Assessment of Uncertainty in High Frequency Data: The Observed Asymptotic Variance*

Per A. Mykland
The University of Chicago

Lan Zhang
University of Illinois at Chicago

This version: May 12, 2014

Abstract

High frequency inference has generated a wave of research interest among econometricians and practitioners, as indicated from the increasing number of estimators based on intra-day data. However, we also witness a scarcity of methodology to assess the uncertainty – standard error– of the estimator. The root of the problem is that whether with or without the presence of microstructure, standard errors rely on estimating the asymptotic variance (AVAR), and often this asymptotic variance involves substantially more complex quantities than the original parameter to be estimated.

Standard errors are important: they are used both to assess the precision of estimators in the form of confidence intervals, to create "feasible statistics" for testing, and also when building forecasting models based on, say, daily estimates.

The contribution of this paper is to provide an alternative and general solution to this problem, which we call *Observed Asymptotic Variance*. It is a general nonparametric method for assessing asymptotic variance (AVAR), and it provides consistent estimators of AVAR for a broad class of integrated parameters $\Theta = \int \theta_t dt$. The spot parameter process θ can be a general semimartingale, with continuous and jump components. The construction and the analysis of $\widehat{AVAR}(\hat{\Theta})$ work well in the presence of microstructure noise, and when the observation times are irregular or asynchronous in the multivariate case. The edge effect – phasing in and phasing out the information on the boundary of the data interval – of any relevant estimator is also analyzed and treated rigorously.

As part of the theoretical development, the paper shows how to feasibly disentangle the effect from estimation error $\hat{\Theta} - \Theta$ and the variation in the parameter θ alone. For the latter, we obtain a consistent estimator of the quadratic variation (QV) of the parameter to be estimated, for example, the QV of the leverage effect.

The methodology is valid for a wide variety of estimators, including the standard ones for variance and covariance, and also for estimators, such as, of leverage effects, high frequency betas, and semi-variance.

KEYWORDS: asynchronous times, consistency, discrete observation, edge effect, irregular times, leverage effect, microscructure, observed information, realized volatility, robust estimation, semimartingale, two scales estimation.

*We would like to thank Jean Jacod, Eric Renault, Olivier Scaillet, and Kevin Sheppard for helpful discussions. Financial support from the National Science Foundation under grant SES 11-24526 is gratefully acknowledged.

1 Introduction

1.1 The Problem of Standard Error

As high frequency data becomes more readily available, the demand for analyzing such big and noisy data is also increasing. Within the recent decade, we have seen the arrival of novel methodologies for using the high frequency data to estimate volatility, to assess the asymmetric information in financial returns via semi-variance, to measure statistical leverage, to make inference on the number of jumps, and many other objects of interest. As financial markets and global economies evolve, we expect an ongoing need to estimate new parameters of interest from data of the high-frequency variety. This process will substantially improve the precision with which we can measure financial and economic quantities.

A main hindrance to this development is that it is often very difficult to set the standard errors for estimators. When we are faced with an unknown Θ, we need not only to find an estimator $\hat{\Theta}_n$, but we also need to estimate the size of the estimation error. The latter is usually expressed in the form of a standard error $\mathrm{se}(\hat{\Theta})_n$, by which we here mean a data-based statistic so that

$$\frac{\hat{\Theta}_n - \Theta}{\mathrm{se}(\hat{\Theta})_n} \xrightarrow{\mathcal{L}} N(0,1), \tag{1}$$

where the convergence in law takes place as the number of observations n becomes large. Understanding the estimation uncertainty is as important in high frequency econometrics as in other fields of statistics, because it provides a tool to test hypotheses, to set up confidence intervals, to improve forecasting, and to optimize tuning parameters in finite sample problems. However, delivering a feasible measure for the standard error of an estimator – no matter whether it is volatility, or regression betas, or leverage effect – has shown itself to be particularly hard in high frequency data.

1.2 The Trouble with Standard Errors: Diagnosis

Why is it so hard to set standard errors? We here provide a diagnosis for this phenomenon, and an alternative solution. The diagnosis is that the literature has tended to rely on an approach to the problem which may be called "estimated theoretical asymptotic variance (AVAR)", similar to the "estimated expected information" of likelihood theory. In other words, one calculates the theoretical AVAR, and then one finds an estimator $\widehat{AVAR}_n$ of the theoretical quantity AVAR.[1]

[1] The underlying theory would be that $n^{\alpha}(\hat{\Theta}_n - \Theta)$ converges in law stably (for some $\alpha > 0$) to a normal distribution with a random variance V (see Section 4.1). We use the notation $\mathrm{AVAR} = \mathrm{AVAR}_n = n^{-2\alpha}V$, and $\widehat{\mathrm{AVAR}}_n$ is consistent if $\widehat{\mathrm{AVAR}}_n = \mathrm{AVAR}_n(1 + o_p(1))$. Hence the standard error takes the form (2). This is useful since it clarifies that one does not need to know the rate of convergence n^{α} to implement our results in data. It also leads to less cluttered notation.

Finally, the standard error is

$$\mathrm{se}(\hat{\Theta})_n \;=\; \sqrt{\widehat{AVAR}_n}. \tag{2}$$

The problem with the above procedure is that the implementation is often overwhelmingly difficult. The theoretical asymptotic variance of $\hat{\Theta}$ is frequently complicated, and in addition actually harder to estimate than Θ itself. To corroborate this, we draw attention to the substantial number of cases where one can find an estimator $\hat{\Theta}_n$ of Θ, but feasible (asymptotically pivotal) statistics of the form (1) are not available.

A notable class of examples of this problem is provided by the number of estimators that are documented for the case where there is no microstructure – thus revealing interest in the problem – but little literature on the case where microstructure is present. Anecdotal evidence suggests that this is usually because microstructure makes the problem so forbidding that researchers never get around to it. Also, the main challenge is not in finding $\widehat{\Theta}_n$, but rather the problem of finding AVAR and $\widehat{\mathrm{AVAR}}_n$.[2] Examples in the literature include, but not limited to, semivariance (Barndorff-Nielsen, Kinnebrouck, and Shephard (2009)), nearest neighbor truncation (Andersen, Dobrev, and Schaumburg (2012), see also Section 8), or estimating the rank of the volatility matrix (Jacod and Podolskij (2013)), or the volatility of volatility (Vetter (2011)), high frequency regression, and ANOVA (Mykland and Zhang (2006, 2012)). In all these examples, one can obtain a point estimate, but under microstructure, one does not have ready access to tests, confidence intervals, or other methods that require a standard error. The overall challenge is thus not specific to one estimator, but holds across estimators of various types, which reminds us that we all are in the same boat in searching for how to quantify the uncertainty in the estimators.[3]

The problem of lacking the estimation error is equally pronounced when one seeks to tackle asynchronous observations, or even just irregular observations. Like microstructure, asynchronicity and irregularity in observation times are natural features of the high frequency data. After all, trades (or quote updates) across different securities rarely occur simultaneously. Even for a single security series (say, Google), one can hardly require it to trade on a regular time schedule, or to have bid and ask updates at synchronous times. However, in the setting that takes account of asynchronous and irregular observation times, the literature is meager on the topic of how to assess the estimation uncertainty, the only exception being the covariance estimator. A deeper understanding of many other highly relevant estimators is greatly needed, for example, co-skewness under microstructure, and estimators which involve large dimension p of multivariate processes (usually formalized by $p \to \infty$ with n). Here again, the lack of literature would seem to be due

[2] Researchers frequently also seek solace in a consideration that data sampled every fine minutes doesn't really have miocrostructure. The latter is optimistic, and is rarely checked out properly, and in any case, one loses a lot of data this way. – The other popular slogan is that one can always pre-average. Indeed, this will in many cases yield $\widehat{\Theta}_n$, but it does not solve the standard error problem.

[3] To the best of our knowledge, assuming the presence of microstructure noise, the theoretical AVAR and its estimation have been documented only in the case of variance (volatility), covariance, leverage effect (skewness), and, in some instances, of jumps. See Section 7 for references.

more to a lack of AVAR and $\widehat{\text{AVAR}}_n$ than a lack of $\hat{\Theta}_n$.

1.3 Our Alternative Solution: An Observed Asymptotic Variance (Observed AVAR)

Our proposal in this paper is for an all-purpose high frequency estimator of standard error that *bypasses* the theoretical asymptotic variance. In other words, we provide a statistic $\widehat{\text{AVAR}}_n$ – which we call the *observed asymptotic variance* – which does not rely on finding the theoretical AVAR. The *observed standard error* then follows by taking the square root (2). It remains true that the proposed estimator $\widehat{\text{AVAR}}_n$ is consistent, in other words, $\widehat{\text{AVAR}}_n = \text{AVAR}_n(1 + o_p(1))$. One does not, however, need to derive any formula for the theoretical AVAR_n. The crucial result (1) is stated as Theorem 6 in Section 6.2, with weights chosen as in Section 6.4.

The observed asymptotic variance resembles the observed information in parametric statistical theory, in that there is no need for an intermediate theoretical asymptotic step, involving expectations or similar operations. Just as in likelihood theory, the observed asymptotic variance is easier to use, and it has a more universal form. – In parametric statistic, there has been a lively debate about the relative accuracy properties of observed and estimated expected information.[4] Some of the same issues may pertain to the corresponding two types of $\widehat{\text{AVAR}}$, but we have not investigated this matter.

Apart from regularity conditions, our only assumption is that $\Theta = \int_0^{\mathcal{T}} \theta_t dt$, where the spot parameter process $\{\theta_t\}$ is allowed to be a general semimartingale, hence $\{\theta_t\}$ can have jump or continuous evolution and it can be either an Itô or non-Itô process as in Calvet and Fisher (2008).[5] Allowing non-Itô processes makes the results appropriate to more areas of data applications. We shall see in Sections 7-8 that the conditions for our results are satisfied broadly, including on quite exotic quantities such as leverage effect, and nearest neighbor truncation. See Section 9 for some additional practical guidance to how to use our theory.

As an additional application, many estimators involve one or more "tuning parameters", such as block or subgrid size. Optimizing the estimator $\hat{\Theta}_n$ as a function of these tuning parameters would naturally involve minimizing the asymptotic variance. We shall see that this optimization can be done on the basis of our proposed $\widehat{\text{AVAR}}_n$. See Sections 2.2.2 and 6.5 for references and

---

[4] Originally going back to the debates between Fisher, and Neyman and Pearson. The neo-likelihood wave would seem to have started with Cox (1958, 1980) and Efron and Hinkley (1978), followed by a large literature, including Barndorff-Nielsen (1986, 1991); DiCiccio and Romano (1989); DiCiccio, Hall, and Romano (1991); Jensen (1992, 1995, 1997); McCullagh (1984, 1987); McCullagh and Tibshirani (1990); Pierce and Peters (1994); Reid (1988); Skovgaard (1986, 1991); Mykland (1995a, 1999, 2001). Connoisseurs of the likelihood argument may feel uneasy with the symbol $\widehat{\text{AVAR}}_n$. We have used this notation to emphasize that the observed asymptotic variance is consistent for the theoretical AVAR.

[5] See also Rosenbaum, Duvernet, and Robert (2010) and Aït-Sahalia and Jacod (2013) for recent interest in this type of evolution.

further development.

1.4 Connections

The basic principle behind the observed AVAR is to segment the available time line into sub-periods, and then compare the estimators in successive sub-periods. We show that this difference can be decomposed into two parts. One part reveals the behavior of $\hat{\Theta}$ in the form of its estimation error, and the other part tells us the dynamics of spot parameter process θ alone. We develop estimators to disentangle these two effects and to construct the observed AVAR. A heuristic outline of the principles is given in Section 2.

Our procedure is unlike resampling in that it is not based on the "Russian doll" principle (Hall (1992, Chapter 1.2)), and in particular it does not involve a second level of nesting. The comparison of adjacent estimators, however, is also a feature of the subsampling developed for volatility in the pioneering work of Kalnina and Linton (2007) and Kalnina (2011).

In addition to the overall construction of observed asymptotic variance, there are two other intellectual novelties in the paper. On the one hand, the comparison of adjacent values of the integral of θ is given a highly precise formulation, in the form of the *Integral-to-Spot Device* (Theorem 1 in Section 3) which shows that "realized volatility" of integrals $\int \theta_t dt$ converges to the volatility of the spot parameter process θ_t. The only condition is that the spot process be a semi-martingale.

On the other hand, the estimation of asymptotic variance $\text{AVAR}(\hat{\Theta} - \Theta)$ is reduced to a problem which resembles that of estimating volatility, with edge effects playing the rôle of "microstructure". We can thus adapt known methods to the current problem of estimating asymptotic variance. It is worth to mention that edge effect is estimator-specific. As its name suggests, edge effect shows up in an estimator whenever the estimator under-uses or over-uses the data at the edge of a sampling interval, relative to the middle portion of the data interval. In a sense, this effect permeates in high frequency inference, especially when the inference involves multi-variate, multi-power, or multi-scale estimation, or microstructure noise. In the current paper, edge effect of different magnitudes is explicitly discussed and treated. The effect is also referred to as burn-in time, and border effect. After setting up the statistical structure, we pursue this in Sections 4-6.

We emphasize that our purpose in this paper is to provide a method for getting at observed asymptotic variance, for any estimator of interest. The proposed approach extends broadly to high frequency inference. The contribution of the current paper is to estimators other than volatility of the financial returns. For the latter, much is known, both in terms of asymptotic variances AVAR, and in terms of resampling.[6] Our examples in Section 7 include volatility examples, just to show behavior in the baseline case. Volatility, however, is not our main focus.

[6] The subsampling device mentioned above, and the bootstrapping of Gonçalves and Meddahi (2009) and Gonçalves, Donovon, and Meddahi (2013).

In addition to the main line of argument, we also provide consistent estimators of the quadratic variation of the spot parameter process θ_t (Sections 2.3, 6.3-6.4). The generalization to the multidimensional case is given in Section 6.6. Sections 7 and 8 give examples. Section 9 provides practical guidance to using the theory, Finally, proofs are, for the most part, located in the Appendix.

2 Outline of Observed Asymptotic Variance

2.1 The Apparent Quadratic Variation of a Parameter Process

We here describe the simplest case in heuristic terms, where the parameter process is continuous, and where edge effects are negligible. A formal statement embodying these results is given after the introduction of the relevant definitions, in Theorem 2 in Section 4.2. We treat the general case in Section 6, after building up the technical tools in Sections 3-5.

We observe a semimartingale X_t at high frequency, or an observable Y_t which may be contaminated by microstructure noise. We suppose that we are interested in estimating integrals of a "parameter" spot process θ_t, which also is assumed to be a semimartingale.[7] For example, we can take θ to be the spot variance of the continuous part of the process X_t: $\theta_t = \sigma_t^2$ where $dX_t = \sigma_t dW_t + dt\text{-terms} + \text{jump terms}$, and W is a Brownian motion. In the multivariate case, θ_t can be a function of the instantaneous covariance. The development, however, holds more generally, such as for the leverage effect where $\theta_t = d[X^c, \sigma^2]_t/dt$, the volatility of volatility where $\theta_t = d[\sigma^2, \sigma^2]_t^c/dt$, or other. The case of multivariate θ_t is considered in Section 6.6.

We here consider B time periods (days, five minutes, or other) $(T_{i-1}, T_i]$ from $T_0 = 0$ to $T_B = \mathcal{T}$. We suppose that we have at hand a consistent estimator $\hat{\Theta}_i$ of

$$\Theta_i = \int_{T_{i-1}}^{T_i} \theta_t dt \ . \tag{3}$$

Even when estimating the spot volatility, one almost invariably estimates such integrals.[8] The estimator $\hat{\Theta}_i$ can depend on tuning parameters, as we shall see in Section 6.5.

The basic insight behind the Observed AVAR is that we can decompose the increment $\hat{\Theta}_{i+1} - \hat{\Theta}_i$

[7] For the definition of semimartingale, see, *e.g.,* Jacod and Shiryaev (1987, Definition I.4.41, p. 43), as well as Protter (2004, Definitions on p. 52, and Definition and Theorem III.1 on p. 102), and also Dellacherie and Meyer (1982). The theory requires the existence of a "spot" θ_t, cf. Section 9.2. – To the extent that the "integral" process has jumps, we assume that they have been suitably removed by the estimation procedure in use, as also discussed at the beginning of Section 7, see also Examples 1 and 8 in the same section. On the other hand, we shall see that the process θ_t can have as many jumps as it wants.

[8] The standard spot estimate is $\hat{\theta}_{T_i} = \hat{\Theta}_i/(T_i - T_{i-1})$ for suitable choice of T_{i-1}. See, for example, Foster and Nelson (1996); Comte and Renault (1998); Mykland and Zhang (2008).

into the parts related to estimator behavior and the part solely tied to parameter behavior:

$$\hat{\Theta}_{i+1} - \hat{\Theta}_i = \underbrace{\hat{\Theta}_{i+1} - \Theta_{i+1}}_{\text{estimation error}} + \underbrace{\Theta_{i+1} - \Theta_i}_{\text{evolution in parameter}} - \underbrace{\hat{\Theta}_i - \Theta_i}_{\text{estimation error}} . \tag{4}$$

We can therefore write the *apparent quadratic variation* of Θ_t[9] as

$$\begin{aligned}\sum_i (\hat{\Theta}_{i+1} - \hat{\Theta}_i)^2 &= 2\sum_i (\hat{\Theta}_i - \Theta_i)^2 + \sum_i (\Theta_{i+1} - \Theta_i)^2 \\ &\quad + \text{ martingale and negligible terms} \\ &= \left(\underbrace{2\sum_i \text{AVAR}(\hat{\Theta}_i - \Theta_i)}_{\text{cumulative AVAR}} + \underbrace{\text{quadratic variation of } \Theta_i}_{\text{parameter behavior}} \right)(1 + o_p(1)) \end{aligned} \tag{5}$$

when $\max_i(T_{i+1} - T_i)$ goes to zero, and the asymptotic variance is accumulated across B time periods.[10]

The problem we seek to solve is to extract the AVARs, that is, to disentangle the AVARs from the parameter behavior in (5). To achieve this, we need to get a handle on the second term in (5). We need to show how the quadratic variation of the integrals $\Theta_i = \int_{T_{i-1}}^{T_i} \theta_t dt$ is tied to the quadratic variation $[\theta, \theta]_t$ of the spot process θ_t. This is most simply obtained when θ_t is continuous; for this case we show in Proposition 4 in Appendix A that if $\Delta T = T_{i+1} - T_i$ is independent of i, then

$$(\Delta T)^{-2} \sum_i (\Theta_{i+1} - \Theta_i)^2 \xrightarrow{p} \frac{2}{3}[\theta, \theta]_{\mathcal{T}} \text{ as } \Delta T \to 0. \tag{6}$$

Here, $[\theta, \theta]_{\mathcal{T}}$ is the total quadratic variation of θ_t over whole interval from 0 to $\mathcal{T}$.[11]

It follows that the quadratic variation term in (5) is, to first order, only tied to the underlying process:

$$\sum_i (\hat{\Theta}_{i+1} - \hat{\Theta}_i)^2 = \left(2\sum_i \text{AVAR}(\hat{\Theta}_i - \Theta_i) + \frac{2}{3}(\Delta T)^2 [\theta, \theta]_{\mathcal{T}} \right)(1 + o_p(1)). \tag{7}$$

The discontinuous θ case is more complicated (Remark 8 in the same appendix), and this forms part of the motivation for the development starting in Section 3. For preliminary illustration, we assume that (6)-(7) is valid.

---

[9] We call $\sum_i (\hat{\Theta}_{i+1} - \hat{\Theta}_i)^2$ the *apparent quadratic variation of* Θ_t *from* 0 *to* $\mathcal{T}$. This is in analogy to the use of *apparent misclassification rate*, see, e.g., Efron and Tibshirani (1991).

[10] See Footnote 1 in the Introduction for the normalization of AVAR. - The statement (5) involves having a small edge effect. We return to this in Sections 4-5.

[11] For notational convenience, we take $[\theta, \theta]_0 = 0$. The development in the paper goes through also for the case when the θ_t process does not start at zero, in which case $[\theta, \theta]_t$ should be replaced by $[\theta, \theta]_t - [\theta, \theta]_0$ for $0 \le t \le \mathcal{T}$.

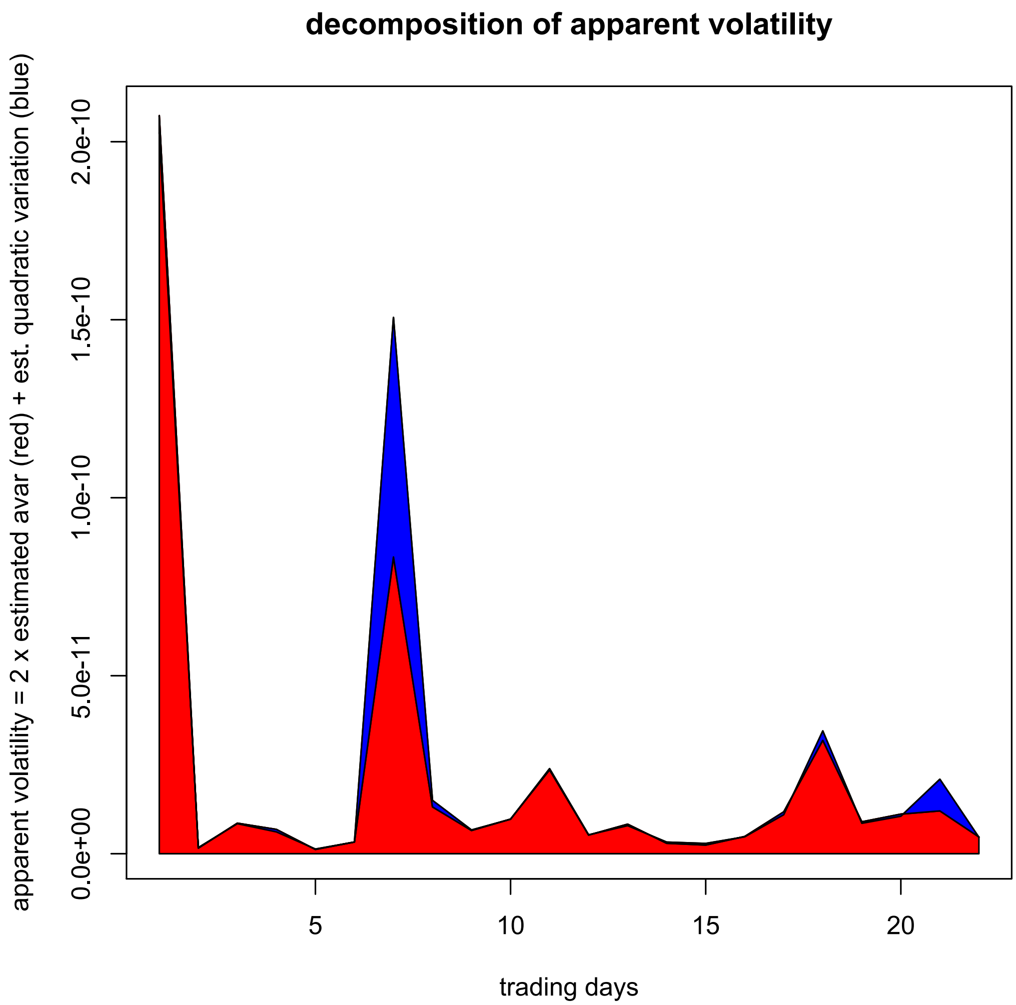

Figure 1: This plot shows the decomposition (5) and (7) in practice, for the S&P E-mini future as traded on the Chicago Mercantile Exchange, for the 22 trading days of May 2007. The total curve is the apparent volatility for each day, the red part is $2 \times \widehat{AVAR}$ for each day, and the blue part is $\frac{2}{3}\Delta T^2 \widehat{[\theta, \theta]}_{\mathcal{T}}$, as given in this Section. The data is preaveraged to fifteen seconds, a (1, 2) TSRV is computed on this basis for each five minute period, using the forward half interval method in Section 4.3. The square root of AVAR is thus of the standard error of the daily TSRV estimators of integrated volatility $\int \sigma_t^2 dt$, and the quadratic variation is the dispersion of the spot volatility: $\theta_t = \sigma_t^2$ and $[\theta, \theta]_{\mathcal{T}} = [\sigma^2, \sigma^2]_{\mathcal{T}}$. The estimation method has low enough edge effect that the "small edge" result in Section 4.2 applies. See also Example 3 in Section 7.

To get a sense of how (5) and (7) play out in real data, we plot the separation of cumulative AVAR and $[\theta,\theta]_{\mathcal{T}}$ using one month of tick-by-tick data from E-mini S&P 500 futures. As shown in Figure 1, cumulative AVAR is the main component in the apparent quadratic variation of Θ, in the meantime we could identify the days that the dispersion $[\theta,\theta]_{\mathcal{T}}$ of the underlying spot parameter moved notably in May 2007.

2.2 Three Applications: the Selection of Tuning Parameters, the Estimation of the Observed AVAR, and the Estimation of the Quadratic Variation of θ

We here give a sketch of the kinds of objects that can be estimated on the basis of (7). The estimators in this section are not final, but intended to convey the spirit of our investigation.

2.2.1 Cumulative Asymptotic Variance

The finding (6) suggests the construction of a two scales estimator in time periods i. From (7) it is easy to see that by taking every two-period interval $(T_i, T_{i+2}]$, and then averaging suitably, we obtain

$$\begin{aligned}\frac{1}{2}\sum_i(\hat{\Theta}_{(T_i,T_{i+2}]}-\hat{\Theta}_{(T_{i-2},T_i]}))^2 &\approx \sum_i \text{AVAR}(\hat{\Theta}_{(T_i,T_{i+2}]}-\Theta_{(T_i,T_{i+2}]})+\frac{2}{3}(2\Delta T)^2[\theta,\theta]_{\mathcal{T}}\\ &\approx 2\sum_i \text{AVAR}(\hat{\Theta}_i-\Theta_i)+\frac{2}{3}(2\Delta T)^2[\theta,\theta]_{\mathcal{T}}, \end{aligned} \tag{8}$$

where $\hat{\Theta}_{(T_{i-2},T_i]}=\hat{\Theta}_{i-1}+\hat{\Theta}_i$ and $\Theta_{(T_{i-2},T_i]}=\Theta_{i-1}+\Theta_i$. Hence we can cancel the $[\theta,\theta]$ term by setting a two-scales estimator for the cumulative asymptotic variance

$$\begin{aligned} TSAVAR &= \frac{2}{3}\sum_i(\hat{\Theta}_{(T_i,T_{i+1}]}-\hat{\Theta}_{(T_{i-1},T_i]})^2-\frac{1}{12}\sum_i(\hat{\Theta}_{(T_i,T_{i+2}]}-\hat{\Theta}_{(T_{i-2},T_i]})^2\\ &= \left(\sum_i \text{AVAR}(\hat{\Theta}_i-\Theta_i)\right)(1+o_p(1)). \end{aligned} \tag{9}$$

The simplest case of this is given in Theorem 2 in Section 4.2. This estimator can be be adapted to jumps in θ_t, and adjusted for edge effects, as we shall see in subsequent sections.[12]

Apart from estimating the asymptotic variance, this estimator is stable under changes in ΔT, and can thus also be used for model selection when ΔT is involved as tuning parameter.

---

[12]This estimator is based on different considerations than the TSRV from Zhang, Mykland, and Aït-Sahalia (2005), while the estimator in Section 2.3 is based on mostly the same considerations. We therefore concentrate on this estimator in the sequel. Also note that if ΔT can be taken to be small enough (cf. the conditions of Theorem 2), one can simply use a one scale estimator for AVAR by ignoring the $[\theta,\theta]_{\mathcal{T}}$ term, cf. Section 9.

2.2.2 Selection of Tuning parameters

Many estimators involve one or more tuning parameters, for example block or subgrid size. The typical situation is that of a tradeoff between two asymptotic variances. This is unlike the more typical situation in statistics, where the bias-variance tradeoff dominates. Variance-variance tradeoff is explicitly carried out in connection with the estimation of integrated volatility in Zhang, Mykland, and Aït-Sahalia (2005); Zhang (2006); Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008); Podolskij and Vetter (2009b,a); Aït-Sahalia, Mykland, and Zhang (2011); Jacod, Li, Mykland, Podolskij, and Vetter (2009a); Jacod and Mykland (2013). The typical question is how many grids to subsample over, or how long a time window to average data over, or how many autocovariances to include. – In a twist of this problem, the adaptive method of Jacod and Mykland (2013) does do local model selection, but there is still a global tuning parameter which is left to be determined. – Similar tuning involving a variance-variance tradeoff occurs in connection with covariance estimation (Zhang (2011); Bibinger and Mykland (2013)), spot volatility estimation (see Mykland and Zhang (2008)), estimation of the leverage effect (Wang and Mykland (2014), Aït-Sahalia, Fan, Wang, and Yang (2013)), estimation of the volatility of volatility (Vetter (2011), Mykland, Shephard, and Sheppard (2012a)). These and other inference situations requiring tuning are described in Section 7.

One can think of the tuning problem as involving a parameter c on which the estimators $\hat{\Theta}_i$ depend. If the choice of c leaves ΔT unchanged, then one can simply chose c to minimize $\sum_i (\hat{\Theta}_{i+1} - \hat{\Theta}_i)^2$. One does not even have to estimate AVAR. The same principle applies to the more complex form of observed AVAR that we shall introduce in the following. See Section 6.5.

2.3 Estimating the Quadratic Variation of θ

We proceed as in Section 2.2.1, but combine (7)-(8) with different weights so as to cancel the asymptotic variances. This gives a two scales estimator of the form

$$\begin{aligned}\widehat{[\theta,\theta]}_{\mathcal{T}} &= (\Delta T)^{-2}\left(\frac{1}{4}\sum_i (\hat{\Theta}_{(T_i,T_{i+2}]} - \hat{\Theta}_{(T_{i-2},T_i]})^2 - \frac{1}{2}\sum_i (\hat{\Theta}_{(T_i,T_{i+1}]} - \hat{\Theta}_{(T_{i-1},T_i]})^2\right)\\ &\xrightarrow{p} [\theta,\theta]_{\mathcal{T}}. \end{aligned} \tag{10}$$

where $\hat{\Theta}_{(T_{i-2},T_i]}$ and $\Theta_{(T_{i-2},T_i]}$ are as in the earlier section. This simple case is stated formally in Theorem 2 in Section 4.2. The more realistic case of edge effect, and of jumps in θ_t is discussed throughout the rest of the paper. – We emphasize that this estimator is different from the ones in Vetter (2011) and Mykland, Shephard, and Sheppard (2012b), which are both focused on data that do not have microstructure.

2.4 Implementation via High Frequency Inference

There are, potentially, two ways of using the decompositions (5) and (7). For these approximations to hold, we need B to be large, which can be achieved by either letting $\Delta T \to 0$, or by keeping ΔT, and sending $\mathcal{T}$ to infinity. This creates two scenarios, high and low frequency, and in this paper we focus on the former.[13]

In our high frequency scenario, $\mathcal{T}$ is a fixed time period, such as a day, and one estimates $\Theta = \int_0^{\mathcal{T}} \theta_t dt$ with the help of $\hat{\Theta} = \hat{\Theta}_1 + \cdots + \hat{\Theta}_B$. The $\hat{\Theta}_i$ can be intra-five-minutes or intra-half-hour estimators. In this case, we naturally suppose that $\Delta T \to 0$.

The main object of interest is $\text{AVAR}(\hat{\Theta} - \Theta)$, that is, the asymptotic variance of $\hat{\Theta}$ over a longer time period $[0, \mathcal{T}]$. One needs this quantity to approximately coincide with the cumulative AVAR from the ΔT-sized segments:

$$\text{AVAR}(\hat{\Theta} - \Theta) = \left(\sum_i \text{AVAR}(\hat{\Theta}_i - \Theta_i) \right) (1 + o_p(1)). \tag{11}$$

The big question in this case is how to implement (11), or something similar. The solution will turn out to depend on the solution to one of two challenges.

2.5 The Two Challenges

There are two main provisos to the presentation above. One is that the spot process θ_t need not be continuous, in which case (6) will fail. The other is that estimators typically have *edge effects*, which can lead the additivity (11) to fail.

As far as θ_t is concerned, we shall see in Section 3 and Appendix A that, on the one hand, the convergence (6) will fail when θ_t is not continuous. On the other hand, we present a powerful result: If we use subsampling and averaging, then a result akin to (6) will hold when θ_t is a *general semimartingale.* In other words, it can have arbitrary jumps, and also other unusual behavior, such as being multi-fractal, or more generally, not an Itô process. This result is the Integral-to-Spot Device, in Theorem 1 in the next section.

Edge effects are a common phenomenon in high frequency estimators, as documented by examples in Section 7. We provide a broad model for such effects in Sections 4-5. It turns out that edge effects are a form of meta-microstructure. There are several candidates for how to handle this meta effect: we have turned to subsampling and averaging since this approach cleanly solved our problems with θ_t.

[13] In terms of the Integral-to-Spot Device (Section 3 and Appendix A), the proof in Equations (A.5)-(A.7) also goes through in the low frequency case, except at the edge. The latter discrepancy would disappear under standard normalization by $\mathcal{T}$. The item that would require extra handling is the much more innocuous looking approximation in Lemma 2.

In other words, subsampling and averaging creates a unified theory that copes with both the edge effects, and the possible discontinuity of θ_t. This also lets us go on to create multi-scale estimators, as in Section 6.1.

3 The Integral-to-Spot Device: A General Result for the Quadratic Variation of Integrals of Semimartingales

The convergence (6) will typically fail when θ_t is not continuous. This is documented in Remark 8 in Appendix A. To carry through (6) in a general setup, we build on the techniques of subsampling and averaging. We extend our notation as follows: for any $S, T \in [0, \mathcal{T}]$, we set

$$\Theta_{(S,T]} = \int_S^T \theta_t dt. \tag{12}$$

Define the K-Averaged Quadratic Variation of Θ by

$$QV_K(\Theta) = \frac{1}{K} \sum_{i=K}^{B-K} (\Theta_{(T_i,T_{i+K}]} - \Theta_{(T_{i-K},T_i]})^2. \tag{13}$$

A main result is then the following, with proof in Appendix A.

THEOREM 1. (THE INTEGRAL-TO-SPOT DEVICE.) *Assume that θ_t is a semimartingale on $[0, \mathcal{T}]$. Set $\Delta T = \mathcal{T}/B$, and assume that $T_i = i\Delta T$. Suppose that $K \to \infty$ and $K\Delta T \to 0$ at the same time. Then*

$$\frac{1}{(K\Delta T)^2} QV_K(\Theta) \xrightarrow{p} \frac{2}{3}[\theta, \theta]_{\mathcal{T}-} \tag{14}$$

where $[\theta, \theta]_{\mathcal{T}-} = \lim_{t \uparrow \mathcal{T}} [\theta, \theta]_t$.

REMARK 1. At the cost of more notation, the result can be generalized to a QV_K process in t. The elements for this are all in Appendix A. – We note that unless one looks into the observations beyond $\mathcal{T}$, one cannot capture the contribution of any jump $\Delta\theta_{\mathcal{T}}$ at the ultimate time $\mathcal{T}$. The possibility of microstructure will also complicate the detection of a jump at the end time $\mathcal{T}$. □

REMARK 2. The result is conjectured to have implications for the consistency of pre-averaging estimators of volatility (Jacod, Li, Mykland, Podolskij, and Vetter (2009b); Podolskij and Vetter (2009b)). Our assumptions are weaker than those required in earlier papers on pre-averaging. In the former paper, for example, one has to think of their X as being similar to our current θ. Since this issue is not the focus of this paper, we have not pursued this matter here. □

4 Intra Day Estimators in High Frequency Data

4.1 Abstract Description of Standard Asymptotics

We continue to use the notation (12) and we suppose that we have at hand estimators $\hat{\Theta}_{(S,T]}$ of $\Theta_{(S,T]}$.

The typical situation is now as follows: there is a semi-martingale M_T and *edge effects* e_S and $\tilde{e}_T$, so that, for $S < T$,

$$\hat{\Theta}_{(S,T]} - \Theta_{(S,T]} = M_T - M_S + \tilde{e}_T - e_S. \tag{15}$$

The edge effect is essentially anything that messes up the martingaleness of the difference $\hat{\Theta}_{(0,T]} - \Theta_{(0,T]}$, and it occurs in many shapes, which we shall document in Section 7.[14] The edge effect has a component e_S relating to phasing in the estimator at the beginning of the time interval, and component $\tilde{e}_T$ for the phasing out at T. For the estimator on the whole interval, we use $\hat{\Theta} = \hat{\Theta}_{(0,\mathcal{T}]}$ from now on. An important construction leading to (15) relates to half-interval estimators (Section 4.3).

Remark 3. (Edge Effect for Bipower Variation) To rephrase, the Edge Effect is the difference in behavior of an estimator between the middle and the edges of the interval on which it is defined. For a preliminary illustration, consider the bi-power estimator (Barndorff-Nielsen and Shephard (2004, 2006)) of the integrated volatility of a process X_t, where X_t is observed (without microstructure) at equidistant times $t_i,\ i = 0, \cdots, n$, spanning $[0, \mathcal{T}]$. The estimator has the form $\hat{\Theta}_{(S,T]} = \frac{\pi}{2} \sum_{S<t_{i-1}\le t_i \le T} |\Delta X_{t_{i-1}}||\Delta X_{t_i}|$. Each absolute return $|\Delta X_{t_i}|$ appears twice in the summation, except the first and the last such return. This is a case of edge effect. The precise form of this effect is given in Equation (68) in Section 7, along with a number of other examples. □

A substantial fraction of the high frequency literature has studied the behavior of $\hat{\Theta}_{(S,T]} - \Theta_{(S,T]}$ for all $S < T \in [0, \mathcal{T}]$, or at least for $S = 0$ and all $T \in [0, \mathcal{T}]$. This is typically required to achieve *stable convergence*.[15]

[14] All of $\hat{\Theta}_{(S,T]}$, M_T, e_S, and $\tilde{e}_T$ will depend on the number of observations n. For the most part, n is omitted from our notation to avoid an excessive number of subscripts, but when crucial for understanding we may write $M_{n,T}$, etc. Equation (15) need only hold for S, T of the form T_i. Normally, these quantities will, in fact, be well defined for all n observation times. One can extend M_t to all $t \in [0, \mathcal{T}]$ as in Section 4.3 below. See also Example 1 in Section 7 for an example of "proactive" interpolation.

[15] Let Z_n be a sequence of random variables. We say that Z_n converges stably in law to Z as $n \to \infty$ if Z is measurable with respect to an extension of $\mathcal{F}$ so that for all $A \in F$ and for all bounded continuous g, $EI_A g(Z_n) \to EI_A g(Z)$ as $n \to \infty$. For background, see 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 $\mathcal{F}$ (Mykland and Zhang (2009, Proposition 1, p. 1408)). – Note that Z_n need not be $\mathcal{F}$-measurable, cf., *inter alia*, Example 4 in Section 7. With this convention, we suppress the need to distinguish between stable and conditional convergence.

The standard asymptotic result in the literature is as follows.[16] We here make this our starting point.

ASSUMPTION 1. (STANDARD CONVERGENCE SITUATION.) *Assume that θ_t is a semimartingale on $[0,\mathcal{T}]$. Assume (15) holds. Stable convergence is defined with reference to a sigma-field $\mathcal{F}$, which contains the information in the underlying processes,[17] including X_t and θ_t, but not necessarily any microstructure noise.[18] There is a convergence rate n^α, a sequence of local martingales $L_{n,t}$ and a finite variation processes $A_{n,t}$, so that*

$$\begin{aligned} n^\alpha M_{n,t} &= L_{n,t} + A_{n,t} \ , \\ TV(A_n - A)_{\mathcal{T}} &\xrightarrow{p} 0 \ , \\ L_{n,t} &\xrightarrow{\mathcal{L}} L_t \text{ stably in law, while} \\ n^\alpha e_{n,S} &\xrightarrow{\mathcal{L}} R_S \text{ and } n^\alpha \tilde{e}_{n,T} \xrightarrow{\mathcal{L}} \tilde{R}_T \text{ stably in law, jointly with } L_{n,t}, \text{ for any } S \text{ and } T \end{aligned} \tag{16}$$

where L_t, A_t, R_S and $\tilde{R}_T$ are limiting quantities. $TV(A_n)_{\mathcal{T}}$ is the total variation of $A_{n,t}$ on $[0,\mathcal{T}]$. L_t is a nonvanishing local martingale, A_t is a continuous process of finite variation,[19] with $L_0 = A_0 = 0$; $(R_T, \tilde{R}_T)$ are conditionally independent given $\mathcal{F}$ for different Ts, and conditionally independent of the process (L_t), also given $\mathcal{F}$.[20] Furthermore,

$$\sup_n E \sup_{0\le t\le \mathcal{T}} |\Delta L_{n,t}| < \infty. \tag{17}$$

□

We recall the basic facts about this situation. For proof and more discussion, see Appendix D.1.

PROPOSITION 1. (QUADRATIC VARIATION AND ASYMPTOTIC VARIANCE.) *Consider a grid $0 = T_0 < T_1 < \cdots < T_B = \mathcal{T}$, and let $\max_i(T_i - T_{i-1}) \to 0$. Under Assumption 1, $n^{2\alpha}\sum_i (M_{n,T_i} - M_{n,T_{i-1}})^2 \xrightarrow{\mathcal{L}} [L,L]_{\mathcal{T}}$ as $n \to \infty$. If L_t is a square integrable martingale conditionally on $\mathcal{F}$, and if $[L,L]_{\mathcal{T}}$ is $\mathcal{F}$-measurable, then*

$$\operatorname{Var}(L_{\mathcal{T}}|\mathcal{F}) \ = \ [L,L]_{\mathcal{T}}. \tag{18}$$

Finally, also suppose that R_0 and $\tilde{R}_{\mathcal{T}}$ have mean zero and finite second moment. Then the asymptotic variance[21] of $\hat{\Theta} - \Theta$ is

$$\operatorname{AVAR}(\hat{\Theta} - \Theta) = n^{-2\alpha}\left([L,L]_{\mathcal{T}} + \operatorname{Var}(R_0) + \operatorname{Var}(\tilde{R}_{\mathcal{T}})\right) + o_p(n^{-2\alpha}). \tag{19}$$

[16] General conditions for this to be true can be found in Hall and Heyde (1980) and Jacod and Shiryaev (2003). This kind of result has also been found in countless articles in specific situation, including those of most researchers in high frequency data.

[17] *I.e.*, with respect to which these will be measurable.

[18] For examples of precise formulations involving partial measurability, see Example 4 is Section 7, and also Zhang, Mykland, and Aït-Sahalia (2005), Zhang (2006), Jacod, Li, Mykland, Podolskij, and Vetter (2009b), and Podolskij and Vetter (2009b).

[19] Often identically zero, but sometimes not, cf. Examples 5 and 7 in Section 7.

[20] By convention, we set $R_{\mathcal{T}} = \tilde{R}_0 = 0$.

[21] See Footnote 1 in the Introduction on normalization of AVAR.

Remark 4. (Asymptotic Normality.) Results like Proposition 1 are most useful when the process L_t is a conditionally Gaussian given $\mathcal{F}$. The proposition then defines the asymptotic distribution of $n^\alpha(\hat{\Theta}-\Theta)$. The $\mathcal{F}$-measurability of $[L,L]_\mathcal{T}$ assures that, *in principle,* this asymptotic quadratic variation can be consistently estimated by data as $n \to \infty$. The current paper, of course, is about how to implement such estimation in practice.

In the conditionally Gaussian case, if there are no jumps to affect the asymptotic distribution, one will normally find that $L_t = \int_0^t f_s dB_s$, where B_t is a Brownian motion independent of the underlying data $\mathcal{F}$, while f is measurable with respect to $\mathcal{F}$, *i.e.,* f can be consistently estimated from the data (*e.g.,* Mykland and Zhang (2012, Theorem 2.28, p. 152)). More generally, conditional Gaussianity can also occur with jumps, as in Examples 5 and 8 in Section 7, and the references mentioned there. See also Jacod and Protter (1998, Section 6, pp. 296-306). □

Remark 5. (AVAR *vs.* AMSE.) There are situations of interest when Assumption 1 is satisfied, but the additional conditions of Proposition 1 are not. Most notably, consider the situation where $[L,L]_\mathcal{T}$ is not $\mathcal{F}$-measurable but instead just integrable. We otherwise continue to suppose that Assumption 1 holds, and that R_0 and $\tilde{R}_\mathcal{T}$ have mean zero and finite second moment. In this case, (19) needs to be replaced by

$$\text{AMSE}(\hat{\Theta}-\Theta) = n^{-2\alpha}\left([L,L]_\mathcal{T} + \text{Var}(R_0) + \text{Var}(\tilde{R}_\mathcal{T})\right) + o_p(n^{-2\alpha}), \tag{20}$$

where AMSE is the asymptotic mean squared error. This situation arises, for example, in the case of endogenous sampling times (Example 1 in Section 7). The same phenomenon occurs under direct estimation of skewness (Kinnebrock and Podolskij (2008, Example 6), Mykland and Zhang (2009, Example 3, p. 1414-1416)). □

4.2 Observed AVAR when the Edge Effect is Small

The simplest case is that of negligible edge effect. This is sufficiently intuitive that we discuss it first. The results promised in Section 2.2 all hold without modification, and we extend them to a more general situation, involving subsampling and averaging.

We consider a grid $0 = T_0 < T_1 < \cdots < T_{B_n} = \mathcal{T}$. We call B_n the total number of basic blocks, and we let K denote the sampling scale over the basic blocks. In analogy with (13), we define the K-averaged apparent quadratic variation of Θ as

$$QV_K(\hat{\Theta}) = \frac{1}{K}\sum_{i=K}^{B_n-K}(\hat{\Theta}_{(T_i,T_{i+K}]} - \hat{\Theta}_{(T_{i-K},T_i]})^2 \tag{21}$$

This generalizes (5) and (7) to K-averaged subsampling and averaging. For comparison, in Section 2.2, we considered sampling scales $K=1$ and $K=2$, as well as the notation $\hat{\Theta}_i = \hat{\Theta}_{(T_{i-1},T_i]}$.

We can then consistently estimate both the asymptotic variance of $\hat{\Theta}$, and the quadratic variation of parameter process θ. We obtain easily that

THEOREM 2. (NEGLIGIBLE EDGE EFFECTS) *Suppose that Assumption 1 holds, with the additional condition that* $\sum_i e^2_{T_i} = o_p(n^{-2\alpha})$ *and* $\sum_i \tilde{e}^2_{T_i} = o_p(n^{-2\alpha})$. *Also assume that* $\max_i(T_{n,i+1} - T_{n,i}) \to 0$. *Let* K, K_1, *and* K_2 *be fixed positive integers,* $K_1 < K_2$. *Then*

$$\frac{1}{K}\sum_i (\hat{\Theta}_{(T_i,T_{i+K}]} - \Theta_{(T_i,T_{i+K}]})^2 = n^{-2\alpha}[L,L]_{\mathcal{T}} + o_p(n^{-2\alpha}). \tag{22}$$

Under the additional assumptions of Proposition 1, the expression in equation (22) equals

$$\frac{1}{K}\sum_i \text{AVAR } (\hat{\Theta}_{(T_i,T_{i+K}]} - \Theta_{(T_i,T_{i+K}]}) = \text{AVAR}(\hat{\Theta} - \Theta) + o_p(n^{-2\alpha}). \tag{23}$$

Also, if θ_t *is a continuous semimartingale and the* T_i*s are equidistant, with* $\Delta T = O(n^{-\alpha})$, *the following results hold, respectively generalizing (7), (9) and (10) in Section 2:*

$$QV_K(\hat{\Theta}) = \frac{2}{3}(K\Delta T)^2[\theta,\theta]_{\mathcal{T}-} + 2n^{-2\alpha}[L,L]_{\mathcal{T}} + o_p(n^{-2\alpha}) \ ,$$

$$TSAVAR = \frac{1}{2}\left(\frac{1}{K_1^2} - \frac{1}{K_2^2}\right)^{-1}\left(\frac{1}{K_1^2}QV_{K_1}(\hat{\Theta}) - \frac{1}{K_2^2}QV_{K_2}(\hat{\Theta})\right) \tag{24}$$

consistently estimates $AVAR(\hat{\Theta} - \Theta)$, *the variation in the estimation error. And provided* ΔT *and* $O(n^{-\alpha})$ *are of the same order, another two scale estimator*

$$\widehat{[\theta,\theta]}_{\mathcal{T}} = \frac{3}{2}(K_2^2 - K_1^2)^{-1}(\Delta T)^{-2}\left(QV_{K_2}(\hat{\Theta}) - QV_{K_1}(\hat{\Theta})\right)$$

consistently estimates $[\theta,\theta]_{\mathcal{T}}$, *the variation in the parameter process.*

The above-mentioned two scale constructions effectively separate the impact of the underlying parameter behavior (as in the quadratic variation of parameter Θ) from estimator behavior (as in $\hat{\Theta} - \Theta$). However, as we shall see in the examples in Section 7, many estimators inherit edge effect of relatively large magnitude and thus require additional treatment. The main instances of small edge effect is that of realized volatility, and bipower variation in the absence of microstructure noise, cf. Examples 1-2 in the examples section. See also Example 3 on a combination of preaveraging and TSRV. Even in these cases, though, one would typically need to go to subsampling and averaging with larger sampling scales K if the spot volatility process can have jumps. See the examples for further discussion. We thus move on to subsampling and averaging with larger sampling scale K, which can deal with much larger edge effects.

4.3 Implementation with Half-interval Estimators

A simple way of obtaining estimators $\hat{\Theta}_{(S,T]}$ is to take half-interval estimators $\hat{\Theta}_{(0,T]}$ and $\hat{\Theta}_{(T,\mathcal{T}]}$ as given, and write, for $S < T$,

$$\begin{aligned} \text{“Forward” estimator: } \hat{\Theta}^{(f)}_{(S,T]} &= \hat{\Theta}_{(0,T]} - \hat{\Theta}_{(0,S]} \text{ and} \\ \text{“Backward” estimator: } \hat{\Theta}^{(b)}_{(S,T]} &= \hat{\Theta}_{(S,\mathcal{T}]} - \hat{\Theta}_{(T,\mathcal{T}]} \end{aligned} \tag{25}$$

with the superscripts "f" for forward and "b" for backward. If we suppose that (15) holds for the two half-interval estimators, then (15) also holds for the estimators in (25), with, in obvious notation,

$$\begin{aligned} \text{Edge effects for forward estimator: } & e_S^{(f)} = \tilde{e}_S \text{ and } \tilde{e}_T^{(f)} = \tilde{e}_T \text{ for } S \in (0, \mathcal{T}] \cap \mathcal{G} \text{ and } T \in [0, \mathcal{T}] \cap \mathcal{G} \\ \text{Edge effects for backward estimator: } & e_S^{(b)} = e_S \text{ and } \tilde{e}_T^{(b)} = e_T \text{ for } S \in [0, \mathcal{T}] \cap \mathcal{G} \text{ and } T \in [0, \mathcal{T}) \cap \mathcal{G} \\ \text{with special cases: } & e_0^{(f)} = e_0 \text{ and } \tilde{e}_{\mathcal{T}}^{(b)} = \tilde{e}_{\mathcal{T}}. \end{aligned} \tag{26}$$

where $\mathcal{G} = \{t_i\}$ are the observation points for data, or, more generally, the points at which the estimator is naturally defined. This suggests that the forward estimator only retains the phasing-out edge effect while the backward estimator is solely subject to the phasing-in edge. One can thus choose half-interval estimators to deal with one side – either the beginning or end part – of the edge effect, instead of two sides.

This eases implementation: the half-interval estimators can easily extend to be defined for all T, by taking a previous/subsequent tick definition

$$\hat{\Theta}^{(f)}_{(0,T]} \triangleq \hat{\Theta}^{(f)}_{(0,T_*]} \text{ and } \hat{\Theta}^{(b)}_{(S,\mathcal{T}]} \triangleq \hat{\Theta}^{(b)}_{(S^*,\mathcal{T}]} \tag{27}$$

where $T_* = \max\{\, t_i \leq T \,,\, t_i \in \mathcal{G}\}$ and $S^* = \max\{\, t_i \geq S \,,\, t_i \in \mathcal{G}\}$. The investigator is thus relieved of making sure that an estimator is defined on any particular interval $(S, T]$: this will always be true for the forward and backward estimator. One can thus take ΔT to be as small as is allowed by our results, cf. Section 9.1 below. The extension (27) has different edge effects, In particular, (26) is generalized by (28) below. The use of half-interval estimators also eases the analysis of the Meta Edge Effects $\mathcal{E}$ introduced in Section 5 (see Remark 6).

We now turn to edge effects for half-interval estimators. In the event that the grid $\{0 = T_0, T_1, \cdots, T_{B_n-1}, T_B = \mathcal{T}\}$ does not coincide with a subset of the grid of observation points t_j, we here give the form of the edge effect for interpolated (or, rather, previous/subsequent tick) values. Suppose (15) and (26) hold for $S, T \in \mathcal{G}$, and suppose that M_t is an interpolation for all t of the martingale M_{t_j}.[22] The edge effect for the forward estimator then gets the following form, generalizing (15) and (26):

$$\begin{aligned} \text{Edge effects for forward estimator: } & e_T^{(f)} = \tilde{e}_T^{(f)} = \tilde{e}_{T_*} + M_T - M_{T_*} + \int_{T_*}^{T} \theta_t dt \text{ for all } T \\ \text{Edge effects for backward estimator: } & e_S^{(b)} = \tilde{e}_S^{(b)} = e_{S_*} + M_{S^*} - M_S + \int_{S}^{S^*} \theta_t dt \text{ for all } S \\ \text{with special cases: } & e_0^{(f)} = e_0 \text{ and } \tilde{e}_{\mathcal{T}}^{(b)} = \tilde{e}_{\mathcal{T}} \end{aligned} \tag{28}$$

Similar extensions apply to other estimators than the half-interval type.

[22] see, *e.g.,* Heath (1977), Mykland (1995b), Mykland and Zhang (2006, 2012). One cannot simply take the martingale M_t to be constant (and càdlàg) on intervals between observations (or between the T_i), because of the presence of the process θ_t.

5 Hard Edge

One cannot always take the edge effect to be negligible. To cope with this, we observe that the situation (15) is rather similar to observing the martingale M_t with microstructure noise e_t and/or $\tilde{e}_t$. This suggests that the current literature on microstructure noise can provide methods for dealing with this problem. We emphasize that the situation is not wholly similar, since there is also the unknown term due to the integral of θ_t. Since our main Theorem 1 in Section 3 relies on a multi-scale construction, it is most convenient to also use this approach on the process (15), thus drawing on Zhang, Mykland, and Aït-Sahalia (2005) and Zhang (2006).[23]

The difficulty under edge effects is that the additivity (11) no longer holds. Take, for example $\hat{\Theta}^{(f)}_{(S,T]}$ from (25). Under Assumption 1, in analogy with Proposition 1,

$$\text{AVAR}(\hat{\Theta}^{(f)}_{(S,T]} - \Theta_{(S,T]}) = n^{-2\alpha}\left([L,L]_T - [L,L]_S) + \text{Var}(\tilde{R}_T) + \text{Var}(\tilde{R}_S)\right) + o_p(n^{-2\alpha}) \tag{29}$$

(unless S and T are very close). We cannot just add up the $\text{AVAR}(\hat{\Theta}_i - \Theta_i)$s and get an overall asymptotic variance for $\hat{\Theta} = \hat{\Theta}_{(0,\mathcal{T}]}$. In the high frequency scenario, the target we are looking for is the full asymptotic variance (19).

To remedy the situation, and also in response to the results in Section 3, we pursue a stronger subsampling and averaging, still using $QV_K(\hat{\Theta})$ from (21), but now letting $K \to \infty$. – We make the following set of assumptions.

Assumption 2. (Hard Edge Assumptions.) *Suppose that there is an integer J for which $e_{T_i} = e'_{T_i} + e''_{T_i}$ and $\tilde{e}_{T_i} = \tilde{e}'_{T_i} + \tilde{e}''_{T_i}$, where there is a filtration*[24] *$(\mathcal{G}_t)$ so that $(e'_{T_i}, \tilde{e}'_{T_i})$ are $\mathcal{G}_{T_{i+J}}$-measurable, and for which $E(e'_{T_i} \mid \mathcal{G}_{T_{i-J}}) = E(\tilde{e}'_{T_i} \mid \mathcal{G}_{T_{i-J}}) = 0$ and where $\sum_i (e''_{T_i})^2 = o_p(n^{-2\alpha})$ and $\sum_i (\tilde{e}''_{T_i})^2 = o_p(n^{-2\alpha})$. Also suppose*[25] *that for some $\beta \geq \alpha$*

$$\sup_n E\ n^{\beta}\left(\max_{0 \le i \le B_n} |e'_{n,T_i}| + \max|\tilde{e}'_{n,T_i}|\right) \ < \ \infty. \tag{30}$$

We finally assume that for $K_n \to \infty$, $K_n = o(B_n)$[26]

$$\frac{1}{K_n}\sum_{i=0}^{K_n-1} n^{2\alpha} u_{T_i} \xrightarrow{p} \mathcal{E}_u(0) \text{ and } \frac{1}{K_n}\sum_{i=B_n-K_n+1}^{B_n} n^{2\alpha} u_{T_i} \xrightarrow{p} \mathcal{E}_u(\mathcal{T}) \tag{31}$$

[23] It is quite possible that other methods for handling microstructure can be adapted to the current problem, but this is beyond the scope of this paper. One would then need to prove results akin to this paper's Theorem 1 (Section 3) and Proposition 5 in Appendix B.1. Since our main goal is to estimate asymptotic variance, we are looking for consistency, and thus the order of convergence is a lesser consideration in this context than in the classical estimation-of-volatility-under-microstructure problem.

[24] For motivation for this filtration, see Appendix C.1.

[25] The default assumption is that $\beta = \alpha$. For the rôle of β, see Section 9.1. See also Examples 2-3 in Section 7 for cases where $\beta > \alpha$.

[26] For connoisseurs: This condition can be modified to requiring that there is B'_n, $B'_n \to \infty$, $B'_n = O_p(B_n)$, so that when $K_n \to \infty$, $K_n \le 2B'_n$, then (31) holds. Equation (33) in Theorem (3) is then changed to "$J \le K_n \le B'_n$" etc

for all of $u = \tilde{e}^2$, $u = e^2$, and $u = (\tilde{e} + e)^2$. We also assume that R_0 and $\tilde{R}_{\mathcal{T}}$ have mean zero and finite second moment, and that

$$\mathcal{E}_{(e^{(b)})^2}(0) = \mathrm{Var}(R_0) \text{ and } \mathcal{E}_{(\tilde{e}^{(f)})^2}(\mathcal{T}) = \mathrm{Var}(\tilde{R}_{\mathcal{T}}), \tag{32}$$

where $e^{(b)}$ and $\tilde{e}^{(f)}$ are as given in (26).[27]

We shall see in our examples that the assumptions above are reasonable. As a complement, they are argued from a mixing perspective in Appendix C.1.

The following is our main result, which provides an alternative to Theorem 2 when the edge effects are not negligible. The subsequent results are, for the most part, corollaries. The proof is in Appendix B.

Theorem 3. (Representation of K-Averaged Apparent Quadratic Variation.) *Suppose that Assumption 1-2 holds. Let $B = B_n$ and $K = K_n$ be a sequence of integers so that*

$$J \leq K_n \leq B_n, \text{ with } K_n B_n^{-1} = O(n^{-\alpha}),\ K_n^{-1} B_n^{1/2} = O(n^{-\alpha+\beta}) \text{ and } K_n \to \infty \text{ as } n \to \infty. \tag{33}$$

Then

$$\begin{aligned} QV_K(\hat{\Theta}) =& \frac{2}{3}(K\Delta T)^2[\theta,\theta]_{\mathcal{T}-} \\ &+ 2n^{-2\alpha}[L,L]_{\mathcal{T}} + \text{ Meta Edge Effect} \\ &+ \frac{1}{K}V_0 + o_p(n^{-2\alpha}). \end{aligned} \tag{34}$$

where V_0 is given by (B.27), (B.66)-(B.67), and (B.76), and does not depend on the choice of K_n sequence. $V_0 = O_p(n^{-2\alpha}B_n)$. The "Meta Edge Effect" (MEE) is given by

$$MEE \ = -n^{-2\alpha}\left(2\mathcal{E}_{\tilde{e}^2}(0) + \mathcal{E}_{(\tilde{e}+e)^2}(0) + \mathcal{E}_{(\tilde{e}+e)^2}(\mathcal{T}) + 2\mathcal{E}_{e^2}(\mathcal{T})\right). \tag{35}$$

If instead K is bounded, $K \geq J$, then $QV_K = MEE \ + \ \frac{1}{K}V_0 + O_p(n^{-2\alpha})$.[28]

Remark 6. (Estimating the Meta Edge Effect.) If $\beta > \alpha$, the following adjustment does not need to be applied (Section 9.1). – With reference to (19), we recall that the $\mathrm{AVAR}(\hat{\Theta} - \Theta) = n^{-2\alpha}\left([L,L]_{\mathcal{T}} + \mathrm{Var}(R_0) + \mathrm{Var}(\tilde{R}_{\mathcal{T}})\right)(1 + o_p(1))$. For the purpose of estimating AVAR we have an interest in modifying the Meta Edge Effect from (35) to $2\left(\mathrm{Var}(R_0) + \mathrm{Var}(\tilde{R}_{\mathcal{T}})\right)$. In other words, we wish to estimate an adjustment of the MEE given by

$$\mathrm{AMEE} \ = \ n^{-2\alpha}\left(2\mathcal{E}_{\tilde{e}^2}(0) + \mathcal{E}_{(\tilde{e}+e)^2}(0) + \mathcal{E}_{(\tilde{e}+e)^2}(\mathcal{T}) + 2\mathcal{E}_{e^2}(\mathcal{T}) + 2\mathrm{Var}(R_0) + 2\mathrm{Var}(\tilde{R}_{\mathcal{T}})\right). \tag{36}$$

[27] This may seem a little odd, but relates to the fact that the edge effects can be moved around depending on whether one uses an estimator $\hat{\Theta}_{(S,T]}$ which is internally defined on the interval $(S,T]$, or alternatively a forward or backward estimator. For comparison, consult equations (26) and (28). The statement (32) covers all these possibilities.

[28] For the relationship to Theorem 2, see Section 9.1.

There are multiple ways of doing this, here is one. Set, for $t > 2$,

$$\widehat{\mathrm{AMEE}}_0 = \frac{1}{K(t-2)} \sum_{i=K}^{\lfloor Kt \rfloor} \left((\hat{\Theta}_{(T_i,T_{i+K}]} - \hat{\Theta}_{(T_{i-K},T_i]})^2 + \frac{1}{3}(\hat{\Theta}^{(f)}_{(T_i,T_{i+K}]} - \hat{\Theta}^{(f)}_{(T_{i-K},T_i]})^2 \right) ,$$

$$\widehat{\mathrm{AMEE}}_{\mathcal{T}} = \frac{1}{K(t-2)} \sum_{i=B-\lfloor Kt \rfloor}^{B-K} \left((\hat{\Theta}_{(T_i,T_{i+K}]} - \hat{\Theta}_{(T_{i-K},T_i]})^2 + \frac{1}{3}(\hat{\Theta}^{(b)}_{(T_i,T_{i+K}]} - \hat{\Theta}^{(b)}_{(T_{i-K},T_i]})^2 \right) \quad (37)$$

and

$$\widehat{\mathrm{AMEE}} = \widehat{\mathrm{AMEE}}_0 + \widehat{\mathrm{AMEE}}_{\mathcal{T}}, \quad (38)$$

where $\hat{\Theta}^{(f)}$ and $\hat{\Theta}^{(b)}$ are defined in Section 4.3. It is clear from (32) and the proof of Theorem 3 that,[29]

$$\widehat{\mathrm{AMEE}} = \mathrm{AMEE}(1 + o_p(1)). \quad (39)$$

Thus one can implement the required adjustment. The finer points of how to optimize t and K are left for another paper. Finally, note that for the forward estimator from Section 4.3, the first line of the expression (37) simplifies to

$$\widehat{\mathrm{AMEE}}_0 = \frac{1}{K(t-2)} \frac{4}{3} \sum_{i=K}^{\lfloor Kt \rfloor} \left(\hat{\Theta}^{(f)}_{(T_i,T_{i+K}]} - \hat{\Theta}^{(f)}_{(T_{i-K},T_i]} \right)^2, \quad (40)$$

and similarly for the backward estimator. □

6 Applications of Intra Day Observed AVAR

6.1 A Multiscale Estimator

We shall in the following harvest the applications of Theorem 3. First, however, note that a two- or multiscale construction (Zhang, Mykland, and Aït-Sahalia (2005), Zhang (2006)) is needed to cancel the really large term, V_0/K, which contains the sum of squares of the edge effects. We consider m scales,

$$J \leq K_1 = K_{n,1} < K_2 = K_{n,2} < \cdots < K_m = K_{n,m} \leq B_n. \quad (41)$$

For simplicity, take the number of scales m to be finite.[30] We consider a multi-scale estimator

$$MSQV = \sum_{l=1}^{m} \gamma_l QV_{K_l}(\hat{\Theta}) \quad (42)$$

[29] If one follows Footnote 26, one needs to require that $K_n \leq tB'_n$ in lieu of $K_n \leq 2B'_n$.

[30] To optimize the convergence rate of the estimator, one may need to follow Zhang (2006) and let $m \to \infty$ with n. However, since the purpose of this paper is to show consistency (which, to first order, is all that is needed for the asymptotic variance), we have deemed this additional complication to be beyond the scope of this paper.

where the $\gamma_l = \gamma_{n,l}$. We wish to to eliminate the V_0/K term. It is easy to see that this requires

$$\sum_{l=1}^{m} \frac{\gamma_l}{K_l} = 0. \tag{43}$$

For the general result, we also need the set $\mathcal{L} = \{l \in [1,m] : \ K_{n,l} \text{ is bounded as } n \to \infty\}$. We then have the following.

THEOREM 4. (MULTISCALE QUADRATIC VARIATION.) *Suppose that Assumptions 1-2 hold. Let m be given. Let $K_{n,l}$ $(l = 1, \cdots m)$ be as in (41) and (43), and suppose that*

$$K_{n,m} B_n^{-1} = O(n^{-\alpha}), \text{ and } K_{n,m} \to \infty \tag{44}$$

Assume that $\sum_{l\in\mathcal{L}} |\gamma_{n,l}| = o(1)$. If $\gamma_n = \sum_{l=1}^{m} \gamma_{n,l}$, then

$$\begin{aligned} MSQV = & \frac{2}{3} \sum_{l=1}^{m} \gamma_{n,l} K_{n,l}^2 (\Delta T)^2 [\theta,\theta]_{\mathcal{T}-} \\ & + \gamma_n \left\{ 2n^{-2\alpha} [L,L]_{\mathcal{T}} + \text{MEE} \right\} \\ & + o_p(\max_l |\gamma_{n,l}| n^{-2\alpha}) + O_p\left(n^{-2\beta} \mathfrak{E}_n^{1/2}\right) + O_p(\gamma_n n^{-\beta} B_n^{-1/2}), \end{aligned} \tag{45}$$

where the MEE is given by (35), and where

$$\mathfrak{E}_n = \sum_{l=1}^{m} \left(\frac{\gamma_{n,l}}{K_{n,l}} \right)^2 (B_n - 2K_{n,l} + 1) \tag{46}$$

The last two orders of error in (45) represent a dissolution and a refinement of the error term $o_p(n^{-2\alpha})$ in Theorem 3 into a variance and a bias term. We can do this since the bias term is from Proposition 5 while the variance terms in from Proposition 6, both in Appendix B. For the bias term, the derivation is in Remark 9 in Appendix B.1, which also has comments on the use of the word "bias". The variance term is handled in Appendix C.2, where we also argue that this in most cases is the exact order of the main error term (Remark 11 in the same Appendix).

This dissolution permits us to accept the things we cannot change, the ability to change the things we can, and the insight to know the difference.[31] It also shows that the choice of $\gamma_{n,l}$ does not influence the bias term, since both for the asymptotic variance and for the quadratic variation of θ_t, the γ_n is fixed, *cf.* equations (47) and (52).

It follows that to optimize the estimator, one would want to minimize $\mathfrak{E}_n$ subject to the relevant linear constraints, such as (43) and (below) either (47) or (52). (The hard constraints of the theorem also have to hold, of course.)

[31] To paraphrase the widely used "Serenity Prayer". One could, of course, seek to estimate what we here call the bias term, but this is beyond the scope of this paper.

6.2 Estimation of Asymptotic Variance

From Theorem 4, we can clearly estimate $\mathrm{AVAR}(\widehat{\Theta}_n)$ by further requiring

$$\sum_{l=1}^{m} \gamma_{n,l} = \frac{1}{2} \text{ and}$$
$$\sum_{l=1}^{m} \gamma_{n,l} K_{n,l}^2 = 0, \tag{47}$$

and by adjusting the edge effect, which is discussed in Remark 6. We thus get an estimator

$$\widehat{\mathrm{AVAR}}(\widehat{\Theta}_n) \;=\; MSQV + \frac{1}{2}\widehat{AMEE}. \tag{48}$$

Theorem 5. (Consistently Estimating the Asymptotic variance.) *Assume the conditions of Proposition 1 and Theorem 4, and also that (47) is satisfied. Also suppose that $\widehat{AMEE}$ is formed with the help of a K_n (which may or may not be a $K_{n,l}$, but which satisfies the conditions of Remark 6. Assume that* $\max_l |\gamma_l| = O(1)$, *and* $B_n^{-1} = o(n^{2\beta-4\alpha})$. *Also assume that* $\mathfrak{E}_n = o(n^{4(\beta-\alpha)})$. *Then*

$$\widehat{\mathrm{AVAR}}(\widehat{\Theta}_n) \;=\; \mathrm{AVAR}(\widehat{\Theta}_n)\,(1 + o_p(1)) \tag{49}$$

We note that the results follow from the previous Theorem and Remark since $\mathrm{AVAR}(\widehat{\Theta}_n) = O_p(n^{-2\alpha})$. If the conditions of Proposition 1 are omitted, then the above result remains valid with AMSE replacing AVAR (cf. Remark 5).

Because of its importance, we here state the main usage as a corollary to the above result:

Theorem 6. (Feasible Estimation.) *Assume the conditions of either Theorem 2 or Theorem 5, and also that $(L_{\mathcal{T}}, R_0, \tilde{R}_{\mathcal{T}})$ is conditionally Gaussian given $\mathcal{F}$. Suppose that $A_{\mathcal{T}} = 0$.*[32] *Then*

$$\frac{\widehat{\Theta}_n - \Theta}{\widehat{AVAR}_n^{1/2}} \xrightarrow{\mathcal{L}} N(0,1) \textit{ stably in law.} \tag{50}$$

Remark 7. (A Three Scales $\widehat{\mathrm{AVAR}}_n$.) There are at two ways of implementing these results. One is to use a three-scales estimator, $m = 3$. In this case, the three $\gamma_{n,l}$ are determined by the three linear equations (43) and (47), the solution being

$$\gamma_{n,1} = -\frac{1}{v_n} K_{n,1}(K_{n,3}^3 - K_{n,2}^3),$$
$$\gamma_{n,2} = \frac{1}{v_n} K_{n,2}(K_{n,3}^3 - K_{n,1}^3), \text{ and}$$
$$\gamma_{n,3} = -\frac{1}{v_n} K_{n,3}(K_{n,2}^3 - K_{n,1}^3), \text{ where}$$
$$v_n = 2(K_{n,1} + K_{n,2} + K_{n,3})(K_{n,2} - K_{n,1})(K_{n,3} - K_{n,1})(K_{n,3} - K_{n,2}). \tag{51}$$

[32] If this latter condition fails, the normal distribution on the r.h.s. will have as its mean the limit of $A_{\mathcal{T}}/n^{\alpha}\mathrm{AVAR}_n^{1/2}$.

In this case, it is easy to see that it is optimal to choose $K_{n,3}$ to be as large as possible, in particular satisfying (44). If one chooses $K_{n,1}/K_{n,3}$ and $K_{n,2}/K_{n,3}$ to be bounded away from each other, and from 0 and 1, then $\mathfrak{E}_n = O(B_n K_{n,3}^{-2}) = O(n^{2\alpha} B_n^{-1})$ since (44) is the exact order. Thus, since $B_n^{-1} = o(n^{2\beta - 4\alpha})$, $\mathfrak{E}_n = o(n^{2(\beta-\alpha)}) = o(n^{4(\beta-\alpha)})$ since $\beta \geq \alpha$. The condition on $\mathfrak{E}_n$ in Theorem 5 is thus satisfied. □

To get an even more efficient estimator for AVAR, we discuss how to proceed with a multi-scale construction ($m > 3$) in Section 6.4.

6.3 Estimating the Quadratic Variation of θ

We can similarly use Theorem 4 to estimate the quadratic variation $[\theta, \theta]_{\mathcal{T}-}$. The side conditions on weights $\gamma_{n,l}$ now become (instead of (47))

$$\sum_{l=1}^{m} \gamma_{n,l} = 0 \text{ and}$$
$$\sum_{l=1}^{m} \gamma_{n,l} K_{n,l}^2 = \frac{3}{2} (\Delta T)^{-2}. \tag{52}$$

There is also no need to worry about edge effect. Our estimator is therefore

$$\widehat{[\theta, \theta]}_{T}^{(n)} = MSQV. \tag{53}$$

which is the same form as (48), except that, of course, the requirement on the γs is different. There is also no need for $\widehat{AMEE}$.

Theorem 7. (Consistently Estimating the Quadratic Variation of θ.) *Assume the conditions of Theorem 4, with the order (44) being exact, and also that (52) is satisfied. Assume that* $\max_l |\gamma_l| = O(n^{2\alpha})$ *and* $\mathfrak{E}_n = o(n^{4(\beta-\alpha)})$. *Then*

$$\widehat{[\theta, \theta]}_{T}^{(n)} \xrightarrow{p} [\theta, \theta]_{\mathcal{T}-} \text{ as } n \to \infty. \tag{54}$$

Once again, the result follows from the previous Theorem 4.

6.4 Optimized Estimators of AVAR and of Quadratic Variation

We now consider the case of general number m of scales. Set

$$\mathbb{A}_n = \begin{pmatrix} K_{n,1}^{-1} & K_{n,2}^{-1} & \cdots & K_{n,m}^{-1} \\ 1 & 1 & \cdots & 1 \\ K_{n,1}^{2} & K_{n,2}^{2} & \cdots & K_{n,m}^{2} \end{pmatrix} \text{ and } \underline{\gamma}_n = \begin{pmatrix} \gamma_{n,1} \\ \cdots \\ \gamma_{n,m} \end{pmatrix} \tag{55}$$

Also let $\mathbb{C}_n = \text{diag}(K_{n,1}^{-2}(B_n - 2K_{n,1}+1), \cdots, K_{n,m}^{-2}(B_n - 2K_{n,m}+1))$. We note that $\mathfrak{E}_n = \underline{\gamma}_n^* \mathbb{C}_n \underline{\gamma}_n$, where "$*$" denotes transpose. Our two optimization problems thus become

$$\min \underline{\gamma}_n^* \mathbb{C}_n \underline{\gamma}_n \text{ subject to } \mathbb{A}_n \underline{\gamma}_n = \underline{b}_n \tag{56}$$

with standard solution (*e.g.,* Boyd and Vandenberghe (2004, p. 304))

$$\underline{\gamma}_n = \mathbb{C}_n^{-1} \mathbb{A}_n^* (\mathbb{A}_n \mathbb{C}_n^{-1} \mathbb{A}_n^*)^{-1} \underline{b}_n. \tag{57}$$

In view of the preceding sections,

- To estimate $\text{AVAR}(\widehat{\Theta}_n)$, and satisfy the conditions of Theorem 5, choose

$$\underline{b}_n = \left(0, \frac{1}{2}, 0\right)^*. \tag{58}$$

- To estimate the quadratic variation $[\theta, \theta]_{\mathcal{T}-}$, and satisfy the conditions of Theorem 7, choose

$$\underline{b}_n = \left(0, 0, \frac{3}{2}(\Delta T_n)^{-2}\right)^*. \tag{59}$$

6.5 Selection of Tuning Parameters

As we discussed in Section 2.2.2, there is a class of estimators for which an optimal choice of tuning parameters will minimize the asymptotic variance. In the earlier section, we have described the situation as a variance-variance tradeoff.

Assumption 3. *Suppose that there is a tuning parameter c (chosen by the econometrician) upon which $\hat{\Theta}_n = \hat{\Theta}_{n,c}$ and $\text{AVAR} = \text{AVAR}_c$ depends.*[33] *Assume (as provided by, say, Theorems 2 and 5) that*

$$\forall c \in \mathcal{C} : \widehat{\text{AVAR}}_{n,c} = \text{AVAR}_c(1 + o_p(1)) \text{ (for fixed) } c. \tag{60}$$

We seek $c^* = \arg\min_c \text{AVAR}_{c \in \mathcal{C}}$, which we for simplicity of discussion take to be unique. $\mathcal{C}$ is a set of values for the tuning parameters within which one wishes to optimize. For the following prima facie discussion, we also take the number of points in $\mathcal{C}$ to be finite.[34]

[33] Observe that Θ does not depend on c, but will normally be (statistically) mutually dependent with c^*. Also, for the purposes of this arguments, we assume that $n^{-2\alpha}\text{AVAR}$ is independent of n (cf. Footnote 1 in the Introduction).

[34] This case if of practical interest. See the example later in this section. In the more general case, one may imagine that there is a finite partition, say, $\mathcal{P}$ of the space of all c's, and that $\mathcal{C}$ has one representative of each element of $\mathcal{P}$. With a well chosen $\mathcal{P}$ and $\mathcal{C}$, this construction will normally achieve approximate optimality. – The consistency part below generalizes straightforwardly to more complex $\mathcal{C}$'s, under, say, uniform convergence conditions. The validity part is best left as a separate paper.

For given number of observations n, our estimate is accordingly $\hat{c}_n = \arg\min_{c\in\mathcal{C}} \widehat{\text{AVAR}}_{n,c}$, where $\widehat{\text{AVAR}}_{n,c}$ is obtained through our proposals in the preceeding sections.

Consistency. Under Assumption 3, automatically

$$\hat{c}_n \to c^*. \tag{61}$$

Validity. This procedure provides an estimator with asymptotic variance AVAR_{c^*}:

$$\text{asymptotic variance of } \hat{\Theta}_{n,\hat{c}_n} - \Theta \ = \ \text{AVAR}_{c^*}. \tag{62}$$

This is the conceptually more complex issue. Since AVAR_c is typically random, so will c^* be random. *A priori,* the insertion of $\hat{c}_n$ into an estimator might in principle create problems for the standard convergence setup discussed in Assumption 1. At least in our simple case, however, this difficulty does not arise. We embody this in a formal result.

Proposition 2. (Optimization commutes with Asymptotic Variance.) *Assume (all) the conditions of Proposition 1, as well as Assumption 3. Also suppose that c^* is $\mathcal{F}$-measurable and that, for each $c \in \mathcal{C}$, $(\hat{\Theta}_{n,c} - \Theta)/\text{AVAR}_c^{1/2}$ converges stably in law to a $N(0,1)$ random variable that is independent of $\mathcal{F}$. Then (62) holds, and also*

$$(\hat{\Theta}_{n,\hat{c}_n} - \Theta)/\text{AVAR}_{c^*}^{1/2} \xrightarrow{\mathcal{L}} N(0,1) \text{ and } (\hat{\Theta}_{n,\hat{c}_n} - \Theta)/\widehat{\text{AVAR}}_{n,\hat{c}_n}^{1/2} \xrightarrow{\mathcal{L}} N(0,1), \text{ both stably.} \tag{63}$$

Proof: With probability one, for n large enough, $n^\alpha(\hat{\Theta}_{n,\hat{c}_n} - \Theta) = \sum_{c\in\mathcal{C}} n^\alpha(\hat{\Theta}_{n,c} - \Theta) I_{\{c=c^*\}}$. We are thus rescued by the stable convergence. ∎

Implementation. In analogy with the discussion in Section 2.2.2, it is not actually necessary to estimate AVAR to find c^*, so long as the $[\theta,\theta]$ component remains stable in c. It is in fact enough to optimize with a single scale $QV_K(\hat{\Theta})$ from Theorems 2 or 3. For maximal efficiency, however, one can use a multiscale estimator from Theorem 4. This time, however, only two constraints are needed. The criterion $MSQV$ is obtained by minimizing $\mathfrak{E}_n$ subject to

$$\sum_{l=1}^{m} \frac{\gamma_l}{K_l} = 0 \text{ and } \sum_{l=1}^{m} \gamma_{n,l} = \frac{1}{2} \ . \tag{64}$$

If one also wishes to know the asymptotic variance, then, obviously, one needs the three conditions of Theorem 5 instead.

Example. Volatility estimation via preaveraging followed by a (J,K) TSRV estimator (Example 3 in Section 7), with J and K finite, provides an example where the action space $\mathcal{C}$ can indeed be taken to be finite. The assumptions of Proposition 2 are satisfied. □

6.6 Several Dimensions

The extension of this theory to several dimensions is straightforward, so long as the grid(s) $\{T_i,\ i = 0, \cdots, B\}$ are the same in all dimensions, or if they are nested.[35] A standard procedure is to use some version of the identity $xy = \frac{1}{2}\left((x+y)^2 - x^2 - y^2\right)$. For typical examples, see the definition of multivariate quadratic variation Jacod and Shiryaev (2003, Eq. (I.4.46), p. 52), or the extension of quasi-likelihood in Aït-Sahalia, Fan, and Xiu (2010).

Since it would be tedious to repreat the entire theory for this case, we state by way of example a generalization of Theorem 4. All the other results in the paper generalize similarly.

THEOREM 8. (MULTIVARIATE MULTISCALE QUADRATIC VARIATION.) *Suppose that θ_t is p-dimensional, for a fixed finite p, and similarly for $\widehat{\Theta}_{(S,T]}$, M_t, e_T, $\tilde{e}_T$, and so on. Quadratic variations and asymptotic variances are, similarly, $p \times p$ matrices. Suppose that the grid $\{T_i,\ i = 0, \cdots, B\}$ is the same in all dimensions, and that QV and $MSQV$ are formed with the same K_i and γ_i. Assume (44), and that that $\sum_{l \in \mathcal{L}} |\gamma_{n,l}| = o(1)$. Then*

(i) Suppose that Assumptions 1-2 are satisfied for the relevant vector and matrix processes. Then the matrix $MSQV$ satisfies the conclusions of Theorem 4.

(ii) Alternatively suppose that Assumptions 1-2 are satisfied marginally in each dimension, by p scalar processes. Then every subsequence of the matrix $MSQV$ has a further subsequence which satisfies (as a matrix) the conclusions of Theorem 4.

The theorem describes two scenarios under which the extension holds. The second scenario may at first sight seem arcane, but is actually the more comfortable case. One only needs to satisfy the conditions on scalar processes, and whatever subsequence one is on, one is headed towards the appropriate limit for that subsequence, thus, say, for the appropriate AVAR. This assures, for example, the stable convergence in law of $\widehat{\text{AVAR}}_n^{-1/2}(\hat{\Theta}_n - \Theta)$ under the assumption of the stable convergence to a limit of (and to the same limit as that of) $\text{AVAR}_n^{-1/2}(\hat{\Theta}_n - \Theta)$.

For earlier use of the subsequence approach, see the discussion in Zhang, Mykland, and Aït-Sahalia (2005, Proof of Theorem 3, p. 1411), as well as the concept of "relatively compact in probability" in Zhang (2011, Definition 3, pp. 35-36). The approach, as well as the proof of Theorem 8*(ii)*, relies on Helly's Theorem (Ash (1972, p. 329)). We omit the proof, which is in the same spirit as in the cited papers.

[35]We conjecture that asynchronous T_i can be handled by the methods in Zhang (2011) or Christensen, Podolskij, and Vetter (2013), but a full exploration of this would be another paper.

7 Examples: Corroboration of Concept

The purpose of this section is to document that the assumptions in this paper are widely satisfied in the existing literature. The relevant papers will typically have expressions for AVAR_n and an estimator thereof. In most cases, however, the alternative observed $\widehat{\text{AVAR}}_n$ is much easier to implement when constructing a feasible statistic of the form (1). For an example of a new analysis where we deliberately do not find the theoretical AVAR, see the next section. – We also in many cases describe carefully the separation into martingale and edge effect, thereby hopefully assisting the understanding of the concept.

Unless the opposite is indicated, we suppose that X_t is an Itô-semimartingale, either with no jumps ($dX_t = \mu_t dt + \sigma_t dW_t$), or with jumps that are removed by truncation (Mancini (2001), Aït-Sahalia and Jacod (2007, 2008, 2009, 2012), Jacod and Todorov (2010), Lee and Mykland (2008, 2012), Jing, Kong, Liu, and Mykland (2012)), or by bi- and multi-power methods (Barndorff-Nielsen and Shephard (2004, 2006), Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006)), as appropriate. See also Zhang (2007), Christensen, Oomen, and Podolskij (2011), and Bajgrowicz, Scaillet, and Treccani (2013). We emphasize that θ can be a general semimartingale,[36] so that, for example, the Lévy driven volatility model in Barndorff-Nielsen and Shephard (2001) is covered by the examples. – We either observe X_{t_i} at times $t_i,\ i = 0, \cdots, n$ spanning $[0, \mathcal{T}]$, or we observe Y_{t_i}, which is a version of X_{t_i} that is contaminated by microstructure noise.

Example 1. (Realized Volatility, No Microstructure.) The parameter is $\theta_t = \sigma_t^2$. The convergence rate is $\alpha = 1/2$. – In the straightforward X-is-continuous case, a popular estimator for the $\int_0^t \theta ds$ is the standard realized volatility (RV), $\sum_{t_{i+1} \le t}(X_{t_{i+1}} - X_{t_i})^2$ (Andersen, Bollerslev, Diebold, and Ebens (2001a); Andersen, Bollerslev, Diebold, and Labys (2001b); Barndorff-Nielsen and Shephard (2002); Jacod and Protter (1998); Zhang (2001)). There is no edge effect if the T_i coincide with observation times. Hence, if the spot volatility is continuous, Theorem 2 applies. If one is not sure of the continuity of σ_t^2, one should use the subsampling and averaging approach to construct the observed AVAR (Theorem 3, and then the methods in Section 6) even if there is no edge effect.

To see the precise correspondence with Assumption 1, one can take $M_{n,t} = \sum_{t_{n,j+1} \le t}(X_{t_{n,j+1}} - X_{t_{n,j}})^2 + (X_t - X_{t_{n,*}})^2 - \int_0^t \sigma_s^2 ds$, where the $t_{n,j}$ are the observation times, with $t_{n,*} = \max_j\{t_{n,j} \le t\}$. This is an additive process (when $\mathcal{T}$ and n is fixed), and it can for simplicity be taken to be a martingale under an equivalent measure.[37]

For slightly more complicated estimators than standard RV, or if the T_i do not coincide with observation times t_j (Section 4.3), one may easily incur edge effects e_{T_i} and $\tilde{e}_{T_i}$ of size $O_p(\Delta t_j^{1/2})$, typically $O_p(n^{-1/2})$, at every T_i. The conditions of Theorem 2 are thus often violated.[38] but those

[36] In all our examples, the spot values of θ exists. See Section 9.2 for further discussion of this.

[37] Footnote 15 in Section 4.1. The interpolation is as in Mykland and Zhang (2012, Chapter 2.3.3 pp. 136).

[38] Since $\sum_i e_{T_i}^2$ would normally be of the same order as $(\mathcal{T}/n)B_n$, which is small but not $O(n^{-1})$. In some cases,

of Theorem 3 are easily verified for X either being a continuous process or having finitely many jumps (Section 9.2). For an intermediate case, we study bipower variation in the next example.

An interesting feature even of this classical example is that even though Assumption 1 holds, the further conditions of Proposition 1 may not hold (Fukasawa (2010a,b), Fukasawa and Rosenbaum (2012), Li, Mykland, Renault, Zhang, and Zheng (2013)): when observation times are endogenous, one obtains a case of martingale bias: an additive component of L_t that is $\mathcal{F}$-measurable. Thus, in this case, the asymptotic quadratic variation of L_t can be larger than the asymptotic variance. The assessment of $\widehat{\text{AVAR}}_n$ is nonetheless useful, as it can be used to compare alternative estimators that seek to remove the bias. (Such as the ones proposed in Li, Mykland, Renault, Zhang, and Zheng (2013)). □

Example 2. (Bipower Variation.). In the absence of microstructure noise in X, the bipower estimator $\hat{\Theta}_{(S,T]} = \frac{\pi}{2}\sum_{S<t_{i-1}\le t_i\le T} |\Delta X_{t_{i-1}}||\Delta X_{t_i}|$ (Barndorff-Nielsen and Shephard (2004, 2006)) estimates the integrated volatility. The convergence rate is $\alpha = 1/2$. To see how the heuristic explanation in Remark 3 interfaces with our definitions, assume for simplicity that $\sigma_t^2 = d[X,X]_t^c/dt$ is continuous, and that the X process has no jumps. Set $\Delta Z_{t_i} = \sqrt{\frac{\pi}{2}}|\Delta X_{t_i}| - \sigma_{t_{i-1}}\sqrt{\Delta t}$. Then

$$\frac{\pi}{2}|\Delta X_{t_{i-1}}||\Delta X_{t_i}| - \int_{t_{i-2}}^{t_{i-1}} \sigma_t^2 dt = \Delta M_{t_i}^{(1)} + \Delta M_{t_{i-1}}^{(2)} \tag{65}$$

where

$$\begin{aligned}\Delta M_{t_i}^{(1)} &= \sqrt{\frac{\pi}{2}}|\Delta X_{t_{i-1}}|\Delta Z_{t_i} \text{ and}\\ \Delta M_{t_i}^{(2)} &= \Delta Z_{t_i}\sigma_{t_{i-1}}\sqrt{\Delta t} + \sqrt{\frac{\pi}{2}}|\Delta X_{t_i}|\Delta\sigma_{t_i}\sqrt{\Delta t} + \sigma_{t_{i-1}}^2\Delta t - \int_{t_{i-1}}^{t_i}\sigma_t^2 dt \end{aligned} \tag{66}$$

Aggregate the increments to get $M_{t_k}^{(1)} = \sum_{i=1}^k \Delta M_{t_i}^{(1)}$, and similarly for $M_{t_k}^{(2)}$. Set $M_{t_k} = M_{t_k}^{(1)} + M_{t_k}^{(2)}$. It is easy to see that while none of $M_{t_k}^{(1)}$, $M_{t_k}^{(2)}$ or M_{t_k} are martingales, they are close enough to satisfy the relevant conditions in Assumption 1.

For notational simplicity, suppose that S, T take values in the grid of observation times $\{t_i, i = 0, \cdots, n\}$, which are in turn assumed to be equidistant. Let $t_k \le t_l$. Then

$$\begin{aligned}\hat{\Theta}_{(t_k,t_l]} - \Theta_{(t_k,t_l]} &= \frac{\pi}{2}\sum_{t_k<t_{i-1}\le t_i\le t_l} |\Delta X_{t_{i-1}}||\Delta X_{t_i}| - \int_{t_l}^{t_k}\sigma_t^2 dt\\ &= M_{t_l}^{(1)} - M_{t_{k+1}}^{(1)} + M_{t_{l-1}}^{(2)} - M_{t_k}^{(2)} - \int_{t_{l-1}}^{t_l}\sigma_t^2 dt\\ &= M_{t_l} - M_{t_k} + \tilde{e}_{t_l} - e_{t_k},\end{aligned} \tag{67}$$

where the edge effects are thus

$$\tilde{e}_{t_l} = -\Delta M_{t_l}^{(2)} - \int_{t_{l-1}}^{t_l}\sigma_t^2 dt \text{ and } e_{t_k} = \Delta M_{t_k}^{(2)}. \tag{68}$$

---

it is possible that a creative interpolation will avoid this. For references on interpolation, see Section 4.3.

Under mild conditions, both $\tilde{e}_{t_l}^2$ and $e_{t_k}^2$ are of exact order $O_p(\Delta t^2)$, uniformly under aggregation. Since in this case, $\alpha = 1/2$, the conditions of Theorem 2 are satisfied if $B_n = o(n)$, and otherwise narrowly missed. In the latter case, the conditions in Theorem 3 still go through. On the other hand, in Assumption 2, $\beta = 1 > \alpha$ under mild regularity conditions (and/or by invoking localization as in Section D.2), and, in particular all the $\mathcal{E}_u(0)$ and $\mathcal{E}_u(\mathcal{T})$ are zero, so there is no need for the asymptotic adjustment in Remark 6. □

Example 3. (Preaveraging followed by TSRV). The parameter remains $\theta_t = \sigma_t^2$. There is microstructure noise. The estimator is constructed as follows. One preaverages observations across blocks of size $O(n^{1/2})$ observations, and then calculates a (J, K) TSRV[39] on the basis of the preaveraged observations, where $1 \leq J < K$ are finite. It is easy to see that this estimator of integrated volatility converges at rate $\alpha = 1/4$, and has particularly benign edge effects, of order $O_p(n^{-1/2})$. The small edge effect condition in Section 4.2 is thus satisfied provided $B_n = o(n^{1/2})$. We have used this in Figure 1. – The conditions for the hard edge effect result in Section 5 is satisfied with $\beta = 1/2$, which is larger than α, so that in Theorem 3, the meta edge effect vanishes. – It is conjectured that the same type of situation pertains to classical preaveraging (Jacod, Li, Mykland, Podolskij, and Vetter (2009b); Podolskij and Vetter (2009b)), but we have not investigated this. □

Example 4. (Multiscale and Kernel Realized Volatility.) The parameter remains $\theta_t = \sigma_t^2$. There is microstructure noise. The convergence rate is $\alpha = 1/4$. – We here show that the Multiscale Realized Volatility (MSRV, Zhang (2006)) is covered by our current development. Following Bibinger and Mykland (2013), the result also covers Realized Kernel estimators (RK, Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008)). Similar considerations will cover the Twoscales Realized Volatility (TSRV, Zhang, Mykland, and Aït-Sahalia (2005)).

We shall go through this case in some detail since it illustrates many of the issues. From equation (15), p. 1024, and eq. (51), p. 1039, in Zhang (2006),

$$M_{n,t} = M_{n,t}^{(1)} + M_{n,t}^{(2)} + M_{n,t}^{(3)}, \tag{69}$$

where[40]

$$\begin{aligned} M_{n,t}^{(1)} &= -2 \sum_{i=1}^{\mathcal{M}_n} a_{n,i} \frac{1}{i} \sum_{t_{i+1} \leq t} \epsilon_{t_{n,j}} \epsilon_{t_{n,j-i}}, \\ M_{n,t}^{(2)} &= \sum_{i=1}^{\mathcal{M}_n} a_{n,i} [X, X]_t^{(n,i)} - \int_0^t \sigma_s^2 ds, \text{ and} \\ M_{n,t}^{(3)} &= 2 \sum_{i=1}^{\mathcal{M}_n} a_{n,i} [X, \epsilon]_t^{(i)}. \end{aligned} \tag{70}$$

[39] For the TSRV, see Zhang, Mykland, and Aït-Sahalia (2005) and Aït-Sahalia, Mykland, and Zhang (2011).

[40] Except that we use $\mathcal{M}_n$ to denote the number of scales (called M_n in the multiscale paper. The square brackets in (70) are discrete sums. The $a_{n,i}$ are given by *Ibid.*, eq. (21)-(22) p. 1026.

The edge effects – e and $\tilde{e}$ – are given by (again, cf. *Ibid.,* eq. (51), p. 1039)

$$e_{n,t_k} = \sum_{i=1}^{\mathcal{M}_n} \frac{a_{n,i}}{i} \sum_{j=0}^{j-1} \epsilon^2_{t_{k+i}} - E\epsilon^2 \text{ and } \tilde{e}_{n,t_k} = \sum_{i=1}^{\mathcal{M}_n} \frac{a_{n,i}}{i} \sum_{j=0}^{j-1} \epsilon^2_{t_{k-i+1}} - E\epsilon^2. \tag{71}$$

up to $O_p(n^{-1/2})$. Under the conditions of *Ibid.,* Theorem 4 (p. 1031), including $\mathcal{M}_n/n^{1/2} \to c$, it is easy to see that the asymptotic variance is retrieved, and also that Assumptions 1-2, as well as the conditions of Proposition 1 are satisfied. – Similar arguments would extend to the dependent but mixing noise in Aït-Sahalia, Mykland, and Zhang (2011). □

Example 5. (Block Estimation of Higher Powers of Volatility.) The parameter is $\theta_t = g(\sigma_t^2)$, with g not being the identity function. In the absence of microstructure noise, the convergence rate is $\alpha = 1/2$. If microstructure noise is present, the convergence rate is $\alpha = 1/4$. We are here concerned with the former case. The estimation of integrals of σ_t^p goes back to Barndorff-Nielsen and Shephard (2002), which showed that the case $g(x) = x^2$ is related to the asymptotic variance of the realized volatility. See also Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006), Mykland and Zhang (2012, Proposition 2.17, p. 138) and Renault, Sarisoy, and Werker (2013) for related developments.

Block estimation (Mykland and Zhang (2009, Section 4.1, p. 1421-1426) has the ability to make these estimators approximately or fully efficient. One path is to keep the block size finite. This avoids bias. When using overlapping blocks, however, the asymptotic variance is hard to compute (Mykland and Zhang (2012, Ch. 2.6.2, pp. 170-172)). This is an instance where the observed AVAR would seem to be particularly appealing, and our Assumptions 1-2 are obviously satisfied.

Another path is to let the block size increase with n. As seen, however, in Mykland and Zhang (2011, Section 5, pp. 224-229), Jacod and Rosenbaum (2013a, Section 2, pp. 2-9), and Jacod and Rosenbaum (2013b, Section 3, pp. 1466-1473), there is a bias that can be corrected for. If one uses the raw estimators, therefore, our asymptotic bias A_t is non-zero.

It nicely illustrates the theory in the current paper to consider the bias terms A_t^1 to A_t^4 in Jacod and Rosenbaum (2013b, p.1468). First of all A_t^1 is edge effect (the authors call it "border terms"). $A_t = A_t^2 + A_t^3$ is asymptotic bias in the sense of our Assumption 1, with requirements on the bias satisfied. The jump term A_t^4 contributes to quadratic variation, and hence one has an interest in classifying it with the martingale part L_t. This can be done and (if desirable) subsequently undone with the help of a measure change, in the spirit of Mykland and Zhang (2009, Section 2.2), which also works for jump processes (Doléans-Dade (1970); Jacod (1975); Gill and Johansen (1990); Andersen, Borgan, Gill, and Keiding (1992); Jacod and Shiryaev (2003)). The martingale part of bias is then viewed the same way as the martingale bias (for endogenous observations) encountered in Example 1 above. – Thus our Assumptions 1-2 are satisfied.

As a final comment, n is typically given for fixed data. When this is the case, it is entirely in the mind of the econometrician whether the block size is finite or not as $n \to \infty$. This raises the question of which asymptotics to use. This conundrum may also be a reason for using the observed asymptotic variance, and other small sample methods. □

Example 6. (Estimation of Co-volatility (Covariance) from Asynchronous Observations.) A popular estimator is due to Hayashi and Yoshida (2005), see also Podolskij and Vetter (2009a), Christensen, Podolskij, and Vetter (2013), and Bibinger and Vetter (2014) for microstructure, jumps, and asymptotic distributions. Alternatives include the Previous-Tick estimator (Zhang (2011), Bibinger and Mykland (2013)), and Quasi-Likelihood (Shephard and Xiu (2012)). The estimator in Mykland and Zhang (2012, Chapter 2.6.3, p. 172-175) is a hybrid of Hayashi-Yoshida and Quasi-Likelihood. In all these cases, it is quite clear that the stable convergence holds, and that the current paper's Assumptions 1-2 are satisfied. The asymptotic distributions, however, are quite complex, and the estimation of AVAR_n is daunting. In comparison, the approach of observed AVAR offers a pleasing alternative to assessing the asymptotic variance of co-volatility. □

Example 7. (High Frequency Regression, and ANOVA.) We are here concerned with systems on the form $dV_t = \beta_t dX_t + dZ_t$, where V_t and X_t can be observed at high frequency, either with or without microstructure. The coefficient process β_t can either be the "beta" from portfolio optimization, with Z_t in the role of idiosyncratic noise, or β_t can be the hedging "delta" for an option, with Z_t as tracking error. Nonparametric estimates can be used directly, or for forecasting, or for model checking. X_t can be multidimensional. The regression problem seeks to estimate or make tests about $\int_0^{\mathcal{T}} \beta_t dt$ (Mykland and Zhang (2009, Section 4.2, pp. 1424-1426), Kalnina (2012), Zhang (2012, Section 4, pp. 268-273), Reiss, Todorov, and Tauchen (2014)). The ANOVA problem seeks to estimate $[Z, Z]_{\mathcal{T}}$ (Zhang (2001) and Mykland and Zhang (2006)). Convergence rates are as for realized or other powers of volatility, with $\alpha = 1/2$ when there is no microstructure noise, and $\alpha = 1/4$ otherwise. Assumptions 1-2 of the current paper are easily seen to be satisfied. □

Example 8. (Continuous Leverage Effect, with or without Microstructure.) The parameter is $\theta_t = d[\sigma^2, X^c]_t/dt$. If there is no microstructure noise, the convergence rate is $\alpha = 1/4$. If microstructure noise is present, the convergence rate is $\alpha = 1/8$. – The estimation of leverage effect is discussed in Wang and Mykland (2014) for the case where X_t is continuous, and in Aït-Sahalia, Fan, Wang, and Yang (2013) for the case where the process X_t can also have jumps.[41] In the latter, more general paper, jumps are removed according to Jacod and Todorov (2010). The asymptotic criteria in (our current) Assumption 1-2 and Proposition 1, are seen to be satisfied through a slight extension of Aït-Sahalia, Fan, Wang, and Yang (2013, Theorem 7.2). □

Example 9. (Volatility of Volatility, no Microstructure.) The process X is assumed to be continuous, and the parameter is $\theta_t = d[\sigma_t^2, \sigma^2]_t^c/dt$. The convergence rate is $\alpha = 1/4$. The results in the literature on this inference problem are Vetter (2011, Theorems 2.1 and 2.5) and Mykland, Shephard, and Sheppard (2012a, Theorem 7 and Corollary 2). The asymptotic criteria in Assumption 1-2 and Proposition 1 hold through a slight extension. □

[41] Both papers study both the case where there is microstructure, and where there is none.

8 A New Application: Nearest Neighbor Truncation

To illustrate the ease with which the current theory can be applied to a new problem, we consider the nearest neighbor truncation of Andersen, Dobrev, and Schaumburg (2012), where estimators are defined and studied for the case where there is no microstructure noise. See also Andersen, Dobrev, and Schaumburg (2014) on quarticity. In both cases, preaveraging is actually used on the data, but not taken account of in the asymptotics.

We here adapt the estimation problem from Andersen, Dobrev, and Schaumburg (2012) to the setting where microstructure noise is present in the model. To get a point estimator, we extend their estimator with the help of pre-averaging and a two scales construction, which is straightforward. We then show that the Observed Asymptotic Variance can be used to assess the statistical error, and hence to create a feasible estimator.

Suppose for simplicity that observations are of the form $Y_{t_j} = X_{t_j} + \epsilon_j$, where the ϵ_j are i.i.d., and the efficient log price process X_t is an Itô semimartingale with finite activity jumps, as assumed by Andersen, Dobrev, and Schaumburg (2012). Using pre-averaging, and in analogy with Equation (4) of their paper, we consider an estimator based on

$$MedRV_{M,n} = \sum_{i=3}^{\lfloor n/M \rfloor - 2} \text{med}(\Delta\bar{Y}_{M,i-2}, \Delta\bar{Y}_{M,i}, \Delta\bar{Y}_{M,i+2})^2 \tag{72}$$

where $\Delta\bar{Y}_{M,i} = \bar{Y}_{M,i} - \bar{Y}_{M,i-1}$ and $\bar{Y}_{M,i} = \frac{1}{M}\sum_{j=(i-1)M_n+1}^{iM_n} Y_j$. For simplicity, suppose that the t_j are equidistant, *i.e.*, $t_j - t_{j-1} = \Delta t = \mathcal{T}/n$ for all j.[42] The statistic $\bar{Y}_{M,i}$ is thus based on observations in the time interval $(\tau_{i-1}, \tau_i]$, where $\tau_i = iM\Delta t$, and $\Delta\tau = M\Delta t$. When taking the median, we have used every second $\Delta\bar{Y}_{M,i}$ to avoid autocorrelation. As $n \to \infty$, we let $M = M_n$, with $M_n/\sqrt{n} \to c$.

To suitably adjust (72), and to verify the conditions of our current theorems, we invoke the results of Mykland and Zhang (2013). Set $Y^c_{t_j} = X^c_{t_j} + \epsilon_j$, and similarly $\bar{Y}^c_i$, where X^c_t is the continuous part of the latent process. Following Mykland and Zhang (2013), there is a contiguous (sequence of) probability measures Q_n, and "super-blocks" of $2\mathcal{M}$ $\bar{Y}^c_i$'s, with starting points $\lambda_{n,l} = 2l\mathcal{M}M_n\Delta t$, so that, conditionally on sigma-field at the start of each block, $\Delta\tau^{-1/2}\Delta\bar{Y}^c_{l\mathcal{M}+1}, \cdots, \Delta\tau^{-1/2}\Delta\bar{Y}^c_{(l+1)\mathcal{M}}$ is a Gaussian MA(1) process with marginal variance $\frac{2}{3}\sigma^2_{\lambda_l} + 2\frac{\nu^2}{c^2\mathcal{T}}$, where $\nu^2 = \text{Var}(\epsilon)$. Thus, if $(\mathcal{F}_t)$ is the filtration generated by the X^c_ts and the ϵs,

$$E_{Q_n}\left\{\sum_{i=2l\mathcal{M}+5}^{2(l+1)\mathcal{M}-4} \text{med}(\Delta\bar{Y}^c_{M_n,i-2}, \Delta\bar{Y}^c_{M_n,i}, \Delta\bar{Y}^c_{M_n,i+2})^2 \,|\mathcal{F}_{\lambda_l}\right\} = (2\mathcal{M}-8)\Delta\tau(\frac{2}{3}\sigma^2_{\lambda_l} + 2\frac{\nu^2}{c^2\mathcal{T}})\frac{6-4\sqrt{3}+\pi}{\pi} \tag{73}$$

in analogy with Andersen, Dobrev, and Schaumburg (2012): if Z_1, Z_2, Z_3 are i.i.d. $N(0,1)$, then $E\text{med}(Z_1, Z_2, Z_3)^2 = (6 - 4\sqrt{3} + \pi)/\pi$. One now needs to dispose of the nuisance parameter ν^2.

[42] Otherwise, a correction factor applies, cf. Mykland and Zhang (2013).

To stay in the sprit of Andersen, Dobrev, and Schaumburg (2012), we adjust by using the MedRV, but doubling the block size: $\Delta\bar{Y}_{2M_n,i} = (\Delta\bar{Y}_{M_n,2i-1} + \Delta\bar{Y}_{M_n,2i})/2$ (which is based on observations in $(\tau_{2i-2}, \tau_{2i}]$. Now observe that, also under Q_n,

$$E_{Q_n}\left\{\sum_{i=l\mathcal{M}+3}^{(l+1)\mathcal{M}-2} \mathrm{med}(\Delta\bar{Y}^c_{2M_n,i-2}, \Delta\bar{Y}^c_{2M_n,i}, \Delta\bar{Y}^c_{2M_n,i+2})^2 \;|\mathcal{F}_{\lambda_l}\right\} = (\mathcal{M}-4)(2\Delta\tau)(\frac{2}{3}\sigma^2_{\lambda_l}+2\frac{\nu^2}{(2c)^2\mathcal{T}})\frac{6-4\sqrt{3}+\pi}{\pi}, \tag{74}$$

where we have in both cases used samples from the time interval $(\tau_{2l\mathcal{M}+4}, \tau_{2(l+1)\mathcal{M}-4}] \subset (\lambda_{n,l}, \lambda_{n,l+1}]$.

$$\begin{aligned}\text{Eq. (74)} \; -\frac{1}{4}\times \; \text{Eq. (73)} \; &= 2(\mathcal{M}-4)\Delta\tau\frac{2}{3}\sigma^2_{\lambda_l}\frac{3}{4}\frac{6-4\sqrt{3}+\pi}{\pi} \\ &= (\tau_{2(l+1)\mathcal{M}-4} - \tau_{2l\mathcal{M}+4})\sigma^2_{\lambda_l}\frac{6-4\sqrt{3}+\pi}{2\pi}. \end{aligned} \tag{75}$$

In view of the development in Mykland and Zhang (2013), the aggregated (over $\mathcal{M}$) terms

$$\begin{aligned}&\sum_{i=l\mathcal{M}+3}^{(l+1)\mathcal{M}-2} \mathrm{med}(\Delta\bar{Y}^c_{2M_n,i-2}, \Delta\bar{Y}^c_{2M_n,i}, \Delta\bar{Y}^c_{2M_n,i+2})^2 - \frac{1}{4}\sum_{i=2l\mathcal{M}+5}^{2(l+1)\mathcal{M}-4} \mathrm{med}(\Delta\bar{Y}^c_{M_n,i-2}, \Delta\bar{Y}^c_{M_n,i}, \Delta\bar{Y}^c_{M_n,i+2})^2 \\ &\qquad - (\tau_{2(l+1)\mathcal{M}-4} - \tau_{2l\mathcal{M}+4})\sigma^2_{\lambda_l}\frac{6-4\sqrt{3}+\pi}{2\pi} \end{aligned} \tag{76}$$

satisfy stable convergence and also the other conditions of Assumptions 1-2 and Proposition 1 under Q_n, with $\alpha = 1/4$. One can take the T_i to be the same as the λ_i. This is easily seen to carry over to the original measure. The left out terms (around the boundaries λ_l) are handled with the big-block-small-block device described in Mykland and Zhang (2012, Chapter 2.6.2, pp. 170-172). Also, the jumps are negligible since assumed to be of finite activity. In conclusion:

PROPOSITION 3. (MEDIAN REALIZED VOLATILITY UNDER MICROSTRUCTURE NOISE.) *Let Θ be the integrated volatility on $[0, \mathcal{T}]$. A pre-averaged extension of the median realized volatility of Andersen, Dobrev, and Schaumburg (2012) is given by*[43]

$$\hat{\Theta} \;=\; \frac{2\pi}{6-4\sqrt{3}+\pi}\left(MedRV_{2M_n,n} - \frac{1}{4}MedRV_{M_n,n}\right), \tag{77}$$

Then, with the T_i taken to be the same as the τ_i, conditions of Assumptions 1-2 and Proposition 1 are satisfied, with $\alpha = 1/4$. Subject to their respective conditions on choice of averaging parameter K, Theorems 3, 4, and 5 are valid. In particular, one can use the observed AVAR to set the standard error, as in Theorem 6.

[43]The estimator can be small sample adjusted as in the original paper, without affecting the conclusion of this proposition. One can also use the average of rolling windows.

9 Some Practical Guidance

9.1 Two or Three Scales? Should one correct for Edge Effect? What choice of B_n?

Three situations. The situation described in Sections 5-6 is a worst case scenario, and should be used by default. For a *prima facie* data analysis, the method can be used without too much mathematics. If the estimator is benign, and one is willing to do a little extra analysis, showing that $\beta > \alpha$ in Assumption 2 relieves one of the need for the meta edge effect correction in the hard edge case. Alternatively, one can verify the "small edge" condition in Section 4.2 and use Theorem 2.

It should be emphasized that when the "small edge" condition prevails, the results of Theorem 2 also hold when $K \to \infty$ (under the precise conditions of Theorem 3), and in that case, θ does not have to be continuous. Thus, a large-ish K_1 and K_2 in Theorem 2 will avoid the problem of this continuity condition. We summarize the situation in the following Table 1.

What is known about the estimator?	How many scales are needed?	Is Meta Edge Effect correction needed?
Default: Only Assumptions 1-2	three or more	yes
Hard Edge, $\beta > \alpha$	two, or three or more; **see "more about β" below**	no
Soft Edge	two (or more)	no

Table 1: Choice of Methodology Depending on Estimator $\hat{\Theta}$

More about β: a choice of theorems, and the choice of B_n. Suppose the conditions of Theorem 3 are satisfied with $\beta > \alpha$ in (30). Then the small edge condition of Theorem 2 is satisfied provided $B_n = o(n^{2(\beta-\alpha)})$ (use Lemma 6 in Appendix D.2). Thus, if $\beta > \alpha$, *one has a choice of whether to use Theorem 2 or Thoerem 3 and its corollaries.* When using Theorem 2 the size of B_n is controlled to not be too large. On the other hand, in the hard edge case, B_n can be as large as one wishes; it is controlled from below by $B_n^{-1} = o(n^{2\beta-4\alpha})$. – It should be clear from the preceding that even a rough analysis of β may yield substantial additional flexibility in how the AVAR is calculated.

Dropping the volatility term? If $K\Delta T = o(n^{-2\alpha})$, then the $[\theta, \theta]$ term in Theorems 2-3 are of lower order than the asymptotic variance. This raises the possibility of a one scale estimator of variance in the soft edge case, and a two scale estimator under hard edge (one may still have to remove the microstructure). This may in some cases work well, for example, for classical realized volatility (Example 1), a one scale estimator is both practicable and, and somewhat similar to quarticity (Barndorff-Nielsen and Shephard (2002), see also Mykland and Zhang (2012, Proposition 2.17, p. 138)). In the more general case, however, disregarding the $[\theta, \theta]$ creates a risk of overestimating AVAR. Imagine, for example, that one were to estimate AVAR as (half) the full amount of apparent quadratic variation Figure 1. Even if a term goes away asymptotically, it doesn't mean that it isn't there for a finite sample size. – If one wishes to use such a reduced scale estimator, it would be wise to also compute the two or three (or more) scales estimator for comparison, as due diligence.

The choice of the K_is. For the hard edge case, a formal analysis is provided in Section 6.4. We conjecture, however, that suitable use of signature plots for the AVAR (as a function of the K_is) will in practice suffice. This is particularly easy to implement for the two scales estimator (24).

9.2 The case where the Spot Process θ_t does not exist

The theory in this paper requires the existence of a "spot" θ_t, and does not apply, say, to estimating the discontinuous part of the quadratic variation. For example, suppose that $\Theta_{(0,T]} = \int_0^T \theta_t dt + \mathfrak{T}_T$, where $\mathfrak{T}_t$ is a process with finitely many jumps in $(0, \mathcal{T}]$. Then, obviously, to first order, $QV_K(\Theta) = [\mathfrak{T}, \mathfrak{T}]_T - [\mathfrak{T}, \mathfrak{T}]_0 + o_p(1)$. The same is true for $QV_K(\hat{\Theta})$ – The situation is not exotic: A simple example would be the estimation of $[X, X]$ when the X process can have jumps. – In our setting, the methodology applies to estimating the continuous part $\int \sigma_t^2$ of this quadratic variation.

For this reason, in our examples (Section 7), we consider that the primary estimating procedure removes anything that can cause $\mathfrak{T}_t$ to be nonzero. In the case that the $\mathfrak{T}_t$ process has finitely many jumps, these can alternatively be removed directly with truncation or bi-/multi-power methods, cf. the references at the beginning of Section 7. We presently show how one can proceed using truncation.

Algorithm 1. (Jump removal in $\hat{\Theta}$.) If there are ν (finitely many) jumps, truncation creates ν removed intervals[44] $(T_{i_j}, T_{i_j+1}]$, $j = 1, \cdots, \nu$. (These intervals are identified with probability one as $n \to \infty$.) One can then proceed as follows. For scale K, omit all $\hat{\Theta}_{(T_i, T_{i+K}]}$ for which $(T_{i_j}, T_{i_j+1}] \subseteq (T_i, T_{i+K}]$ for any of the removed intervals. When $\hat{\Theta}_{(T_i, T_{i+K}]}$ is removed the relevant squares in $QV_K(\hat{\Theta})$ are computed as $(\hat{\Theta}_{(T_{i+K}, T_{i+2K}]} - \hat{\Theta}_{(T_{i-K}, T_i]})^2$. Call this quantity $QV_{K,\text{modified}}(\hat{\Theta})$. Similarly, for the true process θ, denote the modified averaged quadratic variation by $QV_{K,\text{modified}}(\Theta)$. □

The critical piece for analyzing the above construction is then the following, which generalizes Theorem 1 in Section 3, by the same methods.

Theorem 9. (The Integral-to-Spot Device with Removed Intervals.) *Assume that θ_t is a semimartingale on $[0, \mathcal{T}]$. Set $\Delta T = \mathcal{T}/B$, and assume that $T_i = i\Delta T$. Suppose that $K\Delta T \to 0$,*

[44] The method carrying out the truncation may depend on the estimator.

and that either $K \to \infty$ or θ_t is continuous. Suppose that there are stopping times $\tau_1, \cdots, \tau_v \in (0, T)$. Assume that in Algorithm 1 above, $P(\cap_{j=1}^{\nu}\{\tau_j \in (T_{i_j}, T_{i_j+1}]\}) \to 1$ as $B \to \infty$. Then

$$\frac{1}{(K\Delta T)^2} QV_{K,modified}(\Theta) = \left(\frac{2}{3}[\theta,\theta]_{\mathcal{T}-} + \frac{2}{3}\sum_{j=1}^{\nu}([\theta,\theta]_{T_{i_j+1}} - [\theta,\theta]_{T_{i_j}})\right)(1 + o_p(1))$$

$$\xrightarrow{p} \frac{2}{3}[\theta,\theta]_{\mathcal{T}-} + \frac{2}{3}\sum_{j=1}^{\nu}(\Delta\theta_{\tau_j})^2. \tag{78}$$

Thus, if jump times in $\mathfrak{T}_t$ coincide with those of θ_t, the estimation $[\theta,\theta]_{\mathcal{T}-}$ becomes additionally complicated.

The AVAR estimates, however, are not affected. Under the conditions of Theorem 2, the TSAVAR (24) remains consistent for $AVAR(\hat{\Theta} - \Theta)$, and the more complex results in Theorems 5-6 in Section 6.2 are also unchanged. Proposition 2 in Section 6.5 is, of course, also unaffected. Since $QV_{K,\text{modified}}(\hat{\Theta})$ will have lost a fraction ν/B_n of its asymptotic variance component, one can consider a small sample multiplicative adjustment of $(1 - \hat{\nu}/B_n)^{-1}$ to the estimated variances, where $\hat{\nu}$ is the number of removed intervals $(T_{i_j}, T_{i_j+1}]$, but this does not impact the asymptotics.

For reference, note that under the conditions of Theorems 2 and 3, respectively,

$$\begin{aligned} QV_{K,\text{modified}}(\hat{\Theta}) = &\frac{2}{3}(K\Delta T)^2\left([\theta,\theta]_{\mathcal{T}-} + \sum_{j=1}^{\nu}(\Delta\theta_{\tau_j})^2\right) \\ &+ 2n^{-2\alpha}[L,L]_{\mathcal{T}} \\ &+ \text{ Meta Edge Effect } + \frac{1}{K}V_0 \text{ (if in the situation of Theorem 3)} \\ &+ o_p(n^{-2\alpha}). \end{aligned} \tag{79}$$

We note that if jumps are not fully removed, there remain pieces of the process $\mathfrak{T}_t$, and in this case $QV_K(\hat{\Theta})$ would be unusually large. This makes the question of standard error in $QV_K(\hat{\Theta})$ more interesting. – If one splits a jump part of $\mathfrak{T}_T$ into a predictable (often continuous) part, then Theorem 3 remains valid, but with the martingale part attached to L_t. This would create a case for using our procedure in a multi-day setting. – Also note that for the case of many small jumps, the contiguity results of Zhang (2007) may mitigate the problem.

A detailed investigation of these issues is left for another paper.

10 Conclusion

The paper introduces a nonparametric estimator of estimation error which we call the observed asymptotic variance. In analogy with the "observed information" of parametric inference, our statistic estimates the asymptotic variance without needing a formula for the theoretical quantity. As we have seen in our examples, the estimator is consistent in all of them.

We emphasize that the method has a strong applied motivation, and that we think it meets a need. Assessing the standard error of a high-frequency-based estimator is challenging to implement. We hope our proposed methodology will be a useful tool at the disposal of everyone who works with high frequency data.

On the mathematical side the basic insight is Equation (4) in Section 2.1. To operationalize this insight, the two main tools are the Integral-to-Spot Device (Section 3), and the mathematical similarity between edge effects and microstructure noise (Section 5). The estimation of asymptotic variance (AVAR) is implemented with the help of multi-scale methods in Sections 4.2 and 6, and examples are given in Sections 7-8. The observed AVAR can also be used for the selection of tuning parameters, also in the non-obvious case of stable convergence and random variance. As part of the theoretical development, we show how to feasibly disentangle the impact of estimation error $\hat{\Theta}_{(0,\mathcal{T}]} - \Theta_{(0,\mathcal{T}]}$ and the variation $[\theta,\theta]_{\mathcal{T}}$ in the parameter process alone. For the latter, we also obtain a new estimator of quadratic variation of target parameters. The methods generalize readily to several dimensions.

A number of issues have been left for later. Consistency is the only first order requirement on estimators of AVAR, but a main question still remains of how to optimize the number and position of scales K in Section 6. This may involve the convergence rate and the AVAR of the AVAR, and perhaps one can iterate the observed AVAR procedure. As the likelihood movement of the 1980s and 90s has shown, however, statistical accuracy may not only be about the efficiency of estimates of AVAR.[45] There is also room for a more complete theory of tuning parameter selection, of multivariate inference, and of adjusting the meta-edge effects (as in Remark 6). It would also be interesting to extend Observed AVAR to the case where the spot process θ_t is not a semimartingale, and to the case where it does not exist (see Section 9.2).

REFERENCES

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., J. Fan, and D. Xiu (2010): "High Frequency Covariance Estimates with Noisy and Asynchronous Financial Data," *Journal of the American Statistical Association*, 110, 1504–1517.

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

[45] See Footnote 4.

——— (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.

——— (2013): "Is a discretely observed continuous semimartingale Itô or not," Working paper.

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2011): "Ultra high frequency volatility estimation with dependent microstructure noise," *Journal of Econometrics*, 160, 160–175.

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

Andersen, P. K., Ø. Borgan, R. D. Gill, and N. Keiding (1992): *Statistical Models Based on Counting Processes*, New York: Springer.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001a): "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics*, 61, 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001b): "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55.

Andersen, T. G., D. Dobrev, and E. Schaumburg (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.

Ash, R. B. (1972): *Real Analysis and Probability*, Academic Press.

Bajgrowicz, P., O. Scaillet, and A. Treccani (2013): "Jumps in high-frequency data: Spurious detections, dynamics, and news," Working paper.

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. (1986): "Inference on full or partial parameters based on the standardized signed log likelihood ratio," *Biometrika*, 73, 307–322.

——— (1991): "Modified signed log likelihood ratio," *Biometrika*, 78, 557–563.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008): "Designing realized kernels to measure ex-post variation of equity prices in the presence of noise," *Econometrica*, 76, 1481–1536.

Barndorff-Nielsen, O. E., S. Kinnebrouck, and N. Shephard (2009): "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.

——— (2004): "Power and bipower variation with stochastic volatility and jumps (with discussion)," *Journal of Financial Econometrics*, 2, 1–48.

——— (2006): "Econometrics of testing for jumps in financial economics using bipower variation," *Journal of Financial Econometrics*, 4, 1–30.

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.

Bibinger, M. and M. Vetter (2014): "Estimating the quadratic covariation of an asynchronously observed semimartingale with jumps," Discussion paper.

Boyd, S. and L. Vandenberghe (2004): *Convex Optimization*, Cambridge, UK: Cambridge University Press.

Calvet, L. and A. Fisher (2008): *Multifractal Volatility: Theory, Forecasting, and Pricing*, Elsevier - Academic Press.

Christensen, K., R. Oomen, and M. Podolskij (2011): "Fact or friction: Jumps at ultra high frequency," Discussion paper.

Christensen, K., M. Podolskij, and M. Vetter (2013): "On covariation estimation for multivariate continuous Ito semimartingales with noise in non-synchronous observation schemes," *Journal of Multivariate Analysis*, 120, 59–84.

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

Cox, D. R. (1958): "Some Problems connected with Statistical Inference," *Annals of Mathematical Statistics*, 29, 357–372.

——— (1980): "Local Ancillarity," *Biometrika*, 67, 279–286.

Dellacherie, C. and P. Meyer (1982): *Probabilities and Potantial B*, Amsterdam: North-Holland.

DiCiccio, T. J., P. Hall, and J. Romano (1991): "Empirical likelihood is Bartlett-correctable," *Annals of Statistics*, 19, 1053–1061.

DiCiccio, T. J. and J. Romano (1989): "On adjustments based on the signed root of the empirical likelihood ratio statistic," *Biometrika*, 76, 447–456.

Doléans-Dade, C. (1970): "Quelques applications de la formule de changement de variables pour les semimartingales," *Zeitschrift fur Wahrsheinlichkeitstheorie und verwandte Gebiete*, 16, 181–194.

Efron, B. and D. V. Hinkley (1978): "Assessing the Accuracy of the Maximum Likelihood Estimator: Observed Versus Expected Fisher Information," *Biometrika*, 65, 457–482.

Efron, B. and R. Tibshirani (1991): "Statistical Analysis in the Computer Age," *Science*, 253 (July), 390–395.

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

Fukasawa, M. (2010a): "Central limit theorems for the realized volatility based on tick time sampling," *Finance and Stochastics*, 14, 209–233.

——— (2010b): "Realized volatility with stochastic sampling," *Stochastic Processes and Applications*, 120, 829–852.

Fukasawa, M. and M. Rosenbaum (2012): "Central Limit Theorems for Realized Volatility under Hitting Times of an Irregular Grid," To appear in *Stochastic Processes and Applications.*

Gill, R. D. and S. Johansen (1990): "A Survey of Product-Integration with a View toward Application in Survival Analysis," *Annals of Statistics*, 18 (4), 1501–1556.

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.

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.

Hayashi, T. and N. Yoshida (2005): "On Covariance Estimation of Non-synchronously Observed Diffusion Processes," *Bernoulli*, 11, 359–379.

Heath, D. (1977): "Interpolation of martingales," *Annals of Probability*, 5, 804–806.

Jacod, J. (1975): "Multivariate point processes: Predictable projection, Radon-Nikodym derivatives, Representation of martingales," *Zeitschrift fur Wahrsheinlichkeitstheorie und verwandte Gebiete*, 31, 235–253.

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.

——— (2009b): "Microstructure Noise in the Continuous Case: The Pre-Averaging Approach," *Stochastic Processes and Their Applications*, 119, 2249–2276.

Jacod, J. and P. Mykland (2013): "Microstructure Noise in the Continuous Case: Efficiency and the Adaptive Pre-Averaging Method," Tecnical Report, the University of Chicago.

Jacod, J. and M. Podolskij (2013): "A Test for the Rank of the Volatility Process: The Random Perturbation Approach," *Annals of Statistics*, 41, 2391–2427.

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

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

Jacod, J. and M. Rosenbaum (2013a): "Estimation of volatility functionals: the case of a $\sqrt{n}$ window," Discussion paper.

——— (2013b): "Quarticity and other Functionals of Volatility: Efficient Estimation," *Annals of Statistics*, 41, 1462–1484.

Jacod, J. and A. N. Shiryaev (1987): *Limit Theorems for Stochastic Processes*, New York: Springer-Verlag.

——— (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.

Jensen, J. L. (1992): "The modified signed likelihood statistic and saddlepoint approximations," *Biometrika*, 79, 693–703.

——— (1997): "A simple derivation of r^* for curved exponential families," *Scandinavian Journal of Statistics*, 24, 33–46.

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.

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.

Kinnebrock, S. and M. Podolskij (2008): "A Note on the Central Limit Theorem for Bipower Variation of General Functions," *Stochastic Processes and their Applications*, 118, 1056–1070.

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.

Li, Y., P. Mykland, E. Renault, L. Zhang, and X. Zheng (2013): "Realized Volatility when Endogeniety of Time Matters," To appear in *Econometric Theory.*

Mancini, C. (2001): "Disentangling the Jumps of the Diffusion in a Geometric Jumping Brownian Motion," *Giornale dell'Istituto Italiano degli Attuari*, LXIV, 19–47.

McCullagh, P. (1984): "Local sufficiency," *Biometrika*, 71, 233–244.

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

McCullagh, P. and R. Tibshirani (1990): "A simple method for the adjustment of profile likelihoods," *Journal of the Royal Statistcal Society ser. B*, 52, 325–344.

McLeish, D. L. (1975): "A Maximal Inequality and Dependent Strong Laws," *Annals of Probability*, 3 (5), 829–839.

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

——— (1995a): "Dual Likelihood," *Annals of Statistics*, 23, 396–421.

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

——— (1999): "Bartlett identities and large deviations in likelihood theory," *Annals of Statistics*, 27, 1105–1117.

——— (2001): "Likelihood computations without Bartlett identities," *Bernoulli*, 7, 473–485.

Mykland, P. A., N. Shephard, and K. Sheppard (2012a): "Efficient and feasible inference for the components of financial variation using blocked multipower variation," Technical Report, University of Oxford.

——— (2012b): "Efficient and Feasible Inference for the Components of Financial Variation using Blocked Multipower Variation," Discussion paper.

Mykland, P. A. and L. Zhang (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.

——— (2013): "Between Data Cleaning and Inference: Pre-Averaging and other Robust Estimators of the Efficient Price," Working Paper, University of Illinois at Chicago and University of Chicago.

Pierce, D. and D. Peters (1994): "Higher-order asymptotics and the likelihood principle: One-parameter models," *Biometrika*, 81, 1–10.

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.

Protter, P. (2004): *Stochastic Integration and Differential Equations*, New York: Springer-Verlag, second ed.

Reid, N. (1988): "Saddlepoint methods and statistical inference," *Statistical Science*, 3, 213–227.

Reiss, M., V. Todorov, and G. Tauchen (2014): "Nonparametric Test for a Constant Beta over a Fixed Time Interval," Discussion paper.

Renault, E., C. Sarisoy, and B. J. Werker (2013): "Efficient Estimation of Integrated Volatility and Related Processes," Discussion paper.

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

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

Rosenbaum, M., L. Duvernet, and C. Y. Robert (2010): "Testing the type of a semimartingale: Ito against multifractal," *Electronic Journal of Statistics*, 4, 1300–1323.

Shephard, N. and D. Xiu (2012): "Econometric analysis of multivariate realised QML: estimation of the covariation of equity prices under asynchronous trading," .

Skovgaard, I. (1986): "A note on the differentiation of cumulants of log likelihood derivatives," *International Statistical Review*, 54, 29–32.

——— (1991): *Analytic Statistical Models*, Hayward: Institute of Mathematical Statistics.

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

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.

Wu, W. B. and M. Woodroofe (2004): "Martingale approximations for sums of stationary processes. Annals of Probability," *Annals of Probability*, 32, 1674–1690.

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*, 8, 259–275.

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.

APPENDIX: PROOFS AND TECHNICAL ISSUES

A Results on the Quadratic Variation of θ: Tightness and Convergence Properties

Set

$$f_t^{(l,n)} = \frac{1}{K\Delta T} \sum_{K\le i\le B-K; i\equiv l[2K]} \left((T_{i+K}-t)I\{T_{i+K}\ge t > T_i\} + (t-T_{i-K})I\{T_i \ge t > T_{i-K}\}\right). \tag{A.1}$$

where $i \equiv l[2K]$ means that i is on the form $2K + l$. We note that $f_t^{(l)} = f_t^{(l,n)}$ depends on n through ΔT, K, and B. Observe that

$$0 \le f_t^{(l)} \le 1 \text{ for all } t, l, \text{ and } n. \tag{A.2}$$

Define the processes $\theta_t^{(l,n)} = \int_0^t f_s^{(l,n)} d\theta_s$. To motivate the following development, note from Itô's Formula that

$$\frac{1}{K^2(\Delta T)^2} \sum_{K\le i\le B-K, i\equiv l[2K]} (\Theta_{(T_i,T_{i+K}]} - \Theta_{(T_{i-K},T_i]})^2 = \sum_{K\le i\le B-K, i\equiv l[2K]} (\theta^{(l)}_{T_{i+K}} - \theta^{(l)}_{T_{i-K}})^2. \tag{A.3}$$

Lemma 1. *Let θ_t be a semimartingale. Assume that ΔT is independent of i, and that $K\Delta T \to 0$ as $n \to \infty$. Then the collection $\{\theta_t^{(l,n)},$ all $l, n\}$ is tight, with respect to convergence in law with respect to the Skorokhod metric on $\mathbb{D}$, and also P-UT (Jacod and Shiryaev (2003, Chapter VI.3.b, and Definition VI.6.1, p. 377)).*

Lemma 1 follows by invoking (A.2) along with Jacod and Shiryaev (2003, Corollary VI.6.20 p. 381). The corollary can be used since the since the process θ_t, when viewed as a constant sequence of processes, is obviously P-UT. □

Lemma 2. *Let θ_t be a semimartingale. Assume that ΔT is independent of i, and that $K\Delta T \to 0$ as $n \to \infty$. Then*

$$\frac{1}{2K^3(\Delta T)^2} \sum_{i=K}^{B-K} (\Theta_{(T_i,T_{i+K}]} - \Theta_{(T_{i-K},T_i]})^2 = \frac{1}{2K} \sum_{l=1}^{2K} [\theta^{(l)}, \theta^{(l)}]_{\mathcal{T}} + o_p(1). \tag{A.4}$$

Lemma 2 is proved at the end of this section.

Proof of Theorem 1. In view of Lemma 2, we calculate

$$\frac{1}{2K} \sum_{l=1}^{2K} [\theta^{(l,n)}, \theta^{(l,n)}]_{\mathcal{T}} = \int_0^{\mathcal{T}} g_t^{(n)} d[\theta,\theta]_s\ , \tag{A.5}$$

where

$$g_t^{(n)} = \frac{1}{2K}\sum_{l=1}^{2K}(f_t^{(l,n)})^2$$
$$= \frac{1}{2K^3(\Delta T)^2}\sum_{K\leq i\leq B-K}\left((T_{i+K}-t)^2 I\{T_{i+K}\geq t>T_i\}+(t-T_{i-K})^2 I\{T_i\geq t>T_{i-K}\}\right). \quad \text{(A.6)}$$

Under the conditions of the theorem, $K = K_n \to \infty$. Thus. for $t \in (T_{j-1}, T_j] \subseteq (T_K, T_{B-K}]$,

$$g_t^{(n)} = \frac{1}{2K^3(\Delta T)^2}\left(\sum_{j-K\leq i\leq j-1}(T_{i+K}-t)^2+\sum_{j\leq i\leq j+K-1}(t-T_{i-K})^2\right)$$
$$= \frac{1}{2K^3(\Delta T)^2}\left(\sum_{j-K\leq i\leq j-1}(T_{i+K}-T_j)^2+\sum_{j\leq i\leq j+K-1}(T_{j-1}-T_{i-K})^2\right)+O(K^{-1})$$
$$= \frac{1}{K^3}\sum_{k=1}^{K-1}k^2+O(K^{-1}) = \frac{1}{3}+O(K^{-1}), \quad \text{(A.7)}$$

where the $O(K^{-1})$ is uniform in $t \in (T_{K_n}, T_{B_n-K_n}]$, and hence, eventually, on all $[\delta, \mathcal{T}-\delta]$, for any $\delta > 0$. Since, for all $t \in [0, \mathcal{T}]$, $0 \leq g_t^{(n)} \leq 1$ (from (A.2)), and since $g_{\mathcal{T}}^{(n)} = 0$, Theorem 1 follows. □

We then obtain results for the continuous case. Lemmae 1-2 *do* cover the situation where K can be finite, including $K = 1$, which is identical with the situation in the following Proposition and then Remark.

PROPOSITION 4. *Let $\Delta T = T_{i+1} - T_i$ be independent of i. Assume that θ_t is a continuous semi-martingale. Then (6) holds.*

REMARK 8. (CONVERGENCE FAILS FOR FINITE K IN THE PRESENCE OF JUMPS.) The result in Proposition 4 will fail when θ has jumps, and this is one of the reasons why we move to subsampling and averaging. The right hand side of (A.4), however, has different behavior if there is discontinuity.

To see why, suppose for simplicity that θ_t is continuous except for a single jump at (stopping) time $\tau \in (0, \mathcal{T})$. Instead of (6), we get, as $\Delta T \to 0$ ((A.4)-(A.6)),

$$(\Delta T)^{-2}\sum_i(\Theta_{i+1}-\Theta_i)^2 \;=\; \frac{2}{3}\left([\theta^c,\theta^c]_{\mathcal{T}}-[\theta^c,\theta^c]_0\right)+\frac{1}{2}\left((1-U_n)^2+U_n^2\right)\Delta\theta_\tau^2+o_p(1), \quad \text{(A.8)}$$

where $U_n = (\tau - \tau_{n,*})/\Delta T_n$, where $\tau_{n,*} = \max_i\{i\Delta T < \tau\}$. If, for example, the jump happens at a Poisson time independent of the rest of the θ_t process, then one can proceed along the lines of Jacod and Protter (2012, Chapter 4.3) and get that U_n converges in law to a standard uniform random variable.

On the other hand, if τ is a non-random time, such as the time of the news release from a (U.S.) Federal Open Market Committee meeting,[46] the right hand side of (A.8) simply does not converge, in probability or law. – What is arguably even worse is that this effect does not cancel with a two scales construction. □

Before proving Lemma 2, we recall the following useful concept.

DEFINITION 1. THE CANONICAL DECOMPOSITION OF θ.) *We shall be using the canonical decomposition of θ_t (Jacod and Shiryaev (2003, Chapter II.2a pp. 75-76)), which is defined for a general semi-martingale (Ibid. Definition I.4.21, p. 43), by writing*[47]

$$\theta_t = \theta_0 + \theta(h)_t + B(h)_t + \breve{\theta}(h)_t. \tag{A.9}$$

Compared to the notation in our reference work, their X is our θ, their $M(h)$ is our $\theta(h)$, while their $B(h)$ is the same as ours. Also, let $\tilde{C}_t = \langle \theta(h), \theta(h) \rangle$. This is the "second modified characteristic" (Ibid., Definition II.2.16, p. 79). For the case of no truncation function, θ can similarly be decomposed into a local martingale and a finite variation process A_t. See also Ibid, p. 84, for further clarification of the relationship between the untruncated and the truncated processes. We let TV denote total variation,[48] *and set*

$$D(\theta)(h)_t = TV(\breve{\theta})_t - TV(\breve{\theta})_0 + TV(B(h))_t - TV(B(h))_0. \tag{A.10}$$

□

Proof of Lemma 2. Define

$$Z_{n,l}(t) = \sum_{T_{i+K} \le t, i \equiv l[2K]} (\theta^{(l)}_{T_{i+K}} - \theta^{(l)}_{T_{i-K}})^2 + (\theta^{(l)}_t - \theta^{(l)}_{T_{*,L}})^2 - [\theta^{(l)}, \theta^{(l)}]_{T_{*,L}}, \tag{A.11}$$

where $T_{*,L} = \max\{T_{i+K} \le t, i \equiv L[2K]\}$, so that

$$dZ_{n,l}(t) = 2(\theta^{(l)}_{t-} - \theta^{(l)}_{T_{*,L}}) d\theta^{(l)}_t. \tag{A.12}$$

For given truncation function h, define the processes $\theta^{(l)}_t(h) = \int_0^t f^{(l)}_s d\theta(h)_s$, $\breve{\theta}^{(l)}_t(h) = \int_0^t f^{(l)}_s d\breve{\theta}(h)_s$, etc. (The truncation is thus done on the original jumps, and not starting with the process $\theta^{(l)}_t$. This assures uniformity in the following argument.) Similarly, define $Z_{n,L}(h)(t) = 2(\theta^{(l)}_{t-} - \theta^{(l)}_{T_{*,L}}) d\theta^{(l)}(h)_t$, starting at $Z_{n,L}(h)(0) = Z_{n,L}(0) = 0$. Also set

$$Z_n(t) = \frac{1}{2K} \sum_{L=1}^{2K} Z_{n,L}(t) \text{ and } Z_n(h)(t) = \frac{1}{2K} \sum_{L=1}^{2K} Z_{n,L}(h)(t) \tag{A.13}$$

[46]At the time of writing, 2 pm Washington DC time, on the day of the meeting. This time appears to be defined to within single digit milliseconds. See, for example, "Fed probes for leaks ahead of policy news" (*Financial Times,* 24 September 2013).

[47]Note that we use a slightly different decomposition in the latter part of Appendix B, cf. Equation B.41 and Remark 10.

[48]As in Assumption 1 above. Jacod and Shiryaev denotes the total variation by Var.

Observe that $Z_n(\mathcal{T})$ = the left hand side of (A.4).

To bound the difference between $Z_n(t)$ and $Z_n(h)(t)$, note that

$$|Z_{n,L}(h)(t) - Z_{n,L}(t)| \;\le\; 2\int_0^t |\theta_{s-}^{(l)} - \theta_{T_{*,L}}^{(l)}| dD(h)_t \tag{A.14}$$

where $D(h)$ is defined in (A.10), and with the original θ. Also, in the notation of Jacod and Shiryaev (2003, Vi.1.8, p. 326), it follows from (A.2) that for all $t \in [0, \mathcal{T}]$,

$$\begin{aligned} |\theta_{t-}^{(l)} - \theta_{T_{*,L}}^{(l)}| \;&\le\; 2w'_{\mathcal{T}}(\theta^{(l)}(h), 2K\Delta T) \;+\; \sup_{T_{*,L}<s<t} |\Delta\theta_s| \\ &\le\; 2w'_{\mathcal{T}}(\theta^{(l)}(h), 2K\Delta T) \;+\; v_n(t) \end{aligned} \tag{A.15}$$

where $v_n(t) = \sup_{T_{**}<s<t} |\Delta\theta_s|$, with $T_{**} = \max\{T_{i+2K} \le t, \}$, so that

$$\sup_{0\le t\le \mathcal{T}} |Z_n(h)(t) - Z_n(t)| \;\le\; 4\max_{1\le L\le 2K} w'_{\mathcal{T}}(\theta^{(l)}(h), 2K\Delta T) D(h)(\mathcal{T}) \;+\; 2\int_0^{\mathcal{T}} v_n(t) dD(h)_t. \tag{A.16}$$

This is because the right hand side bounds $\sup_{0\le t\le\mathcal{T}} |Z_{n,L}(h)(t) - Z_{n,L}(t)|$ for each L, and thus the average.

Meanwhile, to assess the size of $Z_n(h)_t$, by similar argument,

$$\langle Z_n(h), Z_n(h)\rangle_{\mathcal{T}} \le 8\left(4\max_{1\le L\le 2K} w'_{\mathcal{T}}(\theta^{(l)}(h), 2K\Delta T)^2 \tilde{C}_T \;+\; \int_0^{\mathcal{T}} v_n^2(t) d\tilde{C}_t\right). \tag{A.17}$$

This is because the same bound applies to each $\langle Z_{n,L_1}(h), Z_{n,L_2}(h)\rangle_{\mathcal{T}}$.

Let $\epsilon > 0$ be given. Set

$$D^{\epsilon,d}(h)_t = \sum_{0<t\le t} \Delta D(h)_s I_{\{|\Delta D(h)_s|>\epsilon\}} \tag{A.18}$$

and let D_t^c be the continuous part of D (which is the same as the continuous part of $TV(B)_t$).

Let $N_\epsilon = \#\{t : |\Delta\theta_t|^2 \ge \epsilon \text{ or } \Delta C_t \ge \epsilon\}$. The number N_ϵ is finite with probability one, since $N_\epsilon \epsilon \le [\theta,\theta]_T + \tilde{C}_T < \infty$ a.s. Call these jumps $\tau_1, \cdots, \tau_{N_\epsilon}$. Set $\delta_\epsilon = \min\{\tau_{i+1} - \tau_i,\ 1 \le i \le N_\epsilon - 1\}$. $\delta_\epsilon > 0$ with probability one. When $2K\Delta T < \delta_\epsilon$ (and this does happen eventually, by assumption)

$$\begin{aligned} \int_0^{\mathcal{T}} v_n(t) dD(h)_t &\le \epsilon^{1/2} D(h)_T + \int_0^{\mathcal{T}} v_n(t) dD^{\sqrt{\epsilon},d}(h)_t + \sum_i \int_{\tau_i}^{(\tau_i+2K\Delta T)\wedge\mathcal{T}} v_n(t) dD^c(h)_t(h)_t \\ &\le \epsilon^{1/2} D(h)_T + \sqrt{[\theta,\theta]_{\mathcal{T}}}\left(D^{\sqrt{\epsilon},d}(h)_{\mathcal{T}} + \sum_i (D^c(h)_{(\tau_i+2K\Delta T)\wedge\mathcal{T}} - D^c(h)_{\tau_i})\right) \\ &\to \epsilon^{1/2} D(h)_T + \sqrt{[\theta,\theta]_{\mathcal{T}}} D^{\sqrt{\epsilon},d}(h)_{\mathcal{T}} \text{ as } n \to \infty. \end{aligned} \tag{A.19}$$

Hence $\limsup \int_0^{\mathcal{T}} v_n(t) dD(h)_t$ is bounded by the right hand side of (A.19), w.p. 1. However, since $\epsilon > 0$ was otherwise arbitrary, and since the right hand side tends to zero as $\epsilon \to 0$, it follows that

$$\int_0^{\mathcal{T}} v_n(t) dD(h)_t \to 0 \text{ almost surely as } n \to \infty. \tag{A.20}$$

Similarly but more informally, when $2K\Delta T < \delta_\epsilon$,

$$\begin{aligned} \int_0^{\mathcal{T}} v_n^2(t) d\tilde{C}_t &\leq \epsilon \tilde{C}_{\mathcal{T}} + [\theta,\theta]_{\mathcal{T}} \left(\tilde{C}_{\mathcal{T}}^{\epsilon,d} + \sum_i (\tilde{C}^c_{(\tau_i+2K\Delta T)\wedge\mathcal{T}} - \tilde{C}^c_{\tau_i}) \right) \\ &\to \epsilon \tilde{C}_{\mathcal{T}} + [\theta,\theta]_{\mathcal{T}} \tilde{C}_{\mathcal{T}}^{\epsilon,d} \text{ as } n \to \infty \\ &\to 0 \text{ as } \epsilon \to 0 \text{ a.s., for both convergences.} \end{aligned} \tag{A.21}$$

Finally, by Lemma 1 and by Jacod and Shiryaev (2003, Theorem VI.3.21, p. 350),

$$\max_{1 \leq L \leq 2K} w'_{\mathcal{T}}(\theta^{(l)}(h), 2K\Delta T) \xrightarrow{p} 0 \ \text{ as } n \to \infty. \tag{A.22}$$

Combining (A.16)-(A.17) with (A.20)-(A.22), we obtain, as $n \to \infty$,

$$\begin{aligned} \sup_{0 \leq t \leq \mathcal{T}} |Z_n(h)(t) - Z_n(t)| &\xrightarrow{p} 0 \text{ and} \\ \langle Z_n(h), Z_n(h) \rangle_{\mathcal{T}} &\xrightarrow{p} 0. \end{aligned} \tag{A.23}$$

From the second line in (A.23), by Lenglart's inequality (Jacod and Shiryaev (2003, Lemma I.3.30, p. 35)),

$$\sup_{0 \leq t \leq \mathcal{T}} |Z_n(h)(t)| \xrightarrow{p} 0. \tag{A.24}$$

Combining (A.24) with the first line of (A.23) yields the result of the Lemma, since $Z_n(\mathcal{T})$ = the left hand side of (A.4). □

B Proof of Theorem 3

To help clarify the structure of the proof, we show some intermediate results, and then assemble the pieces. We prove the result for the Case (i) where $K_n \to \infty$. In the Case (ii) where $K_n = O(1)$: For terms that are $o_p(n^{-2\alpha})$ in Case (i), these terms become $O_p(n^{-2\alpha})$ instead in Case (ii) (convergence to zero is replaced by tightness). For terms that have other orders, the situation under Case (ii) is mentioned explicitly.

We use the following notation. For $S < T$, set

$$\Theta'_{(S,T]} = \int_S^T (T-t) d\theta_t \text{ and } \Theta''_{(S,T]} = \int_S^T (t-S) d\theta_t \tag{B.25}$$

By Itô's formula, $\Theta_{(T,T+\delta]} - \Theta_{(T-\delta,T]} = \Theta'_{(T,T+\delta]} + \Theta''_{(T-\delta,T]}$.

B.1 Decomposition of Quadratic Variation into Variance and Quadratic Variation.

The following intermediate result, which generalizes (5) to K scales subsampling and averaging.

PROPOSITION 5. (DECOMPOSITION OF QUADRATIC VARIATION INTO VARIANCE AND QUADRATIC VARIATION.) *Suppose that Assumption 1-2 holds. Let $K = K_n$ be a sequence of integers so that (33) holds. Then*

$$\begin{aligned} QV_K = \frac{1}{K}\sum_{i=K}^{B-K}(\hat{\Theta}_{(T_i,T_{i+K}]} - \hat{\Theta}_{(T_{i-K},T_i]})^2 &= \frac{1}{K}\sum_{i=K}^{B-K}(\Theta_{(T_i,T_{i+K}]} - \Theta_{(T_{i-K},T_i]})^2 \\ &+ \frac{1}{K}\sum_{i=K}^{B-K}\left((\hat{\Theta}_{(T_i,T_{i+K}]} - \Theta_{(T_i,T_{i+K}]}) - (\hat{\Theta}_{(T_{i-K},T_i]} - \Theta_{(T_{i-K},T_i]})\right)^2 \\ &+ \frac{2}{K}V_1 + O_p\left(n^{-(\alpha+\beta)}\frac{K_n - J}{K_n^2}(B - 2K + 1)^{1/2}\right) + o_p(n^{-2\alpha}), \end{aligned} \tag{B.26}$$

where V_1 do not depend on K, specifically

$$V_1 = \sum_{i=2J}^{B}\Theta'_{(T_{i-J},T_i]}\tilde{e}'_{T_i} + \sum_{i=0}^{B-2J}\Theta''_{(T_i,T_{i+J}]}e'_{T_i} - \sum_{i=J}^{B-J}\left(\Theta'_{(T_i,T_{i+J}]} + \Theta''_{(T_{i-J},T_i]}\right)(e'_{T_i} + \tilde{e}'_{T_i}) \tag{B.27}$$

REMARK 9. It is easy to see from the proof that a more precise precise form of the $O_p\left(n^{-(\alpha+\beta)}\frac{K_n-J}{K_n^2}(B-2K+1)^{1/2}\right)$ term is

$$\begin{aligned} &2\frac{\Delta T(K_n - J)}{K_n}V_2 + o_p(n^{-2\alpha}), \text{ where} \\ &V_2 = -\sum_{i=J}^{B-J}(\theta_{T_{i+J}} - \theta_{T_{i-J}})(e'_{T_i} + \tilde{e}'_{T_i}) = O_p(n^{-\beta}B_n^{1/2}). \end{aligned} \tag{B.28}$$

V_2 does also not depend on K. Though this representation does not help with Theorem 3, it will be handy for Theorem 4, were the first order part becomes

$$2\Delta T V_2 \sum_{l=1}^{m}\gamma_{n,l}\frac{(K_{n,l} - J)}{K_{n,l}} \;=\; 2\Delta T V_2 \gamma_n \;=\; O_p(\gamma_n n^{-\beta}B_n^{-1/2}). \tag{B.29}$$

by (43). We treat V_2 as bias, which is a worst case scenario, and a loose choice of words.[49] For example, if the edge effects are independent of the θ_t process, this term will be much smaller, and it is can be handled as a variance type term. □

[49] Our motivation for the term is that is the general case, V_2 has nonzero mean. We also conjecture that it can be consistently estimated from the data.

Proof of Proposition 5. To show the result of the proposition, we need to get rid of the cross term, *i.e.*

$$\frac{1}{K}\sum_{i=K}^{B-K}(\Theta_{(T_i,T_{i+K}]}-\Theta_{(T_{i-K},T_i]})(\hat{\Theta}_{(T_i,T_{i+K}]}-\Theta_{(T_i,T_{i+K}]}-(\hat{\Theta}_{(T_{i-K},T_i]}-\Theta_{(T_{i-K},T_i]})). \tag{B.30}$$

We divide the proof into parts (A)-(C). The overall result follows by combining (B.34), (B.61), and (B.64), below.

(A) We start with the part of (B.30) due to M, which is four times

$$\begin{aligned}&\frac{1}{2K}\sum_{i=K}^{B-K}(\Theta_{(T_i,T_{i+K}]}-\Theta_{(T_{i-K},T_i]})\left((M_{T_{i+K}}-M_{T_i})-(M_{T_i}-M_{T_{i-K}})\right)\\ &\quad=\frac{K\Delta T}{2K}\sum_{l=1}^{2K}\sum_{K\le i\le B-K, i\equiv l[2K]}(\theta^{(l)}_{T_{i+K}}-\theta^{(l)}_{T_{i-K}})\left((M_{T_{i+K}}-M_{T_i})-(M_{T_i}-M_{T_{i-K}})\right),\end{aligned} \tag{B.31}$$

in the notation of Appendix A. We show

LEMMA 3. *Under the conditions of Proposition 5,*

$$\begin{aligned}\textit{l.h.s. of (B.31)} &= \frac{n^{-\alpha}K\Delta T}{2K}\sum_{l=1}^{2K}\sum_{K\le i\le B-K, i\equiv l[2K]}\left(([\theta^{(l)},L_n]_{T_{i+K}}-[\theta^{(l)},L_n]_{T_i})-([\theta^{(l)},L_n]_{T_i}-[\theta^{(l)},L_n]_{T_{i-K}})\right)\\ &\quad+\ o_p(n^{-2\alpha}).\end{aligned} \tag{B.32}$$

The right hand side of (B.32) is a telescope sum. Thus, by the Kunita-Watanabe Inequality (see, *e.g.,* Protter (2004, Theorem II.25 (p. 69))), if $2K\Delta T\le\delta$, (B.32) yields

$$\begin{aligned}\text{l.h.s. of (B.31)}\quad &\le n^{-\alpha}K\Delta T\,(([\theta,\theta]_\delta-[\theta,\theta]_0)[L_n,L_n]_\delta\\ &+([\theta,\theta]_{\mathcal{T}-}-[\theta,\theta]_{\mathcal{T}-\delta})([L_n,L_n]_{\mathcal{T}}-[L_n,L_n]_{\mathcal{T}-\delta}))\ +\ o_p(n^{-2\alpha}).\end{aligned} \tag{B.33}$$

By the stable convergence and the P-UT property of $L_{n,t}$ (Appendix D.1), Jacod and Shiryaev (2003, Theorem VI.6.26, p. 384) yields that $(n^\alpha/K\Delta T)\times$ the right hand side of (B.33) converges in law to $([\theta,\theta]_\delta-[\theta,\theta]_0)[L,L]_\delta+([\theta,\theta]_{\mathcal{T}-}-[\theta,\theta]_{\mathcal{T}-\delta})([L,L]_{\mathcal{T}}-[L,L]_{\mathcal{T}-\delta})$. This expression converges in probability to zero as $\delta\to 0$. Thus (Jacod and Shiryaev (2003, Property VI.3.1, p. 347))

$$\text{l.h.s. of (B.31)}\ = \o_p(n^{-2\alpha}). \tag{B.34}$$

Proof of Lemma 3. To see (B.32), we can proceed as in the proof of Lemma 2. This involves processes of the form $dZ^{\text{cross}}_{n,l}(t)=2(\theta^{(l)}_{t-}-\theta^{(l)}_{T_{*,L}})d(L_n+A_n)_t$ and $dZ^{\text{cross}}_{n,l}(h)(t)=2((L_n+A_n)_{t-}-(L_n+A_n)_{T_{*,L}})d\theta^{(l)}_t$, as well as the jump-truncated local martingale versions $dZ^{\text{cross}}_{n,l}(h)(t)=2(\theta^{(l)}_{t-}-\theta^{(l)}_{T_{*,L}})dL_{n,t}(h)$ and $dZ^{\text{cross}}_{n,l}(h)(t)=2((L_n+A_n)_{t-}-(L_n+A_n)_{T_{*,L}})d\theta^{(l)}(h)_t$. Still in analogy with the proof of the earlier lemma, we need to show properties for the $L_{n,t} + A_{n,t}$ process that mirror those shown there for the $\theta_t^{(l)}$ processes. The $A_{n,t}$ process can be ignored since we have assumed that $TV(A_n)_{\mathcal{T}} \xrightarrow{p} 0$. That leaves us to deal with the elements of the decomposition (A.9)

$$L_{n,t} = L_{n,0} + L(h)_{n,t} + B^L(h)_{n,t} + \breve{L}(h)_{n,t}. \tag{B.35}$$

Inequalities analogous to (A.16)-(A.17) hold in our case. We now show that these bounds go to in probability to something arbitrarily small.

However, $L_{n,t}$ converges to L_t by assumption, and the sequence is P-UT by Appendix D.1. Therefore, in notation similar to the proof of the earlier lemma, and by Jacod and Shiryaev (2003, Theorem VI.3.21, p. 350),

$$w'_{\mathcal{T}}(L(h)_n, 2K\Delta T) \xrightarrow{p} 0 \quad \text{as } n \to \infty. \tag{B.36}$$

while $D_{L_n}(h)_{\mathcal{T}}$ and $\tilde{C}_{\mathcal{T}}^{L_n}$ are tight by Jacod and Shiryaev (2003, Theorem VI.6.16, p. 380). Because of the stable convergence, therefore, terms (in the bounds) involving $\max_{1\le L\le 2K} w'_{\mathcal{T}}(\theta^{(l)}(h), 2K\Delta T)$ and $w'_{\mathcal{T}}(L(h)_n, 2K\Delta T)$ go to zero in probability. This leaves us to deal with with the following four terms:

$$\int_0^{\mathcal{T}} v_n^L(t) dD(h)_t, \quad \int_0^{\mathcal{T}} v_n(t) dD_{L_n}(h)_t, \quad \int_0^{\mathcal{T}} v_n^2(t) d\tilde{C}_t^{L_n}, \text{ and } \int_0^{\mathcal{T}} (v_n^L)^2(t) d\tilde{C}_t \;. \tag{B.37}$$

The convergence in law of $L_{n,t}$ causes slight additional complications. Focus on the first term in (B.37). The three other terms follow analogously. Terms two and four also use Jacod and Shiryaev (2003, Theorem VI.6.15, p. 380), and proceeds through subsequences to convert tightness to convergence in law.

Let $\eta_1 > 0$, and set

$$Y_t^{L_n} \;=\; \sum_{s\le t} g(\Delta L_{n,s}) \text{ and } Y_t^L \;=\; \sum_{s\le t} g(\Delta L_s) \tag{B.38}$$

where g is continuous, nonnegative, and vanishes in a neighborhood of zero, $g(-x) = g(x)$, and $g(x) = x$ for $x \ge \eta$. Also let $\eta_2 > 0$, and let $\mathcal{G}$ is a finite grid on $[0, \mathcal{T}]$, which includes 0 and $\mathcal{T}$, with distance between the grid points no more than η_2. For all $t \in [0, \mathcal{T}]$, let $T_{***} = \max\{T \in \mathcal{G} \,:\, T \le t - \eta_2\} \wedge 0$. When $2K\Delta T \le \eta_2$, $T_{***} \le T_{**}$. We see that $v_n^L(t) \le \eta + Y_{t-} - Y_{T_{***}}$. And so, for $2K\Delta T \le \eta_2$,

$$\int_0^{\mathcal{T}} v_n^L(t) dD(h)_t \le \eta_1 D(h)_{\mathcal{T}} + \int_0^t (Y_{t-}^{L_n} - Y_{T_{***}}^{L_n}) dD(h)_t. \tag{B.39}$$

Because of the stable convergence assumption on the process $L_{n,t}$, it follows that the processes $L_{n,t}$ and $D(h)_t$ converge jointly in law. By Jacod and Shiryaev (2003, Proposition VI.3.16), it follows that the processes $Y_t^{L_n}$ and $D(h)_t$ converge jointly in law. By *Ibid.*, Theorem VI.6.22 (p. 383), the right hand side of (B.39) converges in law to $\eta_1 D(h)_{\mathcal{T}} + \int_0^t (Y_{t-}^L - Y_{T_{***}}^L) dD(h)_t$, so that, for any $\eta_3 > 0$,

$$\limsup_{n\to\infty} P(\int_0^{\mathcal{T}} v_n^L(t) dD(h)_t \ge \eta_3) \le P(\eta_1 D(h)_{\mathcal{T}} + \int_0^t (Y_{t-}^L - Y_{T_{***}}^L) dD(h)_t \ge \eta_3) \tag{B.40}$$

(Jacod and Shiryaev (2003, Property VI.3.1, p. 347)). We have now overcome the problems associated with the convergence in law of $L_{n,t}$, and can proceed on the right hand side of (B.40) as in the proof of Lemma 2, though taking into account the jumps both in L_t and in θ_t. By sending η_2 and then η_1 to zero, we obtain that first term in (B.37) tends to zero in probability.

As the three other terms in (B.37) are dealt with similarly, Lemma 3 is proved. □

With reference to Definition 1, for the purposes of the remainder of this proof, unlike Appendix A, we use (A.10)

$$D(\theta)(h)_t = TV(\breve{\theta})_t - TV(\breve{\theta})_0 + TV(B(h)-A)_t - TV(B(h)-A)_0 \tag{B.41}$$

REMARK 10. (PROPERTIES). We note that the *truncation function* h is bounded, so $\theta(h)_t$ and $B(h)_t$ have bounded jumps, and hence are locally bounded. One can choose $h = h_n$ so that $D(\theta)(h)_{\mathcal{T}}$ is arbitrarily small, *i.e.,* $D(\theta)(h_n)_{\mathcal{T}} \xrightarrow{p} 0$. □

(B) We continue with the part of (B.30) that is due to the $(e'_{T_i}, \tilde{e}'_{T_i})$**.** This part can be written as two times

$$\frac{1}{K}\sum_{i=K}^{B-K}(\Theta'_{(T_i,T_{i+K}]} + \Theta''_{(T_{i-K},T_i]})\left((\tilde{e}'_{T_{i+K}} - e'_{T_i}) - (\tilde{e}'_{T_i} - e'_{T_{i-K}})\right). \tag{B.42}$$

Now again invoke the canonical decomposition of θ_t from Definition 1. The for the part of (B.42) that is due to $\breve{\theta}_{n,t}$, write, in obvious notation, and by invoking (D.81) in Remark 12,

$$\begin{aligned}
&|\frac{1}{K}\sum_{i=K}^{B-K}(\breve{\Theta}'_{(T_i,T_{i+K}]} + \breve{\Theta}''_{(T_{i-K},T_i]})\left((\tilde{e}'_{T_{i+K}} - e'_{T_i}) - (\tilde{e}'_{T_i} - e'_{T_{i-K}})\right)| \\
&\le 4n^{-\alpha}\Gamma\frac{1}{K}\sum_{i=K}^{B-K}|\breve{\Theta}'_{(T_i,T_{i+K}]} + \breve{\Theta}''_{(T_{i-K},T_i]}| \\
&\le 4n^{-\alpha}\Gamma\frac{1}{K}\sum_{i=K}^{B-K}\left(\int_{T_i}^{T_{i+K}}(T_{i+K}-t)dTV(\breve{\theta})_t\int_{T_{i-K}}^{T_i}(t-T_i)dTV(\breve{\theta})_t\right) \\
&\le 16n^{-\alpha}\Gamma K\Delta T(TV(\breve{\theta})_{\mathcal{T}} - TV(\breve{\theta})_0).
\end{aligned} \tag{B.43}$$

For the first term, this is because

$$4n^{-\alpha}\Gamma\frac{1}{K}\sum_{i=K}^{B-K}\int_{T_i}^{T_{i+K}}(T_{i+K}-t)dTV(\breve{\theta})_t \le 8n^{-\alpha}\Gamma K\Delta T(TV(\breve{\theta})_{\mathcal{T}} - TV(\breve{\theta})_0). \tag{B.44}$$

The other term is handled similarly.

In the same way,

$$\begin{aligned}
&|\frac{1}{K}\sum_{i=K}^{B-K}(\Theta^{(B-A)\prime}_{(T_i,T_{i+K}]} + \breve{\Theta}^{(B-A)\prime\prime}_{(T_{i-K},T_i]})\left((\tilde{e}'_{T_{i+K}} - e'_{T_i}) - (\tilde{e}'_{T_i} - e'_{T_{i-K}})\right)| \\
&\le 16n^{-\alpha}\Gamma K\Delta T(TV(B-A)_{\mathcal{T}} - TV(B-A)_0),
\end{aligned} \tag{B.45}$$

We shall use the two bounds in the following. For now, concentrate on the remaining part of (B.42),

$$\frac{1}{K}\sum_{i=K}^{B-K}(\Theta(h)'_{(T_i,T_{i+K}]}+\Theta(h)''_{(T_{i-K},T_i]})\left((\tilde{e}'_{T_{i+K}}-e'_{T_i})-(\tilde{e}'_{T_i}-e'_{T_{i-K}})\right), \tag{B.46}$$

where $\Theta(h)'_{(T_i,T_{i+K}]}$ and $\Theta(h)''_{(T_{i-K},T_i]}$ are defined as $\Theta'_{(T_i,T_{i+K}]}$ and $\Theta''_{(T_{i-K},T_i]}$, but with $\theta(h)+B(h)$ replacing θ. To make things more readable, we split up into more lemmae.

Lemma 4. (The terms in (B.46) that cannot be written on martingale form.) *These terms are the sum of (B.48)-(B.49) below. They satisfy*

$$\frac{1}{K_n}V_1+O_p\left(n^{-(\alpha+\beta)}\frac{K_n-J}{K_n^2}(B_n-2K_n+1)^{1/2}\right)+O_p(n^{-2\alpha}) \tag{B.47}$$

where V_1 is defined in (B.27).

Proof and Elaboration of Lemma 4. We show the result for the case where $\beta=\alpha$. The argument for general β is given in Remark 9. – We are concerned with the part of (B.46) that *cannot* be written on martingale form. Consider first the $\tilde{e}'$ terms, which add up to[50]

$$\begin{aligned}&\frac{1}{K}\sum_{i=K}^{B-K}\Theta(h)'_{(T_{i+K-J},T_{i+K}]}\tilde{e}'_{T_{i+K}}-\frac{1}{K}\sum_{i=K}^{B-K}\left(\int_{T_i}^{T_{i+J}}(T_{i+K}-t)d(\theta+B)(h)_t+\int_{T_{i-J}}^{T_i}(t-T_{i-K})d(\theta+B)(h)_t\right)\tilde{e}'_{T_i}\\ &=\frac{1}{K}\sum_{i=2K}^{B}\Theta(h)'_{(T_{i-J},T_i]}\tilde{e}'_{T_i}-\frac{1}{K}\sum_{i=K}^{B-K}\left(\Delta T(K-J)((\theta+B)(h)_{T_{i+J}}-(\theta+B)(h)_{T_{i-J}})+\Theta(h)'_{(T_i,T_{i+J}]}+\Theta(h)''_{(T_{i-J},T_i]}\right)\tilde{e}'_{T_i}\end{aligned} \tag{B.48}$$

Similarly, the non-martingale terms containing e' add up to

$$\frac{1}{K}\sum_{i=0}^{B-2K}\Theta(h)''_{(T_i,T_{i+J}]}e'_{T_i}-\frac{1}{K}\sum_{i=K}^{B-K}\left(\Delta T(K-J)((\theta+B)(h)_{T_{i+J}}-(\theta+B)(h)_{T_{i-J}})+\Theta(h)'_{(T_i,T_{i+J}]}+\Theta(h)''_{(T_{i-J},T_i]}\right)e'_{T_i} \tag{B.49}$$

We first dispose of (taken from both (B.48)-(B.49))

$$\begin{aligned}&!\frac{1}{K}\sum_{i=K}^{B-K}\left(\Delta T(K-J)((\theta+B)(h)_{T_{i+J}}-(\theta+B)(h)_{T_{i-J}})\right)(e'_{T_i}+\tilde{e}'_{T_i})|\\ &\leq n^{-\alpha}\Delta T(K-J)2\Gamma\frac{1}{K}\sum_{i=K}^{B-K}|(\theta+B)(h)_{T_{i+J}}-(\theta+B)(h)_{T_{i-J}}|\\ &\leq n^{-\alpha}\Delta T(K-J)2\Gamma\frac{1}{K}\left((B-2K+1)^{1/2}\sum_{i=K}^{B-K}(\theta(h)_{T_{i+J}}-\theta(h)_{T_{i-J}})^2+\sum_{i=K}^{B-K}|B(h_{T_{i+J}}-B(h)_{T_{i-J}}|\right)\\ &=O_p\left(n^{-2\alpha}\frac{K_n-J}{K_n^2}(B_n-2K_n+1)^{1/2}\right)\end{aligned} \tag{B.50}$$

---

[50]Since $\int_{T_i}^{T_{i+J}}(T_{i+K}-t)d\theta_t=(T_{i+K}-T_{i+J})(\theta_{T_{i+J}}-\theta_{T_i})+\Theta'_{(T_i,T_{i+J}]}=\Delta T(K-J)(\theta_{T_{i+J}}-\theta_{T_i})+\Theta'_{(T_i,T_{i+J}]}$ and similarly $\int_{T_{i-K}}^{T_i}(t-T_{i-J})d\theta_t=\Delta T(K-J)(\theta_{T_i}-\theta_{T_{i-J}})+\Theta''_{(T_{i-J},T_i]}$, with the same results for $\theta(h)+B(h)$.

The sum of (B.48)-(B.49)) is then equal to

$$\begin{aligned}
&\frac{1}{K}\sum_{i=2J}^{B}\Theta(h)'_{(T_{i-J},T_i]}\tilde{e}'_{T_i} + \frac{1}{K}\sum_{i=0}^{B-2J}\Theta(h)''_{(T_i,T_{i+J}]}e'_{T_i} \\
&-\frac{1}{K}\sum_{i=J}^{B-J}\left(\Theta(h)'_{(T_i,T_{i+J}]} + \Theta(h)''_{(T_{i-J},T_i]}\right)(e'_{T_i} + \tilde{e}'_{T_i}) \\
&+ O_p\left(n^{-2\alpha}\frac{K-J}{K^2}(B-2K+1)^{1/2}\right) + o_p(n^{-2\alpha}). \qquad \text{(B.51)}
\end{aligned}$$

The change of summation limits in the remaining sums is valid because, for example,

$$\begin{aligned}
&|\frac{1}{K}\sum_{i=K}^{B-K}\Theta(h)'_{(T_{i+K-J},T_{i+K}]}\tilde{e}'_{T_{i+K}} - \frac{1}{K}\sum_{i=2J}^{B}\Theta(h)'_{(T_{i-J},T_i]}\tilde{e}'_{T_i}| \leq \frac{1}{K}\sum_{i=2J}^{2K-1}|\Theta(h)'_{(T_{i-J},T_i]}||\tilde{e}'_{T_i}| \\
&\leq n^{-\alpha}\Gamma\frac{1}{K}\sum_{i=2J}^{2K-1}|\Theta(h)'_{(T_{i-J},T_i]}| \\
&\leq n^{-\alpha}\Gamma\frac{1}{K}\sum_{i=2J}^{2K-1}|\int_{T_{i-J}}^{T_i}(T_i-t)d\theta(h)_t| + O_p(n^{-\alpha}\Delta T) \qquad \text{(B.52)}
\end{aligned}$$

by invoking (D.81) in Remark 12. However, the square of the main term in (B.52) is Lenglart dominated by

$$\begin{aligned}
&n^{-2\alpha}\Gamma^2\left(\frac{2(K-J)}{K^2}\right)\sum_{i=2J}^{2K-1}\int_{T_{i-J}}^{T_i}(T_i-t)^2 d\tilde{C}_t \\
&\qquad \leq n^{-2\alpha}\Gamma^2\left(\frac{2(K-J)}{K^2}\right)\sum_{j=1}^{J} j^2(\Delta T)^2(\tilde{C}_{T_{2K-1}} - \tilde{C}_{T_J}) = o_p(n^{-4\alpha}), \qquad \text{(B.53)}
\end{aligned}$$

where we have used a similar kind of bound to that used in (B.43)-(B.45) (but with $(T_i - t)^2$ rather than $(T_i - t)$). Hence, by Lenglart's inequality (Jacod and Shiryaev (2003, Lemma I.3.20, p. 35)), the main term in (B.52) is of order $o_p(n^{-2\alpha})$. The other terms follow similarly, hence (B.48)-(B.49) equals (B.51).

Now write (B.51) as

$$\frac{1}{K}V_1(h) + O_p\left(n^{-2\alpha}\frac{1}{K}(B-2K+1)^{1/2}\right) + o_p(n^{-2\alpha}), \qquad \text{(B.54)}$$

where $V_1(h)$ is as in (B.27), but with the modification that it is based on $\theta(h)$ instead of θ. – Similarly to (B.43)-(B.45), we get that $|V_1(h) - V_1| \leq 16n^{-\alpha}\Gamma K\Delta T(TV(\breve{\theta})_{\mathcal{T}} - TV(\breve{\theta})_0 + TV(B)_{\mathcal{T}} - TV(B)_0)$. This yields that (B.51) can be further written as (B.47) in the Lemma, since $K\Delta T = O(n^{-\alpha})$. □

Lemma 5. (The terms in (B.46) that can be written on martingale form.) *These are the remaining terms in (B.46), minus the sum of (B.48)-(B.49) below. They are of order* $O_p(n^{-2\alpha}K_n^{-1/2})$.

Proof and Elaboration of Lemma 5. For the remaining terms in (B.46), if $K \to \infty$, write them as (B.42) with $\Theta(h)^{R\prime}_{(T_i,T_{i+K}]} + \Theta(h)^{R\prime\prime}_{(T_{i-K},T_i]}$ replacing $\Theta(h)'_{(T_i,T_{i+K}]} + \Theta(h)''_{(T_{i-K},T_i]}$. For the first (the $\tilde{e}'_{T_{i+K}}$) term, proceed as follows. The term (with $I = B$) can be written

$$\frac{1}{K}\sum_{i=K}^{I-K}(\Theta(h)^{R\prime}_{(T_i,T_{i+K}]} + \Theta(h)^{R\prime\prime}_{(T_{i-K},T_i]})\tilde{e}'_{T_{i+K}} \tag{B.55}$$

Use (D.81) in Remark 12. First of all, the finite variation terms (involving $B(h)$) yield a contribution of size $O_p(n^{-\alpha}\Delta T)$. Hence assume that $\Theta(h)^{R\prime}_{(T_i,T_{i+K}]} + \Theta(h)^{R\prime\prime}_{(T_{i-K},T_i]}$ only contains the martingale $\theta(h)$. Hence, by Lemma 7 (choose $N = 2J$), the square of (B.55) is Lenglart dominated by

$$n^{-2\alpha}\Gamma^2(4J-1)\frac{1}{K^2}\sum_{i=K}^{I-K}(\Theta_{(T_i,T_{i+K}]}(h)^{R\prime} + \Theta_{(T_{i-K},T_i]}(h)^{R\prime\prime})^2, \tag{B.56}$$

which in turn (for simplicity) is Lenglart dominated by

$$n^{-2\alpha}\Gamma^2(4J-1)\frac{1}{K^2}\sum_{i=K}^{I-K}(\Theta_{(T_i,T_{i+K}]}(h)' + \Theta_{(T_{i-K},T_i]}(h)'')^2. \tag{B.57}$$

As in (B.53), this expression is in turn Lenglart dominated by

$$\begin{aligned}
&2n^{-2\alpha}\Gamma^2(4J-1)\frac{1}{K^2}\sum_{i=K}^{I-K}\left(\int_{T_i}^{T_{i+K}}(T_{i+K}-t)^2 d\tilde{C}_t + \int_{T_{i-K}}^{T_i}(t-T_{i-K})^2 d\tilde{C}_t\right)\\
&\le 4n^{-2\alpha}\Gamma^2(4J-1)\frac{1}{K^2}\sum_{j=1}^{K} j^2(\Delta T)^2\left(\tilde{C}_{T_I} - \tilde{C}_0\right)\\
&= O_p(n^{-4\alpha}K_n^{-1}).
\end{aligned} \tag{B.58}$$

Hence, by Lenglart's inequality (Jacod and Shiryaev (2003, Lemma I.3.20, p. 35)), (B.55) is of order $O_p(n^{-2\alpha}K_n^{-1/2})$. The other martingale part of terms in (B.46) can be handled similarly. This shows Lemma 5. □

To round up the discussion of (B.42), we see from Lemmae 4-5 that

$$\text{eq. (B.46)} - \frac{1}{K_n}V_1 + O_p\left(n^{-2\alpha}\frac{1}{K}(B-2K+1)^{1/2}\right) + O_p(n^{-2\alpha}K_n^{-1}) \tag{B.59}$$

Meanwhile, from (B.43)-(B.45)

$$n^{2\alpha}|\text{ eq. (B.46)} - \text{eq. (B.42)}| \le 16n^{\alpha}(K\Delta T)\Gamma\left(TV(\breve{\theta})_{\mathcal{T}} - TV(\breve{\theta})_0 + TV(B)_{\mathcal{T}} - TV(B)_0.\right) \tag{B.60}$$

$n^{\alpha}(K\Delta T) = O(1)$ by assumption, and the truncation function h can be chosen so that $TV(\breve{\theta})_{\mathcal{T}} - TV(\breve{\theta})_0 + TV(B-A)_{\mathcal{T}} - TV(B-A)_0 \xrightarrow{p} 0$ (cf. Remark 10). Thus, finally,

$$\text{eq. (B.42)} = n^{-2\alpha}\frac{1}{K_n}V_1 + O_p\left(n^{-2\alpha}\frac{1}{K}(B-2K+1)^{1/2}\right) + O_p(n^{-2\alpha}K_n^{-1}) \tag{B.61}$$

□

(C) We continue with (B.30), now the part due to the $(e''_{T_i}, \tilde{e}''_{T_i})$**.** This part can be written

$$\frac{1}{K}\sum_{i=K}^{B-K}(\Theta'_{(T_i,T_{i+K}]}+\Theta''_{(T_{i-K},T_i]})\left((\tilde{e}''_{T_{i+K}}-e''_{T_i})-(\tilde{e}''_{T_i}-e''_{T_{i-K}})\right) \tag{B.62}$$

Again just focus on one of the sub-terms, say, the one due to $\tilde{e}''_{T_{i+K}}$. From Cauchy-Schwartz,

$$\begin{aligned}
&|\frac{1}{K}\sum_{i=K}^{B-K}(\Theta'_{(T_i,T_{i+K}]}+\Theta''_{(T_{i-K},T_i]})\tilde{e}''_{T_{i+K}}| \\
&\leq \frac{1}{K}\left(\sum_{i=K}^{B-K}(\Theta'_{(T_i,T_{i+K}]}+\Theta''_{(T_{i-K},T_i]})^2\right)^{1/2}\left(\sum_{i=K}^{B-K}(\tilde{e}''_{T_{i+K}})^2\right)^{1/2} \\
&= o_p\left(K^{-1}(K\Delta T)^3B)^{1/2}n^{-\alpha}\right) = o_p\left(K\Delta T n^{-\alpha}\right) \\
&= o_p\left(n^{-2\alpha}\right)
\end{aligned} \tag{B.63}$$

ex. hyp.. The other terms go away similarly. Thus

$$\text{eq. (B.62)} \ = \ o_p(n^{-2\alpha}). \tag{B.64}$$

As indicated at the beginning of the proof, Proposition 5 is thus proved. □

B.2 Representation of the Variance Term

Given Proposition 5, the main outstanding problem is to handle the asymptotic variance term in (B.26).

PROPOSITION 6. (REPRESENTATION OF THE VARIANCE TERM.) *Suppose that Assumption 1-2 holds. Let $K = K_n$ be a sequence of integers so that (33) holds. Then*

$$\begin{aligned}
&\frac{1}{K}\sum_{i=K}^{B-K}\left((\hat{\Theta}_{(T_i,T_{i+K}]}-\Theta_{(T_i,T_{i+K}]})-(\hat{\Theta}_{(T_{i-K},T_i]}-\Theta_{(T_{i-K},T_i]})\right)^2 \\
&= 2n^{-2\alpha}([L,L]_{\mathcal{T}}-[L,L]_0)-n^{-2\alpha}\left(2\mathcal{E}_{\tilde{e}^2}(0)+\mathcal{E}_{(\tilde{e}+e)^2}(0)+\mathcal{E}_{(\tilde{e}+e)^2}(\mathcal{T})+2\mathcal{E}_{e^2}(\mathcal{T})\right) \\
&+\frac{1}{K}\text{TSE}_n+\frac{2}{K}V'_1+o_p(n^{-2\alpha})+O_p(n^{-2\beta}K_n^{-1}(B_n-2K_n+1)^{1/2})
\end{aligned} \tag{B.65}$$

where TSE_n *denotes the total squared error (for the smallest possible K), given by*

$$\text{TSE}_n = 2\sum_{i=0}^{B}(\tilde{e}^2_{T_i}+e^2_{T_i}+\tilde{e}_{T_i}e_{T_i}) \tag{B.66}$$

(with $\tilde{e}_0 = e_{\mathcal{T}} = 0$ by convention), and where

$$V_1' = \sum_{i=J}^{B-J} (M_{T_i} - M_{T_{i-J}})(e_{T_i} + 2\tilde{e}_{T_i}) - \sum_{i=J}^{B-J} (M_{T_i} - M_{T_{i-J}})(2e_{T_i} + \tilde{e}_{T_i}) \tag{B.67}$$

Proof of Proposition 6. Following (15), write

$$\begin{aligned}
&\frac{1}{K}\sum_{i=K}^{B-K} \Big((\hat{\Theta}_{(T_i,T_{i+K}]} - \Theta_{(T_i,T_{i+K}]}) - (\hat{\Theta}_{(T_{i-K},T_i]} - \Theta_{(T_{i-K},T_i]}) \Big)^2 \\
&= \frac{1}{K}\sum_{i=K}^{B-K} \big((M_{T_{i+K}} - M_{T_i}) - (M_{T_i} - M_{T_{i-K}}) + (\tilde{e}_{T_{i+K}} - e_{T_i} - \tilde{e}_{T_i} + e_{T_{i-K}}) \big)^2 \\
&= \frac{1}{K}\sum_{i=K}^{B-K} \Big((M_{T_{i+K}} - M_{T_i})^2 + (M_{T_i} - M_{T_{i-K}})^2 + \tilde{e}^2_{T_{i+K}} + (e_{T_i} + \tilde{e}_{T_i})^2 + e^2_{T_{i-K}} \Big) + \mathrm{crt}_n \\
&= n^{-2\alpha}\frac{1}{K}\sum_{i=K}^{B-K} \big((L_{n,T_{i+K}} - L_{n,T_i})^2 + (L_{n,T_i} - L_{n,T_{i-K}})^2 \big) + \frac{1}{K}\mathrm{TSE}_n + \mathrm{crt}_n \\
&- \frac{1}{K}\left(\sum_{i=0}^{2K-1} \tilde{e}^2_{T_i} + \sum_{i=B-2K+1}^{B} e^2_{T_i} + \sum_{i=0}^{K-1} (\tilde{e}_{T_i} + e_{T_i})^2 + \sum_{i=B-K+1}^{B} (\tilde{e}_{T_i} + e_{T_i})^2 \right) + o_p(n^{-2\alpha}) \\
&= 2n^{-2\alpha}([L,L]_{\mathcal{T}} - [L,L]_0) - n^{-2\alpha} \big(2\mathcal{E}_{\tilde{e}^2}(0) + \mathcal{E}_{(\tilde{e}+e)^2}(0) + \mathcal{E}_{(\tilde{e}+e)^2}(\mathcal{T}) + 2\mathcal{E}_{e^2}(\mathcal{T}) \big) \\
&+ \frac{1}{K}\mathrm{TSE}_n + \mathrm{crt}_n + o_p(n^{-2\alpha})
\end{aligned} \tag{B.68}$$

where crt_n are the cross-terms, and where TSE_n is given by (B.66).

The last transition: for the edge effect erms, this is by (31) in Assumption 2. Meanwhile, to see why, for example, $\frac{1}{K}\sum_{i=K}^{B-K}(L_{n,T_{i+K}} - L_{n,T_i})^2 \xrightarrow{p} ([L,L]_{\mathcal{T}} - [L,L]_0)$, set $L^{(l)}_{n,t} = L_{n,T_{i-L}}$, where $T_i \le t < T_{i+1}$. Also, w.l.o.g., set $[L,L]_0) = 0$. Then

$$\frac{1}{K}\sum_{i=K}^{B-K} (L_{n,T_{i+K}} - L_{n,T_i})^2 = \frac{1}{K}\sum_{L=1}^{K} [L_n^{(l)}, L_n^{(l)}]_{\mathcal{T}} + o_p(1) \tag{B.69}$$

For each l, Jacod and Shiryaev (2003, Proposition VI.6.37, p.387), along with the Assumption (17), assures that $[L_n^{(l)}, L_n^{(l)}]_{\mathcal{T}} \xrightarrow{p} [L,L]_{\mathcal{T}}$. For a given subsequence of n, by iterative further picking of subsequences, one can find a subsequence so that this convergence holds almost surely for all l.

By Toeplitz Lemma, *e.g.,* Hall and Heyde (1980, p. 31), it follows that the resulting subsequence of (B.69) converges a.s. to $[L,L]_{\mathcal{T}}$. Since the initial subsequence was arbitrary, it follows that (B.69) converges in probability to $[L,L]_{\mathcal{T}}$, as required. (One can alternatively proceed along the lines of Lemmae 2-3.)

Cross terms. Without loss of generality we can replace e by e'. – We first handle the terms that only contain e_{T_i} and $\tilde{e}_{T_i}$. One such term (and the others are all handled the same way) is

$\frac{2}{K}\sum_{i=K}^{B-K}\tilde{e}_{T_{i+K}}e_{T_i}$. By Assumption 2, this term has the same limit as $\frac{2}{K}\sum_{i=K}^{B-K}\tilde{e}'_{T_{i+K}}e'_{T_i}$. We then invoke statement (D.81) in Remark 12. Now identify the sum with $S_{n,I}$ in Lemma 7 (with $\mathcal{H}_{n,i}=\mathcal{G}_{T_{i+J}}$, $N=2J$, and $B'_n=B_n-K_n$.). The multi lag angle braket process $\langle S_n,S_n\rangle^{(N)}_{B'_n}\leq(4J-1)4K_n^{-2}(B_n-2K_n+1)\Gamma^2n^{-4\alpha}=O(n^{-4\beta}K_n^{-2}(B_n-2K_n+1))$ since it then follows from Lemma 7 that $\frac{2}{K}\sum_{i=K}^{B-K}\tilde{e}_{T_{i+K}}e_{T_i}=O_p(n^{-2\beta}K_n^{-1}(B_n-2K_n+1)^{1/2})$. This is the uncancellable variance term, which later (Theorem 4) shows up as $O_p\left(n^{-2\beta}\mathfrak{E}_n^{1/2}\right)$, and is therefore included specially in the statement of the Proposition. – The result is clearly the same if $K_n=O(1)$ (Case (B) mentioned at the beginning of Section B).

The cross term on the form $\frac{2}{K}\sum_{i=K}^{B-K}(M_{T_{i+K}}-M_{T_i})(M_{T_i}-M_{T_{i-K}})$ is handled by writing

$$\begin{aligned}\frac{2}{K}\sum_{i=K}^{B-K}(L_{n,T_{i+K}}-L_{n,T_i})(L_{n,T_i}-L_{n,T_{i-K}})&=\frac{1}{K}\sum_{i=K}^{B-K}(L_{n,T_{i+K}}-L_{n,T_{i-K}})^2\\&-\frac{1}{K}\sum_{i=K}^{B-K}(L_{n,T_{i+K}}-L_{n,T_i})^2-\frac{1}{K}\sum_{i=K}^{B-K}(L_{n,T_i}-L_{n,T_{i-K}})^2\\&\xrightarrow{p}(2-1-1)([L,L]_{\mathcal{T}}-[L,L]_0)=0.\end{aligned}\tag{B.70}$$

by the same argument as that surrounding (B.69).

We now approach the cross term on the form

$$\frac{2}{K}\sum_{i=K}^{B-K}(M_{T_{i+K}}-M_{T_i})e_{T_{i-K}}=n^{-2\alpha}\frac{2}{K}\sum_{i=K}^{B-K}(L_{n,T_{i+K}}-L_{n,T_i})n^{\alpha}e_{T_{i-K}}.\tag{B.71}$$

Write

$$\frac{2}{K}\sum_{i=K}^{I-K}(L_{n,T_{i+K}}-L_{n,T_i})n^{\alpha}e_{T_{i-K}}=\sum_{j=K}^{I}(L_{n,T_j}-L_{n,T_{j-1}})H_{n,j}=\int_{T_K}^{T_I}H_n(t-)dL_{n,t}\tag{B.72}$$

where $H_n(t)=H_{n,j}$ for $t\in[T_{j-1},T_j]$, and where $H_j=2K^{-1}\sum n^{\alpha}e_{T_{i-K}}$, the sum being over $i\in[j-K,j-1]\cap[K,I-K]$. By invoking (D.81) in Remark 12, we see that we can take $|H_j|\leq 2\Gamma$. Also, as in the proof of the proof of Lemma 7, we see that $E(H_j^2)\leq(2J-1)K^{-1}\Gamma^2$. Thus, following Lenglart's inequality (Jacod and Shiryaev (2003, Lemma I.3.20, p. 35)), $\sup_{0\leq t\leq\mathcal{T}}|H_t|\xrightarrow{p}0$. By combining *Ibid.*, Lemma VI.3.31 (p. 352), Corollary 3.33 (p. 353), and Theorem VI.6.22(b) (p. 383), and since $L_{n,t}$ is P-UT (Appendix D.1), it follows that B.72 goes to zero in probability. Hence the expression in (B.71) is of order $o_p(n^{-2\alpha})$.

The cross term on the form $\frac{2}{K}\sum_{i=K}^{B-K}(M_{T_i}-M_{T_{i-K}})e_{T_{i+K}}$ is handled much the same way.

The remaining cross terms can be decomposed as $\text{crt}'_n + \text{crt}''_n$, where

$$\begin{aligned}\text{crt}'_n &= \frac{2}{K}\sum_{i=K}^{B-K}\big((M_{T_{i+K}} - M_{T_{i+K-J}})\tilde{e}_{T_{i+K}} - (M_{T_{i-K+J}} - M_{T_{i-K}})e_{T_{i-K}}\big)\\ &- \frac{2}{K}\sum_{i=K}^{B-K}\big((M_{T_{i+J}} - M_{T_i}) - (M_{T_i} - M_{T_{i-J}})\big)(e_{T_i} + \tilde{e}_{T_i})\\ &= \frac{2}{K}V_1' + o_p(n^{-2\alpha}) \end{aligned} \tag{B.73}$$

by the same methods as above, where V_1' is given by (B.67) Meanwhile, the residual crt''_n is a sum of square integrable martingales (along with a canonical decomposition term, in analogy with Definition 1 in Appendix A), and are $o_p(n^{-2\alpha})$, also by the same methods as above. □

B.3 Three Represents: Final Arguments to prove Theorem 3

From Theorem 1 in Section 3, we obtain

$$\frac{1}{K}\sum_{i=K}^{B-K}(\Theta_{(T_i,T_{i+K}]} - \Theta_{(T_{i-K},T_i]})^2 = \frac{2}{3}(K\Delta T)^2[\theta,\theta]_{\mathcal{T}-}(1 + o_p(1)). \tag{B.74}$$

In view of Propositions 5-6, we therefore have

$$\begin{aligned} QV_K = \frac{1}{K}\sum_{i=K}^{B-K}(\hat{\Theta}_{(T_i,T_{i+K}]} - \hat{\Theta}_{(T_{i-K},T_i]})^2 &= \frac{2}{3}(K\Delta T)^2[\theta,\theta]_{\mathcal{T}-}\\ &+ 2n^{-2\alpha}[L,L]_{\mathcal{T}} - \big(2\mathcal{E}_{\tilde{e}^2}(0) + \mathcal{E}_{(\tilde{e}+e)^2}(0) + \mathcal{E}_{(\tilde{e}+e)^2}(\mathcal{T}) + 2\mathcal{E}_{e^2}(\mathcal{T})\big)\\ &+ \frac{1}{K}\text{TSE}_n + \frac{2}{K}V_1' + \frac{2}{K}V_1\\ &+ o_p(n^{-2\alpha}) + O_p(n^{-(\alpha+\beta)}K_n^{-1}(B_n - 2K_n + 1)^{1/2}), \end{aligned} \tag{B.75}$$

where the V_i and TSE_n do not depend on K, and are given by (B.27) and (B.66)-(B.67). If we write

$$V_0 = \text{TSE}_n + 2V_1 + 2V_1', \tag{B.76}$$

and because of the three middle assumptions in (33) we obtain the statement of Theorem 3. □

C Properties and Convergence of the Edge Effect

C.1 About Assumption 2 on the Edge Effects

The formulation means that the main edge effect at T_i is allowed to depend on observations in J time periods on each side of T_i.

The specific conditions can be verified under mixing assumptions. The following is a complement to out examples. This is not intended to provide minimal conditions, just to explain why our conditions are reasonable.

The Decomposition $e_{T_i} = e'_{T_i} + e''_{T_i}$ *and* $\tilde{e}_{T_i} = \tilde{e}'_{T_i} + \tilde{e}''_{T_i}$. We have chosen this way of stating the conditions on the edge effect since, in our examples, this is readily verifiable. To tie the condition to the literature, however, we observe that, subject to mixing conditions, we require $(e_{T_i}, \tilde{e}_{T_i})$ to be a *mixingale*, see, *e.g.,* McLeish (1975) and Hall and Heyde (1980, pp. 19-21, 41). As the name suggests, it is tied up with the concept of mixing. See also Wu and Woodroofe (2004).

α- and ϕ- mixing. For a more general treatment, see McLeish (1975, p. 834) and Hall and Heyde (1980, Chapter 5 and Appendix III). For simplicity, we here focus on ϕ-mixing.[51] If $\mathcal{A}$ and $\mathcal{B}$ are two sigma-fields, then the phi-fixing coefficient is

$$\phi(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}, P(A)>0} |P(B|A) - P(B)| \tag{C.77}$$

The Decomposition, again. Set $\tilde{e}''_{T_i} = \tilde{e}_{T_i} - E(\tilde{e}_{T_i} \mid \mathcal{G}_{T_{i-J}})$, and similarly for e''_{T_i}. The difference $\tilde{e}'_{T_i} = \tilde{e}_{T_i} = \tilde{e}''_{T_i}$ will then have the martingale-like properties described, as will e'_{T_i}.

Meanwhile, if we require, say, that $\sup_n E\ \left(\max_{0 \le i \le B_n} |n^\alpha e_{n,T_i}|^{1+\delta} + \max |n^\alpha \tilde{e}_{n,T_i}|^{1+\delta}\right) < \infty$, for some $\delta > 0$, and also that $\sum_i (Ee_{n,T_i})^2 + (E\tilde{e}_{n,T_i})^2 = o(n^{-2\alpha})$, then the lemma on McLeish (1975, p. 834) assures that our conditions on $(e''_{T_i}, \tilde{e}''_{T_i})$ are satisfied provided

$$\sum_i \phi(\mathcal{G}_{T_{i-J}}, \sigma(e''_{T_i}, \tilde{e}''_{T_i}))^{\frac{2\delta}{1+\delta}} = o(n^{-2\alpha}). \tag{C.78}$$

Normally, however, the number of observations in each interval $(T_{i-1}, T_i]$ will go to infinity with n, thus under exponential mixing (in the original microstructure noise), (C.78) will normally hold.

The assumption (31) can be derived from mixing assumptions in much the same way.

C.2 Proof and Comments on Theorem 4

Proof of Theorem 4. In view of Remark 9, it remains to handle the dominating error term, which corresponds to the first cross term in the proof of Proposition 6. For given K, the full and exact expression is of the form $S^{(K)}_{n,I} = \sum_{i=1}^{I} \zeta^{(K)}_{n,i}$, where

$$\zeta^{(K)}_{n,i} = 2K^{-1}\left(\tilde{e}'_{T_i}(e'_{T_{i-2K}} - u'_{T_{i-K}}) I_{\{2K \le j \le B\}} - u'_{T_i} e_{T_{i-K}} I_{\{K \le j \le B-K\}}\right),$$

with $u'_{T_i} = \tilde{e}_{T_i} + e_{T_i}$.

[51] One can do similar things with α-mixing, using the definition and lemma on McLeish (1975, p. 834).

When dealing with multiple K's, we have $S_{n,I} = \sum_{i=1}^{m} \gamma_{n,K} S_{n,I}^{(K)}$, which we can write $S_{n,I} = \sum_{i=1}^{I} \zeta_{n,i}$, with $\zeta_{n,i} = \sum_{l=1}^{m} \gamma_{n,l} 2K_{n,l}^{-1} \left(\tilde{e}'_{T_i} (e'_{T_{i-2K}} - u'_{T_{i-K}}) I_{\{2K \le j \le B\}} - u'_{T_i} e'_{T_{i-K}} I_{\{K \le j \le B-K\}} \right)$. In analogy with the single term case (in the proof of Proposition 6), We then invoke statement (D.81) in Remark 12, and then apply Lemma 7 (with the same elections $\mathcal{H}_{n,i} = \mathcal{G}_{T_{i+J}}$, $N = 2J$, and $B'_n = B_n - K_n$.). From the Lemma, the multi lag angle bracket process

$$\begin{aligned} \langle S_n, S_n \rangle_{B'_n}^{(N)} &\le (4J-1) 4 \sum_{l=1}^{m} \left(\frac{\gamma_{n,l}}{K_{n,l}} \right)^2 (B_n - 2K_n + 1) \Gamma^2 n^{-4\alpha} \\ &= O_p \left(n^{-4\alpha} \sum_{l=1}^{m} \left(\frac{\gamma_{n,l}}{K_{n,l}} \right)^2 (B_n - 2K_{n,l} + 1) \right). \end{aligned} \tag{C.79}$$

In particular, by Lenglart's Inequality (and our Lemma), S_n has the order in probability as the square root of (C.79). This is used in eq. (45) in Theorem 4. □

Remark 11. Sharpness of the Bound in (C.79) and Theorem 4.) To get a sense of what is the best possible rate, assume that the $(\tilde{e}_{T_i}, e_{T_i})$ are iid with second moments.. In this case, $S_{n,I}$ is a martingale (in I, for fixed n), with predictable quadratic variation $\langle S_n, S_n \rangle_I = \sum_{i=1}^{I} E(\zeta_{n,i}^2 \mid \mathcal{G}_{T_{i-1}})$. By the law of large numbers, subject to moment conditions,

$$\langle S_n, S_n \rangle_B = 4 \sum_{l=1}^{m} \left(\frac{\gamma_{n,l}}{K_{n,l}} \right)^2 (B_n - 2K_{n,l} + 1) \left(E\tilde{e}^2 (Eu^2 + Eu^2) + Eu^2 Eu^2 \right) (1 + o_p(1)). \tag{C.80}$$

A more realistic scenario is to impose the kind of mixing condition discussed in Appendix C.1. We assume second moments, again. For simplicity, consider only one term, say, $\zeta_{n,i} = \sum_{l=1}^{m} \gamma_{n,l} 2K_{n,l}^{-1} \tilde{e}'_{T_i} e'_{T_{i-2K_{n,l}}} I_{\{2K \le j \le B\}}$. One can then decompose $S_{n,I} = \sum_{j=1}^{N} S_{n,I}^{(L)}$ as in the Proof of Lemma 7, with $N = 2J$. The individual $S_{n,I}^{(L)}$ is a martingale, with predictable quadratic variation, $\langle S_n^{(L)}, S_n^{(L)} \rangle_I = \sum_{i \in [1,I], i \equiv L} = \sum_{i=1}^{I} \sum_{l=1}^{m} \gamma_{n,l} 2K_{n,l}^{-1} E((\tilde{e}'_{T_i})^2 \mid \mathcal{G}_{T_{i-J}})(e'_{T_{i-2K_{n,l}}})^2 I_{\{2K \le i \le B\}}$, By the Lemma in McLeish (1975, p. 834), sufficient mixing will assure that this can be well approximated by $\sum_{i=1}^{I} \sum_{l=1}^{m} \gamma_{n,l} 2K_{n,l}^{-1} E((\tilde{e}'_{T_i})^2)(e'_{T_{i-2K_{n,l}}})^2 I_{\{2K \le i \le B\}}$. *If the $K_{n,l}$ are at least J apart,* one can then in turn use the martingale property and the mixing to see that $\langle S_n^{(L)}, S_n^{(L)} \rangle_I = 4 \sum_{l=1}^{m} \left(\frac{\gamma_{n,l}}{K_{n,l}} \right)^2 \left(\sum_{i=1}^{I} E((\tilde{e}'_{T_i})^2) E(e'_{T_{i-2K_{n,l}}})^2 I_{\{2K \le i \le B\}} \right) (1 + o_p(1))$, which is again of the same order as (C.79). Since $S_{n,I}$ is a finite sum of $S_{n,I}^{(L)}$, this order will be preserved except in truly exceptional circumstances.

In fact, given (30), each $S_{n,I}^{(L)}$ is asymptotically normal with the relevant variance.[52] – A similar but more elaborate analysis will give the exact asymptotic variance (and normality) of $S_{n,I}^{(L)}$. □

[52]See Mykland (1994, Section 4) for the relevant regularity condition.

D Odds and Ends

D.1 About the P-UT Property, and Proof of Proposition 1

Proposition 6 uses (17) as well as Jacod and Shiryaev (2003, Corollary 6.30, p 385 and Proposition VI.3.37, p. 387). More generally, (17) can be replaced in Assumption 1 (for our entire development) by a requirement that $L_{n,t}$ be "Predictably Uniformly Tight" (P-UT), in the sense of *Ibid.,* Definition VI.6.1 (p. 377), cf. also their Theorem VI.6.26 (p. 384). The proof of their Corollary 6.30 is, in fact, a proof that the sequence of local martingales is P-UT. The condition (17) is weaker that what is usually required for a central limit theorem, and it does not assure asymptotic negligibility. – The contribution of $A_{n,t} - A_t$ to the asymptotics is negligible by *Ibid.,* Proposition I.3.3 (p. 27), Corollary VI.3.33 (p. 353), and Theorem VI.3.37 (p. 354).

To see equation (18), use that $L_t^2 - [L, L]_t$ is a local martingale w.r.t. filtration $\mathcal{F} \vee \mathcal{F}_t^L$, hence $E(L_\mathcal{T} \mid \mathcal{F}) = 0$ and $E(L_\mathcal{T}^2 - [L, L]_\mathcal{T} \mid \mathcal{F}) = 0$.

D.2 Technical Lemmae

To handle general moments, we shall use the following.

Lemma 6. (Truncating the Edge Effects.) *Suppose Assumption 2. Then, for any $\delta > 0$, there exists (possibly on an extension of the space) e^{tr}_{n,T_i} and $\tilde{e}^{\text{tr}}_{n,T_i}$, and a nonrandom constant Γ, so that*

1. *For each n $e^{\text{tr}}_{n,T_i} = e'_{n,T_i}$ and $\tilde{e}^{\text{tr}}_{n,T_i} = \tilde{e}'_{n,T_i}$ for all $i \in [0, B_n]$, on a measurable set A_n, and $P(A_n) < \delta$;*
2. *e^{tr}_{n,T_i} and $\tilde{e}^{\text{tr}}_{n,T_i}$ satisfy the conditions in Assumption 2 in lieu of e'_{n,T_i} and $\tilde{e}'_{n,T_i}$; and*
3. *$|e^{\text{tr}}_{n,T_i}| \leq \Gamma n^{-\alpha}$ and $|\tilde{e}^{\text{tr}}_{n,T_i}| \leq \Gamma n^{-\alpha}$ for all i and n.*

Remark 12. (Using Lemma 6.) We shall use the lemma to assert, in various places, that

$$|n^\alpha e'_{n,T_i}| \text{ and } |n^\alpha \tilde{e}'_{n,T_i}| \text{ can without loss of generality be taken to be bounded by a constant } \Gamma. \quad (D.81)$$

Here is the specific mechanism that we refer to.

Let Y_n be a sequence of random variables, involving a functional form of e'_{n,T_i} and $\tilde{e}'_{n,T_i}$ (as well as any of the other random quantities in our setup). Let D be a countable set, $D \subset (0, 1)$, with a limit point at zero.

For given $\delta \in D$, create $Y_{n,\delta}$ by replacing the e'_{n,T_i} and $\tilde{e}'_{n,T_i}$ by the e^{tr}_{n,T_i} and $\tilde{e}^{\text{tr}}_{n,T_i}$ as described by Lemma 6. Then $Y_n = Y_{n,\delta}$ on the set A_n. Suppose one can show that there is a random variable

Y (independent of δ) so that $Y_{n,\delta} \xrightarrow{p} Y$ as $n \to \infty$. Then, for any $\epsilon > 0$, and since $P(A_n) < \delta$,

$$\begin{aligned} P(|Y_n - Y| > \epsilon) &\leq P(\{|Y_{n,\delta} - Y| > \epsilon\} \cap A_n^c) + P(A_n) \\ &\leq P(|Y_{n,\delta} - Y| > \epsilon) + \delta \\ &\to \delta \text{ as } n \to \infty. \end{aligned} \tag{D.82}$$

Since D has limit point at zero, it follows that $Y_n \xrightarrow{p} Y$ as $n \to \infty$. □

Proof of Lemma 6. For $L = 1, \cdots 2J$, set $S_{n,I}^{(L)} = \sum_{i \in [1,I] \text{ and } i \equiv L[2J]} e'_{n,T_i}$, where $i \equiv L[N]$ means that i is of the form $i = L + jN$ for some integer j. Then for each L and n, $S_{n,I}^{(L)}$ is a martingale with respect to the filtration $\mathcal{H}_{n,i} = \mathcal{G}_{T_{i+J}}$. We now invoke the construction from Mykland (1994, eq. (4.8), p. 27), which produces e_{n,T_i}^{tr} $(i \equiv L[2J])$. satisfying items (1), (2) and (3) in the Lemma, with, say $A_{n,L,1}$ and $\Gamma_{L,1}$, and with $P(A_{n,L,1}) < \delta/4J$. We repeat this construction for all L, and similarly for $\tilde{e}'_{n,T_i}$, in the latter case giving rise to $A_{n,L,2}$ and $\Gamma_{L,2}$. By construction, the whole set of e_{n,T_i}^{tr} and $\tilde{e}_{n,T_i}^{\text{tr}}$ satisfy items (1), (2) and (3) in the Lemma, with $A_n = \cup A_{n,L,r}$ and $\Gamma = \max \Gamma_{L,r}$. □

To handle cross-terms, we use the following.

Lemma 7. (Negligibility of Multi-lag martingales.) *Let $S_{n,I} = \sum_{i=1}^{I} \zeta_{n,i}$, where we suppose that $\zeta_{n,i}$ is $\mathcal{H}_i^n$-measurable and satisfies that $E(\zeta_i^n \mid \mathcal{H}_{i-N}) = 0$.*[53] *Define $\langle S_n, S_n \rangle_I^{(N)} = \sum_{i=1}^{I} E((\zeta_{n,i})^2 \mid \mathcal{H}_{i-N})$. (It's an N'th lag angle bracket process.) Let α_n be a nonrandom sequence so that $\langle S_n, S_n \rangle_{B'_n}^{(N)} = o_p(\alpha_n)$. Then* $\sup_{1 \leq | \leq B'_n} |S_{n,I}| = o_p((N\alpha_n)^{1/2})$.

Proof of Lemma 7. For $0 \leq L \leq N - 1$, let $S_{n,I}^{(L)} = \sum_{i \in [1,I] \text{ and } i \equiv L[N]} \zeta_{n,i}$, where $i \equiv L[N]$ means that i is of the form $i = L + jN$ for some integer j.

Thus, $S_{n,I} = \sum_{j=1}^{N} S_{n,I}^{(L)}$. Since no two different $S_{n,I}^{(L)}$ change value for the same I, we also get that $[S_n, S_n]_I = \sum_{j=1}^{N} [S_n^{(L)}, S_n^{(L)}]_I$. Meanwhile,

$$\begin{aligned} E(S_{n,I})^2 &= E \sum_{i=K}^{I} (\zeta_{n,i})^2 + 2E \sum_{i=K}^{I} \sum_{j=1}^{N-1} \zeta_{n,i}\zeta_{n,i-j} \\ &= E \sum_{i=K}^{I} (\zeta_{n,i})^2 + 2E \sum_{j=1}^{N-1} \sum_{i=K}^{I} \zeta_{n,i}\zeta_{n,i-j} \\ &\leq E \sum_{i=K}^{I} (\zeta_{n,i})^2 + 2(N-1)E[S_n, S_n]_I \text{ (Cauchy-Schwarz)} \\ &= (2N-1)E[S_n, S_n]_I. \end{aligned} \tag{D.83}$$

[53] As convenient, we can take some ζ's in the beginning to be zero if the sum starts at K or similar. Definitely $\zeta_{n,i} = 0$ for $i < N$. – For an example of such a structure, one can take $\zeta_{n,i} = e'_{n,T_i}$ or $= \tilde{e}'_{n,T_i}$, with $\mathcal{H}_{n,i} = \mathcal{G}_{T_{i+J}}$ and $N = 2J$. This construction is also used in Lemma 6.

Hence, $(S_{n,I})^2$ is Lenglart-dominated (Jacod and Shiryaev (2003, Section I.3c, pp. 35-36), Jacod and Protter (2012, Section 2.1.7, p. 45)) by $(2N-1)[S_n, S_n]_I$, and hence also by $(2N-1)\langle S_n, S_n\rangle_I^{(N)}$. By the same reasoning as in the proof of Jacod and Protter (2012, Proposition 2.2.5, p. 574), the result follows. □