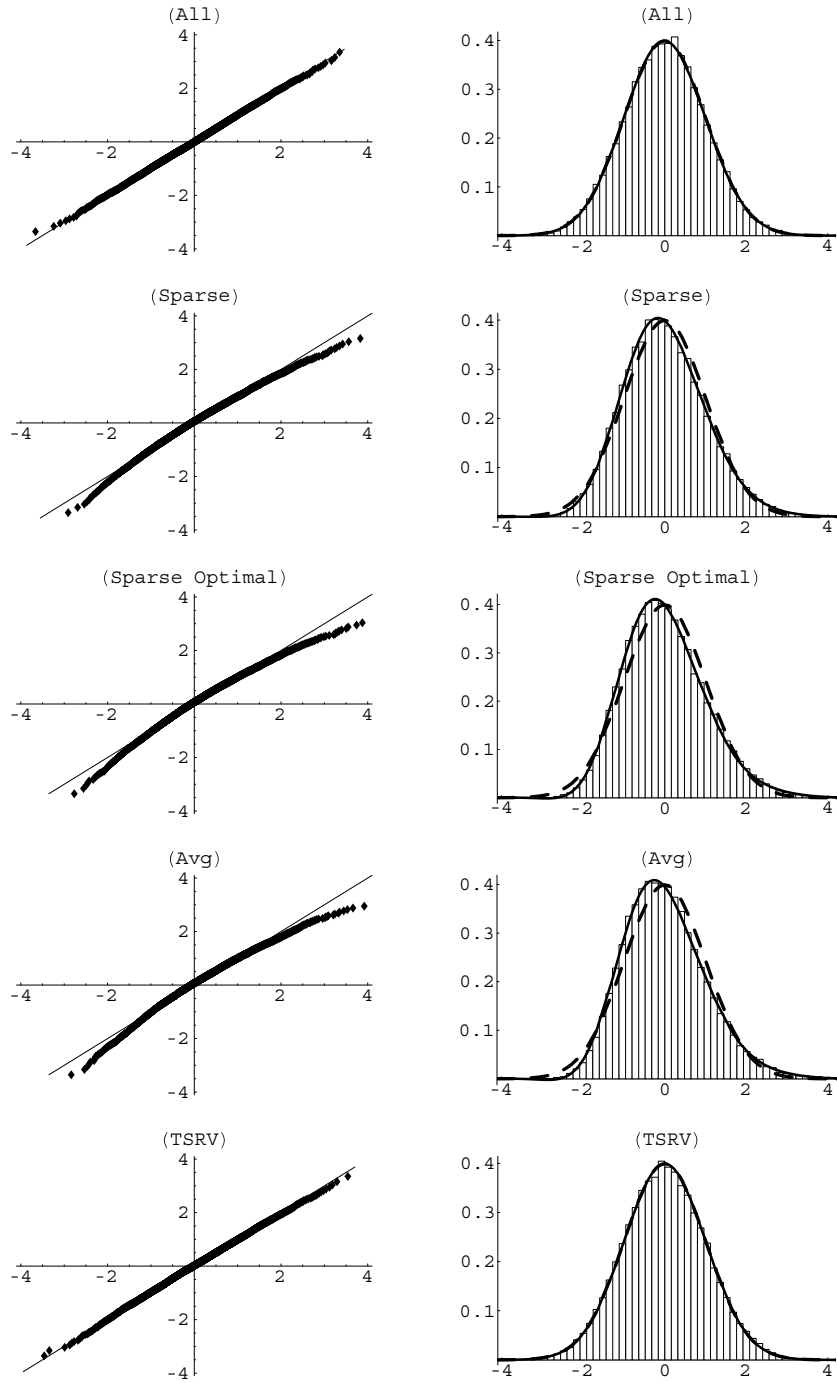


Fig. 1. Left panel: QQ plot for the five estimators based on the asymptotic Normal distribution. Right panel: Comparison of the small sample distribution of the estimator (histogram), the Edgeworth-corrected distribution (solid line) and the standard Normal distribution (dashed line).