The Double Gaussian Approximation for High Frequency Data

PER A. MYKLAND

Department of Statistics, The University of Chicago

LAN ZHANG

Department of Finance, University of Illinois at Chicago

This version: March 5, 2011

Abstract

High frequency data have become an important feature of many areas of research. They permit the creation of estimators in highly non-parametric classes of continuous-time models. In the context of continuous semi-martingale models, we here provide a locally parametric "double Gaussian" approximation, to facilitate the analysis of estimators. As in Mykland and Zhang (2009), the error in the approximation can be offset with a post-asymptotic likelihood correction. The current approximation is valid in large neighborhoods, permitting a sharp analysis of estimators that use local behavior over asymptotically increasing numbers of observations.

KEYWORDS: consistency, cumulants, contiguity, continuity, discrete observation, efficiency, equivalent martingale measure, Itô process, leverage effect, likelihood inference, partial likelihood, quarticity, realized beta, realized volatility, stable convergence, volatility of volatility.

1   Introduction

The purpose of this paper is to investigate the size of neighborhoods under which a discretely observed semimartingale can be taken to have a simple parametric form. The problem arises in high frequency data, financial or otherwise. We have previously investigated this problem in Mykland and Zhang (2009), where it is shown that the data can be taken to be conditionally normal over windows of finitely many observations, with a post-asymptotic likelihood ratio adjustment. The background is the rapidly growing literature on high frequency data in econometrics, and we refer to our earlier paper for a review of the literature. See also Sections 2.2 and 2.4 for further background.

In this paper we show that a more complex parametric structure, the *Double Gaussian* model, can approximate the distribution of the data in much larger sets of observations. Specifically, if there are n observations, the approximation is good in neighborhoods of size $O_p(n^{1/2})$ observations. It also provides a structure which is conditional on the volatility, which is desirable in a number of settings.

The need for easy calculation in neighborhoods of increasing size occurs because many estimators are either not efficient (integrals of powers of volatility, ANOVA, realized betas) or even not consistent (leverage effect, volatility of volatility (see Section 2.4), and most estimators when there is microstructure[1] in the data) in finite neighborhoods, cf. the discussion in our earlier paper. Neighborhoods of size $O_p(n^{1/2})$ often provide the critical trade-off, or the emergence of otherwise unseen bias and variance terms. For references in this direction, see the literature on microstructure (such as Zhang, Mykland, and Aït-Sahalia (2005), Zhang (2006), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), Jacod, Li, Mykland, Podolskij, and Vetter (2009), Podolskij and Vetter (2009), and Reiss (2010)), as well as Mykland, Shephard, and Sheppard (2011) and Wang and Mykland (2011). When neighborhoods of observations are finite or of size $o_p(n^{1/2})$, there is no contiguity adjustment. When using blocks of size $O_p(n^{1/2})$, the post-asymptotic adjustment is quite simple. A worked example is provided in Section 5.

As in Mykland and Zhang (2009), the setting for the approximation is partial likelihood (Cox (1975), Wong (1986)), cf. Remark 4 in Section 4.2. To make for greater comparability, we have kept the notation from the earlier paper whenever possible. The model is defined in Section 2, the double Gaussian approximation is introduced and discussed in Section 3, and the main theorem is given in Section 4. We apply the results to the estimation of quarticity in Section 5.

2 Setting

2.1 The Data Generating Mechanism and its Representations

We shall work with a broad class of continuous semimartingales, namely second order *Itô processes*. This is to say that the primary observed process X_t satisfies

$$dX_t = \mu_t dt + \sigma_t dW_t, \text{ where}$$
$$d\sigma_t = \sigma_t^{\mathrm{DR}} dt + f_t dW_t + g_t dB_t, \tag{1}$$

and where μ_t, f_t, g_t, and σ_t^{DR} are adapted locally bounded random processes, and W_t and B_t are independent Wiener processes. The underlying filtration will be called $(\mathcal{F}_t)$. The probability distribution will be called P.

It is handy to immediately rewrite the mechanism (1) in a couple of ways. First of all, by standard orthogonal transformation, an equivalent system is given by

$$dX_t = \mu_t dt + \sigma_t \left(a_t dZ_t + b_t dU_t \right)$$
$$d\sigma_t = \sigma_t^{\mathrm{DR}} dt + c_t dZ_t, \tag{2}$$

where Z_t and U_t are, again, independent Wiener processes, and where

$$c_t = (f_t^2 + g_t^2)^{1/2} \ , \quad a_t = f_t/c_t \ , \text{ and } b_t = g_t/c_t. \tag{3}$$

Also, under regularity conditions[2], there is an equivalent "statistical risk neutral measure" (Section 2.2 in Mykland and Zhang (2009)) P^* under which

$$Z_t^* = Z_t + \int_0^t \frac{\sigma_s^{\mathrm{DR}}}{c_s} ds \text{ and } U_t^* = U_t + \int_0^t \frac{1}{b_s} \left(\mu_s - \frac{a_s \sigma_s \sigma_s^{\mathrm{DR}}}{c_s} \right) ds$$

are independent Wiener processes. System (2) can thus further be written as

$$dX_t = \sigma_t \left(a_t dZ_t^* + b_t dU_t^* \right),$$
$$d\sigma_t = c_t dZ_t^*. \tag{4}$$

The representation (4) is the central one for our development. Since all convergence in this paper will be of the stable type,[3] one can show that all our asymptotic results will apply with suitable modification to the systems (2) and (1).

2.2 What Kind of Processes are Covered by this Development?

Continuous time processes are widely used in economics and finance. A main milestone was the development of the Black and Scholes (1973)-Merton (1973) options pricing (and trading) formula,

and subsequent developments in this direction, which have practically become a separate area of study. Another main application of such processes is the theory of optimal consumption, and market equilibrium, see for example Merton (1971, 1992), He and Leland (1993), and Hansen and Sargent (2007).

In the earliest models, σ is constant, but over time, both (longitudinal) econometrics and (cross-sectional) empirical options pricing found that σ_t will typically be time varying. An early deviation from constant σ was provided by Heston (1993), with the model

$$dX_t = \left(\nu - \frac{\sigma_t^2}{2}\right) dt + \sigma_t dW_t$$
$$d\sigma_t^2 = \kappa(\alpha - \sigma_t^2)dt + \gamma\sigma_t d(\rho W_t + (1 - \rho^2)^{1/2} B_t) \ . \tag{5}$$

This model has the emblematic feature that σ_t can vary, but it is mean reverting (asymptotically stationary). This is a common feature of most models for volatility. More general models of this type would have the form of a two variable (X, σ) Markov process, driven by two Brownian motions. This can be extended, of course, to arbitrary dimension. Such models are popular for explaining the prices of derivative securities.

At the same time, on the statistical and econometric side, it was also found that σ could be time varying. This was first introduced in a time series setting by Engle (1982, 2000), and Bollerslev (1986). Gradually, the interface with continuous processes was developed. Particularly influential were Andersen and Bollerslev (1998), Barndorff-Nielsen and Shephard (2001, 2002), Meddahi (2001), and subsequent work by these and other authors. Unlike the options pricing and equilibrium literature, the econometric literature has been more agnostic when it comes to specific parametric form of the price process, but a substantial amount of work falls within the framework of model (1).

A popular (but mostly unrealistic) model is the pure "state space" model $dX_t = \nu(X_t)dt + \gamma(X_t)dW_t$. This model falls under description (1), but is only driven by one Brownian motion. In this case $g_t \equiv 0$, and hence the measure change which makes both X_t and σ_t into martingales will typically fail. In a sense, this is a singular point of our model, and results for this model have to be argued directly. In high frequency analysis, however, such a direct analysis typically yields that the non-martingale terms in system (1)-(2) are negligible. See Jacod and Protter (1998) and Mykland and Zhang (2006) for examples of direct argument.

A class not covered by model (1) is where σ_t has long range dependence, for example, being driven by a fractional Brownian motion, as in, *e.g.* Comte and Renault (1998). We conjecture that a similar theory can be built up for this situation, but this is beyond the scope of this paper.

Another class of models not covered by (1) are systems that allow jumps. For the kind of computation discussed here, however, it means that the part of the system with jumps is studied separately. Systems that are pure jump (see, for example, Barndorff-Nielsen and Shephard (2001) and Carr, Geman, Madan, and Yor (2003)) fall outside our model.

2.3 Sampling Times, and High Frequency Data Asymptotics

We shall suppose that the process X_t is observed at times $0 = t_0 < t_1 < ... < t_n = T$.

In asymptotic analysis, we suppose that $t_j = t_{n,j}$ (the additional subscript will sometimes be suppressed). The grids $\mathcal{G}_n = \{0 = t_{n,0} < t_{n,1} < ... < t_{n,n} = T\}$ are *not nested* when n varies. We then do asymptotics as $n \to \infty$, while T is fixed. The basic assumption is that

$$\max_{1 \le i \le n} |t_{n,j} - t_{n,j-1}| = o_p(1). \tag{6}$$

We also suppose that the observation times $t_{n,j}$ are independent of the X_t process. An alternative formulation is thus that the times are conditionally nonrandom, but they are allowed to be irregularly spaced.

We thus preclude dependence between the observation times and the process. Such dependence does appear to exist in some cases, cf. Renault and Werker (2009) and Li, Mykland, Renault, Zhang, and Zheng (2009), but is beyond the scope of this paper. For further discussion of conditions on sampling times, see Sections 3-4 of Mykland and Zhang (2010).

2.4 Quadratic Variation and Covariation; Connection to Volatility and Leverage Effect

For continuous semimartingales $X^{(1)}$, and $X^{(2)}$, their quadratic covariation is given as

$$\begin{aligned}
\langle X^{(1)}, X^{(2)} \rangle_t &= \lim \sum_{t_{n,j+1} \le t} \Delta X^{(1)}_{t_{n,j+1}} \Delta X^{(2)}_{t_{n,j+1}} \\
&= \lim \sum_{t_{n,j+1} \le t} \mathrm{Cov}(\Delta X^{(1)}_{t_{n,j+1}}, \Delta X^{(2)}_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}}),
\end{aligned}$$

where the latter equality depends on regularity conditions, and where the limit is as in (6). If there is discontinuity, the two limits will normally be different. If $X^{(1)} = X^{(2)}$, one refers to the quadratic variation of $X^{(1)}$.[4]

If X_t is given as in (1), then

$$\langle X, X \rangle_t = \int_0^t \sigma_u^2 du.$$

One usually refers to either σ_t or σ_t^2 as the (spot, or instantaneous) volatility of X_t, while $\langle X, X \rangle_t$ is variously called the integrated volatility or integrated variance of X_t. It is customary to use an annualized and square root scale to quote actual numbers. If $[0, T]$ refers to one trading day, then the integrated volatility would usually be given as[5] $\sqrt{250 \times \langle X, X \rangle_T}$, often multiplied by 100 and quoted as "percent volatility". The volatility of a firm's stock is measured in this fashion, where X is the log stock price.

In this parlance, the volatility of volatility is $\langle \sigma, \sigma \rangle_t$, or its derivative, perhaps on the square root scale. The *leverage effect* is some version of the covariation of X_t and σ_t, in spot form ranging from $\langle X, \sigma^2 \rangle_t'$, via $\langle X, \sigma \rangle_t'$ and $\langle X, \log \sigma \rangle_t'$, to a correlation measure $\langle X, \sigma \rangle_t' / \sigma_t \sqrt{\langle \sigma, \sigma \rangle_t'}$. In the Heston model (system (5) in Section 2.2), for example, the correlation measure of the leverage effect is ρ. Tne extreme case of leverage effect is $b_t \equiv 0$, which in the Heston model translates into $\rho = \pm 1$. Both the volatility of volatility and the leverage effect are part of what is normally referred to as volatility risk.

The term leverage effect goes back at least to Black (1976), and originally describes asymmetry in volatility due to financial leverage. A firm can increase the leverage of its share capital by issuing bonds. Even if the value of the firm has a constant volatility, borrowing will induce time varying volatility which is (normally) negatively correlated with the log stock price (see, for example, Section 2.5.4 of Mykland and Zhang (2010)). In more recent studies, the term leverage effect is generally taken to refer to this negative relationship, which can also be explained by, in particular, fear on the part of investors. For a further discussion, see Wang and Mykland (2011).

3 The Double Gaussian Approximation

3.1 Definition

We approximate over K_n windows of the form $(\tau_{n,i-1}, \tau_{n,i}]$, where

$$\mathcal{H}_n = \{0 = \tau_{n,0} < \tau_{n,1} < ... < \tau_{n,K_n} = T\} \subseteq \mathcal{G}_n. \tag{7}$$

Also set

$$M_{n,i} = \#\{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}]\} = \text{ number of intervals } (t_{n,j-1}, t_{n,j}] \text{ in } (\tau_{n,i-1}, \tau_{n,i}]. \tag{8}$$

To see how the approximation works, consider the system (4). In the earlier paper (Mykland and Zhang (2009)), σ_t was approximated by a constant over windows of finitely many observation points $t_{n,j}$, *i.e.*, $\max_i M_{n,i} = O_p(1)$ as $n \to \infty$. Our proposal here is instead to go one level deeper, and hold f_t and g_t constant over windows $(\tau_{n,i-1}, \tau_{n,i}]$. Specifically, define the measure on $C^2((\tau, \tau'])$ by $Q((X_t, \tilde{\sigma}_t), t \in (\tau, \tau']) | \tau, \tau', x, s, f, g)$, where

$$dX_t = \tilde{\sigma}_t \left(a dZ_t^* + b dU_t^*\right), \text{ and}$$
$$d\tilde{\sigma}_t = c dZ_t^* \text{ for } t \in (\tau, \tau'],$$
with window initial values:
$$\lim_{t \downarrow \tau} X_t = x \text{ and } \lim_{t \downarrow \tau} \tilde{\sigma}_t = s. \tag{9}$$

and where a, b, and c are given by

$$c = (f^2 + g^2)^{1/2}, \quad a = f/c, \text{ and } b = g/c. \tag{10}$$

We propose to approximate the distribution of X and σ from (4) on $(\tau_{n,i-1}, \tau_{n,i}]$, and given $\mathcal{F}_{\tau_{n,i-1}}$, by the distribution of X and $\tilde{\sigma}$ under

$$Q(\cdot|\tau_{n,i-1}, \tau_{n,i}, X_{\tau_{n,i-1}}, \tilde{\sigma}_{\tau_{n,i-1}}, f_{\tau_{n,i-1}}, g_{\tau_{n,i-1}}). \tag{11}$$

The reason for putting a tilde on σ and not on X is given in Section 4.1.

3.2 The Representation of Observables

To see why this is a double Gaussian approximation, set

$$dm_t = \tilde{\sigma}_t d\tilde{\sigma}_t \text{ for } t \in (\tau_{n,i-1}, \tau_{n,i}] \tag{12}$$

Under (11), conditionally on $\tilde{\sigma}_t$, and for $(t_{n,j}, t_{n,j+1}] \subseteq (\tau_{n,i-1}, \tau_{n,i}]$, we have, with $\tau = \tau_{n,i-1}$, $a = a_\tau$ and similarly for b and c,

$$\Delta X_{t_{n,j+1}} = \frac{a}{c}\Delta m_{t_{n,j+1}} + \frac{b}{c}U_{n,j+1}(\Delta\langle m, m\rangle_{t_{n,j+1}})^{1/2} \tag{13}$$

where the $U_{n,j}$, $j = 1, ..., n$ are iid standard normal, and where $\Delta\langle m, m\rangle_{t_{n,j+1}}$ is as defined in Section 2.4. In other words,

$$\Delta X_{t_{n,j+1}} = N\left(\frac{a}{c}\Delta m_{t_{n,j+1}}, \left(\frac{b}{c}\right)^2 \Delta\langle m, m\rangle_{t_{n,j+1}}\right) \tag{14}$$

Thus, the approximate X process is Gaussian given the $\tilde{\sigma}$ process, which itself is Gaussian. Hence "double Gaussian" as our term for the approximation. Note also that, under Q,

$$\Delta m_{t_{n,j+1}} = \frac{1}{2}\left(\tilde{\sigma}^2_{t_{n,j+1}} - \tilde{\sigma}^2_{t_{n,j}} - c^2\Delta t_{n,j+1}\right). \tag{15}$$

Necessarily, inference is based on estimation in each window, and then aggregation across windows.

REMARK 1. (CUMULANTS). Here and in later sections, we shall sometimes use cumulants. These were originally introduced under the name of semi-invariants by Thiele (see Lauritzen (2002)), and later by Fisher at the suggestion of Hotelling. See Stigler (2007), p. 611, for some of the colorful discussion surrounding the history of this and related concepts. For a general review of cumulant based methods in statistics, see McCullagh (1987). The first cumulant is the expectation, and the second cumulant is the variance. If random variables U_i have expectation zero, then the third and fourth cumulants are given by

$$\text{cum}(U_1, U_2, U_3) = E(U_1U_2U_3)$$
$$\text{cum}(U_1, U_2, U_3, U_4) = E(U_1U_2U_3U_4) - E(U_1U_2)E(U_3U_4)[3]$$
$$= E(U_1U_2U_3U_4) - E(U_1U_2)E(U_3U_4) - E(U_1U_3)E(U_2U_4) - E(U_1U_4)E(U_2U_3),$$

where the notation "[3]" denotes the sum over all three permutations, as spelled out in the subsequent line. For random variables with non-zero expectation, the definition is extended by noting that all cumulants (except the expectation) are invariant to the addition of constants. We write $\text{cum}_3(U) = \text{cum}(U, U, U)$, and so on. Subscript Q means that the relevant expectations are taken w.r.t. probability Q. □

To see how complex calculations become possible in this framework, let

$$\Delta S_{n,j+1} = \Delta X_{t_{n,j+1}}^2 - \int_{t_{n,j}}^{t_{n,j+1}} \tilde{\sigma}_t^2 dt, \tag{16}$$

and consider the computation of $\text{Var}_Q(\Delta S_{n,j+1} \mid \mathcal{F}_{t_{n,j}})$. Let $\mathcal{A}_{t_{n,j}}$ be the smallest sigma-field containing $\mathcal{F}_{t_{n,j}}$ and under which $(\tilde{\sigma}_t, t \leq t_{n,j+1})$ is measurable. From (C.24) in Appendix C,

$$E_Q\left(\Delta S_{n,j+1}^2 \mid \mathcal{A}_{t_{n,j}}\right) = \left(\frac{a}{c}\right)^4 (\Delta m_{t_{n,j+1}})^4 + \left(4\left(\frac{a}{c}\right)^2\left(\frac{b}{c}\right)^2 - 2\left(\frac{a}{c}\right)^4\right)(\Delta m_{t_{n,j+1}})^2 \Delta\langle m, m\rangle_{t_{n,j+1}}$$

$$+ \left(\left(\frac{a}{c}\right)^4 + 2\left(\frac{b}{c}\right)^4\right)\Delta\langle m, m\rangle_{t_{n,j+1}}^2. \tag{17}$$

Now use that $\text{Var}_Q(\Delta S_{n,j+1} \mid \mathcal{F}_{t_{n,j}})$ is the expectation of (17) under Q.[6] Since $\text{Var}_Q(\Delta m_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}}) = E_Q(\Delta\langle m, m\rangle_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}})$ and

$$E_Q(\Delta\langle m, m\rangle_{t_{n,j+1}} (\Delta m_{t_{n,j+1}})^2 | \mathcal{F}_{t_{n,j}})$$
$$= E_Q(\Delta\langle m, m\rangle_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}})^2 + \text{cum}_{3,Q}(\Delta\langle m, m\rangle_{t_{n,j+1}}, \Delta m_{t_{n,j+1}}, \Delta m_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}}), \tag{18}$$

we can rewrite on cumulant form:

$$\text{Var}_Q(\Delta S_{j+1} \mid \mathcal{F}_{t_{n,j}}) = 2c^{-4} E_Q(\Delta\langle m, m\rangle_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}})^2$$

$$+ \left(\frac{a}{c}\right)^4 \text{cum}_{4,Q}(\Delta m_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}})$$

$$+ \left(4\left(\frac{a}{c}\right)^2\left(\frac{b}{c}\right)^2 - 2\left(\frac{a}{c}\right)^4\right)\text{cum}_{3,Q}(\Delta\langle m, m\rangle_{t_{n,j+1}}, \Delta m_{t_{n,j+1}}, \Delta m_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}})$$

$$+ \left(\left(\frac{a}{c}\right)^4 + 2\left(\frac{b}{c}\right)^4\right)\text{Var}_Q(\Delta\langle m, m\rangle_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}}) \tag{19}$$

For a more detailed calculation, see (C.25)-(C.26) in Appendix C.

For further analysis, we use the following lemma, which is derived in Appendix C:

LEMMA 1. *The following identities are valid:*

$$E_Q(\Delta\langle m, m\rangle_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}}) = c^2\left(\frac{c^2}{2}(\Delta t_{n,j+1})^2 + \tilde{\sigma}_{t_{n,j}}^2 \Delta t_{n,j+1}\right) \text{ and}$$

$$\text{Var}_Q(\Delta\langle m, m\rangle_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}}) = \frac{4}{3}c^6\left(\frac{c^2}{4}(\Delta t_{n,j+1})^4 + \tilde{\sigma}_{t_{n,j}}^2(\Delta t_{n,j+1})^3\right). \tag{20}$$

Furthermore, with higher order cumulants also computed under Q,

$$\text{cum}_{3,Q}(\Delta\langle m, m\rangle_{t_{n,j+1}}, \Delta m_{t_{n,j+1}}, \Delta m_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}}) = 2\text{Var}_Q(\Delta\langle m, m\rangle_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}}) \text{ and}$$
$$\text{cum}_4(\Delta m_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}}) = 9\text{Var}_Q(\Delta\langle m, m\rangle_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}}). \tag{21}$$

Thus,

$$\text{Var}_Q(\Delta S_{n,j+1} \mid \mathcal{F}_{t_{n,j}}) = 2c^{-4}E_Q(\Delta\langle m, m\rangle_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}})^2 + 2c^{-4}(1 + 2a^2)\text{Var}_Q(\Delta\langle m, m\rangle_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}})$$
$$= 2\tilde{\sigma}_{t_{n,j}}^4 \Delta t_{n,j+1}^2 + \left(7 + 8a^2\right)\left(\frac{2}{3}c^2\tilde{\sigma}_{t_{n,j}}^2 \Delta t_{n,j+1}^3 + \frac{1}{6}c^4\Delta t_{n,j+1}^4\right). \tag{22}$$

The biggest order term is, of course, well known from asymptotics of realized volatility. By looking at higher order terms, however, we note that leverage effect does contribute to variance. The formula is used later on in (48), where the second order term has to be included, though only to vanish because of a martingale argument.

4   Main Theorem

4.1   An Extended System, and the QADD

To look at the quality of the approximation, we need to extend Q to allow both the original and the approximated (Gaussian) volatility processes to live on the same space. Thus define P_n^* as an extension of (11), as follows:

$$\text{For all } t \in [0, T] : d\sigma_t = c_t dZ_t^*;$$
$$\text{for } t \in (\tau_{n,i-1}, \tau_{n,i}] : dX_t = \tilde{\sigma}_t \left(a_{\tau_{n,i-1}} dZ_t^* + b_{\tau_{n,i-1}} dU_t^*\right)$$
$$\text{and } d\tilde{\sigma}_t = c_{\tau_{n,i-1}} dZ_t^*,$$
$$\text{with initial values: } \lim_{t\downarrow\tau_{n,i-1}} \tilde{\sigma}_t = \sigma_{\tau_{n,i-1}}$$
$$\text{and } \lim_{t\downarrow\tau_{n,i-1}} X_t = \int_0^{\tau_{n,i-1}} \sigma_t \left(a_t dZ_t^* + b_t dU_t^*\right) \tag{23}$$

The purpose for the final line is to avoid distinguishing between X and $\tilde{X}$. Apart from complicating notation, such a distinction would obscure that our approximation is a measure change on X_t but a process change on σ_t. P_n^* defines a measure on the whole line $[0, T]$.

The distribution of $(X_t, \tilde{\sigma}_t, t \in (\tau_{n,i-1}, \tau_{n,i}])$ under $P_n^*(\cdot|\mathcal{F}_{\tau_{n,i-1}})$ is the same as that of (11).

DEFINITION 1. *To measure the extent to which we err in approximation under Q, we define the following "Quadratic Asymptotic Decoupling Delay" (QADD) by*

$$K_2(t) = \lim_{n\to\infty} \sum_i \sum_{t_{n,j}\in(\tau_{n,i-1},\tau_{n,i})\cap[0,t]} (t_{n,j} - \tau_{n,i-1})^2, \tag{24}$$

provided the limit exists.

For discussion of existence and form of K_2, see Remark 2 and Section 4.3 after the main theorem.

4.2 The Theorem

The following main result is proved in Appendix B:

THEOREM 1. *(Contiguity of P^* and P_n^* on the Observables $\Delta X_{t_{n,j}}$). Consider the system (23). Assume that the Quadratic Asymptotic Decoupling Delay K_2 exists. Let $V_{n,j+1}$ be given as one of the following*

$$V_{n,j+1} = \frac{\Delta X_{t_{n,j+1}}^2}{\Delta t_{n,j+1}\tilde{\sigma}_{t_{n,j}}^2} - 1$$

$$\text{or } = \frac{\Delta X_{t_{n,j+1}}^2 - \int_{t_{n,j}}^{t_{n,j+1}} \tilde{\sigma}_t^2 dt}{\Delta t_{n,j+1}\tilde{\sigma}_{t_{n,j}}^2} \tag{25}$$

and define

$$M_n^{(0)} = \sum_i \sum_{t_{n,j}\in[\tau_{n,i-1},\tau_{n,i})} V_{n,j+1}\sigma_{\tau_{n,i-1}}^{-2}(\tilde{\sigma}_{t_{n,j}} - \sigma_{t_{n,j}}) \tag{26}$$

as well as

$$\Gamma_0 = \int_0^T \sigma_t^{-2}\langle c,c\rangle_t' dK_2(t). \tag{27}$$

Then, subject to regularity conditions, with the proviso in Remark 5, and with either choice in (25), as $n \to \infty$, $M_n^{(0)}$ converges stably in law under P_n^* to a normal distribution with mean zero and variance Γ_0. Also, under P_n^*,

$$\log\frac{dP^*}{dP_n^*} = M_n^{(0)} - \frac{1}{2}\Gamma_0 + o_p(1). \tag{28}$$

The theorem implies that P^* and the approximation P_n^* are *contiguous* on the observables.[7] This is to say that the likelihood ratio dP_n^*/dP^* is uniformly integrable under P^* (as is dP^*/dP_n^* under P_n^*). In particular, if an estimator is consistent under P_n^*, it is also consistent under P^* and P. Rates of convergence (typically $n^{1/2}$) are also preserved, but the asymptotic distribution may change. For a general definition and discussion of contiguity, see Hájek and Sidak (1967), LeCam (1986), LeCam and Yang (2000), as well as Chapter IV of Jacod and Shiryaev (2003).

REMARK 2. (WINDOW SIZES OF ORDER $o_p(n^{-1/2})$) From our assumptions, if $\sup_i M_{n,i} = o_p(n^{1/2})$, then $K_2 \equiv 0$, and the limiting term in Theorem 1 is zero. There is thus no contiguity adjustment in this case. The results in Mykland and Zhang (2009) can be derived from this. □

REMARK 3. (THE TWO CHOICES OF $V_{n,j+1}$) The $M_n^{(0)}$ in (26) is the end point of a P_n^*-martingale for the second choice in (25). The first option makes $M_n^{(0)}$ the end point of a martingale under measure $\tilde{R}_n^*$, defined in the proof in Appendix B. As is seen in the Appendix, $\tilde{R}_n^*/P_n^* \xrightarrow{p} 1$ as $n \to \infty$. An alternative use of Theorem 1, therefore, is to proceed under $\tilde{R}_n^*$ with the first choice of $V_{n,j+1}$ in (25). □

REMARK 4. (PARTIAL LIKELIHOOD CONSTRUCTION) We have been deliberately opaque about the choice of partial likelihood. There are two ways of setting the likelihod ratio dP^*/dP_n^*, both valid under the proof in Appendix B. One is to use $\prod_i \frac{dP^*}{dP_n^*}(X_t, \sigma_t, \tilde{\sigma}_t; \tau_{n,i-1} < t \le \tau_{n,i}|\mathcal{F}_{\tau_{n,i-1}})$, the other is to instead condition on $(X_{\tau_{n,i-1}}, \sigma_{\tau_{n,i-1}}, a_{\tau_{n,i-1}}, b_{\tau_{n,i-1}}, c_{\tau_{n,i-1}}, \langle c, c \rangle'_{\tau_{n,i-1}})$. The latter approach is in analogy with the discussion surrounding formula (13) (p. 1410) in Mykland and Zhang (2009). □

REMARK 5. (HOW TO HANDLE SMALL VALUES OF σ_t AND $\tilde{\sigma}_t$). We assume that σ_t is continuous and nonzero. From this, the process is locally bounded both above and away from zero. From Section 4.5 in Mykland and Zhang (2010), we can thus assume, by a stopping argument, and without loss of generality, that σ_t is bounded above and away from zero on $[0, T]$. This stopping is convenient both for Theorem 1, and for the application of Girsanov's Theorem in passing from System (2) to System (4) in Section 2.1.

The situation with $\tilde{\sigma}_t$ is more precarious, in that this process can be zero or negative. To obtain boundedness (both ways), we can proceed as follows. Since we can assume that σ_t is suitably bounded, the same stopping argument applies to $\tilde{\sigma}_t$, is view of the modulus of continuity of Brownian Motion (see, *e.g.*, Chapter 2.9.F (p. 114-116) of Karatzas and Shreve (1991)). The extent to which this affects Theorem 1 is discussed presently. (1) For the first choice of $V_{n,j+1}$, there is no need to stop $\tilde{\sigma}$, cf. the statement (for contiguous measure $\tilde{R}_n^*$) just before equation (B.16). (2) For the second choice of $V_{n,j+1}$, stopping due to $\tilde{\sigma}_{t_{n,j}}$ being too small may be needed, but the probability of this occurring is asymptotically negligible. (Instead of stopping, one can alternatively set $V_{n,j+1} = 0$ when $\tilde{\sigma}_{t_{n,j}}$ is small, and continue.) If one wishes to avoid the stopping issue, one can take $V_{n,j+1} = S_{n,j+1}/(\text{Var}_n^*(S_{n,j+1}|\mathcal{F}_{t_{n,j}})/2)^{1/2}$, where this variance is given in formula (16). The difference from the second choice of $V_{n,j+1}$ is negligible in most symbolic calculations. □

4.3   Further study of the QADD

Under conditions discussed in the context of AQVT in Zhang, Mykland, and Aït-Sahalia (2005) (p. 1411) and ADD in Mykland and Zhang (2009) (p. 1418), every subsequence has a further subsequence for which $K_2(\cdot)$ exists and is Lipschitz continuous. Thus one can take the limit K_2 in (24) to exist without any major loss of generality.

Consider the case of equally sized blocks of M_n observations ($M_{n,i} = M_n$ for all i), with

$$M_n/n^{1/2} \to v. \tag{29}$$

In the case where the $t_{n,j}$ have a reasonably regular local dispersion,

$$\sum_{t_{n,j}\in(\tau_{n,i-1},\tau_{n,i})} (t_{n,j} - \tau_{n,i-1})^2 = \Delta\tau_{n,i}^2 \sum_{k=1}^{M_n} \left(\frac{j}{M_n}\right)^2 (1 + o_p(1))$$

$$= \frac{1}{3}\Delta\tau_{n,i}^2 M_n (1 + o_p(1)). \tag{30}$$

In this case, therefore,

$$K_2(t) = \frac{1}{3}v^2 T H(t), \tag{31}$$

where $H(t)$ is the "Asymptotic Quadratic Variation of Time" (AQVT) of the $\tau_{n,i}$'s, given by

$$H(t) = \lim_{n\to\infty} \frac{K_n}{T} \sum_{\tau_{n,i+1}\leq t} (\tau_{n,i+1} - \tau_{n,i})^2, \tag{32}$$

provided that the limit exists (cf. the references at the beginning of this remark). The AQVT concept comes up in a number of our earlier investigations. What is unusual here is that it is based on the window boundaries $\tau_{n,i}$ rather than the underlying observation points $t_{n,j}$.

In the case of equidistant observations, the QADD thus takes the form

$$K_2(t) = \frac{1}{3}v^2 T t. \tag{33}$$

In the case where the observations times follow a Poisson process, then (30) also holds, and it follows from considerations like those in Example 6 in Mykland and Zhang (2010) that the $\tau_{n,i+1} - \tau_{n,i}$ are identically distributed, with law $TU_{(M_n)}$, where $U_{(M_n)}$ is the M_n'th order statistic among n iid standard uniform random variables. Thus $H(t) = tK_n^2 EU_{(M_n)}^2 (1 + o_p(1)) = t$, and so also in this case, the QADD gets the form (33).

4.4 Adjusting for the Change from P^* to P_n^*

THEOREM 2. *Assume the setup in Theorem 1. Suppose that, under P_n^*, $(Z_n, M_n^{(0)},)$ converges stably to a bivariate distribution $b + aN(0, I)$, where $N(0, I)$ is a bivariate normal vector independent of $\mathcal{F}_T$, where the vector $b = (b_1, b_2)^T$ and the symmetric 2×2 matrix a are $\mathcal{F}_T$ measurable. Set $A = aa^T$. Then Z_n converges stably under P^* to $b_1 + A_{12} + (A_{11})^{1/2}N(0, 1)$, where $N(0, 1)$ is independent of $\mathcal{F}_T$.*

Recall that $b_2 = 0$ and $A_{22} = \Gamma_0$. The proof is the same as for Theorems 2 and 4 in Mykland and Zhang (2009). Theorem 2 states that when adjusting from Q_n to P^*, the asymptotic variance of Z_n is unchanged, while the asymptotic bias may change.

5 Worked Example: Estimation of Quarticity

For brevity, we here give a single application of the theory, namely that of estimating quarticity:

$$\theta = \int_0^T \sigma_t^4 dt. \tag{34}$$

The example is worked out in some detail, in the hope that this section can provide guidance on how to use the theory. The quarticity is particularly important in that it is proportional to the asymptotic variance[8] of the realized volatility, which is the standard estimator of $\int_0^T \sigma_t^2 dt$ in the no-microstructure case. The name goes back to Barndorff-Nielsen and Shephard (2002). See also Jacod and Protter (1998) and Mykland and Zhang (2006).

For simplicity, assume in following that sampling is equispaced (so $\Delta t_{n,j} = \Delta t_n = T/n$ for all j). We take the block size $M_n = M_{n,i}$ to be independent of i. See Remark 6 for a closer discussion of this. Similarly, $\Delta \tau_n = \Delta t_n M_n$. Define

$$\hat{\sigma}_{n,i}^2 = \frac{1}{\Delta t_n M_n} \sum_{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}]} (\Delta X_{t_{n,j}})^2 \tag{35}$$

Set $\hat{\sigma}_{n,i}^4 = (\hat{\sigma}_{n,i}^2)^2$. We consider two different estimators:

$$\breve{\theta}_n = \sum_i \hat{\sigma}_{n,i}^4 \Delta \tau_n \text{ and } \hat{\theta}_n = \frac{M_n}{M_n + 2} \sum_i \hat{\sigma}_{n,i}^4 \Delta \tau_n. \tag{36}$$

The M_n finite case is analyzed in Section 4.1 of Mykland and Zhang (2009). In this case, considerations of unbiasedness would lead to the use of $\hat{\theta}_n$, which is consistent and asymptotically mixed normal, but not quite efficient for finite block sizes.

Three questions arise from the earlier treatment: (1) Can the estimator be made efficient by letting $M_n \to \infty$ with n? (2) Does blocking give rise to biases or extra variance, which is unseen in the asymptotics based on $M_n = O_p(1)$? (3) Does the difference between $\breve{\theta}_n$ and $\hat{\theta}_n$ matter when $M_n \to \infty$?

We shall therefore re-investigate this estimation problem from the angle of M_n going to infinity with n. We work in the framework of blocks M_n of size $O_p(n^{1/2})$, specifically (29). This is because this gives rise to the most informative asymptotic results. For M_n going to infinity at rate $o_p(n^{1/2})$, see Remark 7.

The third question is most easily answered, without much technology. The difference between the two estimators from (36) is

$$n^{1/2}(\hat{\theta}_n - \breve{\theta}_n) = -2 \frac{n^{1/2}}{M_n + 2} \sum_i \hat{\sigma}_{n,i}^4 \Delta \tau_n$$

$$= -\frac{2}{v} \int_0^T \sigma_t^4 dt + o_p(1), \tag{37}$$

so at most one of the two estimators can be asymptotically unbiased. In fact, it will turn out that neither of them is asymptotically unbiased.[9] (The result (37) follows from considerations like those below, but much more simply.)

We now approach questions (1) and (2). The final conclusion on quarticity is in Section 5.3.

5.1   Estimation under P_n^*

For questions (1) and (2), we will initially estimate (under P_n^*)

$$\tilde{\theta}_n = \int_0^T \tilde{\sigma}_t^4 dt. \tag{38}$$

We will then do the post-asymptotic adjustment to obtain results for the estimation of θ under P^* and P.

To analyze this estimator, denote

$$\Delta L_i \;=\; \hat{\sigma}_{n,i}^2 - \frac{1}{\Delta \tau_n} \int_{\tau_{n,i-1}}^{\tau_{n,i}} \tilde{\sigma}_t^2 dt. \tag{39}$$

It is well known that

$$\mathrm{Var}_n^*(\Delta L_i | \mathcal{F}_{\tau_{n,i-1}}) \;=\; \frac{2}{M_n} \sigma_{\tau_{n,i-1}}^4 + O_p(M_n^{-3/2}) \tag{40}$$

Meanwhile, from Itô's Formula, we get that

$$\int_{\tau_{n,i-1}}^{\tau_{n,i}} \tilde{\sigma}_t^2 dt = \sigma_{\tau_{n,i-1}}^2 \Delta \tau_n + \frac{1}{2} c_{\tau_{n,i-1}}^2 \Delta \tau_n^2 + 2 \int_{\tau_{n,i-1}}^{\tau_{n,i}} (\tau_i - t) \tilde{\sigma}_t d\tilde{\sigma}_t \;, \tag{41}$$

and similarly for $\int_{\tau_{n,i-1}}^{\tau_{n,i}} \tilde{\sigma}_t^4 dt$, so that

$$\left(\frac{1}{\Delta \tau_n} \int_{\tau_{n,i-1}}^{\tau_{n,i}} \tilde{\sigma}_t^2 dt \right)^2 = \frac{1}{\Delta \tau_n} \int_{\tau_{n,i-1}}^{\tau_{n,i}} \tilde{\sigma}_t^4 dt - \frac{2}{3} \Delta \tau_n \sigma_{\tau_{n,i-1}}^2 c_{\tau_{n,i-1}}^2 - \Delta G_i \;+\; O_p(\Delta \tau_n^{3/2}) \tag{42}$$

where

$$\Delta G_i = \frac{4}{\Delta \tau_n} \int_{\tau_{n,i-1}}^{\tau_{n,i}} (\tau_i - t) \tilde{\sigma}_t (\tilde{\sigma}_t^2 - \sigma_{\tau_{n,i-1}}^2) d\tilde{\sigma}_t.$$

It follows that

$$\hat{\sigma}_{n,i}^4 = \left(\frac{1}{\Delta \tau_n} \int_{\tau_{n,i-1}}^{\tau_{n,i}} \tilde{\sigma}_t^2 dt \right)^2 + \Delta L_i^2 + 2\sigma_{\tau_{n,i-1}}^2 \Delta L_i - \Delta G_i + O_p(n^{-3/4})$$

$$= \frac{1}{\Delta \tau_n} \int_{\tau_{n,i-1}}^{\tau_{n,i}} \tilde{\sigma}_t^4 dt - \frac{2}{3} \Delta \tau_n \sigma_{\tau_{n,i-1}}^2 c_{\tau_{n,i-1}}^2 + \Delta L_i^2 + 2\sigma_{\tau_{n,i-1}}^2 \Delta L_i - \Delta G_i + O_p(n^{-3/4}). \tag{43}$$

Hence, by (29)

$$n^{1/2}(\breve{\theta}_n - \tilde{\theta}_n) = n^{1/2}\Delta\tau_n \sum_i \left(-\frac{2}{3}\Delta\tau_n\sigma^2_{\tau_{n,i-1}}c^2_{\tau_{n,i-1}} + \Delta L_i^2 + 2\sigma^2_{\tau_{n,i-1}}\Delta L_i - \Delta G_i \right) + O_p(n^{-1/4})$$

$$= n^{1/2}\Delta\tau_n \left(-\frac{2}{3}\int_0^T \sigma_t^2 c_t^2 dt \ + \ \frac{2}{M_n}\sum_i \sigma^4_{\tau_{n,i-1}} \ + \ 2\sum_i \sigma^2_{\tau_{n,i-1}}\Delta L_i \right) + O_p(n^{-1/4})$$

(44)

by (40), since $\sum_i(\Delta L_i^2 - \mathrm{Var}(\Delta L_i|\mathcal{F}_{\tau_{n,i-1}})) = O_p(n^{-1})$, and since $\sum_i \Delta G_i = O_p(\Delta\tau^{1/2}) = O_p(n^{-1/4})$. From the first line in (37), and since $n^{1/2}\Delta\tau_n \to vT$,

$$n^{1/2}(\hat{\theta}_n - \tilde{\theta}_n) = vT\left(-\frac{2}{3}\int_0^T \sigma_t^2 c_t^2 dt \ + \ 2\sum_i \sigma^2_{\tau_{n,i-1}}\Delta L_i \right) + o_p(1). \tag{45}$$

For the asymptotics under P_n^*, we thus obtain from (40) and standard martingale central limit considerations (Hall and Heyde (1980), Jacod and Shiryaev (2003)) that

$$n^{1/2}(\hat{\theta}_n - \tilde{\theta}_n) \xrightarrow{\mathcal{L}} -vT\frac{2}{3}\int_0^T \sigma_t^2 c_t^2 dt \ + \ \left(8T\int_0^T \sigma_t^8 dt \right)^{1/2} N(0,1), \tag{46}$$

where the convergence is stable and the $N(0,1)$ random variable is independent of the underlying filtration.

5.2 Adjustment to Measure P^*

On the one hand,

$$\tilde{\theta}_n - \theta = \sum_i \int_{\tau_{n,i-1}}^{\tau_{n,i}} (\tilde{\sigma}_t^4 - \sigma_t^4)dt$$

$$= \sum_i \int_{\tau_{n,i-1}}^{\tau_{n,i}} (\tau_i - t)d(\tilde{\sigma}_t^4 - \sigma_t^4)$$

$$= \sum_i \int_{\tau_{n,i-1}}^{\tau_{n,i}} (\tau_i - t)\left(4\tilde{\sigma}_t^3 d\tilde{\sigma}_t - 4\sigma_t^3 d\sigma_t + 6\tilde{\sigma}_t^2 c^2 dt - 6\sigma_t^2 c_t^2 dt \right)$$

$$= o_p(n^{-1/2}). \tag{47}$$

This is since the martingale part is of order $O_p(n^{-2})$, and the drift part is no bigger than $o_p(n^{-1/2})$.

The other part of the adjustment involves the quadratic covariation between $M_n^{(0)}$ (use the second choice for $V_{n,j+1}$ in (25)) and the P_n^*-martingale term in (45), whose end point is $R_n = 2vT\sum_i \sigma^2_{\tau_{n,i-1}}\Delta L_i$. As in (16), write $\Delta S_{n,j+1} = \Delta X^2_{t_{n,j+1}} - \int_{t_{n,j}}^{t_{n,j+1}} \tilde{\sigma}_t^2 dt$, and note that

$$M_n^{(0)} \ = \ \Delta t_n^{-1} \sum_i \sum_{t_{n,j}\in[\tau_{n,i-1},\tau_{n,i})} \sigma^{-2}_{\tau_{n,i-1}}\tilde{\sigma}^{-2}_{t_{n,j}}(\tilde{\sigma}_{t_{n,j}} - \sigma_{t_{n,j}})S_{n,j+1} \text{ and}$$

$$R_n \ = \ 2vT\Delta\tau_n^{-1}\sum_i \sigma^2_{\tau_{n,i-1}} \sum_{t_{n,j}\in[\tau_{n,i-1},\tau_{n,i})} S_{n,j+1}.$$

Now first let $\langle \cdot, \cdot \rangle^{\mathcal{G}}$ denote quadratic covariation (under P_n^*) with respect to the times $t_{n,j}$. Using (22) in Section 3.2, we obtain (where a and c are the values of the process at $\tau_{n,i-1}$)

$$
\begin{aligned}
\Delta \langle R, M^{(0)} \rangle^{\mathcal{G}}_{t_{n,j+1}} &= 2vT\Delta\tau_n^{-1}\Delta t_n^{-1}\tilde{\sigma}_{t_{n,j}}^{-2}(\tilde{\sigma}_{t_{n,j}} - \sigma_{t_{n,j}})\mathrm{Var}_{P_n^*}(S_{n,j+1}|\mathcal{F}_{t_{n,j}}) \\
&= 2vT\Delta\tau_n^{-1}\Delta t_n^{-1}\tilde{\sigma}_{t_{n,j}}^{-2}(\tilde{\sigma}_{t_{n,j}} - \sigma_{t_{n,j}})\left(2\tilde{\sigma}_{t_{n,j}}^4\Delta t_n^2 + \frac{2}{3}(7+8a^2)c^2\tilde{\sigma}_{t_{n,j}}^2\Delta t_n^3 + O_p(\Delta t_n^4)\right) \\
&= 2vTM_n^{-1}(\tilde{\sigma}_{t_{n,j}} - \sigma_{t_{n,j}})\left(2\tilde{\sigma}_{t_{n,j}}^2 + \frac{2}{3}(7+8a^4)c^2\Delta t_n + O_p(\Delta t_n^2)\right).
\end{aligned}
\tag{48}
$$

If we consider the quadratic variation $\langle \cdot, \cdot \rangle^{\mathcal{H}}$ with respect to the times $\tau_{n,i}$, we see from taking the expectation in the above (w.r.t. $\mathcal{F}_{\tau_{n,i-1}}$) that the second term vanishes because of being a P_n^*-martingale, whence

$$
\Delta \langle R, M^{(0)} \rangle^{\mathcal{H}}_{\tau_{n,i}} = 4vTM_n^{-1}\sum_{t_{n,j}\in[\tau_{n,i-1},\tau_{n,i})} E_n^* \left\{\tilde{\sigma}_{t_{n,j}}^2(\tilde{\sigma}_{t_{n,j}} - \sigma_{t_{n,j}})|\mathcal{F}_{\tau_{n,i-1}}\right\} + O_p(n^{-2}).
\tag{49}
$$

To finalize this calculation, note that by a stopping argument along the lines of Remark 5, we can (without loss of generality) take $|c_t|$, $\langle c, c \rangle_t' \leq C_2$, where C_2 is a nonrandom constant. With this, $E^*((\tilde{\sigma}_{t_{n,j}} - \sigma_{t_{n,j}})^2|\mathcal{F}_{\tau_{n,i-1}}) = E^*(\langle \tilde{\sigma} - \sigma, \tilde{\sigma} - \sigma \rangle_{t_{n,j}} - \langle \tilde{\sigma} - \sigma, \tilde{\sigma} - \sigma \rangle_{\tau_{n,i-1}}|\mathcal{F}_{\tau_{n,i-1}}) \leq (C_2/2)(t_{n,j} - \tau_{n,i-1})^2$. Thus

$$
\begin{aligned}
|E_n^* \left\{\tilde{\sigma}_{t_{n,j}}^2(\tilde{\sigma}_{t_{n,j}} - \sigma_{t_{n,j}})|\mathcal{F}_{\tau_{n,i-1}}\right\}| &= |2\sigma_{\tau_{n,i-1}}E_n^* \left\{(\tilde{\sigma}_{t_{n,j}} - \sigma_{\tau_{n,i-1}})(\tilde{\sigma}_{t_{n,j}} - \sigma_{t_{n,j}})|\mathcal{F}_{\tau_{n,i-1}}\right\} \\
&\quad + E_n^* \left\{(\tilde{\sigma}_{t_{n,j}} - \sigma_{\tau_{n,i-1}})^2(\tilde{\sigma}_{t_{n,j}} - \sigma_{t_{n,j}})|\mathcal{F}_{\tau_{n,i-1}}\right\}| \\
&\leq C_3(t_{n,j} - \tau_{n,i-1})^{3/2}
\end{aligned}
$$

where C_3 is a nonrandom constant. With this bound, (49) yields

$$
\begin{aligned}
|\Delta \langle R, M^{(0)} \rangle^{\mathcal{H}}_{\tau_{n,i}}| &\leq 4vTM_n^{-1}C_3\Delta t_n^{3/2}\sum_{j=1}^{M_n-1} j^{3/2} + O_p(n^{-2}) \\
&= 4vTM_n^{3/2}C_3\Delta t_n^{3/2}\int_0^1 x^{3/2}dx\ (1+o_p(1)) \\
&= O_p(n^{-3/4}).
\end{aligned}
$$

Hence, in the end,

$$
\langle R, M^{(0)} \rangle^{\mathcal{H}}_T = O_p(n^{-1/4}).
$$

By Theorem 2, there is thus no adjustment due to contiguity when passing from P_n^* to P^*.

5.3 Final Result on Quarticity

Since Section 5.2 yielded no adjustment, the result from (46) remains valid under P^*, and with θ replacing $\tilde{\theta}_n$. By measure change, the same result also holds for P. Thus:

PROPOSITION 1. *Under both P and P^*, given (29), and subject to regularity conditions,*

$$n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} -vT\frac{2}{3}\int_0^T \sigma_t^2 c_t^2 dt + \left(8T\int_0^T \sigma_t^8 dt\right)^{1/2} N(0,1), \tag{50}$$

where the convergence is stable and the $N(0,1)$ random variable is independent of the underlying filtration.

There is, in other words, an asymptotic bias when estimating quarticity with $\hat{\theta}_n$ in block sizes of $O_p(n^{1/2})$. From equation (37), the alternative estimator $\breve{\theta}_n$ would be even worse, so the $M/(M+2)$ adjustment for unbiasedness does still have meaning when the window size increases at rate $O_p(n^{1/2})$.

In principle, it is possible to estimate the bias, or to eliminate it using a two scales construction. Also, following Section 4.1.2 of Mykland and Zhang (2009), the asymptotic variance is most likely the best possible. The investigation of these issues is, however, beyond the scope of this paper.

REMARK 6. (BLOCK SIZE). In most cases, n/M_n is not an integer, and $n - M_n[n/M_n] = O_p(n^{1/2})$. One can thus not just let all blocks be equal except the last block, because the edge effect in estimation may be of the same order as the asymptotic normal distribution. The best solution is to have blocks of size $M_n + 1$ scattered around the interval $[0,T]$ about evenly. The asymptotics above is then not impacted. $\square$

REMARK 7. (SMALLER WINDOW SIZES). We here consider the case where $M_n \to \infty$, but at rate $o_p(n^{1/2})$. (The finite M case was discussed in the earlier paper.) In this case, there is no contiguity adjustment, and the asymptotic bias is zero. In other words, we then have

$$n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \left(8T\int_0^T \sigma_t^8 dt\right)^{1/2} N(0,1). \tag{51}$$

The bias is gone! There are two interpretations of this, both valid. One is that by using smaller block size, one avoids the bias problem. The other is that this is a way of putting one's head in the sand: the bias is there, but it is just of slightly lower order. According to this viewpoint, one never really knows what is the limit v in (29), since n is finite, and one ought to do inference replacing a hypothetical vT by an actual $M_n/n^{1/2}$. $\square$

REMARK 8. (ANOTHER APPROACH TO ANALYSIS). The above is not the only path one can take. For example, denote by $\mathcal{Y}_{n,i}$ the information at time $\tau_{n,i-1}$, along with the information in $\tilde{\sigma}_t$, $\tau_{n,i-1} < t \le \tau_{n,i}$. By the conditional independence in Section 3.2, we obtain for cumulants under P_n^* (for all integer p) that

$$\text{cum}_p(\hat{\sigma}_{n,i}^2 | \mathcal{Y}_{n,i}) = \frac{1}{(\Delta t_n M_n)^p} \sum_{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}]} \text{cum}_p((\Delta X_{t_{n,j}})^2 | \mathcal{Y}_{n,i}). \tag{52}$$

From this, one can easily see that

$$\text{cum}_p(\hat{\sigma}_{n,i}^2 | \mathcal{Y}_{n,i}) = O_p(M_n^{1-p}), \tag{53}$$

and one can then build an analysis on this. $\square$

6 Conclusion

The paper has shown that locally parametric approximations can be extended to windows of $O_p(n^{1/2})$ observations. This yields easier calculations than working with the original system, and the post-asymptotic contiguity adjustment is also relatively straightforward.

We have seen in Section 5 that this method can find asymptotic (higher order) biases in consistent estimators, and we conjecture that the same will be true if investigating other such estimators, whether of integrated powers of volatility, realized regressions, or ANOVA (in the latter case, we know the bias is there from Zhang (2001)).

Another application of these results arises in the case when estimators are consistent only when the block size is increasing with sample size. This includes the estimation of leverage effect and of the volatility of volatility, and estimation when microstructure is present in the data (cf. the Introduction and Section 2.4 for references).

We conjecture that the results in this paper extend to the multivariate case. System (1) is then replaced with definitions that are given on p. 1406 and 1431 (equation (A.1)) of Mykland and Zhang (2009), but theorems in this direction are left for another day.

We finally note that by investigating each estimator individually, one can obtain precise and weakest possible regularity conditions. This was not the aim here; our goal was to provide an analytic tool for finding asymptotic behavior without calculations being too cumbersome. In fact, from the development in Sections 3.2 and 5, many of the techniques can be used in symbolic calculation programs.

Endnotes

[1]Microstructure for this purpose means measurement error. The term also has a wider usage; see, for example, O'Hara (1995) and Hasbrouck (1996).

[2]The regularity conditions are mostly about integrability in Girsanov's Theorem. See, for example, Chapter 5.5 of Karatzas and Shreve (1991). They also, however, preclude the leverage effect from being so big that $g_t = 0$. For these cases, results similar to the current ones can be derived in a direct (but more tedious) fashion. See Sections 2.2 and 2.4 for further discussion of this.

[3]Let Z_n be a sequence of $\mathcal{F}_T$-measurable random variables. We say that Z_n *converges stably in law to Z* as $n \to \infty$ if Z is measurable with respect to an extension of $\mathcal{F}_T$ so that for all $A \in \mathcal{F}_T$ and for all bounded continuous g, $EI_A g(Z_n) \to EI_A g(Z)$ as $n \to \infty$. I_A denotes the indicator function of A, and $= 1$ if A and $= 0$ otherwise. The same definition applies to triangular arrays. In

the context of inference, $Z_n = n^{1/2}(\hat{\theta}_n - \theta)$, for example, and $Z = N(b, a^2)$. For further discussion of stable convergence, and for the relationship to measure change, see Section 2.2 of Mykland and Zhang (2009), which draws on Rootzén (1980).

[4]One can also define quadratic (co-)variations for a fixed grid $\{0 = t_{n,0} < t_{n,1} < ... < t_{n,n} = T\}$, in which case there is no taking of limits. We use this kind of quadratic variation twice, in Section 5.2 and Appendix B, and in these cases with reference to the (co-)variance based definition ("predictable quadratic variation").

[5]There are approximately 250 trading days in a year.

[6]$\mathrm{Var}(\cdot \mid \mathcal{F}_{t_{n,j}}) = E(\mathrm{Var}(\cdot \mid \mathcal{A}_{t_{n,j}}) \mid \mathcal{F}_{t_{n,j}}) + \mathrm{Var}(E(\cdot \mid \mathcal{A}_{t_{n,j}}) \mid \mathcal{F}_{t_{n,j}})$, since $\mathcal{F}_{t_{n,j}} \subseteq \mathcal{A}_{t_{n,j}}$.

[7]Of course, they are not even equivalent on the whole process $(X_t, 0 \leq t \leq T)$.

[8]The expression asymptotic variance is in this case functionally accurate, even though the quarticity can be random.

[9]We emphasize that asymptotic bias is different from inconsistency. The term usually (and here) refers to a bias of the same size as the asymptotic (mixed) normal distribution.

Acknowledgements

REFERENCES

ANDERSEN, T. G. AND T. BOLLERSLEV (1998): "Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts," *International Economic Review*, 39, 885–905.

BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2008): "Designing realized kernels to measure ex-post variation of equity prices in the presence of noise," *Econometrica*, 76, 1481–1536.

BARNDORFF-NIELSEN, O. E. AND N. SHEPHARD (2001): "Non-Gaussian Ornstein-Uhlenbeck-Based Models And Some Of Their Uses In Financial Economics," *Journal of the Royal Statistical Society, B*, 63, 167–241.

——— (2002): "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society, B*, 64, 253–280.

BLACK, F. (1976): "Studies of Stock Price Volatility Changes," in *Proceedings of the 1976 Meetings of the American Statistical Association*, 171–181.

BLACK, F. AND M. SCHOLES (1973): "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637–654.

BOLLERSLEV, T. (1986): "Generalized Autorgeressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

CARR, P., H. GEMAN, D. B. MADAN, AND M. YOR (2003): "Stochastic volatility for Lévy processes," *Mathematical Finance*, 13, 345–342.

COMTE, F. AND E. RENAULT (1998): "Long memory in continuous-time stochastic volatility models," *Mathematical Finance*, 8, 291–323.

COX, D. R. (1975): "Partial Likelihood," *Biometrika*, 62, 269–276.

ENGLE, R. F. (1982): "Autogregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica*, 50, 987–1008.

——— (2000): "The Econometrics of Ultra-High Frequency Data," *Econometrica*, 68, 1–22.

HÁJEK, J. AND Z. SIDAK (1967): *Theory of Rank Tests*, New York: Academic Press.

HALL, P. AND C. C. HEYDE (1980): *Martingale Limit Theory and Its Application*, Boston: Academic Press.

HANSEN, L. AND T. SARGENT (2007): *Robustness*, New Jersey: Princeton University Press.

HASBROUCK, J. (1996): "Modeling Market Microstructure Time Series," in *Handbook of Statistics, volume 14*, ed. by C. R. Rao and G. S. Maddala, Amsterdam: North-Holland, 647–692.

HE, H. AND H. LELAND (1993): "On Equilibrium Asset Price Processes," *Review of Financial Studies*, 6, 593–617.

HESTON, S. (1993): "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bonds and Currency Options," *Review of Financial Studies*, 6, 327–343.

JACOD, J., Y. LI, P. A. MYKLAND, M. PODOLSKIJ, AND M. VETTER (2009): "Microstructure Noise in the Continuous Case: The Pre-Averaging Approach," *Stochastic Processes and Their Applications*, 119, 2249–2276.

JACOD, J. AND P. PROTTER (1998): "Asymptotic Error Distributions for the Euler Method for Stochastic Differential Equations," *Annals of Probability*, 26, 267–307.

JACOD, J. AND A. N. SHIRYAEV (2003): *Limit Theorems for Stochastic Processes*, New York: Springer-Verlag, second ed.

KARATZAS, I. AND S. E. SHREVE (1991): *Brownian Motion and Stochastic Calculus*, New York: Springer-Verlag.

LAURITZEN, S. L. (2002): *Thiele: Pioneer in Statistics*, Oxford: Oxford Univ. Press.

LECAM, L. (1986): *Asymptotic Methods in Statistical Decision Theory*, New York: Springer-Verlag.

LECAM, L. AND G. YANG (2000): *Asymptotics in Statistics: Some Basic Concepts*, New York: Springer-Verlag, second ed.

LI, Y., P. MYKLAND, E. RENAULT, L. ZHANG, AND X. ZHENG (2009): "Realized Volatility when Endogeniety of Time Matters," Working Paper, University of Chicago.

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

MEDDAHI, N. (2001): "An Eigenfunction Approach for Volatility Modeling," Tech. rep., Université de Montréal.

MERTON, R. C. (1971): "Optimum consumption and portfolio rules in a continuous-time model," *Journal of Economic Theory*, 3, 373–413.

——— (1973): "The Theory of Rational Option Pricing," *Bell journal of Economics and Management Science*, 4, 141–183.

——— (1992): *Continuous Time Finance*, New York, N.Y.: Basil Blackwell.

MYKLAND, P. A. (1994): "Bartlett type identities for martingales," *Annals of Statistics*, 22, 21–38.

MYKLAND, P. A., N. SHEPHARD, AND K. SHEPPARD (2011): "Econometric analysis of financial jumps using efficient multipower variation," In preparation.

MYKLAND, P. A. AND L. ZHANG (2006): "ANOVA for Diffusions and Itô Processes," *Annals of Statistics*, 34, 1931–1963.

——— (2009): "Inference for continuous semimartingales observed at high frequency," *Econometrica*, 77, 1403–1455.

——— (2010): "The Econometrics of High Frequency Data," (to appear in *Statistical Methods for Stochastic Differential Equations*, M. Kessler, A. Lindner, and M. Sørensen, eds., Chapman and Hall/CRC Press).

O'HARA, M. (1995): *Market Microstructure Theorys*, Cambridge, MA: Blackwell Publishers.

PODOLSKIJ, M. AND M. VETTER (2009): "Bipower-type estimation in a noisy diffusion setting," *Stochastic Processes and Their Applications*, 119, 2803–2831.

REISS, M. (2010): "Asymptotic equivalence and sufficiency for volatility estimation under microstructure noise," ArXiv:1001.3006.

RENAULT, E. AND B. J. WERKER (2009): "Causality effects in return volatility measures with random times," *Journal of Econometrics* (forthcoming).

ROOTZÉN, H. (1980): "Limit Distributions for the Error in Approximations of Stochastic Integrals," *Annals of Probability*, 8, 241–251.

STIGLER, S. M. (2007): "The Epic Story of Maximum Likelihood," *Statistical Science*, 22, 598–620.

WANG, D. C. AND P. A. MYKLAND (2011): "The Estimation of Leverage Effect with High Frequency Data," Working paper, University of Oxford.

WONG, W. H. (1986): "Theory of Partial Likelihood," *Annals of Statistics*, 14, 88–123.

ZHANG, L. (2001): "From Martingales to ANOVA: Implied and Realized Volatility," Ph.D. thesis, The University of Chicago, Department of Statistics.

———— (2006): "Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach," *Bernoulli*, 12, 1019–1043.

ZHANG, L., P. A. MYKLAND, AND Y. AÏT-SAHALIA (2005): "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394–1411.

Per A. Mykland, Department of Statistics, The University of Chicago, Chicago, IL 60637, U.S.A.; mykland@galton.uchicago.edu; http://galton.uchicago.edu/~mykland.

Lan Zhang, Department of Finance, The University Illinois at Chicago, Chicago, IL 60607, U.S.A.; lanzhang@uic.edu; http://tigger.uic.edu/~lanzhang/.

APPENDIX: PROOFS

A   Approximation in Windows of one Observation

We here discretize to the nearest $t_{n,j}$ instead of to $\tau_{n,i-1}$.

Consider the the cumulative distribution function $F(x|s, f, g; \Delta t_{n,j+1})$ of $\Delta X_{t_{n,j+1}}/(s\Delta t_{n,j+1}^{1/2})$ under $Q(\cdot|t_{n,j}, t_{n,j+1}, X_{t_{n,j}}, s, f, g)$. The partial log likelihood ratio with P^* is given by

$$LR_{1,n} = \sum_{j=1}^{n} \log \frac{dF(\Delta X_{t_{n,j+1}}/(\sigma_{t_{n,j}}\Delta t_{n,j+1}^{1/2})|\sigma_{t_{n,j}}, f_{t_{n,j}}, g_{t_{n,j}}; \Delta t_{n,j+1})}{dP^*(\Delta X_{t_{n,j+1}}/(\sigma_{t_{n,j}}\Delta t_{n,j+1}^{1/2})|\mathcal{F}_{t_{n,j}})} \tag{A.1}$$

We also consider discretization of f and g to

$$t'_{n,j} = \max\{\tau_{n,i} \leq t_{n,j}\}. \tag{A.2}$$

The approximating measure is then formed with the help of

$$LR_{2,n} = \sum_{j=1}^{n} \log \frac{dF(\Delta X_{t_{n,j+1}}/(\sigma_{t_{n,j}}\Delta t_{n,j+1}^{1/2})|\sigma_{t_{n,j}}, f_{t'_{n,j}}, g_{t'_{n,j}}; \Delta t_{n,j+1})}{dP^*(\Delta X_{t_{n,j+1}}/(\sigma_{t_{n,j}}\Delta t_{n,j+1}^{1/2})|\mathcal{F}_{t_{n,j}})} \tag{A.3}$$

LEMMA 2. *(1) Subject to regularity conditions,*[10]

$$LR_{1,n} \xrightarrow{p} 0 \ as \ n \to \infty. \tag{A.4}$$

The result remains valid if $P^*(\Delta X_{t_{n,j+1}}/(\sigma_{t_{n,j}}\Delta t_{n,j+1}^{1/2})|\mathcal{F}_{t_{n,j}})$ *is replaced by*
$P^*(\Delta X_{t_{n,j+1}}/(\sigma_{t_{n,j}}\Delta t_{n,j+1}^{1/2})|\sigma_{t_{n,j}}, f_{t_{n,j}}, g_{t_{n,j}})$ *in (A.1).*

(2) If f_t *and* g_t *are themselves continuous semimartingales,*

$$LR_{2,n} \xrightarrow{p} 0 \ as \ n \to \infty. \tag{A.5}$$

The result remains valid (and the limit of $LR_{2,n}$ is unchanged) if
$P^*(\Delta X_{t_{n,j+1}}/(\sigma_{t_{n,j}}\Delta t_{n,j+1}^{1/2})|\mathcal{F}_{t_{n,j}})$ *is replaced by* $P^*(\Delta X_{t_{n,j+1}}/(\sigma_{t_{n,j}}\Delta t_{n,j+1}^{1/2})|\sigma_{t_{n,j}}, f_{t'_{n,j}}, g'_{t_{n,j}})$ *in (A.3).*
In either case, call the resulting probability P'_n, *and to sum up,*

$$\log \frac{dP'_n}{dP^*} \xrightarrow{p} 0 \ as \ n \to \infty. \tag{A.6}$$

Proof of Lemma 2. This follows from the proof of Theorem 1 in Mykland and Zhang (2009). Compare the martingale $\check{M}_n^{(0)}$ in (A.20) (p. 33) for the numerator and the denominator in the likelihood ratio, and obtain, in the case (2) that

$$\check{M}_n^{(0)\text{num}} - \check{M}_n^{(0)\text{denom}} = \sum_{j=0}^{n-1} \frac{1}{2}\left(\Delta t_{n,j+1}^{1/2}\frac{f_{t'_{n,j}} - f_{t_{n,j}}}{\sigma_{t_{n,j}}} + O_p(\Delta t_{n,j+1})\right)h_3(\Delta X_{t_{n,j+1}}/(\sigma_{t_{n,j}}\Delta t_{n,j+1})^{1/2}) \tag{A.7}$$

where h_3 is the third Hermite polynomial. (Note that in the notation of the earlier paper, $\Delta \check{W}_{t_{n,j+1}} = \Delta X_{t_{n,j+1}}/\sigma_{t_{n,j}}$). The discrete time quadratic variation of this difference goes to zero. In the case (1), $\check{M}_n^{(0)\text{num}} - \check{M}_n^{(0)\text{denom}} = 0$. ∎

B Proof of Theorem 1

Let R_n^* be the probability distribution formed by further replacing
$F(\Delta X_{t_{n,j+1}}/(\sigma_{t_{n,j}}\Delta t_{n,j+1}^{1/2})|\sigma_{t_{n,j}}, f_{t'_{n,j}}, g_{t'_{n,j}}; \Delta t_{n,j+1})$ by $\Phi(\Delta X_{t_{n,j+1}}/(\sigma_{t_{n,j}}\Delta t_{n,j+1}^{1/2}))$, *i.e.*, one further uses the likelihood ratio

$$LR_{3,n} = \sum_{j=1}^{n} \log \frac{d\Phi(\Delta X_{t_{n,j+1}}/(\sigma_{t_{n,j}}\Delta t_{n,j+1}^{1/2}))}{dF(\Delta X_{t_{n,j+1}}/(\sigma_{t_{n,j}}\Delta t_{n,j+1}^{1/2})|\sigma_{t_{n,j}}, f_{t'_{n,j}}, g_{t'_{n,j}}; \Delta t_{n,j+1})} \tag{B.8}$$

R_n^* is thus the probability distribution given in Section 2.3 of Mykland and Zhang (2009), but with $f_{t'_{n,j}}$ replacing $f_{t_{n,j}}$. By the proof of Theorem 1 in Mykland and Zhang (2009), $\frac{dR_n^*}{dP'_n} = \exp\{LR_{3,n}\}$ converges stably in law and is uniformly integrable (under P'_n), and the same applies to $\frac{dP'_n}{dR_n^*} = \exp\{-LR_{3,n}\}$ (under R_n^*).

If $\tilde{\sigma}_t$ replaces σ_t in the above, we can form $\tilde{P}'_n$ and $\tilde{R}^*_n$, as well as let $\widetilde{LR}_{3,n}$ denote the expression (B.8) with $\tilde{\sigma}_{t_{n,j}}$ replacing $\sigma_{t_{n,j}}$. Note that

$$P^*_n = \tilde{P}'_n. \tag{B.9}$$

We now focus on

$$LR_{4,n} = \sum_{j=1}^{n} \log \frac{d\Phi(\Delta X_{t_{n,j+1}}/(\tilde{\sigma}_{t_{n,j}} \Delta t_{n,j+1}^{1/2}))}{d\Phi(\Delta X_{t_{n,j+1}}/(\sigma_{t_{n,j}} \Delta t_{n,j+1}^{1/2}))} \tag{B.10}$$

Our aim is to show that $\frac{d\tilde{R}^*_n}{dR^*_n} = \exp\{LR_{4,n}\}$ is uniformly integrable and converges stably in law under R^*_n, and similarly for $\frac{dR^*_n}{d\tilde{R}^*_n}$ under $\tilde{R}^*_n$. It follows from this that all of R^*_n, $\tilde{R}^*_n$, $\tilde{P}_n$, $\tilde{P}'_n$, and P^* are contiguous. In particular, convergence in probability has the same meaning under all these measures. Further inspection of the proof of Theorem 1 in Mykland and Zhang (2009) then yields that

$$LR_{3,n} - \widetilde{LR}_{3,n} \xrightarrow{p} 0. \tag{B.11}$$

Thus, from (A.6),

$$\begin{aligned} \log \frac{d\tilde{P}'_n}{dP^*} &= \log \frac{d\tilde{P}'_n}{d\tilde{R}^*_n} + \log \frac{d\tilde{R}^*_n}{dR^*_n} + \log \frac{dR^*_n}{dP_n} + \log \frac{dP'_n}{dP^*} \\ &= \log \frac{d\tilde{R}^*_n}{dR^*_n} + o_p(1). \end{aligned} \tag{B.12}$$

It therefore remains to study the properties of (B.10).

Following Proposition 2 (p. 1417) of Mykland and Zhang (2009),

$$LR_{4,n} = Z_n^{(1)} + \frac{1}{2} \sum_i \sum_{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}]} (\log \sigma^2_{t_{n,j}} - \log \tilde{\sigma}^2_{t_{n,j}}) \tag{B.13}$$

where $Z_n^{(1)} = \sum_j \Delta Z_{n,t_{n,j+1}}^{(1)}$, and

$$\Delta Z_{n,t_{n,j+1}}^{(1)} = \frac{1}{2}(\sigma^{-2}_{t_{n,j}} - \tilde{\sigma}^{-2}_{t_{n,j}})\Delta X^2_{t_{n,j+1}}.\Delta t^{-1}_{n,j+1} \tag{B.14}$$

Set $A_{n,j} = \sigma^{-2}_{t_{n,j}} \tilde{\sigma}^2_{t_{n,j}} - 1$ and let $V_{n,j+1}$ be defined as the first option in (25) in the Theorem,[11] so that

$$\Delta Z_{n,t_{n,j+1}}^{(1)} = \frac{1}{2}(V_{n,j+1} + 1)A_{n,j}. \tag{B.15}$$

Note that $V_{n,j+1} + 1$ is conditionally χ^2_1 under $\tilde{R}^*_n$. Set

$$B_{n,j} = \frac{2}{\sigma_{\tau_{n,i-1}}}(\tilde{\sigma}_{t_{n,j}} - \sigma_{t_{n,j}}) \tag{B.16}$$

If $p \geq 1$ and C_i are a sequence of constants, we obtain, for p-norms under $\tilde{R}_n^*$

$$||A_{n,j} - B_{n,j}||_p \leq \left(||\frac{\tilde{\sigma}_{t_{n,j}} - \sigma_{t_{n,j}}}{\sigma_{t_{n,j}}^2}||_{2p} + 2||\frac{1}{\sigma_{t_{n,j}}^2} - \frac{1}{\sigma_{t'_{n,j}}^2}||_{2p} \right) ||\tilde{\sigma}_{t_{n,j}} - \sigma_{t_{n,j}}||_{2p}$$

$$\leq C_1 \left| ||\langle \tilde{\sigma} - \sigma, \tilde{\sigma} - \sigma \rangle_{t_{n,j}}||_p ||\langle \sigma, \sigma \rangle_{t_{n,j}} - \langle \sigma, \sigma \rangle_{t'_{n,j}}||_p \right) ||\langle \tilde{\sigma} - \sigma, \tilde{\sigma} - \sigma \rangle_{t_{n,j}}||_p$$

$$\leq C_2 (t_{n,j} - t'_{n,j})^3. \tag{B.17}$$

Hence, if we set $Z_n^{(0)} = \sum_j \Delta Z_{n,t_{n,j+1}}^{(0)}$, with

$$\Delta Z_{n,t_{n,j+1}}^{(0)} = \frac{1}{2}(V_{n,j+1} + 1)B_{n,j} = (V_{n,j+1} + 1)\sigma_{\tau_{n,i-1}}(\tilde{\sigma}_{t_{n,j}} - \sigma_{t_{n,j}}), \tag{B.18}$$

we obtain

$$E|Z_n^{(1)} - Z_n^{(0)}|^2 \leq \sum_j ||V_{n,j+1} + 1||_{2q} ||A_{n,j} - B_{n,j}||_{2p}$$

$$\leq C_3 \sum_j (t_{n,j} - t'_{n,j})^3$$

$$\to 0 \text{ as } n \to \infty.$$

Now note that

$$E_{\tilde{R}_n^*}(\Delta Z_{n,t_{n,j+1}}^{(0)} | \mathcal{F}_{t_{n,j}}) = \frac{1}{2}B_{n,j} \tag{B.19}$$

and

$$\text{conditional variance of } \Delta Z_{n,t_{n,j+1}}^{(0)} = \frac{1}{2}B_{n,j}^2 \tag{B.20}$$

Finally, let $M_n^{(0)}$ be the (end point of the) martingale part (under $\tilde{R}_n^*$) of $Z_n^{(0)}$, so that

$$M_n^{(0)} = Z_n^{(1)} - (1/2)\sum_j B_{n,j} = \frac{1}{2}\sum_j V_{n,j+1}B_{n,j}. \tag{B.21}$$

This coincides with the definition (26) in the statement of the theorem. If $\langle \cdot, \cdot \rangle^{\mathcal{G}}$ represents discrete

time predictable quadratic variation on the grid $\mathcal{G}$ (under $\tilde{R}_n^*$), then equation (B.20) yields

$$
\begin{aligned}
\langle M_n^{(0)}, M_n^{(0)} \rangle^{\mathcal{G}} &= \frac{1}{2} \sum_j B_{n,j}^2 \\
&= 2 \sum_j \sigma_{\tau_{n,i-1}}^{-2} (\tilde{\sigma}_{t_{n,j}} - \sigma_{t_{n,j}})^2 \\
&= 2 \sum_j \sigma_{\tau_{n,i-1}}^{-2} \langle \tilde{\sigma} - \sigma, \tilde{\sigma} - \sigma \rangle_{t_{n,j}} + o_p(1) \\
&= 2 \sum_j \sigma_{\tau_{n,i-1}}^{-2} \int_{t'_{n,j}}^{t_{n,j}} (c_t - c_{t'_{n,j}})^2 dt + o_p(1) \\
&= 2 \sum_j \sigma_{\tau_{n,i-1}}^{-2} \int_{t'_{n,j}}^{t_{n,j}} (\langle c, c \rangle_t - \langle c, c \rangle_{t'_{n,j}} dt) + o_p(1) \\
&= \sum_j \sigma_{\tau_{n,i-1}}^{-2} (t_{n,j} - t'_{n,j})^2 \langle c, c \rangle'_{t'_{n,j}}) + o_p(1) \\
&= \int_0^T (\sigma_t^{-2} \langle c, c \rangle'_t) dK_2(t) + o_p(1) \\
&= \Gamma_0 + o_p(1), \tag{B.22}
\end{aligned}
$$

where K_2 is the QADD given by equation (24).

From (B.13), we can now proceed as in the proof of Theorem 3 in Mykland and Zhang (2009); note in particular that $\sup_j B_j^2 \to 0$, and also equation (A.29) in that paper:

$$
\begin{aligned}
LR_{4,n} &= Z_n^{(0)} + \frac{1}{2} \sum_i \sum_{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}]} (\log \sigma_{t_{n,j}}^2 - \log \tilde{\sigma}_{t_{n,j}}^2) + o_p(1) \\
&= Z_n^{(0)} + \frac{1}{2} \sum_j \log(1 - B_j) + o_p(1) \\
&= M_n^{(0)} - \frac{1}{2} \langle M_n^{(0)}, M_n^{(0)} \rangle^{\mathcal{G}} + o_p(1)
\end{aligned}
$$
$$\tag{B.23}$$

The result now follows as in the earlier paper ∎

C Derivations for Section 3.2

We are here wholly under measure Q. Let $\mathcal{A}_{t_{n,j}}$ and $S_{n,j+1}$ be as defined in Section 3.2.

Derivation of (17). This follows from

$$E_Q\left(S_{n,j+1}|\mathcal{A}_{t_{n,j}}\right) = \left(\frac{a}{c}\right)^2\left((\Delta m_{t_{n,j+1}})^2 - \Delta\langle m,m\rangle_{t_{n,j+1}}\right) \text{ and}$$

$$\mathrm{Var}_Q\left(S_{n,j+1}|\mathcal{A}_{t_{n,j}}\right) = 4\left(\frac{a}{c}\right)^2\left(\frac{b}{c}\right)^2\Delta\langle m,m\rangle_{t_{n,j+1}}(\Delta m_{t_{n,j+1}})^2 + 2\left(\frac{b}{c}\right)^4(\Delta\langle m,m\rangle_{t_{n,j+1}})^2. \quad (C.24)$$

∎

More detailed derivation of equation (19):

$$\mathrm{Var}_Q\left(\Delta S_{n,j+1}\mid\mathcal{F}_{t_{n,j}}\right) = E_Q\left(\Delta S_{n,j+1}^2\mid\mathcal{F}_{t_{n,j}}\right)$$

$$= \left(\frac{a}{c}\right)^4 E_Q\left((\Delta m_{t_{n,j+1}})^4\mid\mathcal{F}_{t_{n,j}}\right)$$

$$+ \left(4\left(\frac{a}{c}\right)^2\left(\frac{b}{c}\right)^2 - 2\left(\frac{a}{c}\right)^4\right) E_Q\left((\Delta m_{t_{n,j+1}})^2\Delta\langle m,m\rangle_{t_{n,j+1}}\mid\mathcal{F}_{t_{n,j}}\right)$$

$$+ \left(\left(\frac{a}{c}\right)^4 + 2\left(\frac{b}{c}\right)^4\right) E_Q\left(\Delta\langle m,m\rangle_{t_{n,j+1}}^2\mid\mathcal{F}_{t_{n,j}}\right)$$

$$= \left(\frac{a}{c}\right)^4\left(3\mathrm{Var}_Q\left(\Delta m_{t_{n,j+1}}\mid\mathcal{F}_{t_{n,j}}\right)^2 + \mathrm{cum}_{4,Q}\left(\Delta m_{t_{n,j+1}}\mid\mathcal{F}_{t_{n,j}}\right)\right)$$

$$+ \left(4\left(\frac{a}{c}\right)^2\left(\frac{b}{c}\right)^2 - 2\left(\frac{a}{c}\right)^4\right)$$

$$\times \left(E_Q(\Delta\langle m,m\rangle_{t_{n,j+1}}|\mathcal{F}_{t_{n,j}})^2 + \mathrm{cum}_{3,Q}(\Delta\langle m,m\rangle_{t_{n,j+1}},\Delta m_{t_{n,j+1}},\Delta m_{t_{n,j+1}}|\mathcal{F}_{t_{n,j}})\right)$$

$$+ \left(\left(\frac{a}{c}\right)^4 + 2\left(\frac{b}{c}\right)^4\right)\left(\mathrm{Var}_Q\left(\Delta\langle m,m\rangle_{t_{n,j+1}}|\mathcal{F}_{t_{n,j}}\right) + E_Q\left(\Delta\langle m,m\rangle_{t_{n,j+1}}|\mathcal{F}_{t_{n,j}}\right)^2\right) \quad (C.25)$$

Collecting terms, the coefficient in front of $E_Q\left(\Delta\langle m,m\rangle_{t_{n,j+1}}|\mathcal{F}_{t_{n,j}}\right)^2$ is therefore

$$3\left(\frac{a}{c}\right)^4 + \left(4\left(\frac{a}{c}\right)^2\left(\frac{b}{c}\right)^2 - 2\left(\frac{a}{c}\right)^4\right) + \left(\left(\frac{a}{c}\right)^4 + 2\left(\frac{b}{c}\right)^4\right) = 2\left(\frac{a^2+b^2}{c^2}\right)^2. \quad (C.26)$$

Using $a^2 + b^2 = 1$ yields (19). ∎

Proof of Lemma 1. Let p be a positive integer. By using Itô's formula on $d(t_{n,j+1}-t)^p(\tilde{\sigma}_t^2 - \tilde{\sigma}_{t_{n,j}}^2)$ and collecting terms, we obtain that

$$\int_{t_{n,j}}^{t_{n,j+1}}(t_{n,j+1}-t)^{p-1}d\langle m,m\rangle_t = \frac{c^2}{p}\left(2\int_{t_{n,j}}^{t_{n,j+1}}(t_{n,j+1}-t)^p dm_t + \frac{c^2}{p+1}(\Delta t_{n,j+1})^{p+1} + \tilde{\sigma}_{t_{n,j}}^2(\Delta t_{n,j+1})^p\right).$$

$$(C.27)$$

Hence, using (C.27) with $p = 1$, we get (20), in the case of the variance because

$$\mathrm{Var}_Q(\Delta\langle m,m\rangle_{t_{n,j+1}}|\mathcal{F}_{t_{n,j}}) = 4c^4 E_Q\left(\int_{t_{n,j}}^{t_{n,j+1}}(t_{n,j+1}-t)^2 d\langle m,m\rangle_t|\mathcal{F}_{t_{n,j}}\right) \quad (C.28)$$

where the transition to the second equation in (20) uses (C.27) with $p = 3$. Again using $p = 1$, we also get

$$\text{cum}_{3,Q}(\Delta\langle m, m\rangle_{t_{n,j+1}}, \Delta m_{t_{n,j+1}}, \Delta m_{t_{n,j+1}}|\mathcal{F}_{t_{n,j}})$$
$$= 2c^2\text{cum}_{3,Q}\left(\int_{t_{n,j}}^{t_{n,j+1}}(t_{n,j+1}-t)dm_t, \Delta m_{t_{n,j+1}}, \Delta m_{t_{n,j+1}}|\mathcal{F}_{t_{n,j}}\right). \tag{C.29}$$

By using the Bartlett identities for martingales (Mykland (1994)),

$$\text{cum}_{3,Q}\left(\int_{t_{n,j}}^{t_{n,j+1}}(t_{n,j+1}-t)dm_t, \Delta m_{t_{n,j+1}}, \Delta m_{t_{n,j+1}}|\mathcal{F}_{t_{n,j}}\right)$$
$$= \text{Cov}_Q\left(\int_{t_{n,j}}^{t_{n,j+1}}(t_{n,j+1}-t)dm_t, \Delta\langle m, m\rangle_{t_{n,j+1}}|\mathcal{F}_{t_{n,j}}\right)$$
$$+ 2\text{Cov}_Q\left(\int_{t_{n,j}}^{t_{n,j+1}}(t_{n,j+1}-t)d\langle m, m\rangle_t, \Delta m_{t_{n,j+1}}|\mathcal{F}_{t_{n,j}}\right)$$
$$= \text{Cov}_Q\left(\int_{t_{n,j}}^{t_{n,j+1}}(t_{n,j+1}-t)dm_t, 2c^2\int_{t_{n,j}}^{t_{n,j+1}}(t_{n,j+1}-t)dm_t|\mathcal{F}_{t_{n,j}}\right)$$
$$+ 2\text{Cov}_Q\left(c^2\int_{t_{n,j}}^{t_{n,j+1}}(t_{n,j+1}-t)^2dm_t, \Delta m_{t_{n,j+1}}|\mathcal{F}_{t_{n,j}}\right)$$
$$= 4c^2E_Q\left(\int_{t_{n,j}}^{t_{n,j+1}}(t_{n,j+1}-t)^2d\langle m, m\rangle_t|\mathcal{F}_{t_{n,j}}\right), \tag{C.30}$$

where the second to last transition uses (C.27) with $p = 1$ for the first term, and with $p = 2$ for the second term. Thus, from (C.28),

$$\text{cum}_{3,Q}(\Delta\langle m, m\rangle_{t_{n,j+1}}, \Delta m_{t_{n,j+1}}, \Delta m_{t_{n,j+1}}|\mathcal{F}_{t_{n,j}}) = 2\text{Var}_Q(\Delta\langle m, m\rangle_{t_{n,j+1}}|\mathcal{F}_{t_{n,j}}) \tag{C.31}$$

Next, again by the Bartlett identities,

$$\text{cum}_{4,Q}(\Delta m_{t_{n,j+1}}|\mathcal{F}_{t_{n,j}}) = -3\text{Var}_Q(\Delta\langle m, m\rangle_{t_{n,j+1}}|\mathcal{F}_{t_{n,j}})$$
$$+ 6\text{cum}_{3,Q}(\Delta\langle m, m\rangle_{t_{n,j+1}}, \Delta m_{t_{n,j+1}}, \Delta m_{t_{n,j+1}}|\mathcal{F}_{t_{n,j}})$$
$$= 9\text{Var}_Q(\Delta\langle m, m\rangle_{t_{n,j+1}}|\mathcal{F}_{t_{n,j}}) \tag{C.32}$$

by (C.31). Hence (21) follows. This proves the lemma. ∎