Between Data Cleaning and Inference: Pre-Averaging and Robust Estimators of the Efficient Price*

Per A. Mykland
The University of Chicago

Lan Zhang
University of Illinois at Chicago

This version: 20 January 2015

Abstract

Pre-averaging is a popular strategy for mitigating microstructure in high frequency financial data. As the term suggests, transaction or quote data are averaged over short time periods ranging from 30 seconds to five minutes, and the resulting averages approximate the efficient price process much better than the raw data. Apart from reducing the size of the microstructure, the methodology also helps synchronize data from different securities. The procedure is robust to short term dependence in the noise.

Since averages can be subject to outliers, and since they can pulverise jumps, we have developed a broader theory which also applies to cases where M-estimation is used to pin down the efficient price in local neighbourhoods. M-estimation serves the same function as averaging, but we shall see that it is safer. Good choices of M-estimating function greatly enhances the identification of jumps. The methodology applies off-the-shelf to any high frequency econometric problem.

In this paper, we develop a general theory for pre-averaging and M-estimation based inference. We show that, up to a contiguity adjustment, the pre-averaged process behaves as if one sampled from a semimartingale (with unchanged volatility) plus an independent error.

Estimating the efficient price is a form of pre-processing of the data, and hence the methods in this paper also serve the purpose of data cleaning.

Keywords: consistency, cumulants, contiguity, continuity, discrete observation, efficiency, equivalent martingale measure, Itô process, jumps, leverage effect, M-estimation, microscructure, pre-averaging, realised beta, realised volatility, robust estimation, stable convergence.

*We are grateful to for helpful comments and suggestions from Gib Bassett, Nour Meddahi, Neil Shephard, Kevin Sheppard, and Dan Christina Wang, as well as seminar participants in Chicago, London, Oxford, and Paris. We would also like to thank the Oxford-Man Institute, where the idea for this paper was hatched. Financial support from the National Science Foundation under grants SES 11-24526 amd DMS 14-07812 is gratefully acknowledged.

1 "A tale full of sound and fury"

The recent literature on high frequency financial data has indeed been focused on sound (noise) and fury (jumps). While the tale is significant and important, one of the lessons from it is that both noise and jumps can severely impact *statistical* significance. Especially when they occur in combination.[1]

Unlike Shakespeare's Macbeth, we are fortunately not here faced with ultimate questions, but rather with the more prosaic one of finding a signal – something significant – in the middle of the sound and fury. The purpose of this paper is to introduce two (intertwined) approaches which we believe can be helpful: M-estimation, and contiguity.

The analysis of these data started with the work of Andersen and Bollerslev (1998a,b), Andersen, Bollerslev, Diebold, and Labys (2001, 2003), Barndorff-Nielsen and Shephard (2001, 2002); Barndorff-Nielsen (2004), Jacod and Protter (1998), Zhang (2001), Mykland and Zhang (2006), and the group at Olsen and Associates (Dacorogna, Gençay, Müller, Olsen, and Pictet (2001)), focusing on the concept of *realised volatility* (RV).[2] The work was based on the assumption that log prices follow a semimartingale of the form

$$dX_t = \mu_t dt + \sigma_t dW_t + J_t, \tag{1}$$

where J_t is a process of jumps.[3] W_t is Brownian motion; μ_t and σ_t are random processes that can be dependent with W. We also denote the continuous part of X_t by

$$dX_t^c = \mu_t dt + \sigma_t dW_t. \tag{2}$$

The semimartingale model for prices is required by the no-arbitrage principle in finance theory (Delbaen and Schachermayer (1994, 1995, 1998)).

Somewhat startlingly, the data had feedback to the theory: log prices are not semimartingales after all. The authors found that in actual data, the RV does not, in fact, converge as predicted by theory. This was clarified by the so-called *signature plot* (introduced by Andersen, Bollerslev, Diebold, and Labys (2000), see also the discussion in Mykland and Zhang (2005)). This led researchers to investigate a model where the efficient log price X_t is latent, and one actually observes a contaminated process Y_{t_j}:

$$Y_{t_j} = X_{t_j} + \epsilon_{t_j} \tag{3}$$

[1] See, in particular, the discussions in Jacod and Protter (2012, Chapter 16.5, pp. 521-563) and Aït-Sahalia and Jacod (2014, Appendix A.4, p. 496-502).

[2] An instantaneous version of RV was earlier proposed by Foster and Nelson (1996) and Comte and Renault (1998). Antecedents can be found in Rosenberg (1972), French, Schwert, and Stambaugh (1987), and Merton (1980). For a number of other early papers, see the anthology Shephard (2005). For further references, see the review by Shephard and Andersen (2009).

[3] Some of the cited papers allow for jumps, others not.

The distortion ϵ_{t_j} is called either "microstructure noise" or "measurement error", depending on one's academic field (O'Hara (1995); Hasbrouck (1996)). The t_j can be transaction times, or quote times.

The discovery of the impact of microstructure on inference led researchers to seek methods for high frequency data which allows for such noise. So far, four main approaches have come to light:

- Two- and Multi-scale estimation: weighted subsampled RVs (Zhang, Mykland, and Aït-Sahalia (2005), Zhang (2006, 2011))
- Realised Kernel: weighted autocovariances (Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008))[4]
- Pre-averaging: take weighted local averages before taking squares (Jacod, Li, Mykland, Podolskij, and Vetter (2009a), Podolskij and Vetter (2009b))
- Quasi-likelihood (Xiu (2010))

All methods can achieve up to $O_p(n^{-1/4})$ convergence rate for volatility, which is as good as for parametric inference (σ, μ constant), cf. Gloter (2000), Gloter and Jacod (2000, 2001).[5] The approaches mainly differ in treatment of edge effects. Studies based on different microstructure models are also in development (Robert and Rosenbaum (2009)). A recent, more abstract, line of enquiry is based on equivalence of experiments (Hoffmann (2008), Reiss (2011), Jacod and Reiss (2012), Bibinger, Hautsch, Malec, and Reiss (2013)). The latter path is related to our own; see Example 2 in Section 3.1.

However, existing literature has been confined to estimation of volatility and very closely related objects.[6] Also each estimator has been studied on a case by case basis. This is in contrast to the much greater generality which can be achieved when there is no microstructure, including high frequency regression, analysis of variance, powers of volatility (Mykland and Zhang (2006, 2009), Kalnina (2012), Jacod and Rosenbaum (2013)), empirically based trading strategies (Zhang (2012)), semivariances (Barndorff-Nielsen, Kinnebrouck, and Shephard (2009b)), resampling (Kalnina and Linton (2007), Gonçalves and Meddahi (2009), Kalnina (2011), Gonçalves, Donovon, and Meddahi (2013)), volatility risk premia (Bollerslev, Gibson, and Zhou (2005), Bollerslev, Tauchen, and Zhou (2009)), the volatility of volatility (Vetter (2011)), robust approaches to volatility,[7] jump detection and estimation[8], and so on. In other words, the research assuming no microstructure has flourished.

[4]Realised kernel and Multi-scale estimation can be given adjustments to be asymptotically equivalent, see Bibinger and Mykland (2013).

[5]Other earlier methods based on parametric assumptions include, in particular, Zhou (1998) and Curci and Corsi (2005), which uses the famous parameter-free diagonalisation of the covariance matrix.

[6]Specifically bi- and multipower (Podolskij and Vetter (2009a), Jacod, Podolskij, and Vetter (2009b)) and integrated covariance under asynchronicity (Zhang (2011), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009a), Christensen, Kinnebrock, and Podolskij (2008a)). The only other main classes of estimators that have been studied in the presence of noise are jump (see Footnote 8) and leverage effect (Wang and Mykland (2014),Aït-Sahalia, Fan, Wang, and Yang (2013)).

[7]In addition to the other papers cited, see, *e.g.,* Andersen, Dobrev, and Schaumburg (2012, 2014).

[8] References include Barndorff-Nielsen (2004), Aït-Sahalia (2004), Mancini (2004), Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006), Aït-Sahalia and Jacod (2007, 2008, 2009, 2012), Jacod and Todorov (2010),

To some extent, this is legitimate. As an old saying puts it, one has to learn to walk before one learns how to run. Also, there is the hope that either subsampling or pre-averaging can be used to eliminate the microstructure problem, and/or that data can be cleaned so hard that they don't have error any more. Even with this latter strategy, however, it is difficult to assess the impact of microstructure noise without including it in the model. Data processing, such as subsampling or pre-averaging, may also distort the jump characteristics of the data, and thus adversely affect subsequent inference.

This raises the question of whether we as a community will have to redo everything on an estimator-by-estimator basis for more realistic models that allow for microstructure noise and/or jumps.

The purpose of this paper is to find a way around this gargantuan task. We characterize the price process with sound and fury in presence. We develop a general theory that asymptotically separates the impact of the continuous evolution of a signal (i.e. latent efficient price), of the jumps, and of the microstructure. The theory covers both pre-averaging and M-estimation. On the one hand, our theory reduces the impact of microstructure, irrespective of the target of estimation. Our approach will not solve all problems for going between the noise and no-noise cases, but it is a step in the direction of typing these two together. On the other hand, our theory does not truncate jumps before analysis, and we show that we can tightly control the degree of modification of jumps when using a suitable M-estimator preprocessing before analysis. Thus the inference is transparent about how jump characteristics play a role in inference, again regardless of the "parameters".

We have two main clusters of results. One is Theorems 1-3 in Section 2.5, which show that by moving from pre-averaging to pre-M-estimation, one can to a great extent avoid the pulverisation of jumps that is present in pre-averaging. M-estimation also opens the possibility for better efficiency (Section 2.5.4). The other main result is the Contiguity Theorem 6 in Section 4.2, which shows that, under pre-averaging (including pre-M-estimation), one can behave as if there is no pre-processing at all, but that there will appear to be extra micro-structure. This is up to contiguity, which can be corrected for post-asymptotically.

In the next section, we outline the ingredients of our theory in local neighbourhoods. Then in Section 3 we show how local behaviour in neighbourhoods can be converted into a global behaviour using Edgeworth expansions and contiguity. Section 4 then contains our main contiguity results. Examples of application are given in Section 5[9], whereupon we conclude the paper. Proofs are in the Appendix.

Jing, Kong, Liu, and Mykland (2012) Lee (2005), Lee and Mykland (2008). Huang and Tauchen (2005), Fan and Wang (2005), Jacod and Protter (2012), Lee and Mykland (2012), and Aït-Sahalia and Jacod (2014) do consider microstructure in connection with jumps.

[9] Other examples can be found in Mykland, Shephard, and Sheppard (2012), Mykland and Zhang (2014, Section 8), and Mykland and Zhang (2015)

2 The Elements of a General Theory: Local Behavior

2.1 Background and some notation

Our general theory will be based on estimating the efficient price X in small neighbourhoods. Specifically, we assume that observations Y_{t_j} on the form (1)-(3) are made at times

$$0 = t_0 < ... < t_i < ... < t_n = T. \tag{4}$$

The index n represents the total number of observations, and our arguments will be based on asymptotics as $n \to \infty$ while T is fixed. Meanwhile, neighbourhoods or blocks are defined by a much coarser grid of τ_i, also spanning $[0, T]$, so that

$$\text{block } \# \ i = \{\tau_{i-1} \le t_j < \tau_i\} \tag{5}$$

(the last block, however, includes T). We then seek an estimate $\hat{X}_i$ of the efficient price X in the time period $[\tau_{i-1}, \tau_i)$.

By *"local behaviour"* we mean the behavior of a single $\hat{X}_i$ in a single time period $[\tau_{i-1}, \tau_i)$. We show in the later Sections 3-4 how to sew together the local behaviours across all the time periods.

If we define the block size by

$$M_i = \#\{j : \ \tau_{i-1} \le t_j < \tau_i\}, \tag{6}$$

the hope is that substantial precision in the estimation of X is obtained if $M_i \to \infty$ with n, but with M_i increasing sufficiently slowly that the actual time interval $[\tau_{i-1}, \tau_i)$ stays small.[10] After all, the efficient price X is a moving target.

Example 1. (Pre-averaging.) This idea is behind the concept of pre-averaging (Jacod, Li, Mykland, Podolskij, and Vetter (2009a), Podolskij and Vetter (2009a,b), Jacod, Podolskij, and Vetter (2009b)). Define block averages for block i, $[\tau_{i-1}, \tau_i)$:

$$\bar{Y}_i \ = \ \frac{1}{M_i} \sum_{\tau_{i-1} \le t_j < \tau_i} Y_{t_j},$$

and let $\bar{X}_i$ be defined similarly based on X. The averaging yields a reduction of the size of microstructure noise from $O_p(1)$ to $O_p(M_i^{-1/2})$, since, by central limit type considerations,

$$\begin{aligned}
\bar{Y}_i &= \bar{X}_i \ + \ \bar{\epsilon}_i \\
&= \bar{X}_i \ + \ O_p(M_i^{-1/2}) \\
&\overset{?}{\approx} X_{\tau_{i-1}} \ + \ O_p(M_i^{-1/2})
\end{aligned}$$

[10] When reference to the total number n of observations is needed, we write $t_{n,j}$ instead of t_j, $\tau_{i,n}$ instead of τ_i, $M_{n,i}$ instead of $M_{n,i}$, and so on.

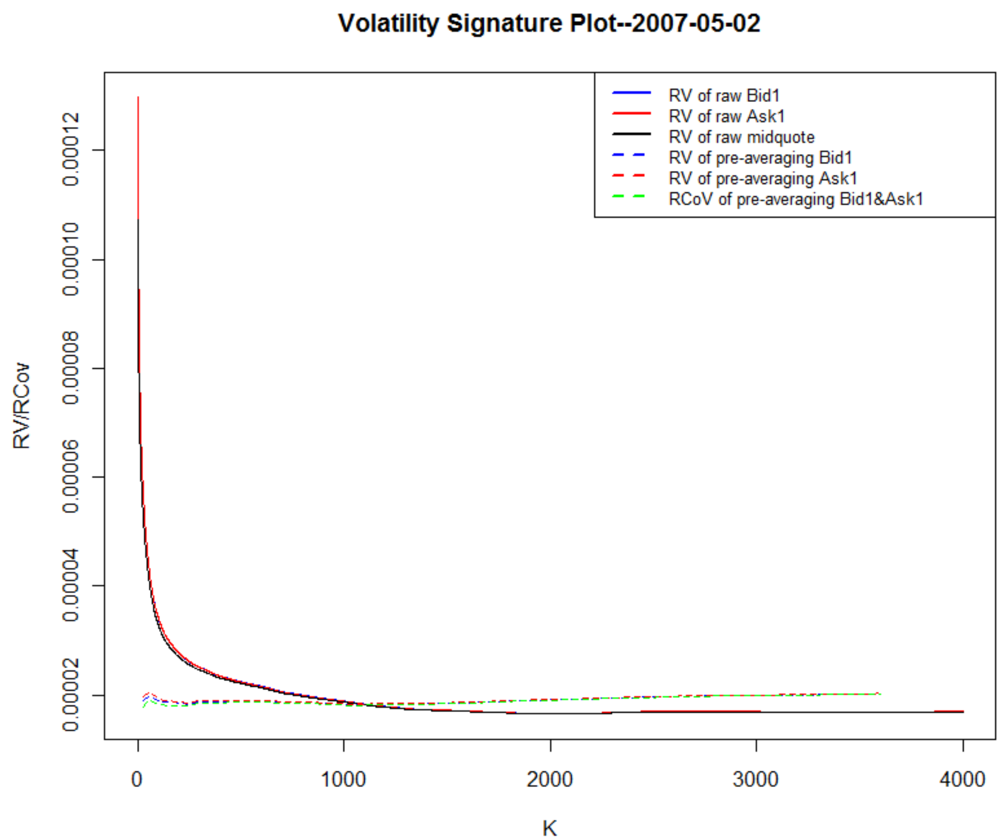

Figure 1: Realised volatility signature plots are given for CME's S&P E-mini quote data (best bid and ask, and midpoint), for both the raw data and for pre-averaged data. The conventional realised variance (RV) on the quotes explodes as the sampling interval K shrinks. This does not occur for the pre-averaged quotes.

The effect is clearly visible in the signature plot in Figure 1. The question, of course, is how to characterize $\bar{X}_i$, and how the averaging procedure impacts overall estimation. The cited papers study this for specific target quantities. We shall give a general form in this paper. □

As is seen from this example, pre-averaging is an appealing way to reduce the size of the noise. It can also be regarded as a form of data cleaning. Arithmetic means, however, are not robust to outliers, and we shall see that they are not robust to jumps. This raises the question of whether other estimators $\hat{X}_i$ can be found that are more robust, while at the same time also reduce the magnitude of the noise. This would be more in the spirit of data cleaning.

2.2 Connection to the Location Problem

To find robust estimators of the efficient price, we seek to emulate the classical problem of estimating location, where observations are i.i.d.,

$$Y_j = \theta + \epsilon_j. \tag{7}$$

This permits us to look in the existing literature for ideas. There is a wide variety of such estimators. The M-, L-, and R-estimators are discussed in the classical book by Huber (1981), P-estimators are due to Johns (1979), estimators that are robust and efficient are displayed by Stone (1974, 1975). Robust estimation include medians and quantiles, see, for example, Bahadur (1966), Koenker and Bassett (1978), Liu (1990), Donoho and Gasko (1992), and Chaudhuri (1996). It should be emphasized that robustness is a large research area, and this is just a small selection of references. In high frequency data, robust methods have been used (somewhat differently than here) by Christensen, Oomen, and Podolskij (2008b) and Andersen, Dobrev, and Schaumburg (2009).

We shall here focus on M-estimation. Similar theory can presumably be developed for other classes of robust estimators (such as L- and R- estimators).

2.3 Classical M-estimation

In the classical setting, M iid observations of the form (7) are made. The goal is to estimate θ. The estimator $\hat{\theta}$ is given as the solution of the estimating equation $\sum_{j=1}^{M} \psi(Y_j - \hat{\theta}_M) = 0$. Here, the estimating function ψ is an anti-symmetric ($\psi(-x) = -\psi(x)$) and usually nondecreasing function. ψ is usually bounded, but doesn't have to be. It is assumed that the noise satisfies $E\psi(\epsilon) = 0$ (more about this in Condition 3). If the ϵ_j are iid: $M^{1/2}(\hat{\theta}_M - \theta) \xrightarrow{\mathcal{L}} N\left(0, a^2\right)$ where, subject to $E\psi(\epsilon)^2 < \infty$,

$$a^2 = \frac{\mathrm{Var}(\psi(\epsilon))}{(E\psi'(\epsilon))^2}. \tag{8}$$

If the iid assumption is weakened to stationarity and exponential strong mixing,[11] then the theory goes through with

$$a^2 = \frac{1}{E(\psi'(\epsilon))^2}\left(\mathrm{Var}(\psi(\epsilon)) + 2\sum_{j=2}^{\infty}\mathrm{Cov}(\psi(\epsilon_1),\psi(\epsilon_j))\right). \tag{9}$$

For bounded ψ, estimation is robust to outliers by truncation: asymptotic variance is minimax in certain set of distributions for ϵ. It also has desirable "breakdown properties" (see the references in the previous section).

2.4 Location of the Efficient Price: Definition and Conditions

In analogy with the classical theory, we define the estimated process $\hat{X}_i$ in block i, $[\tau_{i-1}, \tau_i)$. $\hat{X}_i$ is given by

$$\sum_{\tau_{i-1}\le t_j<\tau_i} \psi(Y_{t_j} - \hat{X}_i) = 0 \tag{10}$$

The "classical" forms of ψ are graphed in Figure 2, and are symbolically given as

1. For $\psi(x) = x$, (10) yields pre-averaging: $\hat{X}_i = \bar{Y}_i$;
2. For $\psi(x) = \mathrm{sign}(x)$, (10) yields pre-medianisation: $\hat{X}_i = \mathrm{median}(Y_{t_j})$ in block i;[12]
3. An intermediate solution, the typical M-estimator form (Huber (1981), see the blue curve in Figure 2), lets c a positive constant and sets ψ to be

$$\psi_c(x) = \begin{cases} x \text{ for } |x| \le c \\ c \times \mathrm{sign}(x) \text{ otherwise} \end{cases} \tag{11}$$

This form represents a compromise: it behaves like the mean for small observations, and like the median for large observations. We shall see that for $\hat{X}_i$, it means treating the jumps and the microstructure robustly, while averaging the part of the returns that come from the continuous X^c. The estimating function ψ_c can be smoothed around $\pm c$ if desirable.

For the purposes of this paper, we make slightly more restrictive assumptions than what is common in the iid setting, as follows:

Condition 1. *The M-estimating function ψ is anti-symmetric ($\psi(-x) = -\psi(x)$), nondecreasing in x, strictly increasing in a neighbourhood of $x = 0$, with a bounded and continuous derivative ψ' which is absolutely continuous. Also, ψ'' is bounded.*

[11] We shall not pursue this possibility in this paper, but it is conjectured that the methods used will go through also in this case.

[12] Pre-medianization needs special theory development, and is for the most part beyond the scope of this paper.

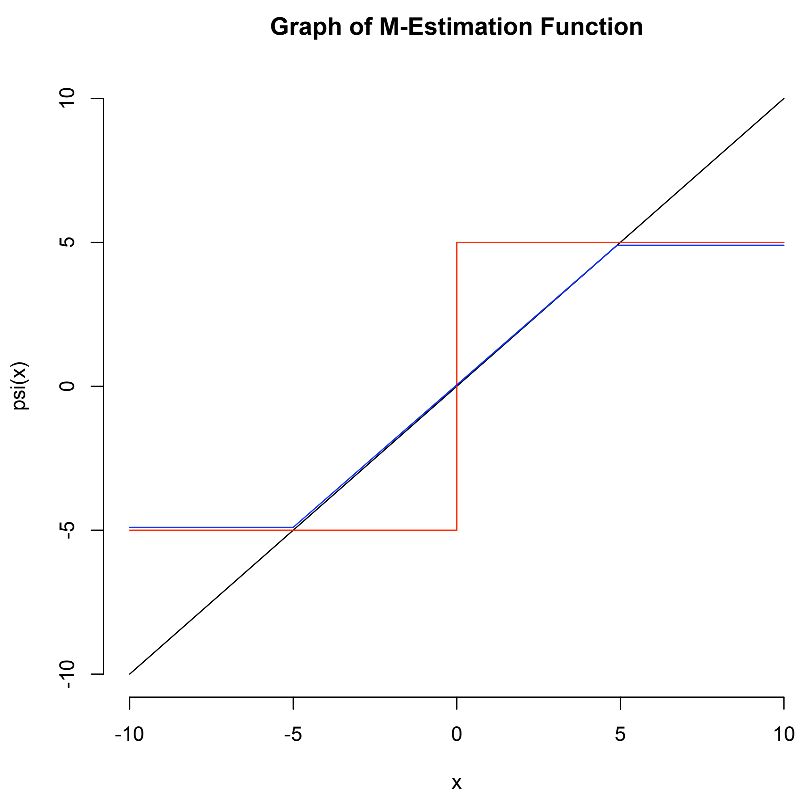

Figure 2: M-estimation function ψ for the three classical forms. Black corresponds to the mean, red to the median, and blue is the typical Huber (1981) form, which is linear in the middle and flat on the edge.

As a warmup, we here show how $\hat{X}_i$ relates to the classical M-estimator. We make assumptions here that are stronger than what is used in this section, but they will be needed in later sections.

Condition 2. The Process. *The observables Y_{t_j} are given by (1)-(3). The X process is a semimartingale, and μ_t and σ_t are random processes; μ_t is locally bounded, and σ_t is a continuous process. $(J_t)_{0\le t\le T}$ is a process of finitely many jumps, which is independent of the continuous part X_t^c of X_t.*[13] *We assume that the X process and all its components (such as σ_t, μ_t, W_t, and J_t) are adapted to a filtration $(\mathcal{F}_t)_{0\le t\le T}$.*

Condition 3. The Microstructure. *We assume that the ϵ_{t_j} are i.i.d.,*[14] *with $E(\psi(\epsilon))^2 < \infty$. We also suppose that $E\psi(\epsilon) = 0$, but this latter assumption is only pro forma.*[15] *The ϵ_{t_j} are assumed to be independent of $\mathcal{F}_T$ (in particular, of the X process) and of the observation times.*

Condition 4. The Observation Times. *The observation times (4) are independent of $\mathcal{F}_T$ (the filtration where X lives), and of the microstructure noise. Suppose that, as $n \to 0$, $\max(t_{n,j+1} - t_{n,j}) = o_p(1)$, and*

$$\sum_{j=0}^{n-1}(t_{n,j+1} - t_{n,j})^3 = O_p(n^{-2}). \tag{12}$$

Let K_n be the number of blocks in $[0,T]$. In terms of the relationship between the $\Delta\tau_i$'s, the M_i's, K_n, and n, we note that in an average sense $\bar{M} = n/K_n$, while at the same time, $\overline{\Delta\tau} = T/K_n$. This means that $\bar{M} = n\overline{\Delta\tau}/T$. We shall assume that this condition holds for each block in an order sense, which motivates the following:

Condition 5. (Orders of M_i and $\Delta\tau_i$.) *We assume that*[16]

$$\begin{aligned} M_{n,i} &= O_p(n\Delta\tau_{n,i}) \ \textit{exactly} && (13)\\ \Delta\tau_{n,i} &= O_p(n^{-1/2}) \ \textit{or smaller} && (14)\\ \Delta\tau_{n,i}^{-1} &= o_p(n^{3/5}) \ \textit{or smaller} && (15)\end{aligned}$$

We note that the framework permits us to work with equisized blocks in clock time, *i.e.,* $\Delta\tau_{n,i} = \Delta\tau_n = T/K_n$ independently of i. It also permits us to work with equisized blocks in transaction time, *i.e.,* $M_{n,i} = M_n = n/K_n$, independently of i. Or something more complicated. This choice is controlled by the econometrician.

[13] We have omitted the infinitely many jumps case since small jumps can in many cases be absorbed into the continuous part via contiguity (Zhang (2007)).

[14] The theory is conjectured to remain valid when the ϵ_{t_j} are stationary and strong mixing, with exponential decay of the mixing coefficients. (See Hall and Heyde (1980), p. 132 for discussion of mixing concepts.)

[15] If $E\psi(\epsilon) \neq 0$ there will be a nonrandom bias in $\hat{X}_i$ which is constant as a function of i. Since most estimators only depend on increments $\Delta\hat{X}_i = \hat{X}_i - \hat{X}_{i-1}$, this bias disappears in application.

[16] A consequence of (13)-(14) is that $M_i\Delta\tau_i = O_p(1)$. On the other hand, from (13) and (15), we obtain $M_i^{-1} = o_p(\Delta\tau_i^{2/3})$. Finally, if one wishes to think of $\Delta\tau_i = O_p(n^{-\alpha/2})$ (which is not required), then (13) means that $M_i = O_p(n^{1-\alpha/2})$ exactly. Meanwhile, (14)-(15) is the same as $1 \le \alpha < 6/5$.

CONVENTION. *Almost all times and observations have a double subscript, e.g., $\tau_{n,i}$. When there is no room for confusion, we occasionally suppress the first subscript n.*

2.5 Location of the Efficient Price: Decomposition Theorems, and How to Avoid the Pulverisation of Jumps

We now obtain the characterization of the estimate $\hat{X}_i$ of the latent efficient price process in block i. The following theorem suggests that, to first order, the M-estimation averages the continuous part of the signal X, but treats the jumps and the noise ϵ_{t_j} robustly.

2.5.1 A First Decomposition Theorem

THEOREM 1. *(Fundamental Decomposition of Estimator of Efficient Price.) Let $\hat{X}_i$ be the M-estimator in block i, defined by (10) and let $\bar{X}_i$ be the block average. Assume Conditions 2-5. Instead of Condition 1, we assume only that the M-estimating function ψ is anti-symmetric ($\psi(-x) = -\psi(x)$), nondecreasing in x, strictly increasing in a neighbourhood of $x = 0$. As above, Let M_i be the number of observations in block i. Finally, let $\hat{\theta}_i$ be the M-estimator based on the $\epsilon'_{t_j} = \epsilon_{t_j} + J_{t_j} - J_{\tau_{i-1}}$, i.e.,*

$$\sum_{\tau_{i-1} \le t_j < \tau_i} \psi(\epsilon_{t_j} + J_{t_j} - J_{\tau_{i-1}} - \hat{\theta}_i) = 0. \tag{16}$$

Then

$$\hat{X}_i = \hat{\theta}_i + X_{\tau_{i-1}} + O_p(\Delta\tau_i^{1/2}). \tag{17}$$

If we also assume Condition 1, then

$$\hat{X}_i = \hat{\theta}_i + X_{\tau_{i-1}} + \frac{\sum_{\tau_{i-1} \le t_j < \tau_i} (X^c_{t_j} - X^c_{\tau_{i-1}})\psi'(\epsilon'_{t_j} - \hat{\theta}_i)}{\sum_{\tau_{i-1} \le t_j < \tau_i} \psi'(\epsilon'_{t_j} - \hat{\theta}_i)} + O_p(\Delta\tau_i). \tag{18}$$

The above result shows that when there are no jumps in interval $[\tau_{i-1}, \tau_i)$, then $\hat{X}_i = \bar{X}_i + \hat{\theta}_i + O_p(\Delta\tau_i)$, thus cleanly decomposing the effect in M-estimation as robust for the noise, but averaging for the continuous part of the signal, *i.e.,* X^c. In the case where there are jumps in $[\tau_{i-1}, \tau_i)$, the noise *and* the jumps are to first order subject to M-estimation, *cf.* (17). On the other hand, the continuous part of the signal is subject to a weighted averaging, *cf.* (18), and also (29) below.

2.5.2 Noise and Jumps: Behaviour of $\hat{\theta}_i$, and a Second Decomposition Theorem

With Theorem 1 in hand, the behaviour of $\hat{\theta}_i$ achieves some importance. In intervals where there are no jumps, we are back to the situation of Section 2.3, with $\theta = 0$. If there are jumps, we can proceed as follows.

Definition 1. (Formal Strategy for handling Jumps, and Observation Times.) *Define* $\mathcal{G}_t = \mathcal{F}_t \vee \sigma((J_t)_{0\leq t\leq T}) \vee \sigma(t_{n,j},\ \text{all } (n,j))$. *In other words, we condition on the jump process and on the times. They can still, however, have a probability distribution. If we need a full filtration, including the noise, we use* $\mathcal{H}_{n,t} = \mathcal{G}_t \vee \sigma(\epsilon_{t_{n,j}}, t_{n,j} \leq t)$. *Stable convergence*[17] *is defined with respect to the filtration* $(\mathcal{G}_t)_{0\leq t\leq T}$. *Noise related items will converge conditionally on* $\mathcal{G}_T$.[18]

Remark 1. From Conditions 2-4, the ϵ_{t_j} are independent of $\mathcal{G}_T$. Also, $(X_t^c)_{0\leq t\leq T}$ remains a semimartingale with respect to filtration $(\mathcal{G}_t)_{0\leq t\leq T}$. □

Definition 2. (The Meaning of an Interval having Jumps.) *The intention of the following is to deal with the problem that a small number of jumps can occur anywhere in a large number of intervals, albeit with small probability.*[19] *Define, as a function of the underlying* $\omega \in \Omega$,

$$i_{n,k} = i_{n,k}(\omega) = \text{ the } k^{\text{th}} \ \ i \text{ so that } \Delta J_{\tau_{n,i}}(\omega) > 0 \tag{19}$$

Suppose that there are N *jumps in total in* $[0,T]$, *then there are at most* N' *such* $i_{n,k}$, *with* $N' \leq N$. *Set*

$$\mathcal{J}_n \ = \ \{i_{n,k} \ : \ k = 1,...,N'\} \tag{20}$$

These are the intervals with jumps. The set $\mathcal{J}_n^c = \{1,...,K_n\} - \mathcal{J}_n$ *is the set of intervals without jumps.*

Remark 2. (Asymptotically, each interval has at most one jump.) Let ζ_k be the time of the k^{th} jump. There are eventually, for $n \geq n_0$,[20] at most one jump in each interval $[\tau_{i-1}, \tau_i)$. Hence

$$\zeta_k \in [\tau_{i_{n,k}-1}, \tau_{i_{n,k}}). \tag{21}$$

For $n \geq n_0$, equation (21) can serve as definition of $i_{n,k}$, *in lieu* of (19).

We emphasize that there is an ambiguity in notation in connection with the symbol ΔJ_{ζ_k}, *which means* $J_{\zeta_k} - J_{\zeta_k -}$. *This is the only instance where we use this meaning of "*Δ*". In all other cases,* Δ *refers to an increment on the grid of the* $\tau_{n,i}$ *or the grid of the* $t_{n,j}$. □

We are now in a position to define what $\hat{\theta}_i$ actually estimates.

[17] Stable convergence is as discussed in Rényi (1963), Aldous and Eagleson (1978), Hall and Heyde (1980, Chapter 3, p. 56), Rootzén (1980). For use in high frequency asymptotics, see Jacod and Protter (1998, Section 2, pp. 169-170), Zhang (2001), and later work by the same authors. Stable convergence commutes with measure change on $\mathcal{G}_T$ (Mykland and Zhang (2009, Proposition 1, p. 1408)). – Note that the converging random variable need not be $\mathcal{G}_T$-measurable, cf. Zhang (2006). With this convention, we suppress the need to distinguish between stable and conditional convergence. For discussions of stable convergence of instantaneous quantities, see Zhang (2001) and Mykland and Zhang (2008).

[18] The is similar to the dichotomy in Zhang, Mykland, and Aït-Sahalia (2005) and Zhang (2006).

[19] This can occur, for example, if the jumps come from a Poisson process, and the intervals $[\tau_{i-1}, \tau_i)$ are equidistant. In this case, conditional on the total number of jumps N, the probability of having at least one jump in any nonrandom interval i is easily seen to be $1 - K_n^{-N}$, *cf.* Ross (1996, Chapter 2.3).

[20] where n_0 can depend on ω

Definition 3. (Fraction of Observations before a Jump, and Target for $\hat{\theta}$.) *If $i = i_{n,k} \in \mathcal{J}_n$, we proceed as follows. By Remark 2, there is, for $n \geq n_0$, only one jump in each such interval $i_{n,k}$. When this happens, let $M'_{n,i_{n,k}} = \#\{\ t_{n,j} \in [\tau_{n,i_{n,k}-1}, \zeta_k)\}$ and $M''_{n,i_{n,k}} = M_{n,i_{n,k}} - M'_{n,i_{n,k}}$. Set*

$$\alpha_{n,k} \ = \ \frac{M'_{n,i_{n,k}}}{M_{n,i_{n,k}}}. \tag{22}$$

Also let

$$\theta_{n,i_{n,k}} \ = \ h(\Delta J_{\zeta_k}; \alpha_{n,k}) \tag{23}$$

where the function $(\delta, \alpha) \to h(\delta; \alpha)$ is implicitly defined as h in the form

$$\alpha f(h) + (1-\alpha) f(h-\delta) = 0 \text{ where } f(x) = E\psi(\epsilon - x). \tag{24}$$

We can thus characterize the behaviour of $\hat{\theta}_{n,i}$.

Theorem 2. ($\hat{\theta}_i$ in all intervals, including those containing jumps.) *Assume the first set of conditions in Theorem 1. Let i_n be a sequence of indices ($1 \leq i_n \leq M_n$) as $n \to \infty$. Then*

$$\hat{\theta}_{n,i_n} \ = \ \theta_{n,i_n} + o_p(1). \tag{25}$$

where

$$\theta_{n,i} \ = \ \begin{cases} 0 \text{ for } i \in \mathcal{J}_n^c \\ \theta_{n,i} \text{ given by (23) for } i \in \mathcal{J}_n \end{cases} \tag{26}$$

Also, conditionally on $\mathcal{G}_T$,

$$M_{n,i_n}^{1/2}(\hat{\theta}_{i_n} - \theta_{i_n}) \overset{\mathcal{L}}{\approx} N(0, a^2_{n,i_n}) \tag{27}$$

where[21]

$$a^2_{n,i} \ = \ \begin{cases} \frac{f_2(0)}{f'(0)^2} \text{ for } i \in \mathcal{J}_n^c \\ \\ \frac{\alpha_{n,k} f_2(\theta_{n,i_{n,k}}) + (1-\alpha_{n,k}) f_2(\theta_{n,i_{n,k}} - \Delta J_{\zeta_k})}{\left(\alpha_{n,k} f'(\theta_{n,i_{n,k}}) + (1-\alpha_{n,k}) f'(\theta_{n,i_{n,k}} - \Delta J_{\zeta_k})\right)^2} \text{ for } i = i_{n,k} \in \mathcal{J}_n \end{cases} \tag{28}$$

and where $f_2(x) = \mathrm{Var}(\psi(\epsilon - x))$. Furthermore, if we also assume Condition 1, then the decomposition (18) can be sharpened to

$$\hat{X}_{n,i_n} = \hat{\theta}_{n,i_n} + X_{\tau_{n,i_n-1}} + \Delta\tau_{n,i}^{1/2} T_{n,i_n} + O_p(\Delta\tau_{n,i_n}), \text{ where} \tag{29}$$

$$T_{n,i} \ = \ \begin{cases} \Delta\tau_{n,i}^{-1/2}(\bar{X}_{n,i} - X_{\tau_{n,i-1}}) \text{ for } i \in \mathcal{J}_n^c \\ \Delta\tau_{n,i}^{-1/2}\left(\frac{\sum_{\tau_{n,i-1} \leq t_j < \zeta_k}(X^c_{t_j} - X^c_{\tau_{i-1}}) f'(\theta_{n,i_{n,k}}) + \sum_{\zeta_k \leq t_j < \tau_{n,i}}(X^c_{t_j} - X^c_{\tau_{i-1}}) f'(\theta_{n,i_{n,k}} - \Delta J_{\zeta_k})}{M'_i f'(\theta_{n,i_{n,k}}) + M''_i f'(\theta_{n,i_{n,k}} - \Delta J_{\zeta_k})}\right) \text{ for } i = i_{n,k} \in \mathcal{J}_n \end{cases} \tag{30}$$

[21] For the case $i \in \mathcal{J}^c$, we are in conformity with the discussion in Section 2.3 and also our Condition 3. The definition of a^2 is as in (8). The same applies to (26).

We see that in all of (26), (28), and (30), the expressions for the jump case ($i \in J$) reduces to those of the no-jump case ($i \in J^c$) by setting $\Delta J = 0$. To see why (29) is an improvement on (18), observe that while the former expression has $M_{n,i}$ different weights for the $X^c_{t_j} - X^c_{\tau_{i-1}}$, the formula (30) has only one ($i \in J^c$) or two ($i \in J$) such weights. This makes it clear that the main remainder term $T_{n,i}$ is a (possibly two-weighted) average of the continuous evolution of the process X. This sets the stage for analysing $T_{n,i}$ in Section 2.7, from which we can obtain a synthesis for the M-estimation method in Section 2.8.

Remark 3. (The Form of our Central Limit Theorems.) The equation (27) is a *bona fide* central limit theorem, as follows. When we say that $Z_{n,1} \overset{\mathcal{L}}{\approx} Z_{n,2}$, we mean that the two probability distributions are close in the sense of a metric that corresponds to convergence in law, such as the Prokhorov metric (Billingsley (1995)). We resort to this formulation because both sides in (27) are moving with n. Not only is the left hand side a triangular array, but the right hand side is also a moving target. The latter is the case both because i_n moves, but also because, when i_n is of the form $i_{n,k} \in \mathcal{J}_n$, then $\alpha_{n,k}$ is also not necessarily convergent. For similar reasons, we shall resort to this formulation in all our limit theorems. □

For the case where there is no jump in an interval, an even sharper decomposition is needed for our global results in Section 4. Such a result is developed in Appendix B.

2.5.3 Going beyond Pre-averaging avoids the Pulverisation of Jumps

As a corollary to Theorem 2, we can define the *effective*[22] *jump signal process* as

$$J^{\mathrm{e}}_{n,i} = \theta_{n,i} + J_{\tau_{n,i-1}}. \tag{31}$$

A first order consequence of (29) is that

$$\hat{X}_{n,i} = J^{\mathrm{e}}_{n,i} + X^c_{\tau_{n,i_n-1}} + \text{ higher order terms,} \tag{32}$$

and the theorem provides the higher order terms.

From (23), we now see that in the case of pre-averaging, $\psi(x) = x$, the jump ΔJ_{ζ_k} is pulverised: $\theta_{n,i} = (1-\alpha_{n,k})\Delta J_{\zeta_k}$, so that (asymptotically) a fraction of $(1-\alpha_{n,k})$ of ΔJ_{ζ_k} is allocated to $J^{\mathrm{e}}_{i_{n,k}}$, while the remaining (fraction $\alpha_{n,k}$) is allocated to $J^{\mathrm{e}}_{i_{n,k+1}}$.[23] In other words, fraction $(1-\alpha_{n,k})$ of the jump is allocated to time τ_{i-1}, while the rest is allocated to time τ_i.[24] The implication is that pre-averaged data dampen the size of a jump by a substantial fraction, and this may further affect a wide range of statistics.[25]

[22] As opposed to "efficient".

[23] This is in view of (23).

[24] This is an asymptotic consideration, but it will be approximately true for finite n since $\theta_{n,i}$ is the limit of $\hat{\theta}_{n,i}$ in (29).

[25] Pre-averaging followed by TSRV may be an exception to this. We shall also see in Section 5 another example of a construction which is immune to jump-pulverisation. However, even in that example, one cannot set standard errors under pulverised jumps

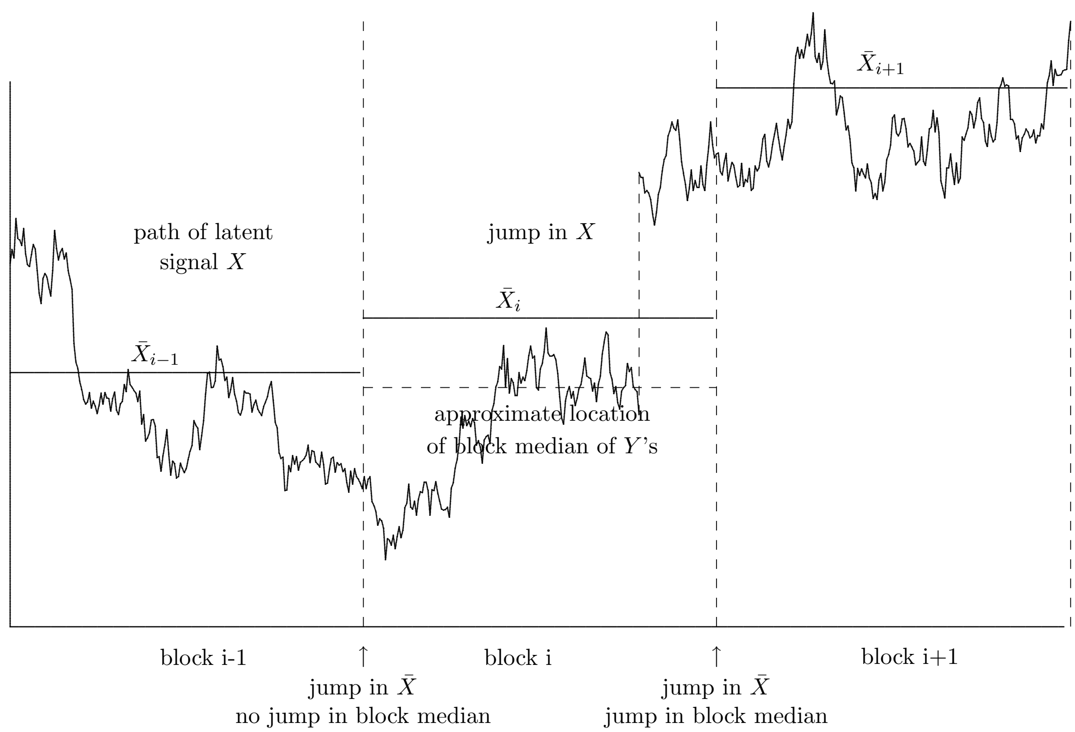

Figure 3: Intuition about the effect of ψ: We have here graphed a Brownian motion in three blocks, with a jump in the latter half of the second block. We assume that observations are made at equidistant times, and that the microstructure is negligible. The solid horizontal line is the mean in each block. For blocks $i-1$ and $i+1$ this line is also the approximate median. In the middle block, however, the median is indicated by the horizontal dashed line. – Because the majority of observations in the middle block is before the jump, the median places itself based on the before-jump observations. Thus the entire jump is allocated to the end time of the middle block. In the opposite case, if a majority of observations in the middle block were after the jump, the jump would be allocated to the starting point of block $\#i$. This is what we mean by the jump being allocated by *majority voting* when one uses the median. – As one can see, the mean tries to strike a compromise, and thereby pulverises the jump by putting it partly at the beginning and partly at the end of the middle block.

As a contrast to pre-averaging, we now consider the case where ψ has a more general form. $f(x) = E(\psi(\epsilon - x))$ now depends on the distribution of ϵ. Since the size of the noise is presumably small, one can consider the case where ϵ has cumulative distribution function $G(\cdot/v)$, and see what happens to $f(x)$ when $v \to 0$. Obviously, $f(x) = -\psi(x) + o(1)$ as $v \to 0$. A deeper investigation might take the form of an expansion in v, but is beyond the scope of this paper.[26] We shall here use a crude (but easy-to-see) bound:

PROPOSITION 1. (CRUDE BOUND ON THE EFFECT OF NOISE.) *Suppose that M-estimating function ψ is anti-symmetric ($\psi(-x) = -\psi(x)$), nondecreasing in x, strictly increasing in a neighbourhood of $x = 0$. Suppose that $|\epsilon| \leq v$. Suppose that h_v is given by (24). Then $|h_v(\delta) - h_0(\delta)| \leq v$.*[27]

We note that $h_0(\delta)$ is obtained by solving (24) with ψ is lieu of f, *i.e.,*

$$\alpha\psi(h_0) + (1-\alpha)\psi(h_0 - \delta) = 0 \tag{33}$$

We now consider the Huber form ψ_c, including $c = 0$ (the median) (Options 2 and 3 in Section 2.3; $c = +\infty$ corresponds to the mean). It is easy to see that if $|\delta| > 2c$ (δ is a largeish jump, in other words), then the solution $h_{c,0}$ of (33) with ψ_c is

$$h_{c,0}(\delta) \;=\; \begin{cases} \delta - c\ \mathrm{sign}(\delta)\frac{\alpha}{1-\alpha} \text{ for } \alpha < \frac{1}{2} \\ c\ \mathrm{sign}(\delta)\frac{1-\alpha}{\alpha} \text{ for } \alpha > \frac{1}{2} \end{cases} \tag{34}$$

From (34) we see that the ideal estimator is thus the median, ψ_0. It is worth noting that this is not only a large sample result. When using the median, it is easy to see that the allocation to $[\tau_{i-1}, \tau_i)$ or $[\tau_i, \tau_{i+1})$ will happen by majority voting, *cf.* Figure 3 and its caption. Unfortunately, for our global (contiguity) results in Section 4, the median does not satisfy the regularity conditions of the theorems.[28] However, since one is most worried about large jumps (Zhang (2007)), an estimating function of the form ψ_c for some $c > 0$ will, for small noise, be adequate.

Also for $c > 0$, there is an aspect of majority voting. If $\alpha < \frac{1}{2}$, the majority of the observations in the interval happen after the jump. The contamination is then limited by c in the direction away from δ. On the other hand, $\alpha > \frac{1}{2}$, the absolute value of the estimate $|h(\delta)|$ is maximally c. Similarly, if $h_{c,v}$ if formed from (24) with a contaminated ψ_c, and the contamination ϵ has absolute value bounded by v, if follows from Proposition 1 that

[26]A more incisive investigation would presumably include the confinement to large jumps, and an expansion of the error term $f(x) + \psi(x)$. This can presumably be carried out with a combination of contiguity (Zhang (2007)) and Laplace type methods for the asymptotic expansion of integrals, see, for example Jensen (1995, Chapter 3).

[27]For symmetric ϵ, the approximation will in most cases be of order $O(v^2)$.

[28]Since the noise can have a suggestion of discrete distribution to it, it may be as well to avoid the median for this reason.

Theorem 3. *Assume the conditions of Proposition 1. Also assume that $|\delta| > 2c$. Then*

$$
\begin{aligned}
|h_{c,v}(\delta) - \delta| &< c + v \text{ for } \alpha < \frac{1}{2} \\
|h_{c,v}(\delta)| &< c + v \text{ for } \alpha > \frac{1}{2}
\end{aligned} \tag{35}
$$

To summarize, (34)-(35) says that, by majority decision, the main part of a large jump in interval i will be allocated to one interval, either interval i or interval $i+1$. (Again, this is to say that the jump will be recorded as having happened at either $\tau_{n,i}$ or $\tau_{n,i+1}$). The amount of jump allocated to the other interval is maximally c or $c+v$, respectively. Under pre-averaging, on the other hand, up to half the jump ($\delta/2$) can be allocated to the other interval.

When there is noise, M-estimation is thus not perfect. But it pulversises large jumps much less than does pre-averaging.

2.5.4 M-estimation and Efficiency

Apart from a potentially better treatment of jumps, M-estimation also offers the possibility of greater efficiency. A main difference between general ψ and pre-averaging, however, lies in the behaviour of $Z_i = M_i^{1/2}(\hat{\theta}_i - \theta_i)$, and here the choice of ψ may affect the asymptotic variance of estimators. If the noise is Gaussian, the asymptotic variance of Z_i itself is, of course, minimized by pre-averaging, but this will not be the case for other noise distributions (Huber (1981)). For iid data, ψ can be chosen as the derivative of the log density of the data (Stone (1974, 1975)). We conjecture that this methodology can apply here as well, though such a development would be beyond the scope of this paper.

2.6 Intra-block behaviour

To find a compact characterization of the error in M-estimation, we shall use the following concept.

Definition 4. Intra-block behaviour *Define the random variable $I_i = I_{n,i}$ inside each block i as follows. Let $t_{j_0} = t_{j_{n,0}}$ be the first $t_j \in [\tau_{n,i-1}, \tau_{n,i})$, and set*

$$
I_{n,i} = \begin{cases} \frac{M_{n,i}-j}{M_{n,i}} \text{ with probability } \frac{\Delta t_{j_0+j}}{\Delta \tau_i} \\ 1 \text{ with probability } \frac{t_{j_0}-\tau_{i-1}}{\Delta \tau_i} \\ 0 \text{ with probability } \frac{\tau_i - t_{j_0+M_{n,i}-1}}{\Delta \tau_i} \end{cases} \tag{36}
$$

We shall see various moments of I appearing in the theorems below. There are two strategies for how to handle these moments. One is to plug in the actual times (in a data analysis). For theoretical or applied purposes, one can alternatively impose the condition that the times are approximately equispaced within blocks $[\tau_{n,i-1}, \tau_{n,i})$. This can take the following three forms:[29]

Definition 5. Regular Times. *A sequence of times $t_{n,j}$ will be said to be "regular" if it has one of the following generating processes:*

T1. Equidistant times. *This is where $\Delta t_{n,j} = T/n$. There is no reason to use anything but equisized blocks, and here clock time and transaction time coincide. This is a common assumption in the literature.*

T2. Mildly irregular Times. *This is where $t_{n,j} = f(j/n)$. We shall for simplicity assume that f is continuously differentiable and increasing, and nonrandom. This assumption (or variants thereof) has been used by Zhang (2006) and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008).*

T3. Time varying Poisson Process Times. *This is where $t_{n,j}$ is the jth observation from a Poisson process with intensity $\lambda_n(t)$. We shall for simplicity assume that the function $t \to \lambda_n(t)$ is continuously differentiable, and nonrandom. In order to make points denser as $n \to \infty$, we impose $n\lambda_- \leq \lambda_n(t) \leq n\lambda_+$.*[30]

Behavior of $\Delta\tau_{n,i}$, $M_{n,i}$ and $I_{n,i}$ under Regular Time Assumptions			
Assumptions		Effect	
T1	$M_{n,i}$ fixed $= M_n$	$\Delta\tau_{n,i} = M_n \Delta t$	$I_{n,i}$ is approximately uniformly distributed in all these cases
T1	$\Delta\tau_{n,i}$ fixed $= \Delta\tau_n$	$M_{n,i} = \Delta\tau_n/\Delta t$	
T2	$M_{n,i}$ fixed $= M_n$	$\Delta\tau_{n,i} \approx M_n f'(f^{(-1)}(\tau_{n,i-1}))$	
T2	$\Delta\tau_{n,i}$ fixed $= \Delta\tau_n$	$M_{n,i} \approx f'(f^{(-1)}(i\Delta\tau_n))/\Delta\tau_n$	
T3	$M_{n,i}$ fixed $= M_n$	$\Delta\tau_{n,i}$ is approximately Erlang distributed with parameters $(M_n, \lambda_n(\tau_{n,i-1}))$	
T3	$\Delta\tau_{n,i}$ fixed $= \Delta\tau_n$	$M_{n,i}$ is Poisson distributed with parameter $\Delta\tau_n^{-1} \int_{(i-1)\Delta\tau_n}^{i\Delta\tau_n} \lambda_n(t)dt \approx \lambda_n(i\Delta\tau_n)$	

Table 1: Behavior of block lengths $M_{n,i}$ and $\Delta\tau_{n,i}$, and of intra-block descriptor $I_{n,i}$ under various regular time assumptions. (For the distribution of $I_{n,i}$ in the Poisson case, see Mykland and Zhang (2012, Example 2.19(ii), p. 139).)

We note that Assumption T3 is quite different from Assumption T2, in that, for example, the asymptotic quadratic variation of time doubles under T3 relative to T2 (Mykland and Zhang (2012, Example 2.24, p. 148)). Note that all of conditions T1-T3 satisfy Condition 4 (*ibid,* Example 2.19, p. 138-139).

[29] Condition T1 is, of course, a special case of Condition T2, but is worth stating separately because of its ubiquity.

[30] As seen in Zhang (2011), such an assumption also permits useful subsampling arguments.

2.7 After the Noise and the Jumps: Averaging the Continuous Part of Signal gives rise to a Form of Microstructure

Section 2.5 details the estimation error $\hat{\theta}_i - \theta_i$ from the microstructure noise and the jump component J of the efficient price. We now investigate the estimation error from the continuous evolution of the efficient price X^c.

We shall here see that the error which comes from estimating the mean of the efficient price is asymptotically normal.

Definition 6. *Define the returns of the continuous part of the efficient price in block # i by*

$$R_{n,i} = \Delta\tau_{n,i}^{-1/2}(X^c_{\tau_{n,i}} - X^c_{\tau_{n,i-1}}) \tag{37}$$

Meanwhile, the part of the estimation error which is due to continuous evolution of the signal is

$$S_{n,i} = \Delta\tau_i^{-1/2}(\hat{X}_{n,i} - X_{\tau_{n,i-1}} - \hat{\theta}_i). \tag{38}$$

Recall from the development in Section 2.5 that

$$S_{n,i} = T_{n,i} + O_p(\Delta\tau_{n,i}^{1/2}). \tag{39}$$

where $T_{n,i}$ is the weighted mean of the $X^c_{t_{n,j}} - X^c_{\tau_{n,i-1}}$ given in (30) in Section 2.5.2. In the case where there is no jump in the interval $[\tau_{n,i-1}, \tau_{n,i})$, one retrieves straight pre-averaging of the signal:

$$T_{n,i} = \Delta\tau_i^{-1/2}(\bar{X}^c_i - X^c_{\tau_{n,i-1}}). \tag{40}$$

From standard martingale central limit considerations, $R_{n,i}/\sigma_{\tau_{n,i-1}}$ is asymptotically $N(0,1)$. We further obtain

Theorem 4. (Asymptotic Regression and Asymptotic Variance.) *Assume Conditions 1-5. Then there is a coefficient $\beta_{n,i}$ and a covariance matrix C_i, both of which only depends on the structure of the times $t_{n,j}$ and on the jump process J_t so that*

$$\tilde{T}_{n,i} = T_{n,i} - \beta_{n,i}R_{n,i} \text{ and } \tilde{S}_{n,i} = S_{n,i} - \beta_{n,i}R_{n,i} \tag{41}$$

(which are identical up to $O_p(\Delta\tau_{n,i}^{1/2})$) so that $T_{n,i} - \beta_{n,i}R_{n,i}$ is asymptotically independent of $R_{n,i}$ given $\mathcal{G}_T$, and so that $(R_{n,i}, \tilde{T}_{n,i})/\sigma_{\tau_{n,i-1}}$ are asymptotically independent, specifically $N(0, C_{n,i})$,[31] *where*

$$C_{n,i} = \begin{pmatrix} 1 & 0 \\ 0 & v_{n,i}^2 \end{pmatrix}. \tag{42}$$

[31] Recall Remark 3.

The convergence in law is "stable".[32] *When there is no jump in* $[\tau_{i-1}, \tau_i)$,

$$\beta_{n,i} \;=\; E(I_{n,i}) \text{ and } v_{n,i}^2 \;=\; \mathrm{Var}(I_{n,i}). \tag{43}$$

When there is one jump[33] *in the interval* $[\tau_{i-1}, \tau_i)$, $\beta_{n,i}$ *and* $v_{n,i}^2$ *are given in equation (C.7)-(C.8) in Appendix C.2. For regular times, the expression for* $\beta_{n,i}$ *in a jump interval is given by (C.10).*

For regular times (Section 2.6) it is easy to see that,

$$E(I_{n,i}) = \frac{1}{2}\ , \quad E(I_{n,i}^2) = \frac{1}{3}\ , \text{ and } \mathrm{Var}(I_{n,i}) = \frac{1}{12}\ , \text{ up to } o_p(1). \tag{44}$$

Remark 4. (Asymptotic Regressions, and the Effective Price.) Apart from providing the asymptotic distribution, Theorem 4 means that (41) represent the *asymptotic regressions* of $T_{n,i}$ and $S_{n,i}$ on $R_{n,i}$. This matters because $R_{n,i}$ is part of the return of the efficient log price, while the remainders in the regression ($\tilde{T}_{n,i}$ and $\tilde{S}_{n,i}$, respectively) are asymptotically (conditionally) independent of the return $R_{n,i}$.

In analogy with (31) in Section 2.5.3, we define the *effective* (still as opposed to "efficient") continuous signal process

$$X_{n,i}^{c,\mathrm{e}} = X_{\tau_{n,i-1}}^c + \Delta\tau_{n,i}^{1/2}\beta_{n,i}R_{n,i}. \tag{45}$$

For regular times,

$$X_{n,i}^{c,\mathrm{e}} = X_{\tau_{n,i-1}}^c + \frac{1}{2}(X_{\tau_{n,i}}^c - X_{\tau_{n,i-1}}^c) = \frac{1}{2}(X_{\tau_{n,i}}^c + X_{\tau_{n,i-1}}^c). \tag{46}$$

For the continuous part of the signal, therefore, sanity prevails, no matter how one removes the jump in Sections 2.5.2-2.5.3. □

2.8 Synthesis for the M-Estimator: Estimation Error as a Form of Microstructure

If we combine Theorems 2 (in Section 2.5.2) and 4 (in Section 2.7), we obtain the following decomposition of our estimated price:

$$\hat{X}_{n,i} \;=\; \underbrace{X_{n,i}^{c,\mathrm{e}} + J_{n,i}^{\mathrm{e}}}_{\text{“\textit{effective}” signal}} \;+\; \underbrace{\hat{\theta}_{n,i} - \theta_{n,i} + \Delta\tau_{n,i}^{1/2}\tilde{S}_{n,i}}_{\text{noise}}, \tag{47}$$

where we recall that $J_{n,i}^{\mathrm{e}}$ is the effective jump signal process defined in (31) in Section 2.5.3. The effective continuous signal process is given by (45) in the previous section.

We think of the terms

$$\eta_{n,i} = \hat{\theta}_{n,i} - \theta_{n,i} + \Delta\tau_{n,i}^{1/2}\tilde{S}_{n,i} \tag{48}$$

as being noise because, having conditioned on $\mathcal{G}_0$,

[32] Recall Footnote 17.

[33] The sequence of intervals may then follow a scheme akin to the one described in Section 2.5.2.

1. $M_{n,i}^{1/2}(\hat{\theta}_{n,i}-\theta_{n,i})$ is asymptotically normal and independent of the X^c process. The asymptotic variance is a^2 (from (8)) where there are no jumps, and given in Theorem 2 otherwise;

2. $\tilde{S}_{n,i} = \tilde{T}_{n,i} + O_p(\Delta\tau_{n,i}^{1/2})$ is also asymptotically stably normal, and independent of the continuous returns $R_{n,i}$. The (random) asymptotic variance is $\sigma^2_{\tau_{n,i-1}}\text{Var}(I_i)$ when there are no jumps, and given in Theorem 4 otherwise, *cf.* (C.8) in Appendix C.2.

The two sources of noise are also independent (conditionally on $\mathcal{G}_0$). One can therefore, think of the asymptotic variances as additive. In particular, when there is no jump in the interval $\#i$,

$$\text{AVAR}(\eta_{n,i}) = M_{n,i}^{-1}a^2 + \Delta\tau_{n,i}\sigma^2_{\tau_{n,i-1}}\text{Var}(I_{n,i}). \tag{49}$$

Remark 5. (Fixed spacings and balanced case.) In addition to assuming that $\Delta t_{n,j} = \Delta t_n$, we also assume that we have equispaced blocks in both transaction and clock time, *i.e.,*

$$\Delta\tau_n = M_n\Delta t_n. \tag{50}$$

We here also consider that we are also in the *balanced case*. This is to say that both sources of noise contribute to the asymptotic variance in (49). To achieve this, M_n^{-1} and $\Delta\tau_n$ must be of the same order, whence $M_n = cn^{1/2}$ (up to rounding to nearest integer), so that

$$\Delta\tau_n = M\Delta t_n = cn^{1/2}\frac{T}{n} = cTn^{-1/2}. \tag{51}$$

Here c is a tuning parameter determined by the econometrician. Fixed spacings is a special case of regular times, whence $\hat{X}_{n,i}$ has asymptotic mean (latent value) (46)-(47). If there are no jumps in interval $\#i$, the asymptotic variance becomes

$$M_n^{-1}a^2 + \Delta\tau_n\sigma^2_{\tau_{n,i-1}}\text{Var}(I_{n,i}) = n^{-1/2}\left(c^{-1}a^2 + \frac{1}{12}cT\sigma^2_{\tau_{n,i-1}}\right). \tag{52}$$

□

3 The Elements of a General Theory: Global Behavior

3.1 Strong Contiguity

Section 2 is entirely about the estimated efficient price process $\hat{X}_i$ on a local block i, *viz.* $[\tau_{i-1}, \tau_i)$. Various statistics will then be built by aggregating functions of $\hat{X}_i$ across blocks. We shall use the machinery of contiguity to study the behaviour of our aggregated estimators. This section explains our theoretical device of contiguity. We shall move to the global results in Section 4.

In order to clarify the structure of results, it is often helpful to move to an alternative but closely related probability distribution. Specifically begin by calling the original probability P. This is the one under which (1)-(3) holds. As discussed in Section 2.2 of Mykland and Zhang (2009), one can with little loss of generality move to an equivalent statistical martingale measure P^* where (1) is replaced by[34]

$$dX_t = \sigma_t dW_t + J_t. \tag{53}$$

This is because measure change commutes with stable convergence (*ibid*, same section, which also defines stable convergence). Note that we shall not change measure on the pure jump process J_t.

This simplification increases the transparency of arguments. We will now define a slight generalization of this concept. We shall consider approximate probabilities P_n under which the observations (and possible also auxiliary variables) have *exactly* (and not asymptotically) the simplified structure displayed in Sections 2.5 and 2.7-2.8, and at the same time provide for P_n to be close to P (and P^*) in a way that permits easy analysis. This is accomplished by the concept of *strong contiguity.*

Definition 7. *(Strong contiguity.) Let P_n be a sequence of probability distributions on a set of random variables (containing the relevant observables) $\mathcal{Z}_n = \{U_{n,1}, \cdots, U_{n,n}\}$. This set can be, $\hat{X}_i, i = 1, \ldots.$, but is typically richer, see the scenarios in Section 3.2.2. Then P_n is strongly contiguous relative to P provided that:*

1. *P_n and P are mutually absolutely continuous on the random variables $\mathcal{Z}_n$.*
2. *There is a a representation*

$$\log \frac{dP}{dP_n}(\mathcal{Z}_n) = L_n - \frac{1}{2}\eta^2 + o_p(1) \tag{54}$$

where L_n is the endpoint of a P_n martingale, and where the quadratic variation of L_n converges in probability to η^2, while L_n itself converges in law stably to $\eta N(0,1)$, where $N(0,1)$ is independent of the underlying data.

We refer to the martingale L_n in (54) as the martingale associated with $\log \frac{dP}{dP_n}$. Symbolically, we write $P_n \sim P$ when the two measures are mutually strongly contiguous. More generally, both probabilities can depend on n. Also, more generally, L_n can be of the form $L_n' + B_n$, where L_n' is a P_n martingale, and B_n is the endpoint of a continuous finite variation process of order $o_p(1)$. The quadratic variation process is unchanged between L_n and L_n'.

The statements about L_n and its quadratic variation are almost equivalent, see Jacod and Shiryaev (2003), and also Mykland and Zhang (2012). It follows from the definition that $\frac{dP}{dP_n}(\mathcal{Z}_n)$ converges in law stably to likelihood ratio $\exp(\eta N(0,1) - \frac{1}{2}\eta^2)$.

It will turn out that process structure can often be much more succinctly described under a strongly contiguous approximation. Meanwhile, the change of probability measure hardly affects

[34]We abuse notation by using the same symbol W in both (1) and (53). Our apologies.

inferential results. Specifically, consistency, rate of convergence, and asymptotic variance are unaffected. For example, if $n^{1/4}(\hat{\gamma}_n - \gamma)$ converges stably in law to $N(b, a^2)$ under P_n, then $n^{1/4}(\hat{\gamma}_n - \gamma)$ converges stably in law to $N(b', a^2)$ under P. The only alteration is therefore a possible change of b to b'. Often there is no change (and $b = b' = 0$), but to work out the change, one uses $b' = b+$ the asymptotic covariance of L_n and $n^{1/4}(\hat{\gamma}_n - \gamma)$. Post-asymptotic likelihood ratio correction is then carried out as in Theorems 2 or 4 of Mykland and Zhang (2009).

The background for these statements is discussed in Section 2.3-2.4 of Mykland and Zhang (2009), and this former paper implicitly uses the strong contiguity concept. We have here proceeded with a formal definition because greater complexity of the problem in the current paper requires more transparent notation and terminology.

As the name suggests, strong contiguity implies the usual statistical concept of contiguity (Hájek and Sidak (1967); LeCam (1986); LeCam and Yang (2000); Jacod and Shiryaev (2003)). The stronger version is suitable for our purposes.

Example 2. (Relationship to equivalence of experiments.) Our strong contiguity implies that P_n is an equivalent experiment to P (and P^*), cf. LeCam (1986); LeCam and Yang (2000). Our analysis therefore ties in with the recent literature on equivalence of experiments for high frequency data, see, in particular, Hoffmann (2008), Reiss (2011), Hoffmann (2008) Jacod and Reiss (2012), and Bibinger, Hautsch, Malec, and Reiss (2013). □

3.2 Contiguity from Edgeworth Expansion

We here see how strong contiguity, and properties such as approximate normality, can be derived from local Edgeworth expansions.

3.2.1 Orders of cumulants, and the (κ_i) process

Let V_i be a vector such as $(\tilde{S}_i, R_i)^T$, reflecting information which accrues in the interval # i. We also permit "state variable processes" (κ_i) that are adapted to the filtration $(\mathcal{G}_t)$ given by Definition 1. The (κ_i) process does not have to exhaust the information in the filtration. However, (κ_i) does need to contain all the information required for the asymptotics of the first few cumulants of V_i as $\Delta\tau_i \to 0$. Specifically, we suppose that under P^*, and as $\Delta\tau_i \to 0$, and in tensor notation

Condition 6. (Behavior of Cumulants)

$$\begin{aligned}
E(V_i^r \mid \mathcal{G}_{\tau_{n,i-1}}) &= \Delta\tau_i^{1/2}\kappa^r(\tau_{n,i-1}) + o_p(\Delta\tau_{n,i}^{1/2}) \\
\mathrm{Cov}(V_i^r, V_i^s \mid \mathcal{G}_{\tau_{n,i-1}}) &= \kappa^{r,s}(\tau_{n,i-1}) + o_p(\Delta\tau_{n,i}^{1/2}) \\
\mathrm{cum}(V_i^r, V_i^s, V_i^t \mid \mathcal{G}_{\tau_{n,i-1}}) &= \Delta\tau_{n,i}^{1/2}\kappa^{r,s,t}(\tau_{n,i-1}) + o_p(\Delta\tau_{n,i}^{1/2}) \qquad (55) \\
\mathrm{cum}(V_i^r, V_i^s, V_i^t, V_i^u \mid \mathcal{G}_{\tau_{n,i-1}}) &= \Delta\tau_{n,i}\kappa^{r,s,t,u}(\tau_{n,i-1}) + o_p(\Delta\tau_{n,i}).
\end{aligned}$$

In general, we assume that there is a process $\kappa_{n,i-1} = (\kappa^r(\tau_{n,i-1}), \kappa^{r,s}(\tau_{n,i-1}), \kappa^{r,s,t}(\tau_{n,i-1}), \kappa^{r,s,t,u}(\tau_{n,i-1});$ *all indices) which is of order* $O_p(1)$*. In the current paper, these processes take the form*

$$(\kappa^r(t), \kappa^{r,s}(t), \kappa^{r,s,t}(t), \kappa^{r,s,t,u}(t);\ \textit{all indices}), \tag{56}$$

but this is not required in general.

Remark 6. (How can we assume this?) For the first moment, one can often set it to zero by using an equivalent martingale measure. We shall see, however, that M-estimation will force a non-zero κ^r.

For the second moment: In the case where $V_i = (R_i, \tilde{T}_i)^T$, we can take the matrix $\kappa^{r,s}_{\tau_{i-1}} = \sigma^2_{\tau_{i-1}} \times C_i$ (cf. 42).[35]

For the third and fourth conditional cumulants, behaviour as in Condition 6 can typically be obtained in terms of underlying semimartingales from the Bartlett-type identities for martingales (Mykland (1994)). Results of the form (55) are derived in Appendix C-D. Similar results for the fourth conditional cumulant are easily derived by the same methods. □

3.2.2 Auxiliary random variables, the Scenarios, Partial likelihood, and the Target Approximation

We partition the variable $U_i = (A_i, V_i)$, where A_i are auxiliary random variables, and V_i are related to observables. We here consider three situations, which will be useful also in combination:

The Three Scenarios.

1. κ and $\hat{\theta}_i$ as auxiliary variable, so that, for example[36]

$$A_{i-1} = (\kappa(\tau_{i-1}), \hat{\theta}_i) \text{ and } V_i = (R_i, \tilde{S}_i) \tag{57}$$

   This corresponds to Theorem 5 in Section 4.1. The theorem will provide a one-step Gaussianity result along the lines of Mykland and Zhang (2009, Theorem 1).

2. All process variables are auxiliary (and have unchanged distribution), except $\tilde{S}_i$:

$$A_{i-1} = (\kappa_{i-1}, \hat{\theta}_i, R_i) \text{ and } V_i = \tilde{S}_i \tag{58}$$

   This corresponds to our main case: Theorem 6 in Section 4.2.

3. A reduced form problem where $\tilde{S}_i$ is not present: $A_{i-1} = (\kappa_{i-1}, \hat{\theta}_i)$ or $A_{i-1} = (\kappa_{i-1})$, in either case with $V_i = R_i$. This scenario plays a back stage rôle in Appendix E.

[35] In this case, the covariance matrix given $\mathcal{G}_{\tau_{n,i-1}}$, is of the form $\xi_{\tau_{i-1}} + E(\Delta\tau_i^{-1}\int_{\tau_{i-1}}^{\tau_i}(\tau_i - t)d\xi_t \mid \mathcal{G}_{\tau_{n,i-1}}) = \xi_{\tau_{i-1}} + O_p(\Delta\tau_i)$ when ξ_t is a continuous semimartingale, as required.

[36] We use $\tilde{S}_i$ rather than S_i since their asymptotic covariance with R_i is zero, thus simplifying expressions. Also, we use $\tilde{S}_i$ rather than $\tilde{T}_i$ because because of the exactness of the decomposition (47).

Note that if we set

$$\mathcal{Z}_{n,i-1} = \{U_1, \cdots, U_{i-1}\}, \tag{59}$$

then in all three scenarios, the sigma-field generated by $\mathcal{Z}_{n,i-1}$ is contained in $\mathcal{H}_{\tau_{i-1}}$, given in Definition 1.

The distinction between the scenarios has to do with what one is trying to achieve in any given situation. Scenario 1 is similar to Theorem 1 in Mykland and Zhang (2009). For the example in Section 5 of the current paper, Scenario 2 is the most interesting, as it reduces the pre-M-estimation effect to (shrinking) noise on the original process. Scenario 3 is exactly like Theorem 1 in Mykland and Zhang (2009), where it plays a main rôle, but in this paper, it is only present behind the scenes. $\square$

We shall alter the measure on V_i given the past, while the conditional measure of A_i stays unchanged, and thereby obtaining a measure P_n. As in Mykland and Zhang (2009), we have the likelihood decomposition[37]

$$f(U_1 ..., U_i, ..., U_K | U_0) = \underbrace{\prod_{i=1}^{K} f(V_i | U_0, ..., U_{i-1})}_{\text{altered from } P^* \text{ to } P_n} \quad \underbrace{\prod_{i=1}^{K} f(A_i | U_0, ..., U_{i-1}, V_i)}_{\text{unchanged from } P^* \text{ to } P_n} \tag{60}$$

We thus obtain, in obvious notation, the *partial likelihood* (Cox (1975); Wong (1986))

$$\log \frac{dP^*}{dP_n}(\mathcal{Z}_n) = \sum_i \log \left(\frac{f(V_i | U_0, ..., U_{i-1})}{f_n(V_i | U_0, ..., U_{i-1})} \right). \tag{61}$$

The above informs our definition of an approximate measure P_n which is conditionally normal:

Definition 8. (Target Approximation) *Define P_n to be the measure on the random variables*

$$\mathcal{Z}_n \;=\; \{U_i,\ i = 1, , K_n\}$$

for which,

$\mathcal{L}_{P_n}(V_i \mid U_0, ..., U_{i-1}) =$ exactly Gaussian with mean zero and covariance matrix $\kappa^{r,s}(\tau_{i-1})$, while

$$\mathcal{L}_{P_n}(A_i \mid U_0, ..., U_{i-1}, V_i) = \mathcal{L}_P(A_i \mid U_0, ..., U_{i-1}, V_i). \tag{62}$$

Since P_n is uniquely defined, we shall refer to this measure as the "canonical normal approximation" corresponding to the sequence $U_i = (V_i, A_i)$.

In other words, the requirement in (61) is that $f_n(\cdot | U_0, ..., U_{i-1})$ be a normal density with mean zero and covariance matrix $\kappa^{r,s}(\tau_{n,i-1})$.

---

[37] U_0 can vary according to scenario, but will typically only contain A_0 type information. In this paper, we shall also take the process $(J_t)_{0 \le t \le T}$ and the observation times as part of U_0. This is most convenient since $(J_t)_{0 \le t \le T}$ is independent of X^c and the ϵ_{t_j}s. We recall that the J process and the observation times are $\mathcal{G}_0$ measurable (Definition 1).

3.2.3 Edgeworth expansion

Now set L_n as the end point of the P_n-martingale

$$L_n \;=\; \sum_i \Delta\tau_i^{1/2}\kappa^r(\tau_{i-1})h_r(V_i) + \frac{1}{3!}\Delta\tau_i^{1/2}\kappa^{r,s,t}(\tau_{i-1})h_{rst}(V_i), \tag{63}$$

where the Hermite polynomials for interval # i are given by $h_r = h_r(v) = \kappa_{r,s}(\tau_{i-1})(v^s - \kappa^s(\tau_{i-1}))$ and $h_{rst} = h_{rst}(v) = h_r h_s h_t - h_r\kappa_{s,t}(\tau_{i-1})[3]$. We here use the notation of McCullagh (1987, Chapter 5), including the summation convention.[38]

Following the description on p. 147 in McCullagh (1987).[39]

$$\log\frac{dP}{dP_n} = \log\frac{dP}{dP_n}(\mathcal{Z}_n) = L_n - \frac{1}{2}\text{ q.v. of } L_n + o_p(1), \tag{64}$$

where "q.v." is the discrete time predictable quadratic variation process (with a suitably chosen time grid). The likelihood ratio in (64) converges stably to $\eta Z - \eta^2/2$, where Z is a random variable independent of the underlying process, and η^2 is the limit of the q.v. of L_n.

To calculate the P_n-predictable quadratic variation of L_n, note that

$$\mathrm{Cov}_{P_n}(h_{rst}, h_{abc} \mid \mathcal{Z}_{n,i-1}) = \kappa_{r,a}(\tau_{i-1})\kappa_{s,b}(\tau_{i-1})\kappa_{t,c}(\tau_{i-1})[3!] \tag{65}$$

(McCullagh (1987), p. 156). Hence,

$$\begin{aligned}&\mathrm{Var}_{P_n}(\Delta\tau_i^{1/2}\kappa^r(\tau_{i-1})h_r(V_i) + \frac{1}{3!}\Delta\tau_i^{1/2}\kappa^{r,s,t}(\tau_{i-1})h_{rst}(V_i)|\mathcal{Z}_{n,i-1})\\ &\quad= \Delta\tau_i\left\{\kappa^r(\tau_{i-1})\kappa^s(\tau_{i-1})\kappa_{r,s}(\tau_{i-1}) + \left(\frac{1}{3!}\right)^2\kappa^{r,s,t}(\tau_{i-1})\kappa^{a,b,c}(\tau_{i-1})\kappa_{r,a}(\tau_{i-1})\kappa_{s,b}(\tau_{i-1})\kappa_{t,c}(\tau_{i-1})[3!]\right\}.\end{aligned} \tag{66}$$

Under (56), we thus obtain

$$\eta^2 = \int_0^T \kappa^r(t)\kappa^s(t)\kappa_{r,s}(t)dt \;+\; \left(\frac{1}{3!}\right)^2\int_0^T \kappa^{r,s,t}(t)\kappa^{a,b,c}(t)\kappa_{r,a}(t)\kappa_{s,b}(t)\kappa_{t,c}(t)dt[3!]. \tag{67}$$

In the more general case (where the process (56) may not exist), it will normally be the case that (66) has a limit in probability. In the worst case, if one is willing to proceed through subsequences, our conditions assure that the limit of (66) will always exist, cf. the argument on p. 1411 of Zhang, Mykland, and Aït-Sahalia (2005).

[38] We use the further definitions: (1) $\kappa_{s,t}(\tau_{i-1})$ is the matrix inverse of $\kappa^{s,t}(\tau_{i-1})$; (2) "[3]" is the sum over the three possible combinations; (3) $h_r = h_r(v) = \kappa_{r,s}(\tau_{i-1})v^s$, the latter because the first conditional moment zero.

[39] See Remark 7 below.

Remark 7. (Regularity Conditions.) The L_n terms describe to main order the behaviour of $\log \frac{dP^*}{dP_n}$ by the same arguments that take you from (A.13) to (A.21) (pp. 1434-5) in Mykland and Zhang (2009). Orders of $O_p(\Delta t^{p/2})$ are replaced by orders of the form $O_p(\Delta\tau_i^{p/2})$, but in compensation, there are much fewer terms in the sum that makes up (63). Note that we require σ_t to be a continuous semimartingale, but this assumption can be lifted at the cost of more technicalities.

We have here followed an approach which does seek to determine the conditions under which the relevant expansions hold. This would massively expand the paper, and is beyond its scope. For references on rigorous conditions, see Wallace (1958), Bhattacharya and Ghosh (1978), Bhattacharya and Rao (1976), Hall (1992), and Jensen (1995). We also take intellectual refuge in the preface of Aldous (1989). For specific references concerning expansions of semimartingales, consult the new results in Li (2012), as well as the references in Remark 12 in Mykland and Zhang (2009). For the Edgeworth expansion of moments, see the proofs or Theorem 19.2 and Theorem 22.1 in Bhattacharya and Rao (1976), cf. also Jensen (1995, pp. 21-22). □

Remark 8. (Interpretation of summations.) Since this notation is exotic to many readers, here are some guidelines. First of all, the "[3!]" does not necessarily cancel against one of the 1/3!. Also, if the coefficients are not distinct, "one has to pretend that they are in order to get the correct coefficients in the sum."[40] In general, the [3!] combinations in equation (65)-(67) are given in Table 2. □

Combinations					
permutations			other		
ra sb tc	rb sc ta	rc sa tb	ra sc tb	rb sa tc	rc sb ta

Table 2: The 3! terms in equations (65)-(67).

Value of $\kappa^{r,s,t}\kappa^{a,b,c}\kappa_{r,a}\kappa_{s,b}\kappa_{t,c}[3!]$ when $\kappa_{r,a}$ is diagonal	
Configuration of indices	Value of $\kappa^{r,s,t}\kappa^{a,b,c}\kappa_{r,a}\kappa_{s,b}\kappa_{t,c}[3!]$
$\{abc\} \neq \{rst\}$	0
$\{abc\} = \{rst\} = \{123\}$ (all indices are different)	$\left(\kappa^{1,2,3}\right)^2 \kappa_{1,1}\kappa_{2,2}\kappa_{3,3}$
$\{abc\} = \{rst\} = \{112\}$	$2\left(\kappa^{1,1,2}\right)^2 \left(\kappa_{1,1}\right)^2 \kappa_{2,2}$
$\{abc\} = \{rst\} = \{111\}$	$6\left(\kappa^{1,1,1}\right)^2 \left(\kappa_{1,1}\right)^3$

Table 3: Value of the sum $\kappa^{r,s,t}\kappa^{a,b,c}\kappa_{r,a}\kappa_{s,b}\kappa_{t,c}[3!]$, when $\kappa_{r,a}$ is diagonal, in (66)-(67).

[40]See the analogous discussion in Mykland (1994), p. 22.

To concretely see how this works out, consider the case of a two dimensional V_i, where the components are asymptotically independent. This would be the case for $(R_i, \tilde{S}_i)$ from Section 2.7. The effect is shown in Table 3

4 The main one-step contiguity results.

4.1 Contiguity to one-step normal distribution for $(R_i, \tilde{S}_i)$

THEOREM 5. (CONTIGUITY TO ONE-STEP NORMAL DISTRIBUTION FOR $(R_i, \tilde{S}_i)$.) *Let $P_{n,1}$ be the canonical normal approximation corresponding to the sequence (57), so that $A_{i-1} = (\kappa_{i-1}, \hat{\theta}_i)$ and $V_i = (R_i, \tilde{S}_i)$. Use convention $V_i^0 = R_i$ and $V_i^1 = \tilde{S}_i$. Then*

1. *$P_{n,1}$ is strongly contiguous with respect to P and P^*, and relative to the set $\mathcal{Z}_n$;*
2. *under $P_{n,1}$, $Z_i = M_i^{1/2}(\hat{\theta}_i - \theta_i)$ are independent with the same distribution as under P, and Z_i is independent of X^c;*[41]
3. *under $P_{n,1}$, $\tilde{S}_i/\sigma_{\tau_{i-1}} v_{n,i}$ are iid normal N(0,1), and independent of the X^c and the Z processes, where $v_{n,i}$ is given in Theorem 4. In intervals i with no jumps, $v_{n,i}^2 = \mathrm{Var}(I_i)$;*[42]
4. *under $P_{n,1}$, $R_i/\sigma_{\tau_{i-1}}$ is normal N(0,1) and independent of $\mathcal{F}_{\tau_{i-1}}$;*
5. *Condition 6 is satisfied; $L_{n,1}$, given as in (63), satisfies (64), and for intervals with no jumps*[43] *the cumulant processes are of the form (for $r, s, t, = 0, 1$)*

$$\begin{aligned}
\kappa^r(\tau_{i-1}) &= \sigma^2_{\tau_{i-1}} b_i^r \\
(\kappa^{r,s}(\tau_{i-1})) &= \sigma^2_{\tau_{i-1}} \times C_i \\
\kappa^{r,s,t}(\tau_{i-1}) &= \sigma_{\tau_{i-1}} \langle \sigma, X^c \rangle'_{\tau_{i-1}} a_i^{r,s,t} \ + \ \sigma^4_{\tau_{i-1}} b_i^{r,s,t}
\end{aligned} \tag{68}$$

where C_i is the matrix from (42), and where

$$\begin{aligned}
a_i^{r,s,t} &= 2E\left((I_i - E(I_i))^{r+s}((I_i \wedge I_i') - E(I_i))^t\right)[3] - 3E\left(((I_i \wedge I_i') - E(I_1))^{r+s+t}\right) \\
b_i^0 &= 0 \text{ and } b_i^1 = \frac{1}{2}\frac{E\psi''(\epsilon)}{E\psi'(\epsilon)} E(I_i(1 - I_i))
\end{aligned} \tag{69}$$

$$b_i^{r,s,t} = 2\frac{E\psi''(\epsilon)}{E\psi'(\epsilon)} \left\{ \mathrm{cum}_{s+1}(I_i)\mathrm{cum}_{t+1}(I_i) + E\left((I_i \wedge I_i')(I_i - E(I_i))^s (I_i' - E(I_i'))^t\right) \right\} \delta_{\{r=1\}}[3]$$

where I_i' is an independent copy of I_i, and where cum_1 *is the expectation and* cum_2 *is the variance.*

---

[41] $\tilde{\theta}_i$ is zero in intervals with i with no jumps, and defined in (23) in Section 2.5.2 otherwise.

[42] Properties 3 and 4 are restatements of $P_{n,1}$ be the canonical normal approximation corresponding to the sequence (57), and are included for clarity.

[43] Apart from the variance, the values of the cumulants in intervals with jumps is asymptotically negligible, so long as Condition 6 remains satisfied. This is because we assume finitely many jumps.

The Theorem is derived in Appendix C-D and with reference to Remark 7 in Section 3.2.3. Note that $a_i^{r,s,t} = 2\tilde{\omega}^{k_1 k_2, k_3}[3]$ in the notation of Appendix C.3.

For regular times (Section 2.6), we obtain that for intervals with no jumps that [44]

$$C_i = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{12} + o_p(1) \end{pmatrix} \text{ and } b_i^1 = \frac{1}{12}\frac{\psi''(\epsilon)}{\psi'(\epsilon)} + o_p(1), \tag{70}$$

while the three dimensional tensors $a_i^{r,s,t}$ and $b_i^{r,s,t}$ are given in Table 4.

Three Dimensional Tensors under Regular Time Assumptions		
$\{r,s,t\}$	$a_i^{r,s,t}$	$b_i^{r,s,t} = 2\frac{E\psi''(\epsilon)}{E\psi'(\epsilon)}\times$
$\{0,0,0\}$	$-\frac{3}{2}$	0
$\{1,0,0\}$, $\{0,1,0\}$, $\{0,0,1\}$	$\frac{11}{12}$	$\frac{5}{24}$
$\{1,1,0\}$, $\{1,0,1\}$, $\{0,1,1\}$	$-\frac{1}{24}$	$\frac{1}{24}$
$\{1,1,1\}$	$\frac{199}{960}$	$\frac{1}{60}$

Table 4: Behavior of $a_i^{r,s,t}$ of $b_i^{r,s,t}$ under regular time assumptions (Section 2.6).

4.2 M-estimation as additional noise (Scenario 2)

Let $P_{n,2}$ be the canonical normal approximation corresponding to the sequence (58), where $A_{i-1} = (\kappa_{i-1}, \hat{\theta}_i, R_i)$ and $V_i = \tilde{S}_i$. We obtain from Appendix E that

Theorem 6. (M-estimation as additional noise.) *Let $P_{n,2}$ be the canonical normal approximation corresponding to the sequence (58).*

1. *$P_{n,2}$ is strongly contiguous with respect to P and P^*, and relative to the set $\mathcal{Z}_n$;*

[44] See also (42) and (44).

2. *under* $P_{n,2}$, $Z_i = M_i^{1/2}(\hat{\theta}_i - \theta_i)$ *are independent with the same distribution as under* P, *and* Z_i *is independent of* X^c;

3. *under* $P_{n,2}$, $\tilde{S}_i/\sigma_{\tau_{i-1}} v_{n,i}$ *are iid normal N(0,1), and independent of the* X^c *and the* Z *processes, where* $v_{n,i}$ *is given in Theorem 4. In intervals* i *with no jumps,* $v_{n,i}^2 = \mathrm{Var}(I_i)$;

4. *under* $P_{n,2}$, X *has the same distribution as under* P^*;

5. *Let* $L_{n,2}$ *be given as*

$$L_{n,2} \;=\; \sum_i \Delta\tau_i^{1/2} \left\{ \breve{\kappa}^1(\tau_{i-1})\breve{h}_1(\breve{V}_i) \;+\; \sum_{(r,s,t)\neq(0,0,0)} \frac{1}{3!}\breve{\kappa}^{r,s,t}(\tau_{i-1})\breve{h}_{rst} \right\} \tag{71}$$

where $\breve{V}_i$, $\breve{h}$, *and* $\breve{\kappa}$ *are the quantities from Theorem 5.* $L_{n,2}$ *satisfies (64).*

Remark 9. $L_{n,2}$ does not satisfy (63). This is consistent with the development of conditional cumulants in McCullagh (1987, Section 5.5.1, pp. 159-161). The current result could alternatively be stated in terms of conditional cumulants, but will indeed involve R_i. □

Remark 10. Note that because of asymptotic independence, there is no asymptotic adjustment to $L_{n,2}$ due to change of measure from $P_{n,1}$ to P^* (Mykland and Zhang (2009, Theorem 2, p. 1412)). The exact martingale would be[45] $L_{n,2} - 3\sum_i \Delta\tau_i^{1/2}\Delta < X^c, \sigma^2 >_{\tau_i}$. The correction term, however, is negligible and thus $L_{n,2}$ conforms with Definition 7. □

5 Examples of Application

We here present one example of application, namely the estimation of even functions of returns. Other examples of application can be found (with reference to this current paper) in the following locations: (1) Mykland, Shephard, and Sheppard (2012) which addresses bi-and multi-power estimators, (2) Mykland and Zhang (2014, Section 8) which adds microstructure to the estimator of Andersen, Dobrev, and Schaumburg (2012, 2014), and (3) Mykland and Zhang (2015), which addresses efficiency, and shows that one can think of $\hat{X}_i$ as having an MA(1)-process structure.

5.1 Functions of Returns

We here consider estimators of the "parameter"

$$\gamma \;=\; \sum_{k=1}^{N} h(\Delta J_{\zeta_k}) \tag{72}$$

[45] See Mykland (1994, p. 23) and Wang and Mykland (2014, p. 205).

where N is the number of jumps of the process J, ζ_k are the actual jump times, and ΔJ_{ζ_k} is the size of the jump of J at ζ_k. We take the function $x \to h(x)$ to be even and such that $h(x) = o(x^3)$ as $x \to 0$. This is a problem which is well understood in the absence of microstructure (Jacod and Protter (2012, Chapter 5.1, pp. 125-133)).

When adding microstructure, however, the problem is substantially more difficult. We refer to the treatment for the case where $\hat{X}_i$ is handled by pre-averaging (Jacod and Protter (2012, Chapter 16.5, pp. 521-563), Aït-Sahalia and Jacod (2014, Appendix A.4, p. 496-502)). We emphasize that, of course, the cited works deal with a much more complicated underlying process, infinitely many jumps. Also, they use overlapping blocks.

To otherwise be on the same ground as the cited authors, we assume that we are in the equispaced and balanced case, *i.e.,* we are in the situation from Remark 5 in Section 2.8. This is only to make expressions simpler, as the equation (76) below does not depend on spacings or blocks.

Recall the representations (46)-(47):

$$\hat{X}_{n,i} \;=\; J^{\mathrm{e}}_{n,i} + \frac{1}{2}(X^c_{\tau_{n,i}} + X^c_{\tau_{n,i-1}}) + \eta_{n,i} \tag{73}$$

where $\eta_{n,i}$ is given by (48) in Section 2.8, so that

$$\Delta\hat{X}_{n,i} \;=\; \Delta J^{\mathrm{e}}_{n,i} + \frac{1}{2}(X^c_{\tau_{n,i}} - X^c_{\tau_{n,i-2}}) + \Delta\eta_{n,i}. \tag{74}$$

We now position ourselves in the situation of Remark 2, and we shall strengthen the earlier statement to say that n_0 is such that for $n \geq n_0$ not only is there only one jump in each interval, but there are no other jumps within three intervals on each side. Because expressions of the form $h(\Delta J^{\mathrm{e}}_{\tau_{n,i-1}})$ will provide the dominating terms in an estimator of (72), we shall need some peace and quiet in the neighbourhood to investigate each jump with due diligence.

As in Remark 2, we study the k^{th} jump of J, at time $\zeta_k \in [\tau_{i_{n,k}-1}, \tau_{i_{n,k}})$. Note that the k^{th} jump takes place at the $i^{\text{th}}_{n,k}$ block. The situation is then as in Table 5. Summing over one and two

$\cdots$	$\Delta J^{\mathrm{e}}_{i_{n,k}-1}$	$\Delta J^{\mathrm{e}}_{i_{n,k}}$	$\Delta J^{\mathrm{e}}_{i_{n,k}+1}$	$\Delta J^{\mathrm{e}}_{i_{n,k}+2}$	$\cdots$
0	0	θ_i	$\Delta J_{\zeta_k} - \theta_i$	0	0

Table 5: Values of ΔJ^{e} around jump at ζ_k

scales in a small neighbourhood of ζ_k then gives

$$\sum_{i=i_{n,k}}^{i_{n,k}+1} h(\Delta J_i^{\mathrm{e}}) = h(\theta_i) + h(\Delta J_{\zeta_k} - \theta_i) \text{ and}$$
$$\sum_{i=i_{n,k}}^{i_{n,k}+2} h(J_i^{\mathrm{e}} - J_{i-2}^{\mathrm{e}}) = h(\theta_i) + h(\Delta J_{\zeta_k}) + h(\Delta J_{\zeta_k} - \theta_i) \tag{75}$$

so that, whether or not one pulverises one's jumps, one gets a two scale construction. One can sum over all $k \in [1, N]$, *i.e.*, over the jumps, and exploit that the ΔJ_i^{e} are zero except when i or $i-1 \in \mathcal{J}_n$. We obtain,

$$\sum_{i=3}^{K_n} h(J_i^{\mathrm{e}} - J_{i-2}^{\mathrm{e}}) - \sum_{i=2}^{K_n} h(\Delta J_i^{\mathrm{e}}) = \sum_{k=1}^{N} \left(2^{\text{nd}} - 1^{\text{st}} \text{ line in (75)}\right)$$
$$= \sum_{k=1}^{N} h(\Delta J_{\zeta_k}) \; = \; \gamma \tag{76}$$

Our proposed estimator of (72) is, therefore,

$$\hat{\gamma}_n \;=\; \sum_i h(\hat{X}_{n,i} - \hat{X}_{n,i-2}) - \sum_i h(\Delta \hat{X}_i) \tag{77}$$

Because of the balanced case assumption, $\frac{1}{2}(X_{\tau_i}^c - X_{\tau_{i-2}}^c) + \Delta\eta_i = O_p(\Delta\tau_i^{1/2})$. Hence,

$$\sum_{i=2}^{K_n} h(\Delta \hat{X}_i) = \sum_{i=2}^{K_n} h(\Delta J_{n,i}^{\mathrm{e}}) + \underbrace{\sum_{i=2}^{K_n} h'(\Delta J_{n,i}^{\mathrm{e}}) \left(\frac{1}{2}(X_{\tau_{i_{n,k}}}^c - X_{\tau_{i_{n,k}-2}}^c) + \Delta\eta_{i_{n,k}}\right)}_{\text{error term (i)}} + o_p(\Delta\tau^{1/2}), \text{ and} \tag{78}$$

$$\sum_{i=3}^{K_n} h(\hat{X}_{n,i} - \hat{X}_{n,i-2}) = \sum_{i=3}^{K_n} h(J_i^{\mathrm{e}} - J_{i-2}^{\mathrm{e}})$$
$$+ \underbrace{\sum_{i=3}^{K_n} h'(J_i^{\mathrm{e}} - J_{i-2}^{\mathrm{e}}) \left(\frac{1}{2}\left((X_{\tau_{i_{n,k}}}^c + X_{\tau_{i-1}}^c) - (X_{\tau_{in,k-2}}^c + X_{\tau_{in,k-3}}^c)\right) + \eta_{i_{n,k}} - \eta_{i_{n,k}-2}\right)}_{\text{error term (ii)}} + o_p(\Delta\tau^{1/2}). \tag{79}$$

There are only finitely many terms in the two sums on the r.h.s. of (78) - (79), and we can write the difference between the error term in (79) and the one in (78) as

error term (ii) $-$ error term (i)

$$= \sum_{k=1}^{N} \Bigg\{ h'(\theta_{n,i_{n,k}}) \left(\frac{1}{2}(X_{\tau_{i_{n,k}-1}}^c - X_{\tau_{i_{n,k}-3}}^c) + \Delta\eta_{i_{n,k}-1}\right) + h'(\Delta J_{\zeta_k} - \theta_{n,i}) \left(\frac{1}{2}(X_{\tau_{i_{n,k}+2}}^c - X_{\tau_{i_{n,k}}}^c) + \Delta\eta_{i_{n,k}+2}\right)$$
$$+ h'(\Delta J_{\zeta_k}) \left(\frac{1}{2}\left((X_{\tau_{i_{n,k}}}^c + X_{\tau_{i_{n,k}-1}}^c) - (X_{\tau_{i_{n,k}-2}}^c + X_{\tau_{i_{n,k}-3}}^c)\right) + \eta_{i_{n,k}} - \eta_{i_{n,k}-2}\right) \Bigg\} \tag{80}$$

We now invoke the contiguity of Theorem 6 in Section 4.2 to say that under $P_{n,2}$, the $\Delta X^c_{\tau_i}$ and η_i processes are independent of each other and of the J and θ_i processes. We shall work with $P_{n,2}$ until further notice.

For given k,

$$\Delta\tau^{-1/2}\Big(X^c_{\tau_{i_{n,k}-1}} - X^c_{\tau_{i_{n,k}-3}} + 2\Delta\eta_{i_{n,k}-1}, X^c_{\tau_{i_{n,k}+2}} - X^c_{\tau_{i_{n,k}}} + 2\Delta\eta_{i_{n,k}+2}, (X^c_{\tau_{i_{n,k}}} + X^c_{\tau_{i_{n,k}-1}}) - (X^c_{\tau_{i_{n,k}-2}} + X^c_{\tau_{i_{n,k}-3}}) + 2\eta_{i_{n,k}} - 2\eta_{i_{n,k}-2}\Big)$$
$$\stackrel{\mathcal{L}}{\approx} (Y_{i_{n,k}-1} + Y_{i_{n,k}-2}, Y_{i_{n,k}+1} + Y_{i_{n,k}+2}, Y_{i_{n,k}} + Y_{i_{n,k}-2} + 2\Delta\tau^{-1/2}\Delta X^c_{\tau_{i_{n,k}-1}}), \tag{81}$$

where the symbol $\stackrel{\mathcal{L}}{\approx}$ means that the two expressions have the same asymptotic limit, in this case under $P_{n,2}$. We have here taken $Y_{n,i} = \tau^{1/2}(\Delta X^c_{\tau_{n,i}} + 2\eta_{n,i})$, and the approximation in law stems from $\Delta\tau^{-1/2}(\Delta X^c_{\tau_{n,i}}, \eta_{n,i}) \stackrel{\mathcal{L}}{\approx} \Delta\tau^{-1/2}(\Delta X^c_{\tau_{n,i}}, -\eta_{n,i})$ by combining Theorems 2 and 6. Under an obvious combination of stable and conditional convergence, the $Y_{n,i_{n,k+j}} \stackrel{\mathcal{L}}{\approx} Y_{k,j}$ jointly (there are only finitely many of them that matter), where the $Y_{k,j}$ are (conditionally on $\mathcal{G}_T$) independent normal with mean zero and variance of the form

$$\mathrm{Var}(Y_{k,j}|\mathcal{G}_T) = \begin{cases} \frac{4}{3}\sigma^2_{\zeta_k} + \frac{4a^2}{c^2T} \text{ for } j \neq 0 \\ (1 + 4v^2_{n,i_{n,k}})\sigma^2_{\zeta_k} + \frac{4a^2_{n,i_{n,k}}}{c^2T} \text{ for } j = 0 \end{cases} \tag{82}$$

We have here again invoked Theorems 2, 4, and 6. The quantities, a^2, $a^2_{n,i_{n,k}}$ and $v^2_{n,i_{n,k}}$ are given in equations (8) (Section 2.3), (28) (Section 2.5.2), and (C.8) (Appendix C.2), respectively. Also, jointly with the above, $2\Delta\tau^{-1/2}\Delta X^c_{\tau_{i_{n,k}-1}} \stackrel{\mathcal{L}}{\approx} Y'_k$ where the Y'_k are conditionally independent (given $\mathcal{G}_T$) of each other, and of $Y_{k,j}$, all $j \neq -1$. $(Y'_k, Y_{k,-1})$ are jointly normal with (conditional) covariance $2\sigma^2_{\zeta_k}$ Meanwhile Y'_k har conditional variance $4\sigma^2_{\zeta_k}$.

From eq. (51), $n^{1/4} = (cT)^{1/2}\Delta\tau^{-1/2}$, hence, in view of the development above,

$$\begin{aligned} n^{1/4}(\hat{\gamma}_n - \gamma) \stackrel{\mathcal{L}}{\approx} \frac{1}{2}(cT)^{1/2}\sum_{k=1}^{N}\big\{&h'(\theta_{n,i_{n,k}})(Y_{k,-1} + Y_{k,-2}) + h'(\Delta J_{\zeta_k} - \theta_{n,i})(Y_{k,2} + Y_{k,1}) \\ &+ h'(\Delta J_{\zeta_k})(Y_{k,0} + Y_{k,-2} + Y'_k)\big\} \end{aligned} \tag{83}$$

This is all under $P_{n,2}$, but it is easy to see that there is no contiguity adjustment (since h is an even function) back to P^* and hence P. The conditional variances and covariance remain the same. This is all in analogy with Mykland and Zhang (2009, Theorem 2, p. 1412).

It is now easy to see that term $\#k$ has conditional variance

$$\begin{aligned} \nu^2_{n,k} = \frac{1}{4}cT\Bigg\{&2(h'(\theta_{n,i_{n,k}})^2 + h'(\Delta J_{\zeta_k} - \theta_{n,i})^2)\left(\frac{4}{3}\sigma^2_{\zeta_k} + \frac{4a^2}{c^2T}\right) \\ &+ h'(\Delta J_{\zeta_k})^2\left((\frac{19}{3} + 4v^2_{n,i_{n,k}})\sigma^2_{\zeta_k} + \frac{4(a^2 + a^2_{n,i_{n,k}})}{c^2T}\right) + 4h'(\Delta J_{\zeta_k})h'(\theta_{n,i_{n,k}})\sigma^2_{\zeta_k}\Bigg\} \end{aligned} \tag{84}$$

Hence, stably in law

$$n^{1/4}(\hat{\gamma}_n - \gamma) \overset{\mathcal{L}}{\approx} \left(\sum_{k=1}^{N} \nu_{n,k}^2 \right)^{1/2} U \tag{85}$$

where U is standard normal, and independent of $\mathcal{G}_T$.

In other words, for this estimator, the potential pulverisation discussed in Section 2.5.3 does not impact the estimator $\hat{\gamma}_n$, or its convergence to the target γ, but it does impact the setting of asymptotic variance. The case for robust estimation thus also occurs in this example.

6 Conclusion

In this paper, we have taken the view that pre-averaging is a way of *estimating* the efficient price under market microstructure noise. This opens the possibility of using other and more robust estimators, and we have here investigated one class of these, namely M-estimators. It turned out that this procedure is robust with respect to the noise and the jumps, while averaging the continuous part of the signal.

We have two main results. One is Theorems 1-3 in Section 2.5, which show that by moving from pre-averaging to pre-M-estimation, one can to a great extent avoid the pulverisation of jumps that is present in pre-averaging. M-estimation also opens the possibility for better efficiency (Section 2.5.4).

The other main result is to analyse estimators globally, as follows. Under a contiguous measure, the estimation error from M-estimation (including pre-averaging) can be seen as an additional component to the microstructure noise. This result is contained in Theorem 6 in Section 4.2. The error due to contiguity can, as usual, be offset with a post-asymptotic likelihood ratio correction. We saw in Section 5 that the result is highly applicable.

As part of the development, Section 3 set up a general framework for finding contiguity results in data systems of this nature.

An issue that has not been addressed in the foregoing is how to handle $\hat{X}$s when blocks are overlapping. We conjecture that the results in the current paper will still provide consistency and the correct convergence rate. One approach may be to combine this with an "observed" standard error, based on the development in Mykland and Zhang (2014). But that is a story for another time.

REFERENCES

Aït-Sahalia, Y. (2004): "Disentangling Diffusion from Jumps," *Journal of Financial Economics*, 74, 487–528.

Aït-Sahalia, Y., J. Fan, C. D. Wang, and X. Yang (2013): "The Estimation of continuous and discontinuous leverage effect," Discussion paper.

Aït-Sahalia, Y. and J. Jacod (2007): "Volatility Estimators for Discretely Sampled Lévy Processes," *Annals of Statistics*, 35, 335–392.

——— (2008): "Fisher's Information for Discretely Sampled Lévy Processes," *Econometrica*, 76, 727–761.

——— (2009): "Testing for jumps in a discretely observed process," *Annals of Statistics*, 37, 184–222.

——— (2012): "Analyzing the Spectrum of Asset Returns: Jump and Volatility Components in High Frequency Data," *Journal of Economic Literature*, 50, 1007–1150.

Aït-Sahalia, Y. and J. Jacod (2014): *High-Frequency Financial Econometrics*, Princeton: Princeton University Press, first ed.

Aldous, D. (1989): *Probability Approximations via the Poisson Clumping Heuristic*, Springer-Verlag.

Aldous, D. J. and G. K. Eagleson (1978): "On Mixing and Stability of Limit Theorems," *Annals of Probability*, 6, 325–331.

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

——— (1998b): "Deutsche Mark-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies," *Journal of Finance*, 53, 219–265.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2000): "Great realizations," *Risk*, 13, 105–108.

——— (2001): "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55.

——— (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

Andersen, T. G., D. Dobrev, and E. Schaumburg (2009): "Jump-Robust Volatility Estimation using Nearest Neighbor Truncation," .

——— (2012): "Jump-Robust Volatility Estimation using Nearest Neighbor Truncation," *Journal of Econometrics*, 169, 75–93.

——— (2014): "A Robust Truncation Approach to Estimation of Integrated Quaticity," *Econometric Theory*, 39, 3–59.

Bahadur, R. (1966): "A note on quantiles in large samples," *Annals of Mathematical Statistics*, 37, 577–580.

Barndorff-Nielsen, O., S. Graversen, J. Jacod, M. Podolskij, and N. Shephard (2006): "A central limit theorem for realised power and bipower variations of continuous semimartingales," in *From Stochastic Calculus to Mathematical Finance, The Shiryaev Festschrift*, ed. by Y. Kabanov, R. Liptser, and J. Stoyanov, Berlin: Springer Verlag, 33–69.

Barndorff-Nielsen, O. E. (2004): "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2, 1–37.

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.

——— (2009a): "Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading," Tech. rep., University of Oxford.

Barndorff-Nielsen, O. E., S. Kinnebrouck, and N. Shephard (2009b): "Measuring downside risk: realised semivariance," in *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, ed. by T. Bollerslev, J. Russell, and M. Watson, 117–136.

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.

Bhattacharya, R. N. and J. Ghosh (1978): "On the validity of the Formal Edgeworth Expansion," *Annals of Statistics*, 6, 434–451.

Bhattacharya, R. N. and R. R. Rao (1976): *Normal Approximation and Asymptotic Expansions*, New York: Wiley.

Bibinger, M., N. Hautsch, P. Malec, and M. Reiss (2013): "Estimating the Quadratic Covariation Matrix from Noisy Observations: LOcal Method of Moments and Efficiency," Preprint.

Bibinger, M. and P. A. Mykland (2013): "Inference for Multi-Dimensional High-Frequency Data: Equivalence of Methods, Central Limit Theorems, and an Application to Conditional Independence Testing," Technical Report, the University of Chicago.

Billingsley, P. (1995): *Probability and Measure*, New York: Wiley, second ed.

Bollerslev, T., M. Gibson, and H. Zhou (2005): "Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities," To appear in *Journal of Econometrics*.

Bollerslev, T., G. Tauchen, and H. Zhou (2009): "Expected Stock Returns and Variance Risk Premia," *Review of Financial Studies*, 22, 4463–4492.

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

Chaudhuri, P. (1996): "On a Geometric Notion of Quantiles for Multivariate Data," *Journal of the American Statistical Association*, 91, 862–872.

Christensen, K., S. Kinnebrock, and M. Podolskij (2008a): "Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data," .

Christensen, K., R. Oomen, and M. Podolskij (2008b): "Realised Quantile-Based Estimation of the Integrated Variance," Tech. rep.

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.

Curci, G. and F. Corsi (2005): "Discrete Sine Transform for Multi-Scales Realized Volatility Measures," Tech. rep., University of Lugano.

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.

Delbaen, F. and W. Schachermayer (1994): "A General Version of the Fundamental Theorem of Asset Pricing," *Mathematische Annalen*, 300, 463–520.

——— (1995): "The existence of absolutely continuous local martingale measures," *Annals of Applied Probability*, 5, 926–945.

——— (1998): "The Fundamental Theorem of Asset Pricing For Unbounded Stochastic Processes," *Mathematische Annalen*, 312, 215–250.

Donoho, D. L. and M. Gasko (1992): "Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness," *Annals of Statistics*, 20, 1803–1827.

Fan, J. and Y. Wang (2005): "Multi-scale Jump and Volatility Analysis for High-Frequency Financial Data," Tech. rep., Princeton University.

Foster, D. and D. Nelson (1996): "Continuous record asymptotics for rolling sample variance estimators," *Econometrica*, 64, 139–174.

French, K. R., G. W. Schwert, and R. F. Stambaugh (1987): "Expected stock returns and volatility," *Journal of Financial Economics*, 19, 3–29.

Gloter, A. (2000): "Estimation des Paramètres d'une Diffusion Cachée," Ph.D. thesis, Université de Marne-la-Vallée.

Gloter, A. and J. Jacod (2000): "Diffusions with Measurement Errors: I - Local Asymptotic Normality and II - Optimal Estimators," Tech. rep., Université de Paris VI.

——— (2001): "Diffusions with Measurement Errors. II - Optimal Estimators," *ESAIM Probab. Statist.*, 5, 243–260.

Gonçalves, S., P. Donovon, and N. Meddahi (2013): "Bootstrapping realized multivariate volatility measures," *Journal of Econometrics*, 172, 49–65.

Gonçalves, S. and N. Meddahi (2009): "Bootstrapping realized volatility," *Econometrica*, 77, 283–306.

Hájek, J. and Z. Sidak (1967): *Theory of Rank Tests*, New York: Academic Press.

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

Hall, P. and C. C. Heyde (1980): *Martingale Limit Theory and Its Application*, Boston: Academic 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.

Hoffmann, M. (2008): "Personal communication," .

Huang, X. and G. T. Tauchen (2005): "The Relative Contribution of Jumps to Total Price Variance," *Journal of Financial Econometrics*, 3, 456–499.

Huber, P. J. (1981): *Robust Statistics*, New York: Wiley.

Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter (2009a): "Microstructure Noise in the Continuous Case: The Pre-Averaging Approach," *Stochastic Processes and Their Applications*, 119, 2249–2276.

Jacod, J., M. Podolskij, and M. Vetter (2009b): "Limit theorems for moving averages of discretized processes plus noise," *Annals of Statistics*, 38, 1478–1545.

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

——— (2012): *Discretization of Processes*, New York: Springer-Verlag, first ed.

Jacod, J. and M. Reiss (2012): "A Remark on the Rates of Convergence for Integrated Volatility Estimations in the presence of jumps," To appear in *The Annals of Statistics*.

Jacod, J. and M. Rosenbaum (2013): "Quarticity and other Functionals of Volatility: Efficient Estimation," *Annals of Statistics*, 41, 1462–1484.

Jacod, J. and A. N. Shiryaev (2003): *Limit Theorems for Stochastic Processes*, New York: Springer-Verlag, second ed.

Jacod, J. and V. Todorov (2010): "Do Price and Volatility Jump Together?" *The Annals of Applied Probability*, 20 (4), 1425–1469.

Jensen, J. (1995): *Saddlepoint Approximations*, Oxford, U.K.: Oxford University Press.

Jing, B.-Y., X.-B. Kong, Z. Liu, and P. A. Mykland (2012): "On the jump activity index for semimartingales," *Journal of Econometrics*, 166, 213–223.

Johns, M. (1979): "Robust Pitman-like estimators," in *Robustness in Statistics*, ed. by Launer and Wilkinson, New York: Academic.

Kalnina, I. (2011): "Subsampling High Frequency Data," *Journal of Econometrics*, 161, 262–283.

——— (2012): "Nonparametric Tests of Time Variation in Betas," Discussion paper.

Kalnina, I. and O. Linton (2007): "Inference about Realized Volatility using Infill Subsampling," Discussion paper.

Karatzas, I. and S. E. Shreve (1991): *Brownian Motion and Stochastic Calculus*, Berlin: Springer-Verlag.

Koenker, R. and G. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

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.

Lee, S. S. and P. A. Mykland (2008): "Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics," *Review of Financial Studies*, 21, 2535–2563.

——— (2012): "Jumps in Equilibrium Prices and Market Microstructure Noise," *Journal of Econometrics*, 168, 396–406.

Lee, S. Y. (2005): "The Detection of Jumps in Financial Data," Ph.D. thesis, The University of Chicago, Graduate School of Business.

Lévy, P. (1948): *Processus Stochastiques et Mouvement Brownien*, Paris: Gauthier-Villars.

Li, C. (2012): "Maximum-likelihood Estimation for Diffusions via Closed-Form Density Estimation," To appear in *Annals of Statistics*.

Liu, R. Y. (1990): "On a Notion of Data Depth Based on Random Simplices," *Annals of Statistics*, 18, 405–414.

Mancini, C. (2004): "Estimating the Integrated Volatility in Stochastic Volatility Models with Lévy Type Jumps," Tech. rep., Universita di Firenze.

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

Merton, R. C. (1980): "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, 8, 323–361.

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

Mykland, P. A., N. Shephard, and K. Sheppard (2012): "Efficient and Feasible Inference for the Components of Financial Variation using Blocked Multipower Variation," Discussion paper.

Mykland, P. A. and L. Zhang (2005): "Discussion of paper "A selective overview of nonparametric metods in financial econometrics" by J. Fan," *Statistical Science*, 20, 347–350.

——— (2006): "ANOVA for Diffusions and Itô Processes," *Annals of Statistics*, 34, 1931–1963.

——— (2008): "Inference for Volatility Type Objects and Implications for Hedging," *Statistics and its Interface*, 1, 255–278.

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

——— (2011): "The Double Gaussian Approximation for High Frequency Data," *Scandinavian Journal of Statistics*, 38, 215–236.

——— (2012): "The Econometrics of High Frequency Data," in *Statistical Methods for Stochastic Differential Equations*, ed. by M. Kessler, A. Lindner, and M. Sørensen, New York: Chapman and Hall/CRC Press, 109–190.

——— (2014): "Assessment of Uncertainty in High Frequency Data: The Observed Asymptotic Variance," In revision for *Econometrica*.

——— (2015): "Superblocks of Pre-Averaged Data: First Aid for Microsctructure, and a Path to Efficiency," Discussion paper.

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

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

——— (2009b): "Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps," *Bernoulli*, 15, 634–658.

Reiss, M. (2011): "Asymptotic equivalence for inference on the volatility from noisy observations," *Annals of Statistics*, 2, 772–802.

Rényi, A. (1963): "On Stable Sequences of Events," *Sankyā Series A*, 25, 293–302.

Robert, C. Y. and M. Rosenbaum (2009): "Volatility and Covariation Estimation when Microstructure Noise and Trading Times are Endogenous," W.P. CMAP, Ecole Polytechnique, Paris.

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

Rosenberg, B. (1972): "The behaviour of random variables with nonstationary variance and the distribution of security prices," Working paper 11, Graduate School of Business Administration, University of California, Berkeley. Reprinted in Shephard (2005).

Ross, S. (1996): *Stochastic Processes*, New York, N.Y.: Wiley, second ed.

Shephard, N., ed. (2005): *Stochastic Volatility: Selected Readings*, Oxford: Oxford University Press.

Shephard, N. and T. G. Andersen (2009): "Stochastic Volatility: Origins and Overview," in *Handbook of Financial Time Series*, ed. by T. G. Andersen, R. Davis, J.-P. Kreiss, and T. Mikosch, 233–254.

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

Stone, C. J. (1974): "Asymptotic Properties of Estimators of a Location Parameter," *Annals of Statistics*, 2, 1127–1137.

——— (1975): "Adaptive Maximum Likelihood Estimators of a Location Parameter," *Annals of Statistics*, 3, 267–284.

Vetter, M. (2011): "Estimation of Integrated Volatility of Volatility with Applications to Goodness-of-fit Testing," Discussion paper.

Wallace, D. L. (1958): "Asymptotic Approximations to Distributions," *Annals of Mathematical Statistics*, 29, 635–654.

Wang, C. D. and P. A. Mykland (2014): "The Estimation of Leverage Effect with High Frequency Data," *Journal of the American Statistical Association*, 109, 197–215.

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

Xiu, D. (2010): "Quasi-Maximum Likelihood Estimation of Volatility WIth High Frequency Data," *Journal of Econometrics*, 159, 235–250.

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.

——— (2007): "What you don't know cannot hurt you: On the detection of small jumps," Tech. rep., University University of Illinois at Chicago.

——— (2011): "Estimating Covariation: Epps Effect and Microstructure Noise," *Journal of Econometrics*, 160, 33–47.

——— (2012): "Implied and realized volatility: Empirical model selection," *Annals of Finance*, to appear.

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.

Zhou, B. (1998): "F-consistency, De-volatization and Normalization of High Frequency Financial Data," in *Nonlinear Modelling of High Frequency Financial Time Series*, ed. by C. L. Dunis and B. Zhou, New York: John Wiley Sons Ltd., 109–123.

APPENDIX: PROOFS

A Proof of Theorem 1

First note that as in discussed in Section 4.5 of Mykland and Zhang (2012), we can assume without loss of generality that σ_t^2 is bounded by a constant σ_+^2 on the whole interval $[0, T]$. Also, as discussed in Section 2.2 of Mykland and Zhang (2009), we can assume that we are under an equivalent martingale measure where $\mu_t \equiv 0$. Set $\epsilon'_{t_j} = \epsilon_{t_j} + J_{t_j} - J_{\tau_{i-1}}$ and $\bar{X}'_i = \bar{X}^c_i + J_{\tau_{i-1}}$.

To first establish the nature of the approximation, let $G_i = \Delta\tau_i^{-1/2} \max_{\tau_{i-1} \le t \le \tau_i} |X^c_t - X^c_{\tau_{i-1}}|$. We note that $G_i = O_p(1)$ (Lévy (1948), see also Karatzas and Shreve (1991, Theorem 3.6.17, pp. 211-212)). Since $Y_{t_j} - (\bar{X}'_i + \epsilon'_{t_j}) = X^c_{t_j} - \bar{X}^c_i$

$$|Y_{t_j} - (\bar{X}'_i + \epsilon'_{t_j})| = |X^c_{t_j} - \bar{X}^c_i| \le |X^c_{t_j} - X^c_{\tau_{i-1}}| + |\bar{X}^c_i - X^c_{\tau_{i-1}}| \le \Delta\tau_i^{1/2} 2G_i$$

Hence,

$$0 = \sum_{\tau_{i-1} \le t_j < \tau_i} \psi(Y_{t_j} - \hat{X}_i) \le \sum_{\tau_{i-1} \le t_j < \tau_i} \psi(\bar{X}'_i + \epsilon'_j - \hat{X}_i + \Delta\tau_i^{1/2} 2G_i).$$

Since ψ is strictly increasing around zero, and since $\sum_{\tau_{i-1} \le t_j < \tau_i} \psi(\epsilon'_{t_j} - \hat{\theta}_i) = 0$, it follows that eventually $\hat{X}_i - (\bar{X}_i + \hat{\theta}_i) \le \Delta\tau_i^{1/2} 2G_i$. Repeating the same argument on the other side yields that eventually

$$|\hat{X}_i - (\bar{X}^c_i + \hat{\theta}_i)| \le \Delta\tau_i^{1/2} 2G_i = O_p(\Delta\tau_i^{1/2}). \tag{A.1}$$

This proves the first part of Theorem 1.

To get a more precise form of the remainder, let

$$\delta_i = \hat{X}_i - (\bar{X}'_i + \hat{\theta}_i). \tag{A.2}$$

In view of (A.1), we can Taylor expand safely. Since

$$Y_{t_j} - \hat{X}_i \;-\; (\epsilon'_{t_j} - \hat{\theta}_i) \;=\; X^c_{t_j} + J_{\tau_{i-1}} - \hat{X}_i + \hat{\theta}_i \;=\; X^c_{t_j} + J_{\tau_{i-1}} - \bar{X}'_i - \delta_i \;=\; X^c_{t_j} - \bar{X}^c_i - \delta_i,$$

we obtain from Taylor's formula that

$$\begin{aligned} 0 &= \sum_{\tau_{i-1} \le t_j < \tau_i} \psi(Y_{t_j} - \hat{X}_i) \\ &= \sum_{\tau_{i-1} \le t_j < \tau_i} \psi(\epsilon'_{t_j} - \hat{\theta}_i) + \sum_{\tau_{i-1} \le t_j < \tau_i} (X^c_{t_j} - \bar{X}^c_i - \delta_i)\psi'(\epsilon'_{t_j} - \hat{\theta}_i) \\ &+ \frac{1}{2} \sum_{\tau_{i-1} \le t_j < \tau_i} \int_0^{X^c_{t_j} - \bar{X}^c_i - \delta_i} (X^c_{t_j} - \bar{X}^c_i - \delta_i - s)\psi''(\epsilon'_{t_j} - \hat{\theta}_i - s)ds \\ &= \sum_{\tau_{i-1} \le t_j < \tau_i} (X^c_{t_j} - \bar{X}^c_i - \delta_i)\psi'(\epsilon'_{t_j} - \hat{\theta}_i) + O_p(M_i \Delta\tau_i) \end{aligned} \tag{A.3}$$

where, in the final step, we have used the definition of $\hat{\theta}_i$, the boundedness of ψ'', as well as the bound (A.1). Hence,

$$\delta_i = \frac{\sum_{\tau_{i-1}\le t_j<\tau_i}(X^c_{t_j} - \bar{X}^c_i)\psi'(\epsilon'_{t_j} - \hat{\theta}_i)}{\sum_{\tau_{i-1}\le t_j<\tau_i}\psi'(\epsilon'_{t_j} - \hat{\theta}_i)} + O_p(\Delta\tau_i). \tag{A.4}$$

In particular,

$$\begin{aligned}\hat{X}_i - \hat{\theta}_i - X_{\tau_{i-1}} &= \bar{X}'_i - X_{\tau_{i-1}} + \delta_i \\ &= \frac{\sum_{\tau_{i-1}\le t_j<\tau_i}(X^c_{t_j} - X^c_{\tau_{i-1}})\psi'(\epsilon'_{t_j} - \hat{\theta}_i)}{\sum_{\tau_{i-1}\le t_j<\tau_i}\psi'(\epsilon'_{t_j} - \hat{\theta}_i)} + O_p(\Delta\tau_i), \end{aligned} \tag{A.5}$$

thus proving the rest of Theorem 1. ∎

B A sharper decomposition of the M-estimator for intervals with no jumps.

For the development in Appendix D, we need a stronger result than those of Section 2.5.

Theorem 7. *(Remainder Term in the Continuous Case in the Fundamental Decomposition of the Estimator of Efficient Price.) Assume Conditions 1-5. Let $[\tau_{i-1}, \tau_i)$ be a block with no jump. Set*

$$D_i = \sum_{\tau_{i-1}\le t_j<\tau_i}(X_{t_j} - \bar{X}_i)(\psi'(\epsilon_{t_j}) - E\psi'(\epsilon)) + \frac{1}{2}s_i^2 E\psi''(\epsilon) \tag{B.6}$$

where $s_i^2 = \sum_{\tau_{i-1}\le t_j<\tau_i}(X_{t_j} - \bar{X}_i)^2$. Then

$$\hat{X}_i - \bar{X}_i = \hat{\theta}_i + M_i^{-1}(E\psi'(\epsilon))^{-1}D_i + O_p(M_i^{-3/2}) + o_p(\Delta\tau_i) \tag{B.7}$$

$$= \hat{\theta}_i + M_i^{-1}(E\psi'(\epsilon))^{-1}D_i + o_p(\Delta\tau_i) \tag{B.8}$$

by the second equation in Footnote 16 to Condition 5, and since $D_i = O_p(1)$. Specifically, given X process,

$$E(D_i|X) = \frac{1}{2}s_i^2 E\psi''(\epsilon).$$

Note that in view of the assumptions, $\hat{\theta}_i$ is an estimator of $\theta_i = 0$ (since there is no jump in the block), so that $M_i^{1/2}\hat{\theta}_i = O_p(1)$.

Proof of Theorem 7. We continue the development from Appendix A, but we now assume that the process X_t is continuous, and will denote X^c by X. δ_i gets the form

$$\delta_i = \hat{X}_i - (\bar{X}_i + \hat{\theta}_i). \tag{B.9}$$

Also, since $\hat{\theta}_i = O_p(M_i^{-1/2})$,

$$\sum_{\tau_{i-1} \le t_j < \tau_i} \psi'(\epsilon_{t_j} - \hat{\theta}_i) = M_i E\psi'(\epsilon) + O_p(M_i^{1/2}) \text{ and}$$

$$\begin{aligned}\sum_{\tau_{i-1} \le t_j < \tau_i} (X_{t_j} - \bar{X}_i)\psi'(\epsilon_{t_j} - \hat{\theta}_i) &= \sum_{\tau_{i-1} \le t_j < \tau_i} (X_{t_j} - \bar{X}_i)\psi'(\epsilon_{t_j}) + O_p(\Delta\tau_i^{1/2}) \\ &= O_p(\Delta\tau_i^{1/2} M_{n,i}^{1/2}). \end{aligned} \tag{B.10}$$

The last transition above comes from an argument similar to (B.14)-(B.16) below. Combining (B.10) with (A.4), we obtain

$$\delta_i = O_p(M_{n,i}^{-1/2} \Delta\tau_i^{1/2}) + O_p(\Delta\tau_i). \tag{B.11}$$

Using the (B.10)-(B.11), we now continue from the exact form of (A.3).

$$\begin{aligned} 0 &= -\delta_i M_i E\psi'(\epsilon) + \sum_{\tau_{i-1} \le t_j < \tau_i} (X_{t_j} - \bar{X}_i)\psi'(\epsilon_{t_j}) \\ &+ \frac{1}{2} \sum_{\tau_{i-1} \le t_j < \tau_i} \int_0^{X_{t_j} - \bar{X}_i} (X_{t_j}^c - \bar{X}_i^c - s)\psi''(\epsilon_{t_j} - \hat{\theta}_i - s)ds + O_p(\Delta\tau_i^{1/2}) \\ &= -\delta_i M_i E\psi'(\epsilon) + D_i + O_p(\Delta\tau_i^{1/2}), \end{aligned} \tag{B.12}$$

where we have used the first equation in Footnote 16 to Condition 5, and where

$$D_i = \sum_{\tau_{i-1} \le t_j < \tau_i} (X_{t_j} - \bar{X}_i)(\psi'(\epsilon_{t_j}) - E\psi'(\epsilon)) + \frac{1}{2} s_i^2 E\psi''(\epsilon). \tag{B.13}$$

The conditional mean and variance of D_i given the X process are

$$E(D_i|X) = \frac{1}{2} s_i^2 E\psi''(\epsilon) \text{ and } \mathrm{Var}(D_i|X) = s_i^2 \mathrm{Var}(\psi'(\epsilon)) \tag{B.14}$$

where

$$s_i^2 = \sum_{\tau_{i-1} \le t_j < \tau_i} (X_{t_j} - \bar{X}_i)^2 \tag{B.15}$$

By standard methods, one can further show that

$$s_i^2 = \sigma_{\tau_{i-1}}^2 \Delta\tau_i M_{n,i} E_n(I_i(1 - I_i)) + O_p(M_i^{-1/2}). \tag{B.16}$$

Hence $D_i = O_p(1)$. In particular, $\delta_i = O_p(M^{-1})$. We can use this to sharpen the error term in (B.12) (when passing from (A.3)) to $O_p(M_i^{-1/2}) + O_p(\Delta\tau_i^{1/2}) + o_p(M_i\Delta\tau_i) = O_p(M_i^{-1/2}) + o_p(M_i\Delta\tau_i)$ by the first equation in Footnote 16 to Condition 5. Rewriting this version of (B.12), we obtain

$$\delta_i \;=\; \frac{1}{M_i E\psi'(\epsilon)} D_i \;+\; O_p(M_i^{-1} \Delta\tau_i^{1/2}). \tag{B.17}$$

This shows (B.7). ∎

C Main calculation: Proofs and formulae for (R_i, T_i)

For simplicity of notation, we assume that τ_{i-1} and τ_i coincide with a t_j; the further generalization is simple but tedious, and does not impact our results to the relevant order of approximation. Set

$$U_i^{(k)} = \Delta\tau_i^{-1/2} \sum_{\tau_{i-1} \le t_j < \tau_i} \left(\frac{M-j}{M} \right)^k \Delta X_{t_j}^c, \tag{C.1}$$

so that

$$R_i = U_i^{(0)} \text{ and } T_i = U_i^{(1)}.$$

C.1 First order behaviour of (R_i, T_i), including proof of Theorem 4 in the continuous case

R_i and T_i are as previously defined in (37) and (40). Consider first the case where there is no jump in $[\tau_{i-1}, \tau_i)$, when $S_i = T_i + O_p(\Delta\tau_i^{1/2})$. For the first part of the result, the form (42) of the asymptotic covariance of $(R_i, T_i)/\sigma_{\tau_{i-1}}$ follows from (C.11) below.

To see stable convergence, let ξ_t be another continuous Itô process, set $\Xi_i = \Delta\xi_{\tau_i}$, and note that

$$\begin{aligned} \langle U^{(k)}, \xi_\cdot/\sqrt{\Delta\tau} \rangle_{\tau_i} &= \langle X, \xi \rangle'_{\tau_{i-1}} E(I_0^k)(1 + o_p(1)) \text{ and} \\ \langle \xi_\cdot/\sqrt{\Delta\tau}, \xi_\cdot/\sqrt{\Delta\tau} \rangle_{\tau_i} &= \langle \xi, \xi \rangle'_{\tau_{i-1}}(1 + o_p(1)). \end{aligned} \tag{C.2}$$

where $U^{(k)}$ is given by (C.1) below. The CLT then yields that (with some abuse of notation)

$$\begin{pmatrix} T_i \\ R_i \\ \Xi_i \end{pmatrix} \overset{\mathcal{L}}{\approx} N \left(0, \begin{pmatrix} \sigma_{\tau_{i-1}}^2 E(I_i^2) & \sigma_{\tau_{i-1}}^2 E(I_i) & \langle X, \xi \rangle'_{\tau_{i-1}} E(I_i) \\ \sigma_{\tau_{i-1}}^2 E(I_i) & \sigma_{\tau_{i-1}}^2 & \langle X, \xi \rangle'_{\tau_{i-1}} \\ \langle X, \xi \rangle'_{\tau_{i-1}} E(I_i) & \langle X, \xi \rangle'_{\tau_{i-1}} & \langle \xi, \xi \rangle'_{\tau_{i-1}} \end{pmatrix} \right). \tag{C.3}$$

A linear transformation yields that

$$\begin{pmatrix} T_i - E(I_i)R_i \\ R_i \\ \Xi_i \end{pmatrix} \overset{\mathcal{L}}{\approx} N \left(0, \begin{pmatrix} \sigma_{\tau_{i-1}}^2 \mathrm{Var}(I_i) & 0 & 0 \\ 0 & \sigma_{\tau_{i-1}}^2 & \langle X, \xi \rangle'_{\tau_i} \\ 0 & \langle X, \xi \rangle'_{\tau_{i-1}} & \langle \xi, \xi \rangle'_{\tau_{i-1}} \end{pmatrix} \right). \tag{C.4}$$

This shows the result of Theorem 4 for intervals with no jump.

C.2 First order behaviour of (R_i, T_i), including proof of Theorem 4 for the discontinuous case

Assume that there is no more than one jump ΔJ_{ζ_k} in interval $[\tau_{n,i_{n,k}-1}, \tau_{n,i_{n,k}})$. This will eventually occur. For notational convenience write i_k for $i_{n,k}$.

Let T_i be as in (30) in Theorem 2 in Section 2.5.2. Because of asymptotic negligibility, we can take $t_{j_0} = \tau_{i_k-1}$ and $t_{j_0+M_i'-1} = \zeta_k$. Rewriting as above,

$$T_i = \Delta\tau_i^{-1/2} D_{n,k}^{-1} \times$$
$$\left(\sum_{j=1}^{M_i'-1} \Delta X_{t_{j_0+j}}^c \left(\frac{M_i' - j}{M_i} f'(\theta_{n,i_k}) + \frac{M_i''}{M_i} f'(\theta_{n,i_k} - \Delta J_{\zeta_k}) \right) + \sum_{j=M_i'+1}^{M_i} \Delta X_{t_{j_0+j}}^c \left(\frac{M_i'' - j}{M_i} \right) f'(\theta_{n,i_k} - \Delta J_{\zeta_k}) \right) \tag{C.5}$$

where $D_{n,k} = \alpha_{n,i} f'(\theta_{n,i_{n,k}}) + (1 - \alpha_{n,i}) f'(\theta_{n,i_{n,k}} - \Delta J_{\zeta_k})$. We obtain in the same way as before the CLT

$$\begin{pmatrix} T_i \\ R_i \end{pmatrix} \overset{\mathcal{L}}{\approx} N\left(0, \sigma_{\tau_{i-1}}^2 \begin{pmatrix} \breve{\upsilon}_{n,i}^2 & \beta_{n,i} \\ \beta_{n,i} & 1 \end{pmatrix}\right), \tag{C.6}$$

where

$$\beta_{n,i} = D_{n,k}^{-1} \left\{ E\chi_{n,k} \left\{ I_{i,k} f'(\theta_{n,i_k}) - (1 - \alpha_{n,i_k})(f'(\theta_{n,i_k}) - f'(\theta_{n,i_k} - \Delta J_{\zeta_k})) + E(1 - \chi_{n,k}) f'(\theta_{n,i_k} - \Delta J_{\zeta_k}) \right\} \right\}. \tag{C.7}$$

Here $\chi_{n,k} = \mathbf{I}\{I_{i_{n,i_k}} > 1 - \alpha_{n,i_k}\}$, where $\mathbf{I}\{\cdot\}$ is the indicator function. Also $\breve{\upsilon}_{n,i}^2 = \upsilon_{n,i}^2 + \beta_{n,i}^2$, where

$$\upsilon_{n,i}^2 = D_{n,k}^{-2} \left\{ w_{n,i,11} f'(\theta_{n,i_k})^2 + 2w_{n,i,12} f'(\theta_{n,i_k}) f'(\theta_{n,i_k} - \Delta J_{\zeta_k} + w_{n,i,22} f'(\theta_{n,i_k} - \Delta J_{\zeta_k})^2 \right\}, \tag{C.8}$$

where

$$\begin{aligned} w_{n,i,11} &= E\left\{ (I_{n,i_k} - 1)^2 \chi_{n,k} \right\} - (E\left\{ (I_{n,i_k} - (1 - \alpha_{n,i_k}))\chi_{n,k} \right\})^2, \\ w_{n,i,12} &= (E\left\{ (I_{n,i_k} - (1 - \alpha_{n,i_k}))\chi_{n,k} \right\})(1 - E(I_{n,i_k}\chi_{n,k})), \text{ and} \\ w_{n,i,22} &= \mathrm{Var}(I_{n,i_k}) - w_{n,i,11} - 2w_{n,i,12}. \end{aligned} \tag{C.9}$$

The first order regressions of T_i and S_i on R_i are given by (41). In this form, $\tilde{T}_i/\sigma_{\tau_{i-1}}$ and $\tilde{S}_i/\sigma_{\tau_{i-1}}$ are asymptotically independent of $R_i/\sigma_{\tau_{i-1}}$, and are stably normal with variance $\upsilon_{n,i}^2$. The stable convergence follows in the same way as before. – In the case of regular times,

$$\beta_{n,i_{n,k}} = \frac{\alpha_{n,i}^2 f'(\theta_{n,i_k}) + (1 - \alpha_{n,i}^2) f'(\theta_{n,i_k} - \Delta J_{\zeta_k})}{\alpha_{n,i} f'(\theta_{n,i_k}) + (1 - \alpha_{n,i}) f'(\theta_{n,i_k} - \Delta J_{\zeta_k})}. \tag{C.10}$$

C.3 Proof of Theorem 5: Second order behaviour of (R_i, T_i)

For calculations, focus on the first block. The later blocks follow by the same method but more notation. For simplicity, write M for M_1 and $\Delta\tau$ for $\Delta\tau_1$. We do not assume equidistant spacings.

Note first that, in obvious notation,

$$\begin{aligned}\langle U^{(k_1)}, U^{(k_2)}\rangle &= \Delta\tau^{-1}\sum_{j=1}^{M-1}\left(\frac{M-j}{M}\right)^{k_1+k_2}\Delta\langle X^c, X^c\rangle_{t_j} \\ &= \Delta\tau^{-1}\sum_{j=1}^{M-1}\left(\frac{M-j}{M}\right)^{k_1+k_2}\int_{t_{j-1}}^{t_j}\sigma_t^2 dt \\ &= \sigma_0^2 E(I_0^{k_1+k_2})(1+o_p(1)). \end{aligned} \tag{C.11}$$

For the non-asymptotic covariance expressions, we also get

$$\mathrm{Cov}(U^{(k_1)}, U^{(k_2)}) = \sigma_0^2 E(I_0^{k_1+k_2})(1+O_p(\Delta\tau)) \tag{C.12}$$

since $E\Delta\langle X^c, X^c\rangle_{t_j} = \Delta t_j\sigma_0^2 + O_p(\int_{t_{j-1}}^{t_j} tdt) = \Delta t_j\sigma_0^2(1+O_p(\Delta\tau))$.

We now turn to the third cumulant.

Remark 11. (Important Notation.) From (29), the cumulant will have the form (55) also when there is a jump in the interval $[\tau_{i-1}, \tau_i)$. However, since there are only a finite number of these by assumption, and since the cumulants (apart from the variance) are of order $O_p(\Delta\tau_i^{1/2})$, they will not impact the overall sums of cumulants that give rise to the martingale L_n in (63). Hence, we shall not calculate these explicitly. We shall also, by convention and to avoid clutter, denote X^c by X for the rest of the Appendix. □

We obtain

$$\begin{aligned}\mathrm{cum}_3(U^{(k_1)}, U^{(k_2)}, U^{(k_3)}) &= \mathrm{Cov}(\langle U^{(k_1)}, U^{(k_2)}\rangle, U^{(k_3)})[3] \text{ (notation of McCullagh (1987))} \\ &= \Delta\tau^{-3/2}\mathrm{Cov}(\sum_{j=1}^{M-1}\left(\frac{M-j}{M}\right)^{k_1+k_2}\int_{t_{j-1}}^{t_j}\sigma_t^2 dt, \sum_{l=1}^{M-1}\left(\frac{M-l}{M}\right)^{k_3}\Delta X_{t_l})[3] \\ &= \Delta\tau^{-3/2}\sum_{j=1}^{M-1}\left(\frac{M-j}{M}\right)^{k_1+k_2}\int_{t_{j-1}}^{t_j} dt\mathrm{Cov}(\sigma_t^2, \sum_{l=1}^{M-1}\left(\frac{M-l}{M}\right)^{k_3}\Delta X_{t_l})[3]\end{aligned}$$

Now note that

$$\begin{aligned}\mathrm{Cov}(\sigma_t^2, \sum_{l=1}^{M-1}\left(\frac{M-l}{M}\right)^{k_3}\Delta X_{t_l}) &= \mathrm{Cov}(\sigma_0^2 + 2\int_0^t \sigma_u d\sigma_u + \langle\sigma,\sigma\rangle_t, \sum_{l=1}^{M-1}\left(\frac{M-l}{M}\right)^{k_3}\Delta X_{t_l}) \\ &\approx \mathrm{Cov}(2\int_0^t \sigma_u d\sigma_u, \sum_{l=1}^{M-1}\left(\frac{M-l}{M}\right)^{k_3}\Delta X_{t_l}) \\ &\qquad \text{[exact in the double Gaussian case (Mykland and Zhang (2011)]} \\ &\approx 2\sigma_0\langle\sigma, X\rangle_0'\sum_{t_{l-1}\le t}\left(\frac{M-l}{M}\right)^{k_3}\min(\Delta t_l, t-t_{l-1})\end{aligned}$$

So that

$$
\begin{aligned}
&\mathrm{cum}_3(U^{(k_1)}, U^{(k_2)}, U^{(k_3)}) \\
&\quad \approx \Delta\tau^{-3/2} 2\sigma_0 \langle \sigma, X\rangle_0' \sum_{j=1}^{M-1} \left(\frac{M-j}{M}\right)^{k_1+k_2} \int_{t_{j-1}}^{t_j} dt \sum_{t_{l-1}\le t} \left(\frac{M-l}{M}\right)^{k_3} \min(\Delta t_l, t - t_{l-1})[3] \\
&\quad = \Delta\tau^{-3/2} 2\sigma_0 \langle \sigma, X\rangle_0' \sum_{j=1}^{M-1} \left(\frac{M-j}{M}\right)^{k_1+k_2} \left(\sum_{l=1}^{j-1} \left(\frac{M-l}{M}\right)^{k_3} \int_{t_{j-1}}^{t_j} dt\Delta t_l + \left(\frac{M-j}{M}\right)^{k_3} \int_{t_{j-1}}^{t_j} (t - t_{j-1}) dt\right)[3] \\
&\quad = \Delta\tau^{-3/2} 2\sigma_0 \langle \sigma, X\rangle_0' \sum_{j=1}^{M-1} \left(\frac{M-j}{M}\right)^{k_1+k_2} \left(\sum_{l=1}^{j-1} \left(\frac{M-l}{M}\right)^{k_3} \Delta t_l \Delta t_j + \left(\frac{M-j}{M}\right)^{k_3} \frac{1}{2}\Delta t_j^2\right)[3] \\
&\quad = \Delta\tau^{1/2} 2\sigma_0 \langle \sigma, X\rangle_0' \omega^{k_1k_2,k_3}[3], \qquad \text{(C.13)}
\end{aligned}
$$

where

$$
\omega^{k_1k_2,k_3} = E\left(I_1^{k_1+k_2}(I_1')^{k_3}\chi)\right) \qquad \text{(C.14)}
$$

where I_1' is an independent copy of I_1, $\mathbf{I}\{\cdot\}$ is the indicator function, and

$$
\chi = \mathbf{I}\{I_1' < I_1\} + \frac{1}{2}\mathbf{I}\{I_1' = I_1\}.
$$

To get a further handle on (C.14), observe that

$$
\begin{aligned}
E\left(I_1^a(I_1')^b\chi\right) &= E\left(I_1^a(I_1 \wedge I_1')^b\chi\right) \\
&= E\left(I_1^a(I_1 \wedge I_1')^b\right) - E\left(I_1^a(I_1 \wedge I_1')^b(1-\chi)\right) \\
&= E\left(I_1^a(I_1 \wedge I_1')^b\right) - E\left((I_1 \wedge I_1')^{a+b}(1-\chi)\right) \\
&= E\left(I_1^a(I_1 \wedge I_1')^b\right) - \frac{1}{2}E\left((I_1 \wedge I_1')^{a+b}\right) \qquad \text{(C.15)}
\end{aligned}
$$

where we have used that, by symmetry,

$$
E\left((I_1 \wedge I_1')^{a+b}(1-\chi)\right) = E\left((I_1 \wedge I_1')^{a+b}\chi\right) \qquad \text{(C.16)}
$$

while the two terms in (C.16) must sum to $E\left((I_1 \wedge I_1')^{a+b}\right)$. From (C.15) we thus obtain that

$$
\omega^{k_1k_2,k_3}[3] = E\left(I_1^{k_1+k_2}(I_1 \wedge I_1')^{k_3}\right)[3] - \frac{3}{2}E\left((I_1 \wedge I_1')^{k_1+k_2+k_3}\right) \qquad \text{(C.17)}
$$

Using (C.17), define $\tilde{\omega}^{k_1k_2,k_3}$ as the quantity which arises when replacing T by $\tilde{T}$. It is easy to see that, giving both moment and cumulant representation,

$$\begin{aligned}
\tilde{\omega}^{k_1k_2,k_3}[3] &= E\left((I_1 - E(I_1))^{k_1+k_2}((I_1 \wedge I_1') - E(I_1))^{k_3}\right)[3] - \frac{3}{2}E\left(((I_1 \wedge I_1') - E(I_1))^{k_1+k_2+k_3}\right) \\
&= \mathrm{cum}_3(I_1^{k_1}, I_1^{k_2}, (I_1 \wedge I_1')^{k_3})[3] + E((I_1 \wedge I_1') - E(I_1))^{k_3}\mathrm{Cov}(I_1^{k_1}, I_1^{k_2})[3] \\
&- \frac{3}{2}\mathrm{cum}_3((I_1 \wedge I_1')^{k_1}, (I_1 \wedge I_1')^{k_2}, (I_1 \wedge I_1')^{k_3}) \\
&- \frac{3}{2}E((I_1 \wedge I_1') - E(I_1))^{k_3}\mathrm{Cov}((I_1 \wedge I_1')^{k_1}, (I_1 \wedge I_1')^{k_2})[3] \\
&- \frac{3}{2}E((I_1 \wedge I_1') - E(I_1))^{k_1}E((I_1 \wedge I_1') - E(I_1))^{k_2}E((I_1 \wedge I_1') - E(I_1))^{k_3} \\
&= 3\delta_{\{(k_1,k_2,k_3)=(1,1,1)\}}\left(\mathrm{cum}_3(I_1, I_1, I_1 \wedge I_1') - \frac{1}{2}\mathrm{cum}_3(I_1 \wedge I_1')\right) \\
&+ E((I_1 \wedge I_1') - E(I_1))^{k_3}\delta_{\{(k_1,k_2)=(1,1)\}}\left(\mathrm{Var}(I_1) - \frac{3}{2}\mathrm{Var}(I_1 \wedge I_1')\right)[3] \\
&+ \frac{3}{2}\left(E((I_1 \wedge I_1') - E(I_1))\right)^{k_1+k_2+k_3}
\end{aligned} \tag{C.18}$$

since the k_i are either 0 or 1.

In the case of equispaced times t_j, we obtain

$$\begin{aligned}
\omega^{k_1k_2,k_3} &= \left(\frac{\Delta t}{\Delta \tau}\right)^2 \sum_{j=1}^{M-1}\left(\frac{M-j}{M}\right)^{k_1+k_2}\left(\sum_{l=1}^{j-1}\left(\frac{M-l}{M}\right)^{k_3} + \frac{1}{2}\left(\frac{M-j}{M}\right)^{k_3}\right) \\
&= \sum_{j=1}^{M-1}\left(\frac{M-j}{M}\right)^{k_1+k_2}\frac{1}{M}\left(\sum_{l=1}^{j-1}\left(\frac{M-l}{M}\right)^{k_3}\frac{1}{M} + \frac{1}{2}\left(\frac{M-j}{M}\right)^{k_3}\frac{1}{M}\right) \\
&\approx \int_0^1 x^{k_1+k_2}dx \int_x^1 y^{k_3}dy \quad \text{(up to asymptotically negligible relative error)} \\
&= \int_0^1 x^{k_1+k_2}(1 - x^{k_3+1})dx(k_3+1)^{-1} \\
&= \left((k_1+k_2+1)^{-1} - (k_1+k_2+k_3+2)^{-1}\right)(k_3+1)^{-1} \\
&= \frac{1}{(k_1+k_2+k_3+2)}\left(\frac{1}{(k_1+k_2+1)}\right).
\end{aligned} \tag{C.19}$$

In particular,

$$\omega^{k_1k_2,k_3}[3] = \frac{1}{(k_1+k_2+k_3+2)}\left(\frac{1}{(k_1+k_2+1)}[3]\right). \tag{C.20}$$

D Proof of Theorem 5: The complete cumulants

Remark 11 continues to apply, and we continue the convention from there. Also, we are here in Scenario 1, i.e., (57).

D.1 Cumulants involving s_i^2

For expressions involving s_i^2, we will use (B.16), and also that

$$\frac{s_i^2}{\Delta\tau_i M_i} = T_i^2 + \frac{1}{\Delta\tau_i M_i} \sum_{\tau_{i-1} \le t_j < \tau_i} (X_{t_j} - X_{\tau_{i-1}})^2 \tag{D.1}$$

and so, for example,

$$\begin{aligned}
\mathrm{Cov}(T_i, \frac{s_i^2}{\Delta\tau_i M_i} \mid \mathcal{Z}_{n,i-1}) &= \mathrm{cum}_3(T_i \mid \mathcal{Z}_{n,i-1}) \\
&+ \frac{1}{M_i} \sum_{\tau_{i-1} \le t_j < \tau_i} \mathrm{cum}_3(T_i, \Delta\tau_i^{-1/2}(X_{t_j} - X_{\tau_{i-1}}), \Delta\tau_i^{-1/2}(X_{t_j} - X_{\tau_{i-1}}) \mid \mathcal{Z}_{n,i-1}) \\
&= O_p(\Delta\tau_i^{1/2});
\end{aligned} \tag{D.2}$$

for the first term, this is explicitly shown in Section C.3, and for the second term, it follows by a very similar calculation (replace R_i by $R_i^{(j)} = \Delta\tau_i^{-1/2}(X_{t_j} - X_{\tau_{i-1}})$ and proceed in the same way). By similar methods,

$$\begin{aligned}
&\mathrm{cum}_3(U_i^{(k_1)}, U_i^{(k_2)}, \frac{s_i^2}{\Delta\tau_i M_i} \mid \mathcal{Z}_{n,i-1}) \\
&= \mathrm{cum}_4(U_i^{(k_1)}, U_i^{(k_2)}, T_i, T_i \mid \mathcal{Z}_{n,i-1}) + \mathrm{cum}_3(U_i^{(k_1)}, U_i^{(k_2)}, T_i^2 \mid \mathcal{Z}_{n,i-1}) \\
&+ \frac{1}{M_i} \sum_{\tau_{i-1} \le t_j < \tau_i} \mathrm{cum}_4(U_i^{(k_1)}, U_i^{(k_2)}, R_i^{(j)}, R_i^{(j)} \mid \mathcal{Z}_{n,i-1}) \\
&+ \frac{1}{M_i} \sum_{\tau_{i-1} \le t_j < \tau_i} \mathrm{cum}_3(U_i^{(k_1)}, U_i^{(k_2)}, (R_i^{(j)})^2 \mid \mathcal{Z}_{n,i-1}) + o_p(1) \\
&= 2\mathrm{Cov}(U_i^{(k_1)}, T_i \mid \mathcal{Z}_{n,i-1})\mathrm{Cov}(U_i^{(k_2)}, T_i \mid \mathcal{Z}_{n,i-1}) \\
&+ 2\frac{1}{M_i} \sum_{\tau_{i-1} \le t_j < \tau_i} \mathrm{Cov}(U_i^{(k_1)}, R_i^{(j)} \mid \mathcal{Z}_{n,i-1})\mathrm{Cov}(U_i^{(k_2)}, R_i^{(j)} \mid \mathcal{Z}_{n,i-1}) + o_p(1) \\
&= 2\sigma_{\tau_{i-1}}^4 \left\{ E(I_i^{k_1+1})E(I_i^{k_2+1}) + E\left((I_i \wedge I_i')I_i^{k_1}(I_i')^{k_2}\right) \right\} + o_p(1)
\end{aligned} \tag{D.3}$$

where I_i' is an independent copy of I_i. (Very similar expressions are given in Section C.3.) Note that the fourth cumulants do not contribute to the expression, and we have used (C.12) in the final transition. If we set

$$\tilde{U}_i^{(1)} = U_i^{(1)} - E(I_i)U_i^{(0)} \text{ and } \tilde{U}_i^{(0)} = U_i^{(0)}, \tag{D.4}$$

we obtain similarly that

$$\begin{aligned}
&\mathrm{cum}_3(\tilde{U}_i^{(k_1)}, \tilde{U}_i^{(k_2)}, \frac{s_i^2}{\Delta\tau_i M_i} \mid \mathcal{Z}_{n,i-1}) \\
&= 2\sigma_{\tau_{i-1}}^4 \left\{ \mathrm{cum}_{k_1+1}(I_i)\mathrm{cum}_{k_2+1}(I_i) + E\left((I_i \wedge I_i')(I_i - E(I_i))^{k_1}(I_i' - E(I_i'))^{k_2}\right) \right\} + o_p(1),
\end{aligned} \tag{D.5}$$

where cum_1 is the expectation and cum_2 is the variance.

D.2 Conditional cumulants of $H_i = \Delta\tau_i^{-1/2}\left(\hat{X}_i - \bar{X}_i - \hat{\theta}_i\right) = \Delta\tau_i^{-1/2}\delta_i$

Set $Z_i = M_i^{1/2}\hat{\theta}_i$, and $\tilde{D}_i = \Delta\tau_i^{-1/2}M_i^{-1/2}(D_i - E(D_i \mid X))$. Also denote $\Theta = (\hat{\theta}_i)_{i=1,2,\dots}$ and $X = (X_t)_{0\le t\le T}$.

First, note that by symmetry,

$$E(D_i \mid X, \Theta_i) = E(D_i \mid X) = \frac{1}{2}s_i^2 E\psi''(\epsilon). \tag{D.6}$$

Meanwhile, from p. 164 in McCullagh (1987), and since the information in $(\hat{\theta}_v)_{v\neq i}$ is negligible,

$$\begin{aligned}\mathrm{Var}(\tilde{D}_i \mid X, \Theta) &= \mathrm{Var}(\tilde{D}_i \mid X, Z_i) + O_p(M_i^{-1}) = (\Delta\tau_i M_i)^{-1}s_i^2 + O_p(M_i^{-1})\\ \mathrm{cum}_3(\tilde{D}_i \mid X, \Theta) &= \mathrm{cum}_3(\tilde{D}_i \mid X, Z_i) + O_p(\Delta\tau_i^{1/2}M_i^{-3/2}) = O_p(\Delta\tau_i^{1/2}M_i^{-3/2})\end{aligned} \tag{D.7}$$

The biggest order terms go away as follows. On the one hand, by construction, $\mathrm{Cov}(Z_i, D_i \mid X) = \mathrm{cum}_3(Z_i, S_i, D_i \mid X) = 0$. On the other hand, we calculate by stochastic expansion. For example, for the second third order cumulant, set $Z_i^{(1)} = M_i^{1/2}E\psi'(\epsilon))^{-1}\sum_j \psi(\epsilon_{t_j})$ so that $Z_i = Z_i^{(1)} + O_p(M_i^{-1/2})$. Then, by stochastic expansion, $\mathrm{cum}_3(Z_i, \tilde{D}_i, \tilde{D}_i \mid X) = \mathrm{cum}_3(Z_i^{(1)}, \tilde{D}_i, \tilde{D}_i \mid X) + O_p(M_i^{-7/2}) = O_p(M_i^{-1})$.

Set $H_i = \Delta\tau_i^{-1/2}\left(\hat{X}_i - \bar{X}_i - \hat{\theta}_i\right)$. Also, since this term will occur a lot, set

$$K_1 = \frac{1}{2}\frac{E\psi''(\epsilon)}{E\psi'(\epsilon)}. \tag{D.8}$$

From Theorem 1, $H_i = M_i^{-1/2}(E\psi'(\epsilon))^{-1} \times \Delta\tau_i^{-1/2}M_i^{-1/2}D_i + o_p(\Delta\tau_i^{1/2})$. Thus

$$\begin{aligned}E(H_i \mid X, Z_i) = E(H_i \mid X) &= \Delta\tau_i^{1/2}\frac{s_i^2}{\Delta\tau_i M_i}K_1 \ + \ o_p(\Delta\tau_i^{1/2})\\ \mathrm{Var}(H_i \mid X, Z_i) &= \Delta\tau_i^{-1}M_i^{-2}s_i^2(E\psi'(\epsilon))^{-2} + O_p(M_i^{-2}) + o_p(M_i^{-1/2}\Delta\tau_i^{1/2}) + o_p(\Delta\tau_i)\\ &= o_p(\Delta\tau_i^{2/3})\\ \mathrm{cum}_3(H_i \mid X, Z_i) &= M_i^{-3/2}O_p(\Delta\tau_i^{1/2}M_i^{-3/2}) + o_p(\Delta\tau_i)\\ &= o_p(\Delta\tau_i),\end{aligned} \tag{D.9}$$

where both the second transitions were due to the second equation in Footnote 16 to Condition 5, as well as the order of s_i^2.

D.3 Conditional cumulants of $S_i = H_i + T_i$ and $\tilde{S}_i$

Recall that $S_i = \Delta\tau_i^{-1/2}\left(\hat{X}_i - X_{\tau_{i-1}} - \hat{\theta}_i\right) = H_i + T_i$. Set

$$\bar{U_i}^{(k)} = R_i \text{ for } k = 0 \text{ and } = S_i \text{ for } k = 1.$$

Thus $\bar{U_i}^{(k)} = U_i^{(k)} + H_i \delta_{\{k=1\}}$. From (D.9),

$$\begin{aligned}
E(S_i \mid X, \Theta) &= T_i + \Delta\tau_i^{1/2} \frac{s_i^2}{\Delta\tau_i M_i} K_1 \;+\; o_p(\Delta\tau_i^{1/2}) \\
\mathrm{Var}(S_i \mid X, \Theta) &= o_p(\Delta\tau_i^{2/3}) \\
\mathrm{cum}_3(S_i \mid X, \Theta) &= o_p(\Delta\tau_i).
\end{aligned} \tag{D.10}$$

Using rules for conditional cumulants (Brillinger (1969); Speed (1983)), and since $E(R_i \mid \mathcal{Z}_{n,i-1}) = E(T_i \mid \mathcal{Z}_{n,i-1}) = 0$, we thus obtain

$$\begin{aligned}
E(S_i \mid \mathcal{Z}_{n,i-1}) &= \Delta\tau_i^{1/2} \frac{E(s_i^2 \mid \mathcal{Z}_{n,i-1})}{\Delta\tau_i M_i} K_1 \;+\; o_p(\Delta\tau_i^{1/2}) \\
&= \Delta\tau_i^{1/2} \sigma^2_{\tau_{i-1}} E(I_i(1-I_i)) K_1 \;+\; o_p(\Delta\tau_i^{1/2}) \text{ [by (B.16)]} \\
\mathrm{Cov}(S_i, \bar{U_i}^{(k)} \mid \mathcal{Z}_{n,i-1}) &= \mathrm{Cov}(E(S_i \mid X, \Theta), E(\bar{U_i}^{(k)} \mid X, \Theta) \mid \mathcal{Z}_{n,i-1}) + E(\mathrm{Cov}(S_i, \bar{U_i}^{(k)} \mid X, \Theta) \mid \mathcal{Z}_{n,i-1}) \\
&= \mathrm{Cov}(T_i, U_i^{(k)} \mid \mathcal{Z}_{n,i-1}) + \Delta\tau_i^{1/2}(k+1)\mathrm{Cov}(\frac{s_i^2}{\Delta\tau_i M_i} K_1, U_i^{(k)} \mid \mathcal{Z}_{n,i-1}) + o_p(\Delta\tau_i^{1/2}) \\
&= \mathrm{Cov}(T_i, U_i^{(k)} \mid \mathcal{Z}_{n,i-1}) + o_p(\Delta\tau_i^{1/2}) \text{ [from (D.2) and (C.12)]} \\
\mathrm{cum}_3(\bar{U_i}^{(k_1)}, \bar{U_i}^{(k_2)}, \bar{U_i}^{(k_3)} \mid \mathcal{Z}_{n,i-1}) &= \mathrm{cum}_3(E(\bar{U_i}^{(k_1)} \mid X, \Theta), E(\bar{U_i}^{(k)} \mid X, \Theta), E(\bar{U_i}^{(k)} \mid X, \Theta) \mid \mathcal{Z}_{n,i-1}) \\
&+ \mathrm{Cov}(E(\bar{U_i}^{(k_1)} \mid X, \Theta), \mathrm{Cov}(\bar{U_i}^{(k_2)}, \bar{U_i}^{(k_3)} \mid X, \Theta) \mid \mathcal{Z}_{n,i-1})[3] \\
&+ E(\mathrm{cum}_3(\bar{U_i}^{(k_1)}, \bar{U_i}^{(k_2)}, \bar{U_i}^{(k_3)} \mid X, \Theta) \mid \mathcal{Z}_{n,i-1}) \\
&= \mathrm{cum}_3(E(\bar{U_i}^{(k_1)} \mid X, \Theta), E(\bar{U_i}^{(k)} \mid X, \Theta), E(\bar{U_i}^{(k)} \mid X, \Theta) \mid \mathcal{Z}_{n,i-1}) + o_p(\Delta\tau_i) \\
&= \mathrm{cum}_3(U_i^{(k_1)}, U_i^{(k_2)}, U_i^{(k_3)} \mid \mathcal{Z}_{n,i-1}) \\
&+ \Delta\tau_i^{1/2} K_1 \mathrm{cum}_3(\frac{s_i^2}{\Delta\tau_i M_i}, U_i^{(k_2)}, U_i^{(k_3)}) \delta_{\{k_1=1\}}[3] \;+\; o_p(\Delta\tau_i^{1/2})
\end{aligned} \tag{D.11}$$

The third cumulant $\mathrm{cum}_3(U_i^{(k_1)}, U_i^{(k_2)}, U_i^{(k_3)} \mid \mathcal{Z}_{n,i-1})$ is given in Section C.3, where it is seen to be of exact order $O_p(\Delta\tau_i^{1/2})$, as required. For expressions involving s_i^2, we have used (B.16), and also the results from Section D.1. The third cumulant $\mathrm{cum}_3(\frac{s_i^2}{\Delta\tau_i M_i} K_1, U_i^{(k_2)}, U_i^{(k_3)})$ is given by (D.5) in Section D.1.

Finally, set $V_i^0 = R_i$ and $V_i^1 = \tilde{S}_i = H_i + \tilde{T}_i$, as in Theorem 5. We obtain, with $\tilde{U}$ given in (D.4),

$$\begin{aligned}
\mathrm{cum}_3(V_i^{k_1}, V_i^{k_2}, V_i^{k_3} \mid \mathcal{Z}_{n,i-1}) &= \mathrm{cum}_3(\tilde{U_i}^{(k_1)}, \tilde{U_i}^{(k_2)}, \tilde{U_i}^{(k_3)} \mid \mathcal{Z}_{n,i-1}) \\
&+ \Delta\tau_i^{1/2} K_1 \mathrm{cum}_3(\frac{s_i^2}{\Delta\tau_i M_i}, \tilde{U_i}^{(k_2)}, \tilde{U_i}^{(k_3)}) \delta_{\{k_1=1\}}[3] \;+\; o_p(\Delta\tau_i^{1/2}) \\
&= \Delta\tau_i^{1/2}(b_i^{k_1k_2k_3} + a_i^{k_1k_2k_3}) + \; o_p(\Delta\tau_i^{1/2})
\end{aligned} \tag{D.12}$$

where $a_i^{k_1k_2k_3}$ and $b_i^{k_1k_2k_3}$ is given in equation (69) in Theorem 5. The expressions for the expectation and variance terms follow similarly.

E Proof of Theorem 6

Remark 11 continues to apply, and we continue the convention from there. – Recall that $P_{n,2}$ is the canonical normal approximation corresponding to the sequence (58), where $A_{i-1} = (\kappa_{i-1}, \hat{\theta}_i, R_i)$ and $V_i = \tilde{S}_i$. Also let $\breve{A}_{i-1} = (\kappa_{i-1}, \hat{\theta}_i)$ and $\breve{V}_i = (R_i, \tilde{S}_i)$ be the partition from Theorem 5.

Observe that under all of P^*, $P_{n,1}$ and $P_{n,2}$,

$$\begin{aligned}\log f(\tilde{S}_i \mid \kappa_{i-1}, \hat{\theta}_i, R_i, \tilde{S}_{i-1}, U_{i-2}, ..., U_0) &= \log f(R_i, \tilde{S}_i \mid \kappa_{i-1}, \hat{\theta}_i, U_{i-2}, ..., U_0) \\ &\quad - \log f(R_i, \mid \kappa_{i-1}, \hat{\theta}_i, U_{i-2}, ..., U_0) \end{aligned} \tag{E.1}$$

while under $P_{n,1}$ and $P_{n,2}$,

$$\begin{aligned}\log f(R_i, \tilde{S}_i \mid \kappa_{i-1}, \hat{\theta}_i, U_{i-2}, ..., U_0) &= \log f(\tilde{S}_i \mid \kappa_{i-1}, \hat{\theta}_i, \tilde{S}_{i-1}, U_{i-2}, ..., U_0) \\ &\quad + \log f(R_i \mid \kappa_{i-1}, \hat{\theta}_i, \tilde{S}_{i-1}, U_{i-2}, ..., U_0). \end{aligned} \tag{E.2}$$

. Thus

$$\begin{aligned}\log \frac{f_{P^*}(\tilde{S}_i \mid R_i, \kappa_{i-1}, \hat{\theta}_i, U_{i-2}, ..., U_0)}{f_{P_{n,2}}(\tilde{S}_i \mid R_i, \kappa_{i-1}, \hat{\theta}_i, U_{i-2}, ..., U_0)} &= \log \frac{f_{P^*}(\tilde{S}_i, R_i \mid \kappa_{i-1}, \hat{\theta}_i, U_{i-2}, ..., U_0)}{f_{P_{n,1}}(\tilde{S}_i, R_i \mid \kappa_{i-1}, \hat{\theta}_i, U_{i-2}, ..., U_0)} \\ &\quad - \log \frac{f_{P^*}(R_i \mid \kappa_{i-1}, \hat{\theta}_i, U_{i-2}, ..., U_0)}{f_{P_{n,1}}(R_i \mid \kappa_{i-1}, \hat{\theta}_i, U_{i-2}, ..., U_0)}. \end{aligned} \tag{E.3}$$

The problem thus reduces to

$$\begin{aligned}\log \frac{dP^*}{dP_{n,2}} \text{ based on } (V, A) \;&=\; \log \frac{dP^*}{dP_{n,1}} \text{ based on } (\breve{V}, \breve{A}) \text{ as in Theorem 5} \\ &\quad - \log \frac{dP^*}{dP_{n,0}}, \text{ where} \end{aligned} \tag{E.4}$$

$$\frac{dP^*}{dP_{n,0}} \;=\; \log \frac{dP^*}{dP_{n,1}} \text{ based on } (V', A') \tag{E.5}$$

where $A'_{i-1} = (\kappa_{i-1}, \hat{\theta}_i)$ and $V'_i = R_i$. Observe that $P_{n,0}$ is the restriction of $P_{n,1}$ to a smaller sigma-field.

$P_{n,0}$ falls under the setup in Section 3.2.2. Because of the independence of the $\hat{\theta}_i$s, $P_{n,0}$ is multiplicatively related to the one step contiguous normal target measure studied in Mykland and Zhang (2009, Sections 2.3-2.4). In particular, the cumulants are, in this case, additively related.

The martingale L_n (under $P_{n,1}$) from (63) corresponding to $\log \frac{dP^*}{dP_{n,1}}$ is $L_{n,1}$ from Theorem 5. Meanwhile, if $L_{n,0}$ is the martingale (also under $P_{n,1}$) corresponding to $\log \frac{dP^*}{dP_{n,0}}$. We obtain in the same way as Theorem 5 that

$$L_{n,0} \;=\; \sum_i \Delta\tau_i^{1/2} \breve{\kappa}^0(\tau_{i-1}) \breve{h}_0(\breve{V}_i) + \frac{1}{3!} \Delta\tau_i^{1/2} \breve{\kappa}^{0,0,0}(\tau_{i-1}) \breve{h}_{000}(\breve{V}_i), \tag{E.6}$$

whence $L_{n,2} = L_{n,1} - L_{n,0}$. The result then follows from the proof of Theorem 5 (in this paper) as well as the proofs of Theorems 1-2 in Mykland and Zhang (2009). ∎