

TECHNICAL WORKING PAPER SERIES

EDGEWORTH EXPANSIONS FOR REALIZED VOLATILITY
AND RELATED ESTIMATORS

Lan Zhang
Per Mykland
Yacine Aït-Sahalia

Technical Working Paper 319
<http://www.nber.org/papers/T0319>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2005

Financial support from the NSF under grants SBR-0350772 (Aït-Sahalia) and DMS-0204639 (Mykland and Zhang) is gratefully acknowledged. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

©2005 by Lan Zhang, Per Mykland, and Yacine Aït-Sahalia. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Edgeworth Expansions for Realized Volatility and Related Estimators
Lan Zhang, Per Mykland, and Yacine Aït-Sahalia
NBER Technical Working Paper No. 319
October 2005
JEL No. C13, C14, C15, C22

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 compute Edgeworth expansions for such estimators. Unlike the usual expansions, we have found that in order to obtain meaningful terms, one needs to let the size of the noise to go zero asymptotically. The results have application to Cornish-Fisher inversion and bootstrapping.

Lan Zhang
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213
lanzhang@uic.edu

Per A. Mykland
Department of Statistics
The University of Chicago
Chicago, IL 60637
mykland@galton.uchicago.edu

Yacine Aït-Sahalia
Bendheim Center for Finance
Princeton University
Princeton, NJ 08544-1021
and NBER
yacine@princeton.edu

1 Introduction

Volatility estimation from high frequency data has received substantial attention in the recent literature: see ,e.g., Dacorogna et al. (2001), Andersen et al. (2001), Zhang (2001), Barndorff-Nielsen and Shephard (2002), Meddahi (2002), and Oomen (2002). 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.

Our interest in this was initially motivated by the apparent inefficiency inherent in throwing away so much data. We formally analyzed the issue in Aït-Sahalia et al. (2005a), where we studied the impact of different types market microstructure noise on the properties of RV estimators and proposed likelihood corrections for (parametric) volatility estimation. As part of our analysis of the properties of RV estimators when market microstructure noise is taken into account, in the nonparametric context, we were led in Zhang et al. (2002) to propose five different RV-like estimation strategies, culminating with an estimator based on combining two time scales, which we called TSRV (two scale realized volatility). TSRV is the first nonparametric volatility estimator in the literature to be consistent. Following this work, other varieties have been introduced to improve the efficiency or deal with more complex noise structures: see Zhang (2004), Aït-Sahalia et al. (2005b) and Barndorff-Nielsen et al. (2004).¹

One thing in common among all these RV-type estimators is that the limit theory pre-

¹Other 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. (2002). None of these estimators are consistent, however.

dicts that these estimators should be asymptotically normal. Without noise, the asymptotic normality of RV estimates dates back to at least Jacod (1994) and Jacod and Protter (1998); see also e.g., Barndorff-Nielsen and Shephard (2002) and Mykland and Zhang (2002). 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. (2002).

As we shall see, however, simulation results do not always agree well with what the theory predicts, even with fairly large sample sizes. The usual remedy in such situations is to use Edgeworth expansions and, in this paper, we will derive such expansions for the volatility estimators when the observations of the price process are noisy. Recently, Goncalves and Meddahi (2005) have developed an Edgeworth expansion for the basic RV estimator when there is no noise. Their expansion applies to the studentized statistic based on the standard RV and it is used for assessing the accuracy of the bootstrap in comparison to the first order asymptotic approach. By contrast, we develop here an Edgeworth expansion for nonstudentized statistics for the standard RV, TSRV and other estimators, but allow for the presence of microstructure noise.

We argue that the lack of normality is caused by the coexistence of a small effective sample size and small noise. 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 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.

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 also 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

$$Y_{t_i} = X_{t_i} + \epsilon_{t_i}. \quad (2.1)$$

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 assume that the volatility process, when stochastic, is orthogonal to the Brownian motion driving the price process.

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

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

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 the integrated volatility $\int_0^T \sigma_t^2 dt$, 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.

In Zhang et al. (2002), we considered five RV-type estimators. Ranked from the statistically least desirable to the most desirable, we started with the “all” estimator $[Y, Y]^{(\text{all})}$, where RV is based on the entire sample and consecutive returns are used; 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; 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); the averaging estimator $[Y, Y]^{(\text{avg})}$, which is constructed by averaging the sparse estimators and thus also utilizes the entire sample, and finally 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

$$[Y, Y]_T^{(\text{all})} = \sum_{t_i \in \mathcal{G}} (Y_{t_{i+1}} - Y_{t_i})^2, \quad (2.4)$$

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,

$$[Y, Y]_T^{(\text{sparse})} = \sum_{t_j, t_{j,+} \in \mathcal{H}} (Y_{t_{j,+}} - Y_{t_j})^2, \quad (2.5)$$

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 (2.5) 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. (2002).)

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

$$[Y, Y]_T^{(\text{avg})} = \frac{1}{K} \sum_{k=1}^K \underbrace{\sum_{t_j, t_{j,+} \in \mathcal{G}^{(k)}} (Y_{t_{j,+}} - Y_{t_j})^2}_{=[Y, Y]_T^{(k)}}, \quad (2.6)$$

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. (2002) for an explicit formula.) A special case of (2.6) 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

$$\widehat{\langle X, X \rangle}_T = \left(1 - \frac{\bar{n}}{n}\right)^{-1} \left([Y, Y]_T^{(\text{avg})} - \frac{\bar{n}}{n} [Y, Y]_T^{(\text{all})} \right) \quad (2.7)$$

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 the returns on a fast time scale. \bar{n} in (2.7) is the average sample size across different grids. (Note that this is what is called the ‘‘adjusted’’ TSRV in Zhang et al. (2002).)

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

$$[Y, Y]_T = [X, X]_T + 2[X, \epsilon]_T + [\epsilon, \epsilon]_T. \quad (2.8)$$

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. (2002); all five have asymptotically 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 in that

$$\begin{aligned}
 [Y, Y]_T^{(\text{sparse})} &\stackrel{\mathcal{L}}{\approx} \langle X, X \rangle_T + \underbrace{2n_{\text{sparse}}E\epsilon^2}_{\text{bias due to noise}} \tag{3.1} \\
 &+ \underbrace{[Var([\epsilon, \epsilon]_T^{(\text{sparse})}) + 8[X, X]_T^{(\text{sparse})}E\epsilon^2]}_{\text{due to noise}} + \underbrace{\frac{2T}{n_{\text{sparse}}} \int_0^T \sigma_t^4 dt}_{\text{due to discretization}}]^{1/2} Z_{\text{total}}, \\
 &\underbrace{\hspace{15em}}_{\text{total variance}}
 \end{aligned}$$

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

If the sample size n_{sparse} is large relative to the noise, the variance due to noise in (3.1) would be dominated by $Var([\epsilon, \epsilon]_T^{(\text{sparse})})$ 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 $Var([\epsilon, \epsilon]_T^{(\text{sparse})})$. 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. (2002).

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}^* , where for equidistant observations,

$$n_{\text{sparse}}^* = (E\epsilon^2)^{-2/3} \left(\frac{T}{4} \int_0^T \sigma_t^4 dt \right)^{1/3}. \quad (3.2)$$

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 (3.1) that 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 (2.8), we have

$$\begin{aligned} [Y, Y]_T^{(\text{avg})} &\stackrel{\mathcal{L}}{\approx} \langle X, X \rangle_T + \underbrace{2\bar{n}E\epsilon^2}_{\text{bias due to noise}} \\ &+ \underbrace{\left[\underbrace{\text{Var}([\epsilon, \epsilon]_T^{(\text{avg})}) + \frac{8}{K} [X, X]_T^{(\text{avg})} E\epsilon^2}_{\text{due to noise}} + \underbrace{\frac{4T}{3\bar{n}} \int_0^T \sigma_t^4 dt}_{\text{due to discretization}} \right]^{1/2}}_{\text{total variance}} Z_{\text{total}}, \end{aligned} \quad (3.3)$$

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 (3.3) 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 (3.3) and then set $K^* \approx n/\bar{n}^*$. In the equidistantly sampled case,

$$\bar{n}^* = \left(\frac{T}{6(E\epsilon^2)^2} \int_0^T \sigma_t^4 dt \right)^{1/3}. \quad (3.4)$$

If one removes the bias in either $[Y, Y]_T^{(\text{avg})}$ or $[Y, Y]_T^{(\text{avg, opt})}$, it follows from (3.3) 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 (3.3). 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.

Figures 1- 5 report the QQ plots of the standardized distribution of the five estimators before any Edgeworth correction is applied. It is clear that the sparse, sparse-optimal and averaging estimators are not symmetrically distributed. Comparing to a normal distribution, these three estimators have thinner tails at large values and fatter tail at low values. 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.

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 only are

used, to capture skewness and kurtosis, as follows:

$$f_n(x) = \phi(z) \left[1 + \frac{1}{6} \text{cum}_3(T_n) h_3(z) + \frac{1}{24} \text{cum}_4(T_n) h_4(z) + \frac{1}{72} \text{cum}_3(T_n)^2 h_5(z) + \dots \right] \quad (4.1)$$

where $z = (x - E(T_n))/\text{Var}(T_n)^{1/2}$, and where the Hermite polynomials h_i are given by

$$\begin{aligned} h_3(z) &= z^3 - 3z \\ h_4(z) &= z^4 - 6z^2 + 3 \\ h_5(z) &= z^5 - 10z^3 + 15z. \end{aligned}$$

The neglected terms are typically of lower order in n than the terms that are included, and we shall refer to this 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. This can be turned into expansions for p -values, and to Cornish-Fisher expansions for critical values, for which 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 (2.8) 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 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 need the form of the cumulants up to order four of T_n . This is what we do in the following for the sparse and average volatility. We assume that the “size” of the law of ϵ goes to zero, formally that $E|\epsilon|^p \rightarrow 0$ for all $p \in (0, 8]$. In particular, say, $O_p(E|\epsilon|^5) = o_p(E|\epsilon|^4)$. If one does not do this, then the expansion will not work as well, as demonstrated in Section 5.3 below.

4.2 Conditional Cumulants

We start by deriving explicit expressions for the conditional cumulants for $[Y, Y]$ and $[Y, Y]^{(\text{avg})}$, given the latent process X . All the expressions we give below about $[Y, Y]$ hold for both $[Y, Y]^{(\text{all})}$ and $[Y, Y]^{(\text{sparse})}$; 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

$$c_3(n) = O_p(nE[\epsilon^6]) \quad (4.3)$$

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

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

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 have:

Proposition 1.

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

where $cum_3([\epsilon, \epsilon]_T)$ is given in (4.3). Also

$$\begin{aligned} cum_3(K[Y, Y]_T^{(avg)}|X) &= cum_3(K[\epsilon, \epsilon]_T^{(avg)}) + 48K[X, X]_T^{(avg)}E\epsilon^4 \\ &\quad + O_p(K\bar{n}^{-1/2}E[|\epsilon|^3]). \end{aligned}$$

4.2.2 Fourth-Order Conditional Cumulants

For the fourth-order cumulant, denote

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

We have that:

Lemma 2.

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

Also here,

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

For the conditional fourth-order cumulant, we know that

$$\begin{aligned} cum_4([Y, Y]|X) &= cum_4([\epsilon, \epsilon]_T) + 24cum([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, [X, \epsilon]_T, [X, \epsilon]_T|X) \\ &\quad + 8cum([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, [X, \epsilon]_T|X) \\ &\quad + 32cum([\epsilon, \epsilon]_T, [X, \epsilon]_T, [X, \epsilon]_T, [X, \epsilon]_T|X) \\ &\quad + 16cum_4([X, \epsilon]|X). \end{aligned} \tag{4.5}$$

Similar argument as in deriving the third cumulant shows that the latter three terms in the right hand side of (4.5) are of order $O_p(n^{-1/2}E[|\epsilon|^5])$. Gathering terms of the appropriate order, we obtain:

Proposition 2.

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

Also, for the average estimator,

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

4.3 Unconditional Cumulants

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)):

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

In what follows, we apply these formulae to derive the unconditional cumulants for our estimators.

4.3.1 Unconditional Cumulants for Sparse Estimators

In Zhang et al. (2002), we showed that

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

and also that

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

This allows us to obtain the unconditional cumulants as:

$$\begin{aligned} cum_3([Y, Y]_T - \langle X, X \rangle_T) &= c_3(n) + 48E(\epsilon^4)E[X, X] \\ &\quad + 24Var(\epsilon)Cov([X, X]_T, [X, X]_T - \langle X, X \rangle_T) \quad (4.6) \\ &\quad + cum_3([X, X]_T - \langle X, X \rangle_T) + O(n^{-1/2} E[|\epsilon|^3]) \end{aligned}$$

and

$$\begin{aligned}
cum_4([Y, Y]_T - \langle X, X \rangle_T) &= c_4(n) + 24 \frac{1}{n} c_3(n) E[X, X]_T \\
&+ 192 E \epsilon^4 Cov([X, X]_T, [X, X]_T - \langle X, X \rangle_T) \\
&+ 192 (Var(\epsilon))^2 Var([X, X]_T) \\
&+ 48 Var(\epsilon) cum_3([X, X]_T, [X, X]_T - \langle X, X \rangle_T, [X, X]_T - \langle X, X \rangle_T) \\
&+ cum_4([X, X]_T - \langle X, X \rangle_T) + O(n^{-1/2} E[|\epsilon|^5])
\end{aligned} \tag{4.7}$$

To calculate cumulants of $[X, X]_T - \langle X, X \rangle_T$, consider now the case where there is no leverage effect. That is to say that one can take σ_t to be (either conditionally or unconditionally) nonrandom. In this case,

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

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

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

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

$p = 1$	$p = 2$	$p = 3$	$p = 4$
$cum_p(\chi_1^2)$	1	2	8
		8	54

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

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

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 ,

$$\begin{aligned}
cum_3([Y, Y]_T - \langle X, X \rangle_T) &= c_3(n) + 48E(\epsilon^4) \int_0^T \sigma_t^2 dt \\
&+ 48Var(\epsilon)n^{-1}T \int_0^T \sigma_t^4 dt \\
&+ 8n^{-2}T^2 \int_0^T \sigma_t^6 dt \\
&+ O(n^{-3/2}E[\epsilon^2]) + O(n^{-1/2}E[|\epsilon|^3]) + O(n^{-5/2})
\end{aligned} \tag{4.8}$$

Similarly for the fourth cumulant

$$\begin{aligned}
cum_4([Y, Y]_T - \langle X, X \rangle_T) &= c_4(n) + 24n^{-1}c_3(n) \int_0^T \sigma_t^2 dt \\
&+ 384(E\epsilon^4 + Var(\epsilon)^2)n^{-1}T \int_0^T \sigma_t^4 dt \\
&+ 384Var(\epsilon)n^{-2}T^2 \int_0^T \sigma_t^6 dt + 54n^{-3}T^3 \int_0^T \sigma_t^8 dt \\
&+ O(n^{-1/2}E[|\epsilon|^5]) + O(n^{-3/2}E[\epsilon^4]) \\
&+ O(n^{-5/2}E[\epsilon^2]) + O(n^{-7/2})
\end{aligned} \tag{4.9}$$

It is clear that one needs $\epsilon_n = o_p(n^{-1/2})$ to keep all the terms in (4.8) and (4.9) non-negligible. In this case, the error term in equation (4.8) is of order $O(n^{-1/2}E[|\epsilon|^3]) + O(n^{-5/2})$, while that in equation (4.9) is of order $O(n^{-1/2}E[|\epsilon|^5]) + O(n^{-7/2})$. In the case of optimal-sparse estimator, (3.2) leads 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

$$n_{\text{sparse}}^* = \left(\frac{T}{2} \frac{\sigma^2}{E\epsilon^2} \right)^{2/3}. \tag{4.10}$$

Also, in this case, it is easy to see (by the same derivation as above) that the error terms in equations (4.8) and (4.9) are, respectively, $O(n^{-1/2}E[|\epsilon|^3])$ and $O(n^{-1/2}E[|\epsilon|^5])$. Plug (4.10) into (4.8) and (4.9) for the choice of n , and it follows that

$$\begin{aligned}
cum_3([Y, Y]_T^{(\text{sparse, opt})} - \langle X, X \rangle_T) &= 48(\sigma^2 T)^{4/3} 2^{2/3} (E\epsilon^2)^{5/3} \\
&+ 8(\sigma^2 T)^{5/3} (2E\epsilon^2)^{4/3} + O(E|\epsilon|^{11/3})
\end{aligned} \tag{4.11}$$

and

$$\begin{aligned}
cum_4([Y, Y]_T^{(\text{sparse, opt})} - \langle X, X \rangle_T) &= 384(E\epsilon^4 + Var(\epsilon^2)(\sigma^2 T)^{4/3}(2E\epsilon^2)^{2/3} \\
&\quad + 384(\sigma^2 T)^{5/3}2^{4/3}(E\epsilon^2)^{7/3} \\
&\quad + 54(\sigma^2 T)^2(2E\epsilon^2)^2 + O(E|\epsilon|^{17/3})
\end{aligned} \tag{4.12}$$

respectively.

But under optimal sampling, we have

$$\begin{aligned}
Var([Y, Y]_T^{(\text{sparse, opt})}) &= E \left(Var([Y, Y]_T^{(\text{sparse, opt})} | X) \right) + Var \left(E([Y, Y]_T^{(\text{sparse, opt})} | X) \right) \\
&= 8 \langle X, X \rangle_T E\epsilon^2 + \frac{2}{n_{\text{sparse}}^*} (\sigma^2 T)^2 \\
&\quad + 4n_{\text{sparse}}^* E\epsilon^4 - 2Var(\epsilon^2) \\
&= 2(\sigma^2 T)^{\frac{4}{3}} (2E\epsilon^2)^{\frac{2}{3}} + O_p(E\epsilon^2),
\end{aligned} \tag{4.13}$$

hence,

$$cum_3 \left((E\epsilon^2)^{-1/3} ([Y, Y]_T^{(\text{sparse, opt})} - \langle X, X \rangle_T) \right) = O((E|\epsilon|)^{2/3}) \tag{4.14}$$

$$cum_4 \left((E\epsilon^2)^{-1/3} ([Y, Y]_T^{(\text{sparse, opt})} - \langle X, X \rangle_T) \right) = O((E|\epsilon|)^{4/3}). \tag{4.15}$$

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

4.3.2 Unconditional Cumulants for the Averaging Estimator

Similarly, for the averaging estimators,

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

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

with

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

Also, from Zhang et al. (2002), for nonrandom σ_t , we have that

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

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

$$\begin{aligned} cum_3([Y, Y]_T^{(avg)} - \langle X, X \rangle_T) &= \frac{1}{K^2} c_3(\bar{n}) + 48 \frac{1}{K^2} E(\epsilon^4) E[X, X]_T^{(avg)} \\ &+ 24 \frac{1}{K} Var(\epsilon) Cov([X, X]_T^{(avg)}, [X, X]_T^{(avg)} - \langle X, X \rangle_T) \quad (4.20) \\ &+ cum_3([X, X]_T^{(avg)} - \langle X, X \rangle_T) + O(K^{-2} \bar{n}^{-1/2} E[|\epsilon|^3]) \end{aligned}$$

and

$$\begin{aligned} cum_4([Y, Y]_T^{(avg)} - \langle X, X \rangle_T) &= \frac{1}{K^3} c_4(\bar{n}) + 24 \frac{1}{K^3} \frac{c_3(\bar{n})}{\bar{n}} E[X, X]_T^{(avg)} \\ &+ 192 \frac{1}{K^2} E\epsilon^4 Cov([X, X]_T^{(avg)}, [X, X]_T^{(avg)} - \langle X, X \rangle_T) \\ &+ 192 \frac{1}{K^2} (Var(\epsilon))^2 Var([X, X]_T^{(avg)}) \quad (4.21) \\ &+ 48 \frac{1}{K} Var(\epsilon) cum_3([X, X]_T^{(avg)}, [X, X]_T^{(avg)} - \langle X, X \rangle_T, [X, X]_T^{(avg)} - \langle X, X \rangle_T) \\ &+ cum_4([X, X]_T^{(avg)} - \langle X, X \rangle_T) + O(K^{-3} \bar{n}^{-1/2} E[|\epsilon|^5]) \end{aligned}$$

To calculate cumulants of $[X, X]_T^{(avg)} - \langle X, X \rangle_T$ 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 $Cov(Z_s, W_t) = \int_0^t f(s, u) du$ for $0 \leq t \leq s \leq T$.

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

$$cov(Z_s, Z_u) = \int_0^u f(s, t) f(u, t) dt. \quad (4.23)$$

Proposition 4. *Under the assumptions above,*

$$\begin{aligned} cum_3(D_T) &= 6 \int_0^T ds \int_0^s cov(Z_s, Z_u) f(s, u) du \\ &= 6 \int_0^T ds \int_0^s du \int_0^u f(s, u) f(s, t) f(u, t) dt \end{aligned} \quad (4.24)$$

$$\begin{aligned} cum_4(D_T) &= -12 \int_0^T ds \int_0^s dt \left(\int_0^t f(s, u) f(t, u) du \right)^2 \\ &+ 24 \int_0^T ds \int_0^s dx \int_0^x du \int_0^u dt (f(x, u) f(x, t) f(s, u) f(s, t) \\ &+ f(x, u) f(u, t) f(s, x) f(s, t) + f(x, t) f(u, t) f(s, x) f(s, u)) \end{aligned} \quad (4.25)$$

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,

$$\begin{aligned} cum_3([X, X]_T^{(\text{avg})}) &= 48 \left(\frac{K}{n} \right)^2 T^2 \int_0^T \sigma_t^6 dt \int_0^1 dy \int_0^1 dx (1-y)(1-x)(1-(x+y))^+ \\ &= 48 \left(\frac{K}{n} \right)^2 T^2 \int_0^T \sigma_t^6 dt \int_0^1 dz \int_{1-z}^1 dv zv(z+v-1) \\ &+ o\left(\left(\frac{K}{n} \right)^2 \right) \\ &= \frac{44}{10} \left(\frac{K}{n} \right)^2 T^2 \int_0^T \sigma_t^6 dt + o\left(\left(\frac{K}{n} \right)^2 \right) \end{aligned} \quad (4.28)$$

and

$$\begin{aligned}
cum_4(\langle X, X \rangle_T^{(avg)}) &= \left(\frac{K}{n}\right)^3 T^3 \int_0^T \sigma_t^8 dt \left\{ -192 \int_0^1 dy \left(\int_0^1 (1 - (x + y))^+ (1 - x) dx \right)^2 \right. \\
&\quad + 384 \int_0^1 dz \int_0^1 dy \int_0^1 dw [(1 - y)^+ (1 - (y + w))^+ (1 - (y + z))^+ (1 - (w + y + z))^+ \\
&\quad + (1 - y)^+ (1 - w)^+ (1 - z)^+ (1 - (w + y + z))^+ \\
&\quad \left. + (1 - (w + y))^+ (1 - w)^+ (1 - z)^+ (1 - (y + z))^+] \right\} \\
&\quad + o\left(\left(\frac{K}{n}\right)^3\right) \\
&= \frac{1888}{105} \left(\frac{K}{n}\right)^3 T^3 \int_0^T \sigma_t^8 dt + o\left(\left(\frac{K}{n}\right)^3\right)
\end{aligned} \tag{4.29}$$

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 ,*

$$\begin{aligned}
cum_3(\langle Y, Y \rangle_T^{(avg)} - \langle X, X \rangle_T) &= \frac{1}{K^2} 8 \left\{ \left(\bar{n} - \frac{3}{4}\right) cum_3(\epsilon^2) - 7\left(\bar{n} - \frac{6}{7}\right) cum_3(\epsilon)^2 \right. \\
&\quad \left. + 6\left(\bar{n} - \frac{1}{2}\right) \text{Var}(\epsilon) \text{Var}(\epsilon^2) \right\} \\
&\quad + 48 \frac{1}{K^2} E(\epsilon^4) \int_0^T \sigma_t^2 dt + \frac{96}{3} \frac{1}{n} E(\epsilon^2) T \int_0^T \sigma_t^4 dt \\
&\quad + \frac{44}{10} \left(\frac{K}{n}\right)^2 T^2 \int_0^T \sigma_t^6 dt + \text{smaller terms}
\end{aligned} \tag{4.30}$$

$$\begin{aligned}
cum_4(\langle Y, Y \rangle_T^{(avg)} - \langle X, X \rangle_T) &= 16 \frac{\bar{n}}{K^3} \{ cum_4(\epsilon^2) + (E\epsilon^4)^2 - 3(E\epsilon^2)^4 + 12\text{Var}(\epsilon^2)E\epsilon^4 \\
&\quad - 32E\epsilon^3 \text{cov}(\epsilon^2, \epsilon^3) + 24E\epsilon^2 (E\epsilon^3)^2 + 12cum_3(\epsilon^2)E\epsilon^2 \} + O\left(\frac{1}{K^3} E|\epsilon|^8\right) \\
&\quad + 192 \frac{1}{K^3} \{ cum_3(\epsilon^2) - 7 cum_3(\epsilon)^2 + 6 \text{Var}(\epsilon) \text{Var}(\epsilon^2) \} \int_0^T \sigma_t^2 dt + O\left(\frac{1}{nK^2} E|\epsilon|^6\right) \\
&\quad + 256 \frac{1}{nK} (E\epsilon^4 + (\text{Var}(\epsilon))^2) T \int_0^T \sigma_t^4 dt + o\left(\frac{1}{nK} E|\epsilon|^4\right) \\
&\quad + \frac{2112}{10} \frac{K}{n^2} \text{Var}(\epsilon) T^2 \int_0^T \sigma_t^6 dt + o\left(\frac{K}{n^2} E|\epsilon|^2\right) \\
&\quad + \frac{1888}{105} \left(\frac{K}{n}\right)^3 T^3 \int_0^T \sigma_t^8 dt + o\left(\left(\frac{K}{n}\right)^3\right) + \text{smaller terms}
\end{aligned} \tag{4.31}$$

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

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

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

$$cum_3([Y, Y]_T^{(\text{avg, opt})} - \langle X, X \rangle_T) = \frac{22}{5} \left(\frac{K}{n} \right)^2 T^2 \int_0^T \sigma_t^6 dt + o\left(\left(\frac{K}{n} \right)^2 \right),$$

and

$$cum_4([Y, Y]_T^{(\text{avg, opt})} - \langle X, X \rangle_T) = \frac{1888}{105} \left(\frac{K}{n} \right)^3 T^3 \int_0^T \sigma_t^8 dt + o\left(\left(\frac{K}{n} \right)^3 \right)$$

respectively.

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

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

hence, we have that

$$cum_3 \left((E\epsilon^2)^{-1/3} ([Y, Y]_T^{(\text{avg, opt})} - \langle X, X \rangle_T) \right) = O((E\epsilon^2)^{1/3}) \rightarrow 0, \quad (4.33)$$

$$cum_4 \left((E\epsilon^2)^{-1/3} ([Y, Y]_T^{(\text{avg, opt})} - \langle X, X \rangle_T) \right) = O((E\epsilon^2)^{2/3}) \rightarrow 0, \quad (4.34)$$

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

By comparing to the expression for the sparse case, it is clear that the average volatility is, in the sense of order of convergence, as close, but no closer, to normal than the sparsely sampled volatility.

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

$$\text{cum}_3(\widehat{\langle X, X \rangle}_T - \langle X, X \rangle_T) = \frac{22}{5} \left(\frac{K}{n}\right)^2 T^2 \int_0^T \sigma_t^6 dt + o\left(\left(\frac{K}{n}\right)^2\right), \quad (4.35)$$

and

$$\text{cum}_4(\widehat{\langle X, X \rangle}_T - \langle X, X \rangle_T) = \frac{1888}{105} \left(\frac{K}{n}\right)^3 T^3 \int_0^T \sigma_t^8 dt + o\left(\left(\frac{K}{n}\right)^3\right). \quad (4.36)$$

However, the value of K is quite different for TSRV than for the averaging volatility estimator. When σ is constant, it is shown in Section 4 of Zhang et al. (2002) that for TSRV, the optimal choice of K is given by

$$K = \left(\frac{16(E\epsilon^2)^2}{T E\eta^2}\right)^{1/3} n^{2/3}. \quad (4.37)$$

As is seen from Table 1, this choice of K gives radically different approximate values 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.

4.5 The Failure of Ordinary Edgeworth Expansions

The development in this paper is based on the assumption that the size of ϵ goes to zero as $n \rightarrow \infty$. This is an unusual assumption. One would normally develop asymptotics as $n \rightarrow \infty$ for fixed size of ϵ . We here demonstrate that when one uses fixed ϵ asymptotics to produce leading terms in cumulants, the resulting expansion will fail to produce an accurate correction to the normal distribution.

Take the sparse case. If ϵ is fixed, one obtains in analogy with Proposition 3 that

$$\text{cum}_3([Y, Y]_T - \langle X, X \rangle_T) = c_3(n) + 48E(\epsilon^4) \int_0^T \sigma_t^2 dt + O(n^{-1/2}), \quad (4.38)$$

and

$$cum_4([Y, Y]_T - \langle X, X \rangle_T) = c_4(n) + 24n^{-1}c_3(n) \int_0^T \sigma_t^2 dt + O(n^{-1/2}), \quad (4.39)$$

where $c_3(n)$ and $c_4(n)$ are given in Lemmas 1-2. Compare these expressions to formulas (4.11)-(4.12), and note that not the same terms are included. In particular, the leading (order $O(n)$) terms in the above equations is not even present in (4.11)-(4.12). It is easy to see that similar results hold in the average RV case. It is, therefore, as if, in some cases, the asymptotics should naturally be done with the size of ϵ going to zero.

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 basic model $dX_t = (\mu - \sigma^2/2) dt + \sigma dB_t$ at a time interval $\Delta t = 1$ second, with parameter values $\mu = 0.05$ and $\sigma^2 = 0.04$. As for the market microstructure noise ϵ , we assume that it is Gaussian and small. Specifically, we set $(E\epsilon^2)^{1/2} = 0.0005$ (i.e., the standard deviation of the noise is 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, finally, 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 now report our simulation results in Table 1 and Figures 1 -10. 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 (4.13) 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

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

As discussed in Section 3.3, Figures 1- 5 show the QQ plots (against the normal distribution) of the standardized distribution of the five estimators before the Edgeworth correction is conducted.

We now also 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 Figures 6 -10, and in a different form in Table 1. 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 straight (uncorrected) Normal asymptotic distribution; “Asymptotic (Edgeworth)” refers to the value predicted by our theory (the asymptotic cumulants are given up to the approximation given in the previous section).

An inspection of Table 1 suggests that our expansion theory provides a good approximation to all four moments of the small sample distribution in each estimation scheme. Comparing different columns in Table 1, we also do not see substantial differences across estimators. On the other hand, distribution-wise, all five estimators display very different properties relative to the standard normal. This is especially so given the sample size n .

Finally, Figures 6-10 convey the similar message as in Table 1. In each figure, the histogram displays 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 distribution of $N(0,1)$. In the “all” and TSRV cases, these last two curves are indistinguishable.

In summary, examination of these figures confirms that the sample distributions of all five estimators conform to our Edgeworth expansion, while (except in the “all” and TSRV cases), the normal approximation is somewhat off.

6 Conclusions

We have here 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. This is that the better expansion is obtained by using as 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 3rd and 4th cumulant, but the same technology can potentially be used for higher order cumulants.

References

- AÏT-SAHALIA, Y., P. A. MYKLAND, AND L. ZHANG (2005a): “How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise,” *Review of Financial Studies*, 18, 351–416.
- (2005b): “Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise,” Tech. rep., Princeton University.
- ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2001): “The Distribution of Exchange Rate Realized Volatility,” *Journal of the American Statistical Association*, 96, 42–55.
- BANDI, F. M. AND J. R. RUSSELL (2003): “Microstructure Noise, Realized Volatility and Optimal Sampling,” Tech. rep., University of Chicago Graduate School of Business.
- BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2004): “Regular and Modified Kernel-Based Estimators of Integrated Variance: The Case with Independent Noise,” Tech. rep., Department of Mathematical Sciences, University of Aarhus.
- BARNDORFF-NIELSEN, O. E. AND N. SHEPHARD (2002): “Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models,” *Journal of the Royal Statistical Society, B*, 64, 253–280.
- BICKEL, P. J., F. GÖTZE, AND W. R. VAN ZWET (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.
- DACOROGNA, M. M., R. GENÇAY, U. MÜLLER, R. B. OLSEN, AND O. V. PICTET (2001): *An Introduction to High-Frequency Finance*, San Diego: Academic Press.
- FELLER, W. (1971): *An Introduction to Probability Theory and Its Applications, Volume 2*, New York: John Wiley and Sons.
- GONCALVES, S. AND N. MEDDAHI (2005): “Bootstrapping realized volatility,” Tech. rep., Université de Montréal.
- HALL, P. (1992): *The bootstrap and Edgeworth expansion*, New York: Springer.
- HANSEN, P. R. AND A. LUNDE (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. AND P. PROTTER (1998): “Asymptotic Error Distributions for the Euler Method for Stochastic Differential Equations,” *Annals of Probability*, 26, 267–307.
- MCCULLAGH, P. (1987): *Tensor Methods in Statistics*, London, U.K.: Chapman and Hall.
- 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.
- (1994): “Bartlett type identities for martingales,” *Annals of Statistics*, 22, 21–38.
- (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.
- MYKLAND, P. A. AND L. ZHANG (2002): “ANOVA for Diffusions,” *The Annals of Statistics*, forthcoming, –, –.
- 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.
- ZHANG, L. (2001): “From Martingales to ANOVA: Implied and Realized Volatility,” Ph.D. thesis, The University of Chicago, Department of Statistics.
- (2004): “Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach,” Tech. rep., Carnegie-Mellon University.
- ZHANG, L., P. A. MYKLAND, AND Y. AÏT-SAHALIA (2002): “A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data,” *Journal of the American Statistical Association*, forthcoming, –, –.
- ZHOU, B. (1996): “High-Frequency Data and Volatility in Foreign-Exchange Rates,” *Journal of Business & Economic Statistics*, 14, 45–52.

Appendix: Proofs

A Proof of Lemma 1

Let a_i be defined as in (B.1). We can then write

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

where

$$cum\left(\sum_{i=0}^n a_i \epsilon_{t_i}^2, \sum_{j=0}^n a_j \epsilon_{t_j}^2, \sum_{k=0}^{n-1} \epsilon_{t_k} \epsilon_{t_{k+1}}\right) = 2 \sum_{k=0}^{n-1} a_k a_{k+1} cum(\epsilon_{t_k}^2, \epsilon_{t_{k+1}}^2, \epsilon_{t_k} \epsilon_{t_{k+1}}) \tag{A.2}$$

$$= 2(n-1)(E\epsilon^3)^2 \tag{A.3}$$

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,

$$cum\left(\sum_{i=0}^n a_i \epsilon_{t_i}^2, \sum_{j=0}^{n-1} \epsilon_{t_j} \epsilon_{t_{j+1}}, \sum_{k=0}^{n-1} \epsilon_{t_k} \epsilon_{t_{k+1}}\right) = 2 \sum_{j=0}^{n-1} a_j cum(\epsilon_{t_j}^2, \epsilon_{t_j} \epsilon_{t_{j+1}}, \epsilon_{t_j} \epsilon_{t_{j+1}}) \tag{A.4}$$

$$= 2\left(n - \frac{1}{2}\right)(E\epsilon^2)Var(\epsilon^2) \tag{A.5}$$

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,

$$cum\left(\sum_{i=0}^{n-1} \epsilon_{t_i} \epsilon_{t_{i+1}}, \sum_{j=0}^{n-1} \epsilon_{t_j} \epsilon_{t_{j+1}}, \sum_{k=0}^{n-1} \epsilon_{t_k} \epsilon_{t_{k+1}}\right) = \sum_{i=0}^{n-1} cum_3(\epsilon_{t_i} \epsilon_{t_{i+1}}) = n(E\epsilon^3)^2, \tag{A.6}$$

$$cum\left(\sum_{i=0}^n a_i \epsilon_{t_i}^2, \sum_{j=0}^n a_j \epsilon_{t_j}^2, \sum_{k=0}^n a_k \epsilon_{t_k}^2\right) = \sum_{i=0}^n a_i^3 cum_3(\epsilon_{t_i}^2) = \left(n - \frac{3}{4}\right)cum_3(\epsilon^2), \tag{A.7}$$

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

Inserting (A.2)-(A.7) in (A.1) yields (4.3).

B Proof of Proposition 1

To proceed, define

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

and

$$b_i = \begin{cases} \Delta X_{t_{i-1}} - \Delta X_{t_i} & \text{if } 1 \leq i \leq n-1 \\ \Delta X_{t_{n-1}} & \text{if } i = n \\ -\Delta X_{t_0} & \text{if } i = 0 \end{cases} \quad (\text{B.2})$$

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

Then it follows that

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

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

Also

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

Finally,

$$\begin{aligned} cum_3([X, \epsilon]_T | X) &= \sum_{i=0}^n b_i^3 cum_3(\epsilon) \\ &= E(\epsilon^3) \left[-3 \sum_{i=1}^{n-1} (\Delta X_{t_{i-1}})^2 (\Delta X_{t_i}) + 3 \sum_{i=1}^{n-1} (\Delta X_{t_{i-1}}) (\Delta X_{t_i})^2 \right] \\ &= O_p(n^{-1/2} E[|\epsilon|^3]) \end{aligned} \quad (\text{B.5})$$

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

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

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

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 $\sum_{i=0}^{n-1} a_i a_{i+1} = n - 1$.

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

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

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

$$cum_4\left(\sum_{i=0}^n a_i \epsilon_{t_i}^2\right) = \sum_{i=0}^n a_i^4 cum_4(\epsilon^2) = (n - \frac{7}{8}) cum_4(\epsilon^2). \tag{C.5}$$

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

$$\begin{aligned}
c_4(n) &= cum_4\left(2 \sum_{i=0}^n a_i \epsilon_{t_i}^2 - 2 \sum_{i=0}^{n-1} \epsilon_{t_i} \epsilon_{t_{i+1}}\right), \\
&= 16[cum_4\left(\sum_{i=0}^n a_i \epsilon_{t_i}^2\right) + cum_4\left(\sum_{i=0}^{n-1} \epsilon_{t_i} \epsilon_{t_{i+1}}\right) \\
&\quad - \binom{4}{1} cum\left(\sum_{i=0}^n a_i \epsilon_{t_i}^2, \sum_{j=0}^n a_j \epsilon_{t_j}^2, \sum_{k=0}^n a_k \epsilon_{t_k}^2, \sum_{l=0}^{n-1} \epsilon_{t_l} \epsilon_{t_{l+1}}\right) \\
&\quad - \binom{4}{3} cum\left(\sum_{i=0}^n a_i \epsilon_{t_i}^2, \sum_{j=0}^{n-1} \epsilon_{t_j} \epsilon_{t_{j+1}}, \sum_{k=0}^{n-1} \epsilon_{t_k} \epsilon_{t_{k+1}}, \sum_{l=0}^{n-1} \epsilon_{t_l} \epsilon_{t_{l+1}}\right) \\
&\quad + \binom{4}{2} cum\left(\sum_{i=0}^n a_i \epsilon_{t_i}^2, \sum_{j=0}^n a_j \epsilon_{t_j}^2, \sum_{k=0}^{n-1} \epsilon_{t_k} \epsilon_{t_{k+1}}, \sum_{l=0}^{n-1} \epsilon_{t_l} \epsilon_{t_{l+1}}\right)] \\
&= 16\left\{\left(n - \frac{7}{8}\right) cum_4(\epsilon^2) + n(E\epsilon^4)^2 - 3n(E\epsilon^2)^4 + 12(n-1)var(\epsilon^2)E\epsilon^4 \right. \\
&\quad \left. - 32\left(n - \frac{17}{16}\right)E\epsilon^3 cov(\epsilon^2, \epsilon^3) + 24\left(n - \frac{7}{4}\right)E\epsilon^2(E\epsilon^3)^2 + 12\left(n - \frac{3}{4}\right)cum_3(\epsilon^2)E\epsilon^2\right\}
\end{aligned} \tag{C.6}$$

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

D Proof of Proposition 2

It remains to deal with the second term in equation (4.5),

$$\begin{aligned}
& cum([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, [X, \epsilon]_T, [X, \epsilon]_T | X) \\
&= \sum_{i,j} b_i b_j cum([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, \epsilon_{t_i}, \epsilon_{t_j}) \\
&= \sum_i b_i^2 cum([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, \epsilon_{t_i}, \epsilon_{t_i}) \tag{D.1}
\end{aligned}$$

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

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

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

Now following the two identities:

$$\begin{aligned}
cum([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, \epsilon_i, \epsilon_i) &= cum_3([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, \epsilon_i^2) \\
&\quad - 2(Cov([\epsilon, \epsilon]_T, \epsilon_i))^2 \\
cum([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, \epsilon_i, \epsilon_{i+1}) &= cum_3([\epsilon, \epsilon]_T, [\epsilon, \epsilon]_T, \epsilon_i \epsilon_{i+1}) \\
&\quad - 2Cov([\epsilon, \epsilon]_T, \epsilon_i)Cov([\epsilon, \epsilon]_T, \epsilon_{i+1}),
\end{aligned}$$

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

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

Hence, (D.1) becomes

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

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,

$$\begin{aligned}
 cum_3(D_T) &= 3cov(D_T, \langle D, D \rangle_T) \text{ by the third Bartlett identity} \\
 &= 3cov(D_T, \int_0^T Z_s^2 ds) \\
 &= 3 \int_0^T cov(D_T, Z_s^2) ds
 \end{aligned} \tag{E.1}$$

To compute the integrand,

$$\begin{aligned}
 cov(D_T, Z_s^2) &= cov(D_s, Z_s^2) \text{ since } D_t \text{ is a martingale} \\
 &= cum_3(D_s, Z_s, Z_s) \text{ since } ED_s = EZ_s = 0 \\
 &= cum_3(D_s, Z_s^{(s)}, Z_s^{(s)}) \\
 &= 2cov(Z_s^{(s)}, \langle D, Z^{(s)} \rangle_s) + cov(D_s, \langle Z^{(s)}, Z^{(s)} \rangle_s) \text{ by the third Bartlett identity} \\
 &= 2cov(Z_s, \int_0^s Z_u f(s, u) du) \text{ by (4.22) and since } \langle Z^{(s)}, Z^{(s)} \rangle_s \text{ is nonrandom} \\
 &= 2 \int_0^s cov(Z_s, Z_u) f(s, u) du \\
 &= 2 \int_0^s du \int_0^u f(s, u) f(s, t) f(u, t) dt
 \end{aligned} \tag{E.2}$$

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

$$cum_3(D_T^{(1)}, D_T^{(2)}, D_T^{(3)}) = 2 \int_0^T ds \int_0^s cov(Z_s^{(1)}, Z_u^{(2)}) f^{(3)}(s, u) du [3], \tag{E.3}$$

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

For the fourth cumulant,

$$cum_4(D_T) = -3cov(\langle D, D \rangle_T, \langle D, D \rangle_T) + 6cum_3(D_T, D_T, \langle D, D \rangle_T), \tag{E.4}$$

by the fourth Bartlett identity. For the first term

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

For the other term in (E.4)

$$cum_3(D_T, D_T, \langle D, D \rangle_T) = \int_0^T cum_3(D_T, D_T, Z_s^2) ds$$

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 $t \leq a \leq s$)).

$$\begin{aligned}
cum_3(D_T, D_T, \langle D, D \rangle_T) &= 4 \int_0^T ds \int_0^s dx \int_0^x du \int_0^u dt (f(x, u) f(x, t) f(s, u) f(s, t) \\
&\quad + f(x, u) f(u, t) f(s, x) f(s, t) + f(x, t) f(u, t) f(s, x) f(s, u)) \tag{E.6}
\end{aligned}$$

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 Mean	0.003	0.002	-0.004	-0.005	0.003
Asymptotic Mean	0	0	0	0	0
Sample Stdev	0.9993	1.001	0.997	0.996	1.015
Asymptotic Stdev	1	1	1	1	1
Sample Skewness	0.028	0.3295	0.425	0.453	0.042
Asymp. Skewness (Normal)	0	0	0	0	0
Asymp. Skewness (Edgeworth)	0.025	0.3294	0.427	0.451	0.035
Sample Kurtosis	3.010	3.162	3.256	3.34	2.997
Asymp. Kurtosis (Normal)	3	3	3	3	3
Asymp. Kurtosis (Edgeworth)	3.001	3.169	3.287	3.25	3.003

Table 1: Monte-Carlo simulations: This table reports the first four sample and asymptotic moments for the five estimators, comparing the asymptotics based on the Normal distribution and our Edgeworth correction.

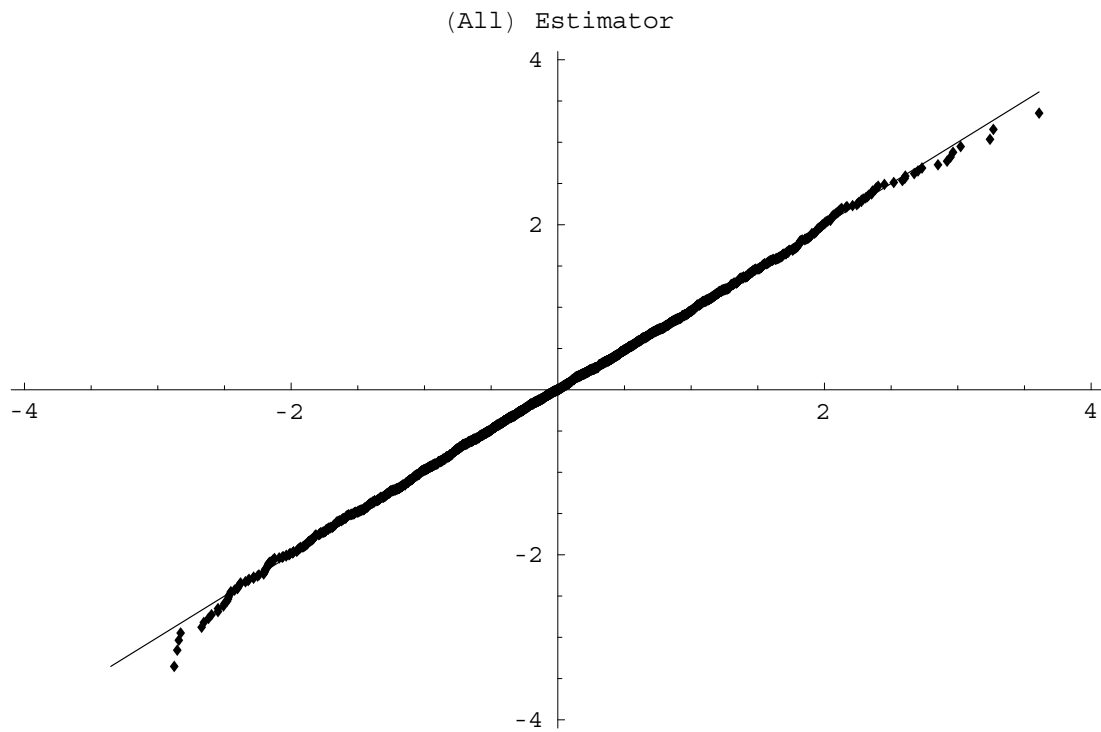


Figure 1: QQ plot for the estimator $[Y, Y]^{(\text{all})}$ based on the asymptotic Normal distribution.

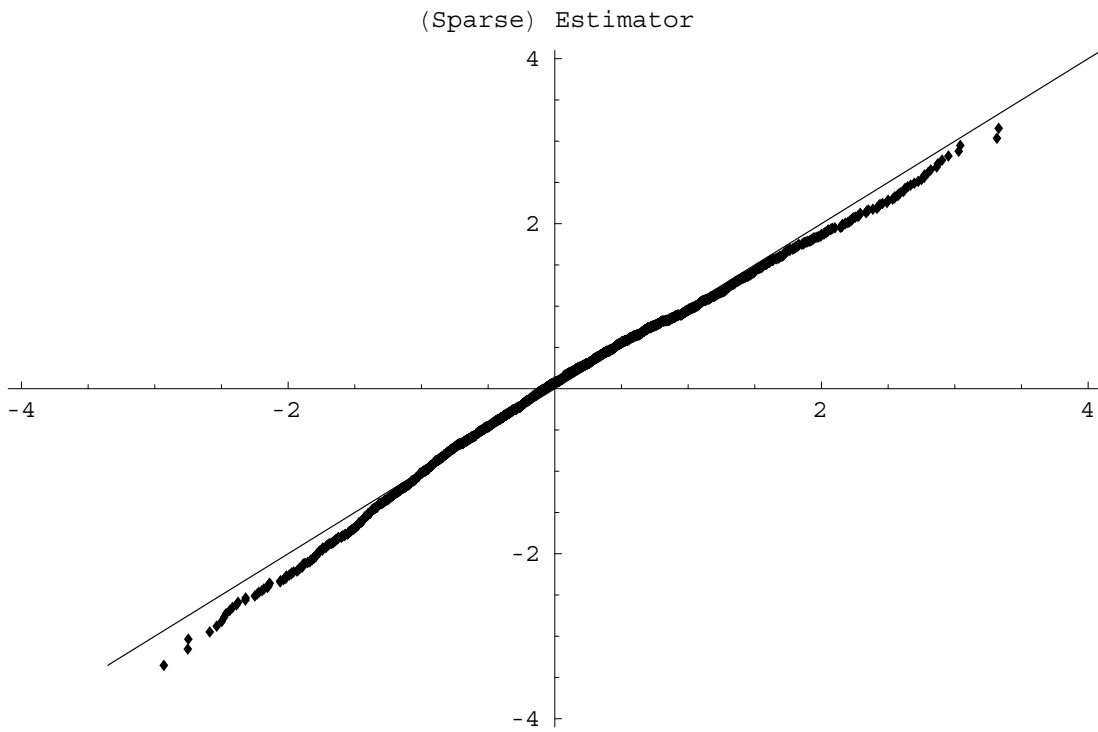


Figure 2: QQ plot for the estimator $[Y, Y]^{(\text{sparse})}$ based on the asymptotic Normal distribution.

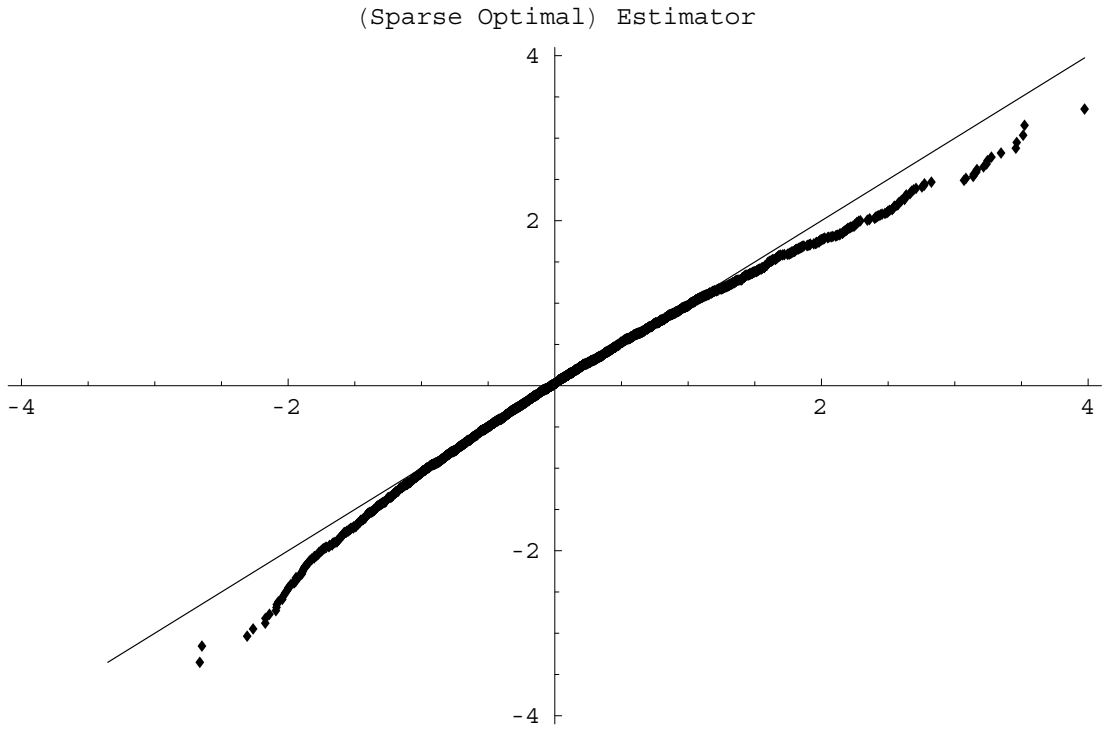


Figure 3: QQ plot for the estimator $[Y, Y]^{(\text{aparse, opt})}$ based on the asymptotic Normal distribution.

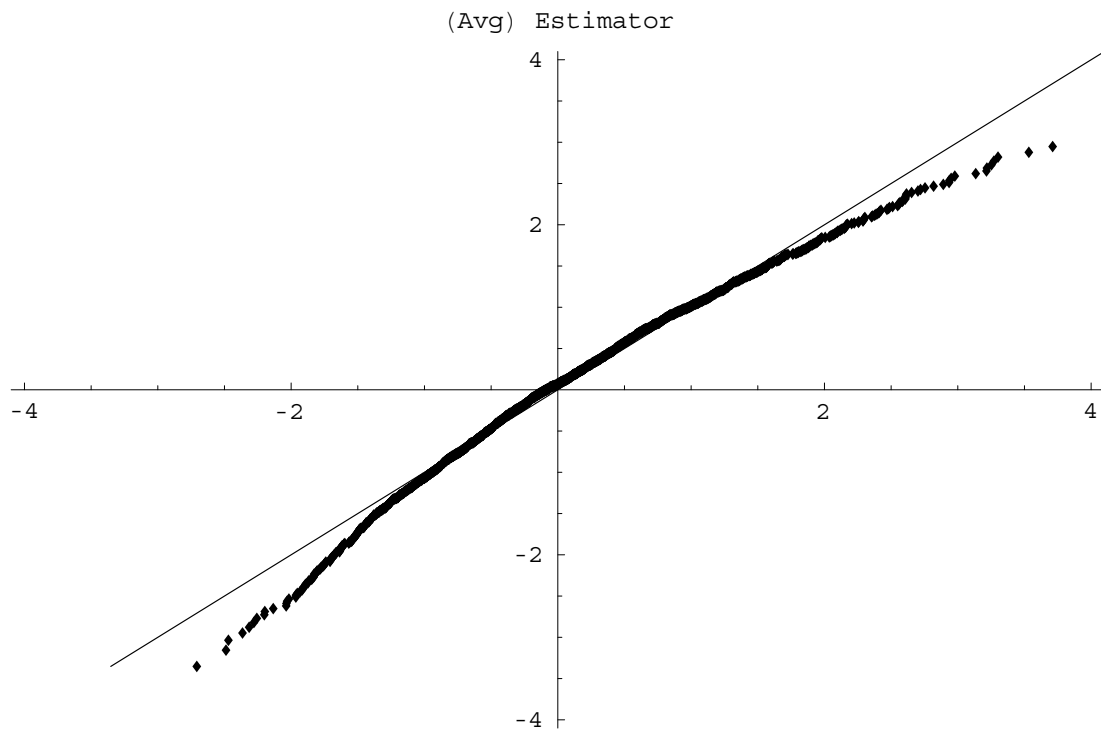


Figure 4: QQ plot for the estimator $[Y, Y]^{(\text{avg})}$ based on the asymptotic Normal distribution.

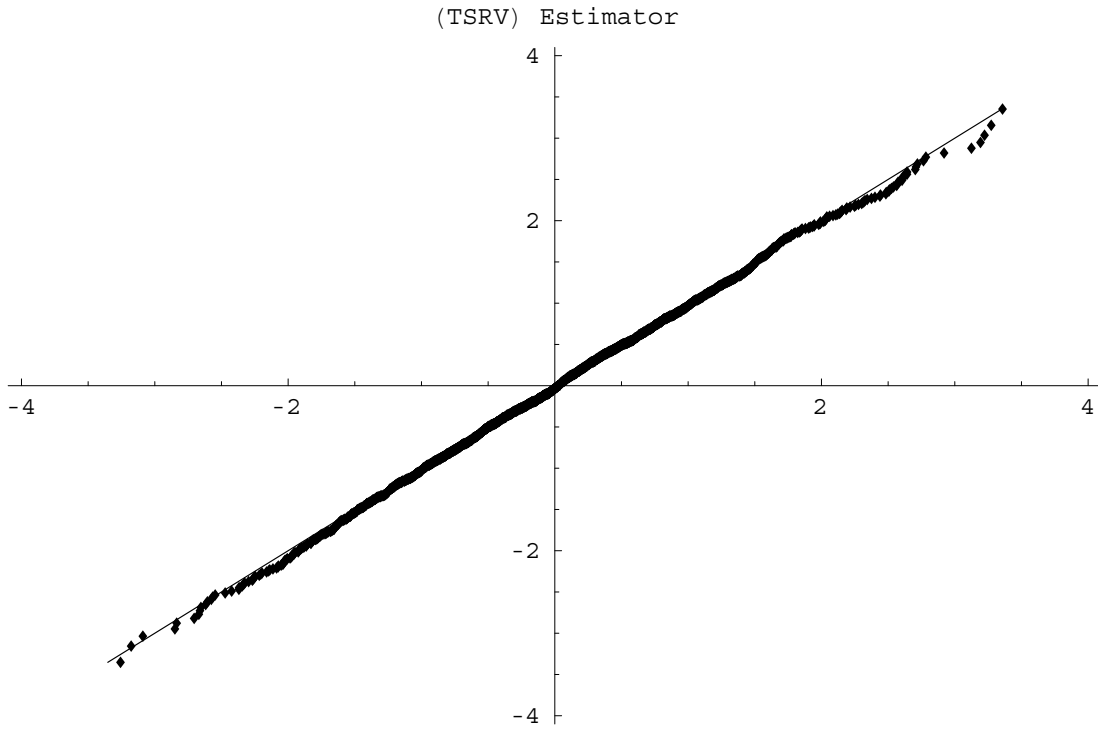


Figure 5: QQ plot for the TSRV estimator $\widehat{\langle X, X \rangle}$ based on the asymptotic Normal distribution.

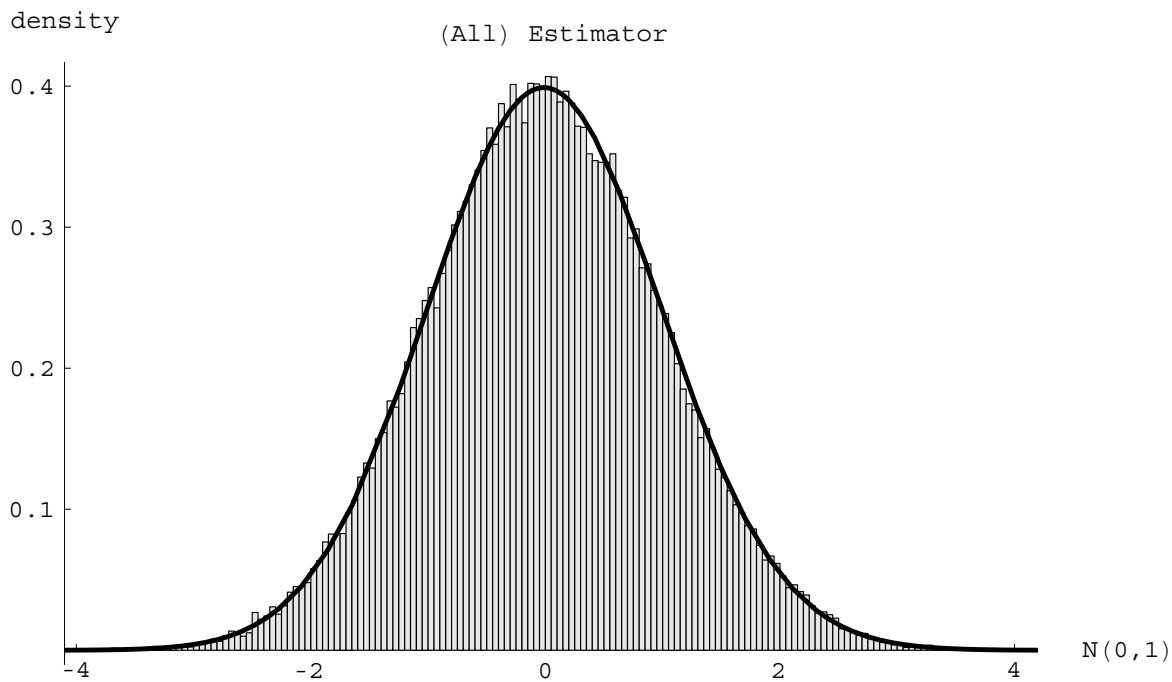


Figure 6: Comparison of the small sample distribution of the $[Y, Y]^{(\text{all})}$ estimator (histogram), the Edgeworth-corrected distribution (solid line) and the asymptotically Normal distribution (dashed line).

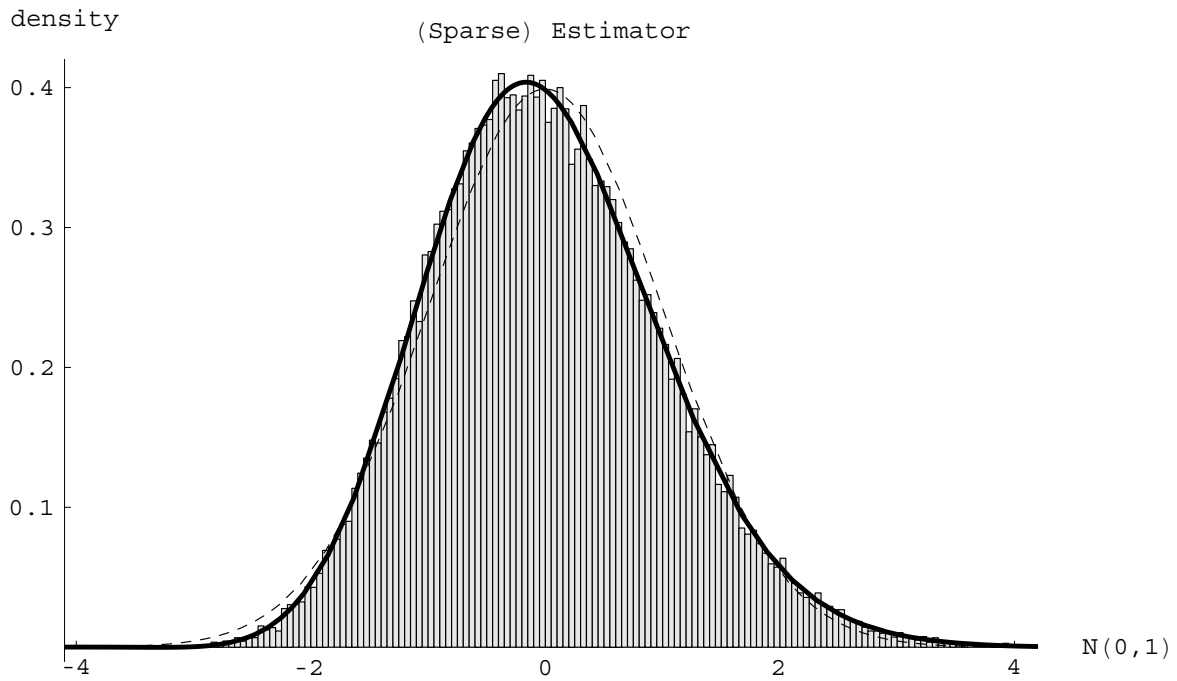


Figure 7: Comparison of the small sample distribution of the $[Y, Y]^{(\text{sparse})}$ estimator (histogram), the Edgeworth-corrected distribution (solid line) and the asymptotically Normal distribution (dashed line).

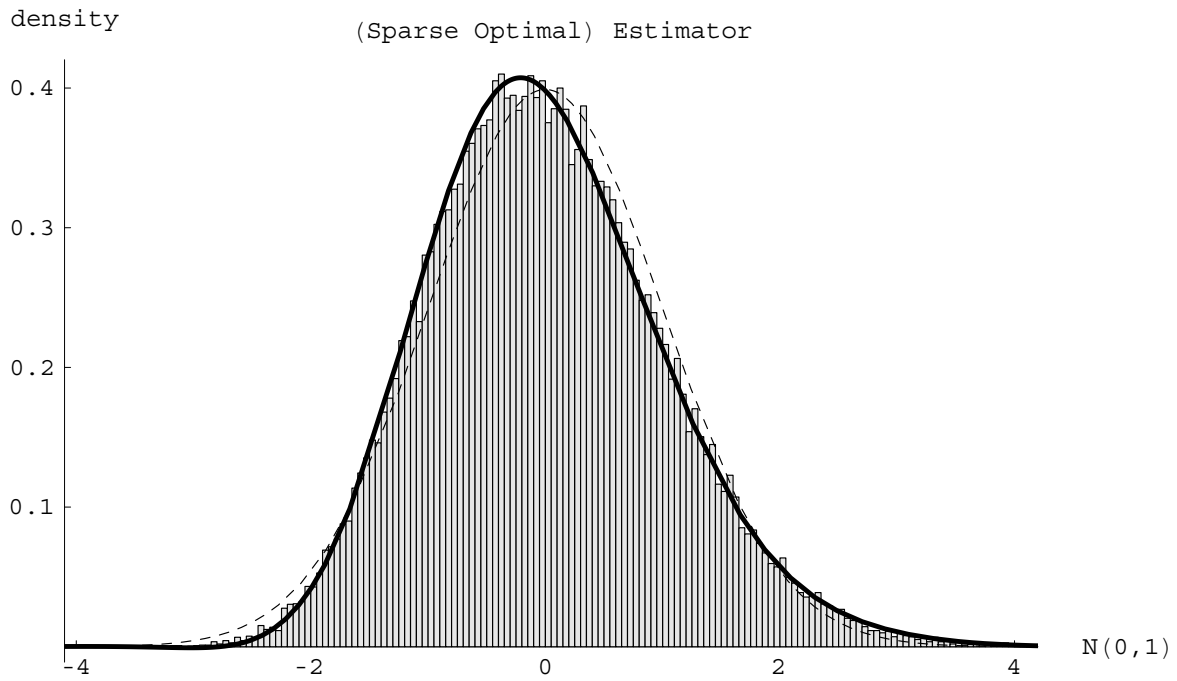


Figure 8: Comparison of the small sample distribution of the $[Y, Y]^{(\text{sparse, opt})}$ estimator (histogram), the Edgeworth-corrected distribution (solid line) and the asymptotically Normal distribution (dashed line).

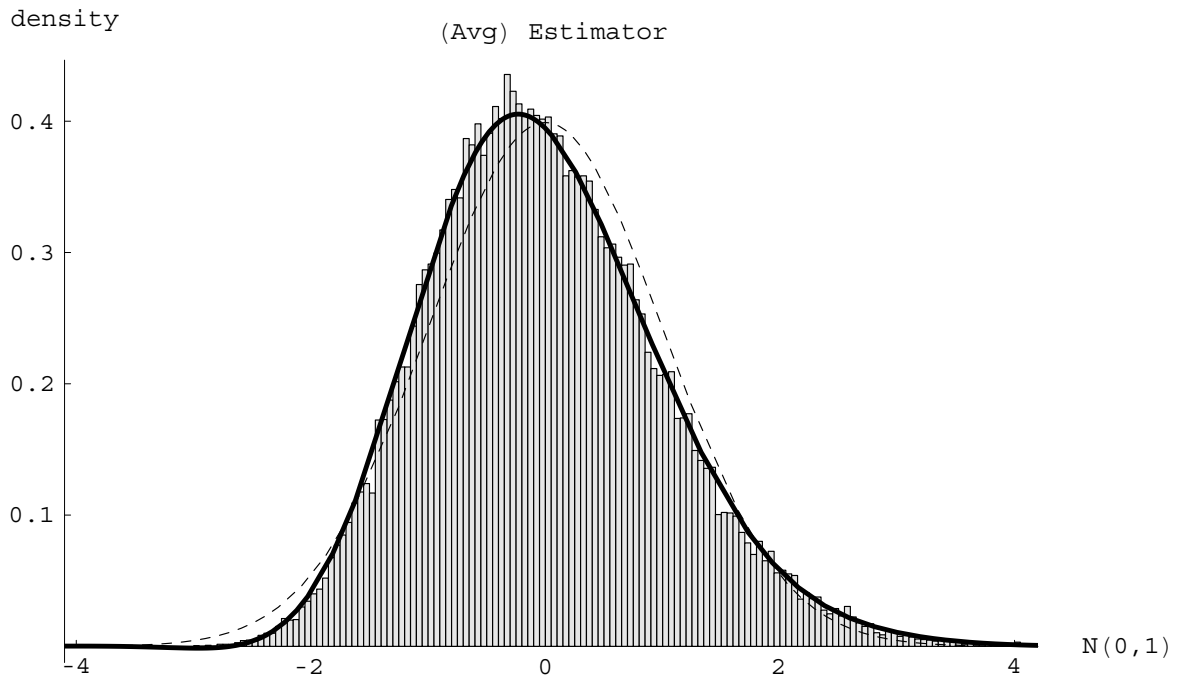


Figure 9: Comparison of the small sample distribution of the $[Y, Y]^{(\text{avg})}$ estimator (histogram), the Edgeworth-corrected distribution (solid line) and the asymptotically Normal distribution (dashed line).

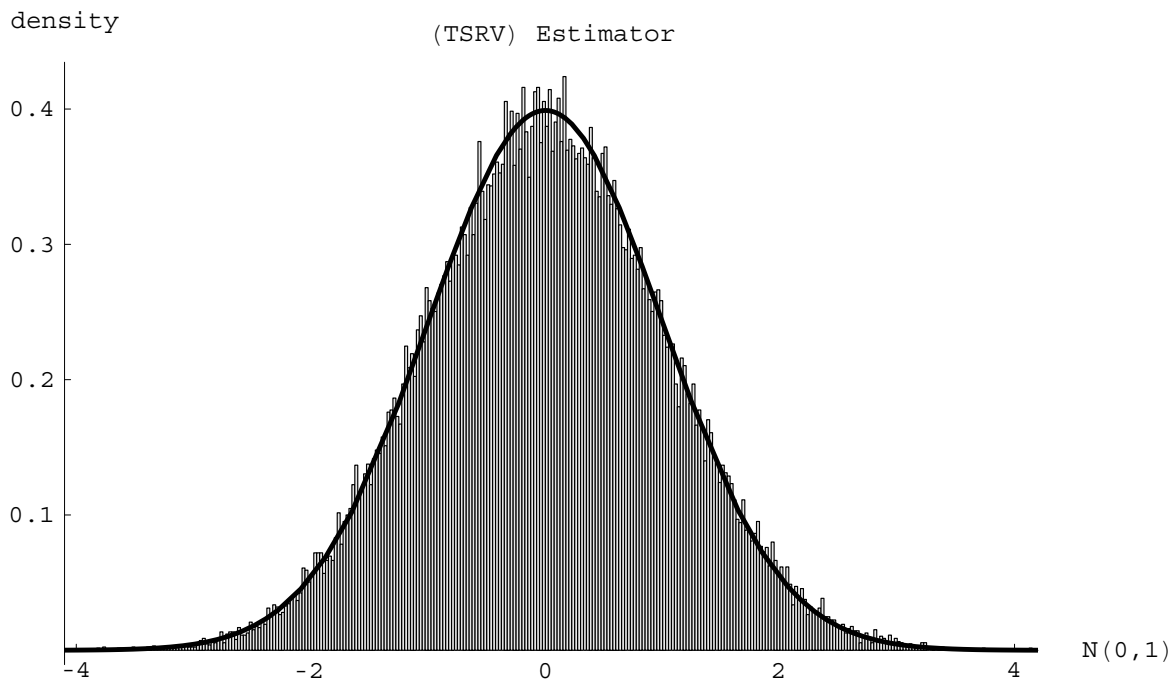


Figure 10: Comparison of the small sample distribution of the TSRV estimator (histogram), the Edgeworth-corrected distribution (solid line) and the asymptotically Normal distribution (dashed line).