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: September 2016

Abstract

The availability of high frequency financial data has generated a series of estimators based on intra-day data, improving the quality of large areas of financial econometrics. However, estimating the standard error of these estimators is often challenging. The root of the problem is that traditionally, standard errors rely on estimating a theoretically derived asymptotic variance, 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, to build forecasting models based on, say, daily estimates, and also to optimize the tuning parameters.

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). It provides consistent estimators of AVAR for a broad class of integrated parameters $\Theta = \int \theta_t dt$, where the spot parameter process θ can be a general semi-martingale, with continuous and jump components. The observed AVAR is implemented with the help of a two-scales method. Its construction works well in the presence of microstructure noise, and when the observation times are irregular or asynchronous in the multivariate case.

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

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

*We would like to thank Ole Barndorff-Nielsen, Dachuan Chen, Jean Jacod, Eric Renault, Mathieu Rosenbaum, Olivier Scaillet, Neil Shephard, Kevin Sheppard, Christina D. Wang, and also Elie Tamer (the editor) and the referees for *Econometrica*, for helpful discussions and advice. We also appreciate their encouragement. Financial support from the National Science Foundation under grants SES 11-24526 and DMS 14-07812 and DMS 14-07820 is gratefully acknowledged.

1 Introduction

1.1 Two Standard Errors

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 and leverage effect, to make inference relating to 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 typical analysis takes the following form. One seeks to estimate an integrated parameter Θ,

$$\Theta = \int_0^{\mathcal{T}} \theta_t dt \tag{1}$$

on the basis of n data points, say, $X_{t_1}, X_{t_2}, \cdots, X_{t_n}$, where $\{\theta_t\}$ is a spot parameter process such as volatility, leverage effect, instantaneous regression coefficients, etc. To arrive at feasible inference, one typically needs:

Theoretical Requirement 1. (Asymptotic Validity of Normal Approximation.)[1] As the number of observations n becomes large,

i. An estimator $\hat{\Theta}_n$ which is consistent
ii. A *standard error* $\mathrm{se}(\hat{\Theta}_\mathrm{n})$, *i.e.,* a data-based statistic for which

$$\frac{\hat{\Theta}_n - \Theta}{\mathrm{se}(\hat{\Theta}_\mathrm{n})} \xrightarrow{\mathcal{L}} N(0,1) \text{ stably.} \tag{2}$$

□

The conventional way to implement Step (ii) is to go through the following additional steps:

Theoretical Requirement 2. (Estimated Asymptotic Variance.)

i. A limit theory: $n^\alpha(\hat{\Theta}_n - \Theta) \xrightarrow{\mathcal{L}} V^{\frac{1}{2}} N(0,1)$ stably in law, where V is an asymptotic variance;[2]
ii. Derive the mathematical expression for V;
iii. Find a consistent estimator $\hat{V}_n$ of V;
iv. Set $\mathrm{se}(\hat{\Theta}_\mathrm{n}) = n^{-\alpha}|\hat{V}_n|^{\frac{1}{2}}$. □

[1]See Proposition 2 in Section 3.1 for precise conditions. Stable convergence is described in Definition 3 in the same section.

[2] For subsequent decluttering of notation, we set $\mathrm{AVAR}_n = n^{-2\alpha}V$. $\widehat{\mathrm{AVAR}}_n$ is consistent if and only if $\widehat{\mathrm{AVAR}}_n = \mathrm{AVAR}_n(1 + o_p(1))$. Formulae for V and AVAR_n are given explicitly in (16)-(17) in Section 3.1.

Our Alternative. Our purpose in this paper is to circumvent Theoretical Requirement 2, by developing general formulae for $\mathrm{se}(\hat{\Theta}_\mathrm{n})$, which do not depend on knowing the convergence rate α or the asymptotic variance V. We call this *the observed standard error*. We can express $\mathrm{se}(\hat{\Theta}_\mathrm{n}) = |\widehat{\mathrm{AVAR}}_n|^{\frac{1}{2}}$, where $\widehat{\mathrm{AVAR}}_n$ is *the Observed Asymptotic Variance.*

Our general formula for observed $\widehat{AVAR}_n$ is a two-scales construction given in Definition 4 (Section 3.2). The estimator is consistent for the asymptotic variance (it satisfies Theoretical Requirement 2) using Theorem 4 (Section 3.2) and Proposition 1 (Section 3.1). Theoretical Requirement 1 is then satisfied via Proposition 2 (Section 3.1).

Apart from regularity conditions, our only assumption is that 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).[3] We shall see in Section 7 that the conditions for our results are satisfied broadly, including on quite exotic quantities such as leverage effect. Additional guidance on theory is provided in Section 6.

Practical guidance to how to use our theory is provided in Section 5. We emphasize that for empirical analysis, one does not need to know the analytical form of V to use the Observed Asymptotic Variance. The technique permits the setting of *prima facie* standard errors by just using our formulae and without any prior theoretical derivation. One can then verify the theoretical conditions afterwards. This is much like the practice in parametric inference (using the observed information) and when bootstrapping.

1.2 Why do we need a Standard Error?

Currently, the main use of standard errors are hypothesis testing and self-contained confidence intervals based on (2). In high frequency econometrics, such intervals go back to Barndorff-Nielsen and Shephard (2002a), where $\widehat{\mathrm{AVAR}}_n$ was set as the $\frac{2}{3}\times$ the *quarticity* (see Section 3.3). Confidence intervals and tests have been the main spur for pursuing the asymptotics described in Theoretical Requirement 2. Other early contributions to this type of asymptotics are those of Foster and Nelson (1996), Comte and Renault (1998), Jacod and Protter (1998), and Zhang (2001). A substantial amount of work on this problem has followed, as described below and throughout the paper.

There are other applications that require using the standard error. For example,

i. Incorporation into forecasting models: see Andersen, Bollerslev, and Meddahi (2005) and Bollerslev, Patton, and Quaedvlieg (2016).

ii. Optimal combination of intraday high frequency estimators, see an early draft of Meddahi (2002), as well as Andreou and Ghysels (2002) and Ghysels, Mykland, and Renault (2012).

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

iii. Model selection in high frequency regression, see Zhang (2012, Section 4, pp. 268-273). As documented in the cited paper, this problem has applications to the estimation of high frequency betas, as well as to non-parametric options trading.

iv. Selection of tuning parameters: 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 $\widehat{\mathrm{AVAR}}_n$. See Section 4 for references and further development.

1.3 Why do we need the Observed Asymptotic Variance?

The rationale behind this paper is that Theoretical Requirement (TR) 2(ii)-(iii) is a main hindrance to the development and use of inference in high frequency data. Recall that (TR) 2(ii) entails deriving a Central Limit Theorem for $\hat{\Theta}_n$ in order to obtain the analytical form of the asymptotic variance V, and then (iii) requires find a consistent estimator for V.

It can already be difficult to construct appropriate estimators $\widehat{\Theta}$, and it is often a substantial work to carry out the steps in TR 2(ii)-(iii). To corroborate this, we draw attention to the large literature that provides estimators $\hat{\Theta}_n$ of Θ, but it lacks feasible (asymptotically pivotal) statistics of the form (2). In particular, it is challenging to build an estimator $\hat{\Theta}_n$ or $\hat{\theta}_n$ that accommodates the presence of microstructure noise and non-synchronousness in observation times. But, the main challenge is to derive the theoretical asymptotic variance AVAR and to find $\widehat{\mathrm{AVAR}}_n$. Examples in the literature include, but are not limited to, semivariance (Barndorff-Nielsen, Kinnebrock, and Shephard (2009)); nearest neighbor truncation (Andersen, Dobrev, and Schaumburg (2012), see Mykland and Zhang (2016c) for applying the observed AVAR to this case); estimating the rank of the volatility matrix (Jacod and Podolskij (2013)); principal component analysis (Aït-Sahalia and Xiu (2015)); the volatility of volatility (Vetter (2015); see Remark 4 in Section 2.3 and Example 10 in Section 7); and high frequency regression, and ANOVA (Mykland and Zhang (2006, 2009, 2012); see Example 7 in Section 7). In all these examples, one can obtain a point estimate in the presence of microstructure noise, but one does not have ready access to tests, confidence intervals, and the other methods discussed in Section 1.2. 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.[4]

It should be emphasized that in many cases, the asymptotic variance is on the form of an integral of a function of volatility. In this case, TR 2(iii) in Section 1.1 can often be met with the theory in Jacod and Protter (2012, Section 16.4-16.5, pp. 512-554), Jacod and Rosenbaum (2013, 2015), and Mykland and Zhang (2009, Section 4.1, p. 1421-1426). These papers are important

[4] 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.

contributions to the AVAR problem. Not all asymptotic variances, however, are on such a form (such as, Examples 5, 9, and 10 in Section 7; and Robert and Rosenbaum (2011, 2012)). Also, even when the AVAR is on this form, it may be difficult to go through step TR 2(ii). In addition, there are cases where the estimation approach may be based on robustness considerations which would make the cited volatility estimators inappropriate (*e.g.*, Andersen, Dobrev, and Schaumburg (2012, 2014)).

1.4 Connections to the Literature

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.

The observed AVAR has a lot in common with quarticity estimate of AVAR in realized volatility, in the seminal work of Barndorff-Nielsen and Shephard (2002a, 2004a). Also, it resembles observed information in likelihood theory. The difference between the observed AVAR and the estimated AVAR (going through Theoretical Requirement 2 in Section 1.1) correspond to the difference between the observed and the estimated expected information in parametric inference. We discuss these connections further in Section 3.3.

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. We emphasize that our block parameter K is (typically) unrelated to any block size used to construct the estimator $\widehat{\Theta}_n$. For a precise discussion of this, see Section 5.4.

The comparison of adjacent estimators, however, is also a feature of the subsampling developed for volatility in the work of Kalnina and Linton (2007) and Kalnina (2011), with an important subsequent study by Christensen, Podolskij, Thamrongrat, and Veliyev (2015). Bootstrapping has been developed in the papers by Gonçalves and Meddahi (2009) and Gonçalves, Donovon, and Meddahi (2013), but is further away from the approach of the current paper.

Apart from the overall construction of observed asymptotic variance, there are two other intellectual novelties in the paper. Firstly, the comparison of adjacent values of the integral of θ is given a precise formulation in the *Integral-to-Spot Device* (Theorem 1 and Corollary 1, in Section 2.3) 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. Secondly, 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 noise". We can thus adapt known methods to the current problem of estimating asymptotic variance. It is worth to mention that edge effects are estimator-specific. As its name suggests, edge effects show 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. As we shall see in our examples (Section 7), edge effects are ubiquitous in high frequency inference. The effect is also referred to as burn-in time, and border effect.

We emphasize that our purpose in this paper is to provide a method for getting at observed asymptotic variance, for any estimator of interest. Our proposed approach extends broadly to high frequency inference. The contribution of the current paper is, in particular, to estimators other than volatility. For the latter, much is known both in terms of asymptotic variances and in terms of resampling as discussed above.

The rest of this paper is organized as follows. Section 2 introduces the intuition behind the observed AVAR and the *Integral-to-Spot Device*. Section 3 develops the general formulae for Observed $\widehat{AVAR}_n$ using a two-scales construction. It also provides consistent estimators of the quadratic variation of the spot parameter process θ_t. The use of the $\widehat{AVAR}_n$ to select tuning parameters is discussed in Section 4. Section 5 provides practical guidance to using the theory. The generalization to the multidimensional case is described in Remark 12 in Section 5.2. Section 6 gives advice on how to verify the conditions of the theory. Section 7 provides examples. Finally, Section 8 concludes. Proofs are located in the Appendix.

2 Finite Sample Quadratic Variations of a Parameter Process.

2.1 Setup

We observe data at high frequency, in a time period from 0 to $\mathcal{T}$. The data will normally take the form of samples from a semimartingale X_t, typically contaminated by microstructure noise. We are interested in estimating integrals of a "parameter" spot process θ_t, which also is assumed to be a semimartingale.

For example, we can take θ_t to be the spot variance of the continuous part X_t^c 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 Remark 12.

Definition 1. (Model Structure and Notation). *The notation $[X, X]_t$ refers to the continuous-time quadratic variation of semimartingale X from time zero to time t (e.g., Jacod and Shiryaev (2003, p. 51-52), Protter (2004, p. 66)). The quadratic variation is also known as (ex-post) inte-*

grated variance (Barndorff-Nielsen and Shephard (2002b)).[5] *Semimartingales are defined in, 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). We assume that all our semimartingales are càdlàg (right continuous with left limits). All data generating and latent (such as X_t and θ_t) processes live on a probability space $(\Omega, \mathcal{F}, P)$.*[6]

We consider integrated parameters and their estimators[7] over time intervals $(S,T] \subset [0,\mathcal{T}]$:

$$\Theta_{(S,T]} = \int_S^T \theta_t dt \text{ and } \hat{\Theta}_{(S,T]} = \text{ a consistent estimator of } \Theta_{(S,T]}. \tag{3}$$

Even when estimating the spot volatility, one almost invariably estimates such integrals.[8]

To get a stab at the asymptotic variance we shall use the following finite sample quantities.

Definition 2. *(Rolling Quadratic Variations of Integrated Processes.) Divide the time interval $[0,\mathcal{T}]$ into B basic blocks of time periods (days, five minutes, thirty seconds, or other) $(T_{i-1}, T_i]$, with $T_0 = 0$ and $T_B = \mathcal{T}$. The blocks are assumed to be of equal size: Set $\Delta T = \mathcal{T}/B$, and assume that $T_i = i\Delta T$. We shall permit rolling overlapping intervals, and so let K be an integer no greater than $B/2$. We define*

$$\text{The quadratic variation of } \Theta\text{: } QV_{B,K}(\Theta) = \frac{1}{K}\sum_{i=K}^{B-K}(\Theta_{(T_i,T_{i+K}]} - \Theta_{(T_{i-K},T_i]})^2, \text{ and}$$
$$\text{The quadratic variation of } \widehat{\Theta}\text{: } QV_{B,K}(\hat{\Theta}) = \frac{1}{K}\sum_{i=K}^{B-K}(\hat{\Theta}_{(T_i,T_{i+K}]} - \hat{\Theta}_{(T_{i-K},T_i]})^2 \tag{4}$$

We emphasize that the above discretized quadratic variations are defined on the discrete grid $\{0, \Delta T, 2\Delta T, \cdots, \mathcal{T}\}$, as opposed to the continuous-time quadratic variation $[X,X]_t$ discussed above.

Later on, from Section 3 onwards, B, ΔT, and K will depend (explicitly or implicitly) on an index n, which usually denotes the number of observations. We may then write $\Delta T = \Delta T_n$, or

[5]Similarly, $[X,Z]_t$ refers to the continuous-time quadratic co-variation (or integrated covariance) of semimartingales X and Z.

[6]A full specification of the model also involves a filtration $(\mathcal{F}_t)_{0\le t\le\mathcal{T}}$, $\mathcal{F}_\mathcal{T} \subseteq \mathcal{F}$, which we for simplicity shall take to be fixed throughout the paper, until we reach Section 6. Also until then, when we say that X_t is a "semimartingale", we automatically mean a semimartingale relative to $(\mathcal{F}_t)_{0\le t\le\mathcal{T}}$ and P. The "filtered probability space" $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\le t\le\mathcal{T}}, P)$ is also taken to satisfy the "usual conditions" (Jacod and Shiryaev (2003, Definition I.1.2-I.1.3, p. 2)).

[7]All estimators are implicitly or explicitly indexed by the number of observations n. Consistency, convergence in law, etc, refers to behavior as $n \to \infty$,

[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). The theory requires the existence of a "spot" θ_t, cf. Section 5.3. To the extent that the "integral" process has jumps, we assume that such jumps have been suitably removed by the estimation procedure in use, as also discussed at the beginning of Section 7, see also Examples 1 and 9 in the same section. See also Section 5.3. On the other hand, we shall see that the process θ_t can have as many jumps as it wants.

omit the index n if the meaning is obvious.

2.2 The Basic Insight

The basic insight behind the Observed AVAR is that we can decompose the increment $\hat{\Theta}_{(T_i,T_{i+K}]} - \hat{\Theta}_{(T_{i-K},T_i]}$ into the parts related to estimator behavior and the part solely tied to parameter behavior:

$$\hat{\Theta}_{(T_i,T_{i+K}]} - \hat{\Theta}_{(T_{i-K},T_i]} = \underbrace{\left(\hat{\Theta}_{(T_i,T_{i+K}]} - \Theta_{(T_i,T_{i+K}]}\right)}_{\text{estimation error}} + \underbrace{\left(\Theta_{(T_i,T_{i+K}]} - \Theta_{(T_{i-K},T_i]}\right)}_{\text{evolution in parameter}} - \underbrace{\left(\hat{\Theta}_{(T_{i-K},T_i]} - \Theta_{(T_{i-K},T_i]}\right)}_{\text{estimation error}}. \tag{5}$$

In consequence, we can write the quadratic variation of $\hat{\Theta}$ as

$$\begin{aligned} QV_{B,K}(\hat{\Theta}) &= \frac{2}{K}\sum_i (\hat{\Theta}_{(T_i,T_{i+K}]} - \Theta_{(T_i,T_{i+K}]})^2 + \frac{1}{K}\sum_i (\Theta_{(T_i,T_{i+K}]} - \Theta_{(T_{i-K},T_i]})^2 \\ &+ \text{ martingale and negligible terms} \\ &= \left(\underbrace{2 \text{ AVAR}(\hat{\Theta}_{(0,\mathcal{T}]} - \Theta_{(0,\mathcal{T}]})}_{\text{estimation error}} + \underbrace{QV_{B,K}(\Theta)}_{\text{parameter behavior}} \right)(1 + o_p(1)) \end{aligned} \tag{6}$$

when ΔT goes to zero.[9]

To turn this from a heuristic to a rigorous theory, we need to

i. Explain how to go from the first to the second line of (6), and in particular explain how $\frac{1}{K}\sum_i(\hat{\Theta}_{(T_i,T_{i+K}]} - \Theta_{(T_i,T_{i+K}]})^2$ comes to be related to the asymptotic variance of $\hat{\Theta}_{(0,\mathcal{T}]} - \Theta_{(0,\mathcal{T}]}$. We shall do this in Section 3.

ii. Disentangle $\text{AVAR}(\hat{\Theta}_{(0,\mathcal{T}]} - \Theta_{(0,\mathcal{T}]})$ from $QV_{B,K}(\Theta)$. We shall do this by finding that the latter is approximately equal to $\frac{2}{3}(K\Delta T)^2[\theta,\theta]_{\mathcal{T}-}$. We shall then be able to write two (or more) distinct linear equations on the form (6), which we can solve for AVAR.

We start with (ii): the approximation of $QV_{B,K}(\Theta)$.

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

A main result is the following, with proof in Appendix B. The appendix also contains a simplified version of the proof for finite K as $B \to \infty$.

[9] See Footnote 2 in the Introduction for the normalization of AVAR.

Theorem 1. (The Integral-to-Spot Device, General Case.) *Assume that θ_t is a semi-martingale on $[0, \mathcal{T}]$. Also suppose that $K\Delta T \to 0$ (K may have subsequences that are $O_p(1)$ or that go to infinity). Set $t_* = \max\{i\Delta T : \ i\Delta T < t\}$ and $t^* = \min\{i\Delta T : \ i\Delta T \geq t\}$. Then*

$$\frac{1}{(K\Delta T)^2} QV_{B,K}(\Theta) \ = \ \frac{2}{3}\left(1 - \frac{1}{K^2}\right)[\theta,\theta]_{\mathcal{T}-} + \frac{1}{K^2}\int_0^{\mathcal{T}} \left(\left(\frac{t^*-t}{\Delta T}\right)^2 + \left(\frac{t-t_*}{\Delta T}\right)^2\right) d[\theta,\theta]_t \ + \ o_p(1) \tag{7}$$

where $[\theta,\theta]_{\mathcal{T}-} = \lim_{t\uparrow\mathcal{T}}[\theta,\theta]_t$. The convergence in probability is uniform in ΔT, so long as $\Delta T > 0$ and $K\Delta T \to 0$.[10]

Remark 1. (Consistency for Absolutely Continuous $[\theta,\theta]_t$.) If $[\theta,\theta]_t$ is absolutely continuous, the right hand side of (7) equals $\frac{2}{3}[\theta,\theta]_{\mathcal{T}} \ + \ o_p(1)$, also for finite K. The reason is that the limit of the second term in (7) then equals $\frac{2}{3}\frac{1}{K^2}[\theta,\theta]_{\mathcal{T}-}$. □

It would seem from Theorem 1 that much nuisance is created when there are jumps in θ. As further analyzed in Section 3.2, however, it is typically meaningful to add the extra restriction that $K \to \infty$. This solves the discontinuity problem, as follows.

Corollary 1. (The Integral-to-Spot Device, Consistent Case.) *In addition to the Assumptions of Theorem 1, also suppose that $K \to \infty$. Then*

$$\frac{1}{(K\Delta T)^2} QV_{B,K}(\Theta) \ = \ \frac{2}{3}[\theta,\theta]_{\mathcal{T}-} \ + \ o_p(1). \tag{8}$$

Remark 2. (Convergence fails for finite K in the Presence of Jumps.) To understand the impact of jumps on the above Theorem 1, note that for finite K it will not, in general, produce a limit when θ has jumps.

To see why, suppose for simplicity that θ_t is continuous except for a single jump at (stopping) time $\tau \in (0, \mathcal{T})$. Also assume that $[\theta,\theta]^c_t = [\theta^c,\theta^c]_t$ (the continuous part of $[\theta,\theta]_t$) is absolutely continuous. Recall that $\Delta T = \Delta T_B = \mathcal{T}/B$. For $K = 1$, we get from (7) that

$$(\Delta T)^{-2}\sum_i (\Theta_{(T_i.T_{i+1}]} - \Theta_{(T_{i-1}.T_i]})^2 \ = \ \frac{2}{3}\left([\theta^c,\theta^c]_{\mathcal{T}}\right) + \frac{1}{2}\left((1-U_B)^2 + U_B^2\right)\Delta\theta_\tau^2 + o_p(1), \tag{9}$$

where $U_B = (\tau - \tau_{B,*})/\Delta T_B$, where $\tau_{B,*} = \max_i\{i\Delta T < \tau\}$. 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_B converges in law to a standard uniform random variable. Similar considerations apply more generally to Theorem 1 if θ_t is an Itô-seminartingale in the sense of Jacod and Protter (2012, Chapter 4.4, p. 114).

[10]See Remark 15 in Appendix A. The same holds for Theorem 3 in Section 3.2, and Theorem 8 in Appendix C. In other theorems, the uniformity is valid subject to the needs of other assumptions, such as the balance condition (30) in Section 3.2.

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,[11] the right hand side of (9) simply does not converge, in probability or law. □

Remark 3. (Link to Pre-Averaging, and the Factor 2/3.) Think of θ_t as X_t. One can relate Theorem 1 to pre-averaging.[12] An integral is much like a sum, and so we are continuously pre-averaging θ_t, and then using the averaged quantity to find the volatility of θ_t. The factor 2/3 originates from the procedure of pre-averaging, *cf.* the example on p. 2255 in Jacod, Li, Mykland, Podolskij, and Vetter (2009a). A similar factor of 1/2 appears in the estimation of leverage effect, see Mykland and Zhang (2009).[13] This downward bias is typically referred to as "smoothing bias", and is well studied in the literature on nonparametric density estimation (Stoker (1993)). For use of this terminology in the high frequency setting, see Aït-Sahalia, Fan, and Li (2013, Section 4.2, p. 230).

Theorem 1 is concerned with the volatility of a general semimartingale, and this has not been studied in full generality by the pre-averaging literature.[14] It is thus conjectured to have implications for the consistency of pre-averaging estimators of volatility. To see this, consider equidistant discrete observations of θ_{T_i}, and $\overline{\Theta}_{(T_i,T_{i+K}]} = \sum_{j=i+1}^{i+k}\theta_{T_j}\Delta T_n$, and define $QV_{B,K}(\overline{\Theta})$ in analogy with (4). From the proof of Proposition 5 (in Appendix D.2), it is clear that Theorem 1 yields the following corollary:

$$\frac{1}{(K\Delta T)^2}QV_{B,K}(\overline{\Theta}) \;=\; \frac{2}{3}\left(1-\frac{1}{K^2}\right)[\theta,\theta]_{\mathcal{T}-} + \frac{1}{K^2}\int_0^{\mathcal{T}}\left(\left(\frac{t^*-t}{\Delta T}\right)^2 + \left(\frac{t-t_*}{\Delta T}\right)^2\right)d[\theta,\theta]_t \;+\; o_p(1). \tag{10}$$

With standard calculations, one can get similar results when observing data with microstructure noise, $Y_{T_i} = \theta_{T_i} + \epsilon_{T_i}$. What is clear from (10), however, is that pre-averaging (followed by the usual two scales correction) is robust to the most exotic forms of jumps, but with two caveats. One is that (naturally) one cannot capture a jump at the end time $\mathcal{T}$. The other is that if one has reason to scale with a sufficiently small K, one may pick up the $\frac{1}{K^2}$ term in (10) at some point, for example as asymptotic bias. This term would not be a problem with the usual scaling $K = O(B^{1/2})$, but sometimes a smaller K is warranted, see, *e.g.*, Jacod and Rosenbaum (2013, 2015); Mykland and Zhang (2016b), or in the case of models with shrinking size of noise. □

Remark 4. (Link to Volatility of Volatility.) We shall see in Section 3 that Theorem 1 is an ingredient in the estimation of $[\theta,\theta]_{\mathcal{T}-}$. A specific procedure is given in Theorem 4 in Section 3.2. In particular, for $\theta_t = \sigma_t^2$, one retireves an estimator of volatility of volatility. This connects to an earlier estimator of $[\sigma^2,\sigma^2]_{\mathcal{T}}$ by Vetter (2015), which is further discussed in Example 10. An

[11] 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).

[12] Jacod, Li, Mykland, Podolskij, and Vetter (2009a); Podolskij and Vetter (2009b).

[13] For more on the leverage effect, and further references, see Example 9 in Section 7.

[14] The closest we can find is Chapter 16.2-16.3 of Jacod and Protter (2012), which has several important contributions. We see our statement (10) as a complement to their findings.

estimator of volatility that is based on different principles can be found in Mykland, Shephard, and Sheppard (2012, Theorem 7 and Corollary 2). □

We emphasize that the cited papers in Remarks 3-4 also have central limit theorems (CLT), rather than just consistency. Our main focus is asymptotic variance (AVAR), where only consistency is necessary, and we are interested in the weakest possible conditions for such consistency to hold. Higher order properties of the AVAR would be interesting, *cf.* the discussion of likelihood methods in Section 3.3, but this seems beyond the scope of this paper.

The particular sharpness of Theorem 1 is due to the following result. Since it may have other applications, we provide the main building block as a separate result. The proof is also in Appendix B. The result is also true for many other processes than semimartingales.

THEOREM 2. (REWRITING INTEGRAL DIFFERENCES AS SEMIMARTINGALE INCREMENTS.) *Let θ_t be a semimartingale. We use the following notation. For nonrandom times $S < T$, set*

$$\Theta'_{(S,T]} = \int_S^T (T-t)d\theta_t \text{ and } \Theta''_{(S,T]} = \int_S^T (t-S)d\theta_t. \tag{11}$$

Then, if $\delta > 0$ is nonrandom

$$\Theta_{(T,T+\delta]} - \Theta_{(T-\delta,T]} = \Theta'_{(T,T+\delta]} + \Theta''_{(T-\delta,T]}. \tag{12}$$

3 Estimating Asymptotic Variance in High Frequency Data

3.1 General Principles for the Asymptotic Variance

Following the notation (3), we have at hand estimators $\hat{\Theta}_{(S,T]} = \hat{\Theta}^{(n)}_{(S,T]}$ of $\Theta_{(S,T]}$.[15]

The typical statistical situation is now as follows: there is a semimartingale $M_{n,t}$ and *edge effects* $e_{n,S}$ and $\tilde{e}_{n,T}$, so that,

$$\hat{\Theta}^{(n)}_{(S,T]} - \Theta_{(S,T]} = \underbrace{M_{n,T} - M_{n,S}}_{\text{semimartingale}} + \underbrace{\tilde{e}_{n,T} - e_{n,S}}_{\text{edge effects}} \text{ for } S < T \in \mathcal{T}_n, \tag{13}$$

where $\mathcal{T}_n = \{T_{n,i} : \ i = 0, \cdots B_n\}$.[16] The edge effect is essentially anything that messes up the semimartingaleness of the difference $\hat{\Theta}_{(0,T]} - \Theta_{(0,T]}$, and it occurs in many shapes, which we shall document in Section 7.[17] The edge effect has a component e_S relating to phasing in the estimator

[15] See Section 5.1 on how to obtain such estimators from half-interval estimators. The latter are required for stable convergence results, *cf.* the development in this section and in Section 6.

[16] Until we reach Section 6.

[17] 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

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 $\widehat{\Theta}_n = \widehat{\Theta}^{(n)}_{(0,\mathcal{T}]}$ from now on. An important construction leading to (13) relates to half-interval estimators (Section 5.1).

REMARK 5. (EDGE EFFECTS.) To rephrase, the Edge Effect reflects the difference in behavior of an estimator between the middle and the edges of the interval on which it is defined. For a conceptual illustration, consider the bi-power estimator (Barndorff-Nielsen and Shephard (2004b, 2006)) of the integrated volatility of a process X_t, where X_t is observed (without microstructure noise) at equidistant times $t_i,\ i = 0, \cdots, n$, spanning $[0, \mathcal{T}]$. The estimator has the form $\hat{\Theta}_{(S,T]} = \frac{\pi}{2}\sum_{S<t_{i-1}\leq t_i\leq T} |\Delta X_{t_{i-1}}||\Delta X_{t_i}|$. 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 Example 2 in Section 7, along with a number of other examples. In fact, the only estimators that we can identify to not have edge effect, is realized volatility and other power variations absent microstructure noise. □

Meanwhile, we seek an estimator of the asymptotic variance of $\hat{\Theta}^{(n)}_{(0,\mathcal{T}]}$. For a conceptual path, we turn to the substantial fraction of the high frequency literature which has been concerned with the study of the asymptotic behavior of $\hat{\Theta}^{(n)}_{(S,T]} - \Theta^{(n)}_{(S,T]}$ for all $S < T \in [0, \mathcal{T}]$. This is typically required to achieve *stable convergence.*

DEFINITION 3. (STABLE CONVERGENCE.) *Let $L_n = (L_{n,t})_{0\leq t\leq\mathcal{T}}$ be a sequence of semimartingales (Definition 1 in Section 2.1). We say that L_n converges stably in law to $L = (L_t)_{0\leq t\leq\mathcal{T}}$ with respect to a sigma-field $\mathcal{G} \subseteq \mathcal{F}$, and as $n \to \infty$, if (1) L_t is measurable with respect to a sigma-field $\tilde{\mathcal{G}}$ belonging to an extension $(\tilde{\Omega}, \tilde{\mathcal{G}}, \tilde{P})$ of $(\Omega, \mathcal{G}, P)$; and (2) for every $\mathcal{G}$-measurable (real valued) random variable Y, (L_n, Y) converges in law to (L, Y).*

For further explanation of stable convergence, and of the upcoming P-UT condition, see Appendix D.1.

As illustrated by a number of examples in Section 7 below, the standard high frequency asymptotic *result* in the literature is now as follows.

CONDITION 1. (STANDARD CONVERGENCE RESULT IN THE LITERATURE.) *Assume (13), and that one can show the following. There is an $\alpha > 0$ so that as $n \to \infty$,*

$$n^\alpha M_{n,t} \xrightarrow{\mathcal{L}} L_t \text{ stably in law} \tag{14}$$

with respect to a sigma-field $\mathcal{G}$. The quadratic variation $[L, L]_\mathcal{T}$ (Section 2.1) is measurable with respect to $\mathcal{G}$, and L_t is a local martingale conditionally on $\mathcal{G}$. Also, $e_{n,T_n} = o_p(n^{-\alpha})$ and $\tilde{e}_{n,S_n} = o_p(n^{-\alpha})$ for any $S_n, T_n \in \mathcal{T}_n$. Finally, the sequence $n^\alpha M_{n,t}$ is Predictably Uniformly Tight (P-UT)

our notation to avoid an excessive number of subscripts, but when crucial for understanding we may write $M_{n,T}$, etc. By convention, we use superscript "(n)" when it is too unaesthetic to place n as a subscript.

(Appendix D.1; Jacod and Shiryaev (2003, Chapter VI.3.b, and Definition VI.6.1, p. 377)).[18]

We recall the basic facts about this situation. First, Condition 1 assures that, with $\hat{\Theta}_n = \hat{\Theta}^{(n)}_{(0,\mathcal{T}]}$ and $\Theta = \Theta_{(0,\mathcal{T}]}$,

$$n^{\alpha}(\hat{\Theta}_n - \Theta) \xrightarrow{\mathcal{L}} L_{\mathcal{T}} \text{ stably in law .} \tag{15}$$

Also, the asymptotic variance of $n^{\alpha}(\hat{\Theta}_n - \Theta)$ given the underlying data represented by $\mathcal{G}$ is

$$V = \text{AVAR}(n^{\alpha}(\hat{\Theta}_n - \Theta)) = \text{Var}(L_{\mathcal{T}}|\mathcal{G}). \tag{16}$$

To declutter the notation, we shall define $\text{AVAR}_n = \text{AVAR}((\hat{\Theta}_n - \Theta))$, formally[19]

$$\text{AVAR}_n = n^{-2\alpha}\text{Var}(L_{\mathcal{T}}|\mathcal{G}). \tag{17}$$

Second, we have guidance on how to estimate the asymptotic variance:

Proposition 1. (Quadratic Variation and Asymptotic Variance.) *Assume Condition 1. Then the conditional variance* $\text{Var}(L_{\mathcal{T}}|\mathcal{G})$ *exists (is "well defined") and*

$$[M_n, M_n]_{\mathcal{T}} \;=\; \text{AVAR}_n \;(1 + o_p(1)). \tag{18}$$

Proof of Proposition 1. Towards the end of Appendix D.1. *Q.E.D.*

In particular, a necessary and sufficient condition for an estimator $\widehat{\text{AVAR}}_n$ of asymptotic variance to be consistent, *i.e.*, $\widehat{\text{AVAR}}_n \;=\; \text{AVAR}_n \;(1 + o_p(1))$, is that

$$\widehat{\text{AVAR}}_n \;=\; [M_n, M_n]_{\mathcal{T}} \;(1 + o_p(1)). \tag{19}$$

We emphasize that for empirical analysis, one does not need to know the form or value of any of the limiting quantities L_t, $[L, L]_{\mathcal{T}}$, and $\mathcal{G}$ in Condition 1 in order to estimate the asymptotic variance.[20] All one needs is to check the criterion (19). We shall in the sequel use this path to show that our proposed estimator is consistent. The procedure can be used much like observed information or bootstrapping, and recall that practical guidance is provided in Section 5.

Because of its importance, and also to illustrate the simplicity of the approach, we here state the main usage as a corollary to the above development.

Proposition 2. (Feasible Estimation.) *Assume the Condition 1. Also assume that* L_T *is conditionally Gaussian given* $\mathcal{G}$*. Suppose that* $\widehat{\text{AVAR}}_n \;=\; [M_n, M_n]_{\mathcal{T}} \;(1 + o_p(1))$*. Set* $\text{se}(\hat{\Theta}_\text{n}) =$

[18]See Section 6 for further explanation of this condition, as well as some standard methods for how to verify it. Examples of verification are also given in Section 7.

[19]As foreshadowed by Footnote 2. In the notation of this earlier footnote, $V = \text{Var}(L_{\mathcal{T}}|\mathcal{G})$.

[20]In fact, an automatic minimal $\mathcal{G}$ is provided by Proposition 6 in Appendix D.1.

$|\widehat{\mathrm{AVAR}}_n|^{\frac{1}{2}}$. *Then*

$$\frac{\widehat{\Theta}_n - \Theta}{\mathrm{se}(\hat{\Theta}_n)} \xrightarrow{\mathcal{L}} N(0,1) \text{ stably in law.} \tag{20}$$

3.2 Main Findings: A General Expansion Result for $QV_{B,K}(\widehat{\Theta}_n)$, and The Two Scales AVAR and $\widehat{[\theta,\theta]}$

For a given grid, we use the notation

$$\mathrm{ave}(e_{T_i}^2) \triangleq \frac{1}{B_n}\sum_i e_{T_i}^2 \tag{21}$$

and similarly for $\mathrm{ave}(\tilde{e}_{T_i}^2)$. Observe that $\tilde{e}_0 = e_{\mathcal{T}} = 0$ by convention. We obtain:

Theorem 3. (Expansion of $QV_{B,K}(\hat{\Theta})$.) *Assume Condition 1. Let $K = K_n$ be positive integers, and assume that $K_n \Delta T_n \to 0$. Also assume about the averages of the edge effects that*

$$\mathrm{ave}(e_{T_i}^2) \ = o_p(K_n\Delta T_n n^{-2\alpha}) \text{ and } \mathrm{ave}(\tilde{e}_{T_i}^2) = o_p(K_n\Delta T_n n^{-2\alpha}). \tag{22}$$

Then

$$\frac{1}{2K}\sum_{K\le i\le B-K}(\hat{\Theta}_{(T_{i-K},T_{i+K}]} - \Theta_{(T_{i-K},T_{i+K}]})^2 = \mathrm{AVAR}_n\ (1+o_p(1)). \tag{23}$$

Also, if we assume that θ_t is a semimartingale on $[0,\mathcal{T}]$, and that

$$\Delta T_n = o(n^{-\alpha}), \tag{24}$$

then

$$QV_{B,K}(\hat{\Theta}) \ = \ 2AVAR_n \ + \ \frac{2}{3}(K_n\Delta T_n)^2[\theta,\theta]_{\mathcal{T}-} \ + \ o_p((K_n\Delta T_n)^2) + o_p(n^{-2\alpha}) \tag{25}$$

The convergence in probability is uniform in ΔT, so long as $\Delta T > 0$ and $K\Delta T \to 0$.

Proof. See Appendix C, where it is also shown that a related result holds under (occasionally useful) weaker conditions. *Q.E.D.*

On the basis of Theorem 3, we now provide the estimators that we recommend for most situations.

Definition 4. (Two Scales AVAR, and Volatility of Spot θ.) *Let B, K and $QV_{B,K}(\hat{\Theta})$*

be as in Definition 2. Let $K_1 < K_2$ be two distinct values of K. The estimators[21]

$$\mathrm{TSAVAR}_n = \frac{1}{2}\left(\frac{1}{K_1^2} - \frac{1}{K_2^2}\right)^{-1}\left(\frac{1}{K_1^2}QV_{B,K_1}(\hat{\Theta}) - \frac{1}{K_2^2}QV_{B,K_2}(\hat{\Theta})\right) \text{ and} \tag{26}$$

$$\widehat{[\theta,\theta]}_{\mathcal{T}-} = \frac{3}{2}(K_2^2 - K_1^2)^{-1}(\Delta T)^{-2}\left(QV_{B,K_2}(\hat{\Theta}) - QV_{B,K_1}(\hat{\Theta})\right) \tag{27}$$

as well as $\mathrm{se}(\hat{\Theta}_\mathrm{n}) = |\mathrm{TSAVAR}_n|^{\frac{1}{2}}$ *are referred to as two scales asymptotic variance, volatility, and standard error. When $K_2 = 2K_1 = 2K$, we shall refer to $(1,2)$ estimators. Specifically, the (1,2) TSAVAR is*

$$\mathrm{TSAVAR}_n = \frac{2}{3}\left(QV_{B,K}(\hat{\Theta}) - \frac{1}{4}QV_{B,2K}(\hat{\Theta})\right) \tag{28}$$

The consistency of the two scales estimators is given by the following result.

Theorem 4. (Consistency of Two Scales AVAR and Volatility of Spot θ. Feasibility of inference.) *Assume Condition 1, and that θ_t is a semimartingale on $[0, \mathcal{T}]$. Assume that*

$$\mathrm{ave}(e_{T_i}^2) = o_p(n^{-3\alpha}) \text{ and } \mathrm{ave}(\tilde{e}_{T_i}^2) = o_p(n^{-3\alpha}). \tag{29}$$

Assume that $\Delta T_n = o(n^{-\alpha})$. Let $K_{n,1} < K_{n,2}$ be positive integers, and assume that $K_{n,i}\Delta T_n \to 0$ for $i = 1, 2$. Assume that both $K_{n,1}$ and $K_{n,2}$ satisfy the balance condition

$$K_n \Delta T_n \text{ are of the same order as } n^{-\alpha}. \tag{30}$$

with $\liminf_{n\to\infty}(K_{n,2}/K_{n,1}) > 1$, *which assures that neither of main terms in (25) is ignorable.*

Then, TSAVAR_n *and* $\widehat{[\theta,\theta]}_{\mathcal{T}-}$ *are consistent:*

$$\begin{aligned} \mathrm{TSAVAR}_n &= \mathrm{AVAR}_n\ (1 + o_p(1)) \text{ and} \\ \widehat{[\theta,\theta]}_{\mathcal{T}-} &\xrightarrow{p} [\theta,\theta]_{\mathcal{T}-}. \end{aligned} \tag{31}$$

Finally, if L_T is conditionally Gaussian given $\mathcal{G}$, then

$$\frac{\widehat{\Theta}_n - \Theta}{\mathrm{se}(\hat{\Theta}_\mathrm{n})} \xrightarrow{\mathcal{L}} N(0,1) \text{ stably in law.} \tag{32}$$

[21] The TSAVAR ((26) and (28)) does not have similar coefficients to the Two-Scales Realized Volatility (TSRV, Zhang, Mykland, and Aït-Sahalia (2005)). For a heuristic explanation of this, consider the left hand side of (5), and write it as noise + signal - noise from previous interval. This looks like a scene from "inference with micro-structure noise", especially if the noise is shrinking (via, say, pre-averaging). The AVAR problem is different, however, in that the "signal" has different properties. In particular, it shrinks at rate $O_p(K_n\Delta T_n)$ by Theorems 1-2. It also has different dependence structure.

PROOF OF THEOREM 4. Theorem 3 and assumption (30) gives rise to (33), for $K = K_1$ and K_2. Ignoring remainder terms gives rise to estimators defined by a system of two equations and two unknowns by letting $K = K_1$ and $= K_2$ in (34). By linear algebra, this system is equivalent to the formulae for the estimators TSAVAR_n and $\widehat{[\theta,\theta]}_{\mathcal{T}-}$ given in (26)-(27) in Definition 4. The estimators are consistent by substituting (34) into (33) and then using that $\liminf_{n\to\infty}(K_{n,2}/K_{n,1}) > 1$. The last part of the result follows from Proposition 2. *Q.E.D.*

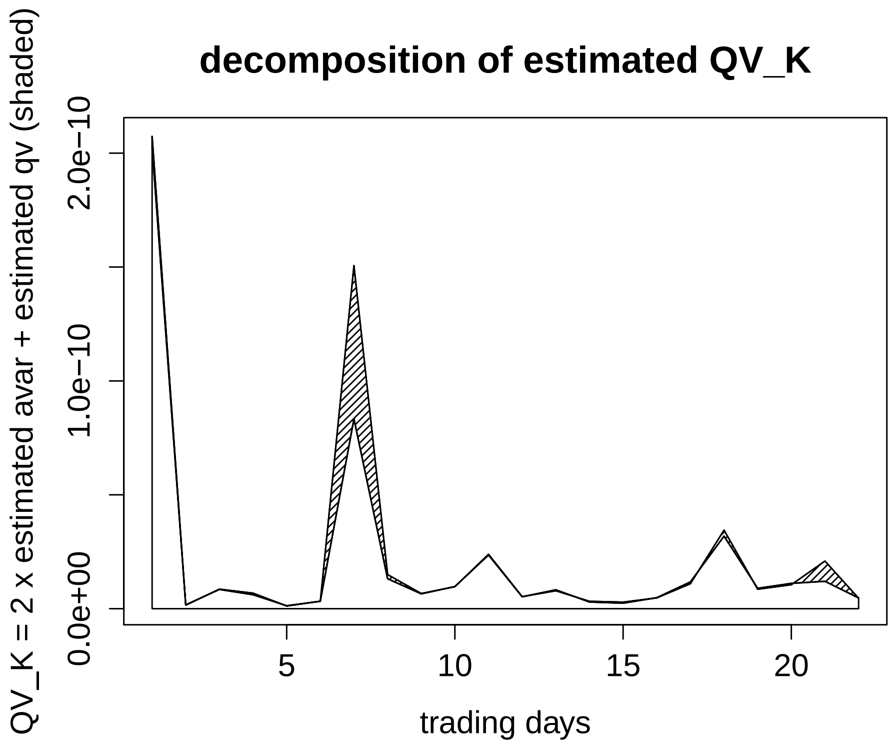

Figure 1: This plot illustrates the empirical decomposition (34) for the S&P E-mini future as traded on the Chicago Mercantile Exchange, across the 22 trading days of May 2007. The top curve is the total volatility $QV_{B,K}(\hat{\Theta})$ for each day, the lower curve is $2 \times$ TSAVAR for each day, and the shaded part is $\frac{2}{3}(K_1\Delta T)^2\widehat{[\theta,\theta]}_{\mathcal{T}-}$. In the estimation, the underlying parameter is the spot volatility: $\theta_t = \sigma_t^2$. $\widehat{\Theta}_{(S,T]}$ is based on first pre-averaging the data to 15 seconds, and then computing a TSRV on these pre-averages with $j = 20$ and $k = 40$ (Mykland and Zhang (2016a), see also Example 4 in Section 7). The estimator is thus of integrated volatility $\Theta_{(S,T]} = \int_S^T \sigma_t^2 dt$, and $[\theta,\theta]_{\mathcal{T}} = [\sigma^2,\sigma^2]_{\mathcal{T}}$. For $QV_{B,K}(\hat{\Theta})$, we take ΔT to be five minutes, and a (1, 2) TSAVAR is computed on this basis for every five minute period, using the forward half interval method in Section 5.1. The estimation method satisfies the edge condition (29) in Theorem 4 (Example 4).

REMARK 6. (THEORETICAL AND EMPIRICAL DECOMPOSITIONS OF $QV_{B,K}$.) Under the assumption (30), for $K = K_1$ or K_2, we have the theoretical decomposition:

$$QV_{B,K}(\hat{\Theta}) \;=\; 2\text{AVAR}_n \;+\; \frac{2}{3}(K\Delta T)^2[\theta,\theta]_{\mathcal{T}-} \;+\; o_p(n^{-2\alpha}). \tag{33}$$

Meanwhile, the two scales estimators TSAVAR_n and $\widehat{[\theta,\theta]}_{\mathcal{T}-}$ satisfy a corresponding empirical decomposition:

$$QV_{B,K_i}(\hat{\Theta}) \;=\; 2\;\text{TSAVAR}_n \;+\; \frac{2}{3}(K_i\Delta T)^2\widehat{[\theta,\theta]}_{\mathcal{T}-}\;,\; i=1,2. \tag{34}$$

One can think of (34) as an empirical decomposition of $QV_{B,K_i}(\hat{\Theta})$ into TSAVAR_n and $\widehat{[\theta,\theta]}_{\mathcal{T}-}$, *cf.* Figure 1. □

To get a sense of how the empirical decomposition (34) plays out in real data, we plot the separation of TSAVAR_n and $\widehat{[\theta,\theta]}_{\mathcal{T}-}$ 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 $QV_{B,K}(\hat{\Theta})$, and on day 7 and day 21, the dispersion $[\theta,\theta]_{\mathcal{T}}$ of the underlying spot parameter moves notably.

Remark 7. (Guidance on ΔT and K, and the choices that lead from Theorem 3 to Theorem 4) Both ΔT and K are under the control of the econometrician, and we offer the following main approach to choosing these two tuning parameters.

i. By linear combination of $QV_{B,K}$ for two or more K's, one can eliminate either the $[L,L]_{\mathcal{T}}$ or the $[\theta,\theta]_{\mathcal{T}-}$ term in (25). We have seen this in Theorem 4 above. This means that the main question is how to optimize Theorem 3 with respect to ΔT and K.

ii. On the one hand, ΔT may be arbitrarily small. ΔT is, therefore, limited only by one's computational power. In particular the assumption (24) is routinely satisfied in practice.

iii. In fact, ΔT ought to be small. In particular, by a sufficiency argument, $QV_{2B,2K}(\hat{\Theta})$ will under mild conditions have less variability (given the data) than $QV_{B,K}(\hat{\Theta})$. This is akin to the desirability of post-averaging after subsampling (Zhang, Mykland, and Aït-Sahalia (2005, Section 3.1, p. 1399)).

iv. On the other hand, $K\Delta T$ ought not to be very small. As a general rule, we recommend to take $K_n\Delta T_n$ to be of the same order as $n^{-\alpha}$, *cf.* Condition (30) in Theorem 4. This reflects the need for $K_n\Delta T_n$ to respect both the lower and the upper bounds imposed by Theorem 3.

   The reason for this is twofold. First, assumption (22) is a lower bound on $K_n\Delta T_n$. This bound is required to guarantee that the remainder term in (25) is no larger than $o_p((K_n\Delta T_n)^2) + o_p(n^{-2\alpha})$. It can be seen in explicit form from Theorem 8 (in Appendix C) that without assumption (22), one can expect the remainder term in (25) to involve $R_{n,K}$ (eq. (C.33) in the Appendix), which is of the same order as $(K_n\Delta T_n)^{-1}(\text{ave}(e_{T_i}^2) + \text{ave}(\tilde{e}_{T_i}^2))$. Hence assumption (22) is necessary.

Second, choosing $K_n\Delta T_n$ to be of larger order than $n^{-\alpha}$ causes AVAR_n to be dwarfed $[\theta,\theta]_{\mathcal{T}-}$. This is an upper bound on $K_n\Delta T_n$.[22]

Under the assumption (29) in Theorem 4, $K_n\Delta T_n$ is thus chosen to be as large as possible while still satisfying Theorem 3.

v. In summary, one should thus think of ΔT as a computational parameter, while $\delta = K\Delta T$ represents an amount of time over which one can reasonably compute estimators $\hat{\Theta}_{(T,T+\delta]}$. □

Remark 8. Finite Sample Adjustment. Without impacting the asymptotics, one can make finite sample adjustments, and use $\frac{B-2K+1}{B}QV_{B,K}(\hat{\Theta})$ in lieu of $QV_{B,K}(\hat{\Theta})$, and $\frac{B-2K-\mathcal{M}'+1}{B}QV_{B,K,\mathcal{M}'}(\hat{\Theta})$ in lieu of $QV_{B,K,\mathcal{M}'}(\hat{\Theta})$ from (47). The adjustment will produce "unbiasedness" in Theorem 1 when $[\theta,\theta]_t$ is absolutely continuous with constant derivative. □

3.3 One Scale Standard Error, Quarticity, and the Likelihood Connection

Tiny Edge Effects. For some estimators, one can choose

$$K_n\Delta T_n = o(n^{-\alpha}). \tag{35}$$

while (22) remains satisfied. This is most often the case for estimators based on data with no microstructure noise, such as Realized Volatility (RV, Example 1), Bipower Variation (Example 2), and some estimator of integrals of functions of volatility (Example 6). (The examples are further discussed in Section 7, where references to the literature is also given.) We emphasize that the choice (35) may not be possible for estimators based on increasing-size blocks, or on data with microstructure noise.

Remark 9. (A One Scale Standard Error). Assume the conditions of Theorem 3 except condition (24). Assume instead (35). Set $\widehat{\text{AVAR}}_n = \frac{1}{2}QV_{B,K}(\hat{\Theta})$. Then $\widehat{\text{AVAR}}_n$ is consistent. □

Quarticity. The quarticity of Barndorff-Nielsen and Shephard (2002a, 2004a) can be viewed as a one scale estimator in our setup. Instead of (5) in Section 2.2, one writes[23]

$$\hat{\Theta}_{(T_i,T_{i+K}]} = \underbrace{\hat{\Theta}_{(T_i,T_{i+K}]} - \Theta_{(T_i,T_{i+K}]}}_{\text{estimation error}} + \underbrace{\Theta_{(T_i,T_{i+K}]}}_{\text{parameter value}}. \tag{36}$$

This consideration leads to a generalized quarticity, on the form $Q_{B,K} = \frac{1}{K}\sum_{i=0}^{B-K}(\hat{\Theta}_{(T_i,T_{i+K}]})^2$.

[22] A more elaborate development may allow for $K_n\Delta T_n$ to be of order larger than $O(n^{-\alpha})$, as with the cancellation of microstructure noise in two- and multi-scale estimation (Zhang, Mykland, and Aït-Sahalia (2005), Zhang (2006)) but such a development is beyond the scope of this paper.

[23] This is close to the argument in Barndorff-Nielsen and Shephard (2004a, Appendix B.1.1, p. 922-923).

THEOREM 5. (EXPANSION OF Q_{K,B_n}.) *Assume Condition 1, and that θ_t is a continuous semi-martingale on $[0, \mathcal{T}]$. Suppose that K is a finite integer, and that $\Delta T_n = O(n^{-2\alpha})$. Also assume (22) about the averages of the edge effects. Then*

$$Q_{B_n,K} \;=\; \mathrm{AVAR}_n + K\Delta T_n \int_0^{\mathcal{T}} \theta_t^2 dt + o_p(n^{-2\alpha}). \tag{37}$$

PROOF. By the same method as the proof of Theorem 3, and by the sample-path continuity of θ_t. *Q.E.D.*

In the case where $\hat{\Theta}$ is realized variance based on the observed process X_t, $\theta_t = \sigma_t^2$ (the volatility of X), $B_n = n$, and $\alpha = \frac{1}{2}$. There is no edge effect (cf. Example 1 in Section 7). Thus, $\mathrm{AVAR}_n = 2\Delta T_n \int_0^{\mathcal{T}} \theta_t^2 dt$. In the case where $K = 1$, one retrieves $Q_{1,B_n} = \frac{3}{2}\mathrm{AVAR}_n(1 + o_p(1))$. We thus retrieve the results of Barndorff-Nielsen and Shephard (2002a, 2004a), also similarly in the case of (synchronous) covariance, correlation, and regression.

A number of estimators have similar behavior in the sense that AVAR is proportional to $\int_0^{\mathcal{T}} \theta_t^2 dt$. These include Bipower and Multipower Variation (Barndorff-Nielsen and Shephard (2004b, 2006)), and estimation of integrals of $\theta = \sigma^p$ with finite blocks (Mykland and Zhang (2009, Section 4.1, p. 1421-1426), Mykland and Zhang (2012, Ch. 2.6.2, pp. 170-172)).

In the more general case where $\int_0^{\mathcal{T}} \theta_t^2 dt$ is not directly related to AVAR_n, one can go to a two scales estimator and obtain that $2Q_{B_n,1} - Q_{B_n,2} = \mathrm{AVAR}_n + o_p(n^{-2\alpha})$. We have not investigated the situation for quarticity where θ_t is discontinuous.

A LIKELIHOOD CONNECTION. We think of the observed AVAR as akin to the observed information in likelihood theory. Barndorff-Nielsen and Shephard have a similar view of quarticity (Barndorff-Nielsen and Shephard (2015)).

The observed asymptotic variance is like 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 statistics, there has been a lively debate about the relative accuracy properties of observed and estimated expected information. In statistics, *accuracy* refers to the the closeness of an approximation to the true distribution of a statistic. For the standard error, accuracy can refer *either* to how close the statistic is to the actual standard deviation of a statistic, *or* to how the $\mathrm{se}(\hat{\Theta}_\mathrm{n})$ best accomplishes the asymptotic approximation of the law of $\hat{\Theta}_n - \Theta/\mathrm{se}(\hat{\Theta}_\mathrm{n})$ to a normal or other reference distribution. Some of the same considerations may apply to the observed AVAR in this paper, but this question is beyond the scope of this paper.

The subject originally goes 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), who demonstrated that the observed information in many cases was a more accurate measure of the variance of an estimator. This breakthrough was followed by a large literature, including Barndorff-Nielsen (1986, 1991); Jensen (1992, 1995, 1997); McCullagh (1984, 1987); Skovgaard (1986, 1991); Mykland (1999, 2001).

4 Application: 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 (2009b); Jacod and Mykland (2015). 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 (2015) does carry out 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); Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011); Bibinger and Mykland (2016)), spot volatility estimation (see Mykland and Zhang (2008)), estimation of the leverage effect (Wang and Mykland (2014), Aït-Sahalia, Fan, Laeven, Wang, and Yang (2016), Kalnina and Xiu (2015)), and estimation of the volatility of volatility (Vetter (2015), Mykland, Shephard, and Sheppard (2012)). 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 $\hat{\Theta}_{n,c}$ depend.

Condition 2. *Suppose that there is a tuning parameter c (chosen by the econometrician) upon which $\hat{\Theta}_n = \hat{\Theta}_{n,c}$ and $\mathrm{AVAR}_n = \mathrm{AVAR}_{n,c}$ depend.*[24] *Assume (as provided by, say, Proposition 1 in Section 3.1, or Theorem 4 in Section 3.2)*

$$\forall c \in \mathcal{C} : \widehat{\mathrm{AVAR}}_{n,c} = \mathrm{AVAR}_c(1 + o_p(1)) \text{ (for fixed } c). \tag{38}$$

We seek $c^* = \arg\min_c \mathrm{AVAR}_{c\in\mathcal{C}}$, 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

[24]Observe that Θ does not depend on c, but will normally be (statistically) mutually dependent with c^*. Recall that we assume that $\mathrm{AVAR}_n = n^{-2\alpha}V$, (cf. (16)-(17) in Section 3.1 as well as Footnote 2 in the Introduction.

facie discussion, we also take the number of points in $\mathcal{C}$ to be finite.[25]

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

Consistency. Under Condition 2, automatically,

$$\hat{c}_n \to c^*. \tag{39}$$

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

$$\text{asymptotic variance of } \hat{\Theta}_{n,\hat{c}_n} - \Theta \;=\; \mathrm{AVAR}_{n,c^*}. \tag{40}$$

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 Condition 1. At least in our simple case, however, this difficulty does not arise. We embody this in a formal result.

Proposition 3. (Optimization commutes with Asymptotic Variance.) *Assume Conditions 1 and 2. Also suppose that c^* is $\mathcal{G}$-measurable, and that, for each $c \in \mathcal{C}$, $(\hat{\Theta}_{n,c} - \Theta)/\mathrm{AVAR}_c^{1/2}$ converges stably in law to a $N(0,1)$ random variable that is independent of $\mathcal{G}$.*[26] *Then (40) holds, and also*

$$(\hat{\Theta}_{n,\hat{c}_n} - \Theta)/\mathrm{AVAR}_{c^*}^{1/2} \xrightarrow{\mathcal{L}} N(0,1) \text{ and } (\hat{\Theta}_{n,\hat{c}_n} - \Theta)/\widehat{\mathrm{AVAR}}_{n,\hat{c}_n}^{1/2} \xrightarrow{\mathcal{L}} N(0,1), \text{ both stably.} \tag{41}$$

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. *Q.E.D.*

Example. A (J,K)-TSRV estimator based on pre-averaged data, with J and K finite, provides an example where the action space $\mathcal{C}$ can indeed be taken to be finite. This estimator is studied in Example 4 in Section 7 below, and it is seen that the two-scale $\widehat{\mathrm{AVAR}}_n$ from Section 3.2 satisfies Theorem 4. The assumptions of Proposition 3 are thus satisfied. □

[25] This case is 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.

[26] In other words, one must check the conditions of Proposition 2 for each $c \in \mathcal{C}$.

5 Guidance: I. Practice

We here give advice on how to practically carry out the estimation of the asymptotic variance. The situation is that one has a data set and wishes an $\widehat{\mathrm{AVAR}}_n$.

5.1 Creating Estimators $\hat{\Theta}^{(n)}_{(S,T]}$ in each subinterval $(S,T]$.

In practice, a simple way to obtain estimators $\hat{\Theta}^{(n)}_{(S,T]}$ is to start with a given collection of half-interval estimators $\hat{\Theta}^{(n)}_{(0,T]}$, $0 < T \leq \mathcal{T}$, and write, for $S < T$,

$$\hat{\Theta}^{(n)}_{(S,T]} \;=\; \hat{\Theta}^{(n)}_{(0,T]} - \hat{\Theta}^{(n)}_{(0,S]} \tag{42}$$

We call estimators of the form (42) *forward estimators*. If the half-interval estimators have representation $\hat{\Theta}^{(n)}_{(0,T]} - \Theta_{(0,T]} = M_{n,T} + \tilde{e}_{n,T} - \tilde{e}_{n,0}$, then obviously the representation (13) continues to hold for the forward estimators, with $e_T = \tilde{e}_T$.

Remark 10. (Additive Estimators.) The forward estimators satisfy

$$\hat{\Theta}^{(n)}_{(S,T]} - \Theta_{(S,T]} + \hat{\Theta}^{(n)}_{(T,U]} - \Theta_{(T,U]} = \hat{\Theta}^{(n)}_{(S,U]} - \Theta_{(S,U]}, \text{ for } S < T < U. \tag{43}$$

Another construction of this type is the *backward estimators*: $\hat{\Theta}^{(n,b)}_{(S,T]} \;=\; \hat{\Theta}^{(n)}_{(S,\mathcal{T}]} - \hat{\Theta}^{(n)}_{(T,\mathcal{T}]}$. The development is analogous to that of forward estimators. If estimators are constructed with hindsight, after time $\mathcal{T}$, one can also average the forward and backward estimator, which has slightly better properties by sufficiency considerations. (Similarly to Remark 7(iii).) □

5.2 Irregular Sampling: Validity of the Previous Tick Approach. Several Dimensions

For simplicity, we discuss this issue for the forward or other additive estimator introduced above. We suppose that data arrives at times $t_{n,i},\ i = 0, \cdots, B'_n$. We shall take this to mean that the underlying half-interval estimator $\hat{\Theta}^{(n)}_{(0,T]}$ changes values at times $T = t_{n,i}$. We then set

$$\hat{\Theta}^{(n)}_{(0,T_i]} \triangleq \hat{\Theta}^{(n)}_{(0,T_{n,i,*}]} \text{ where } T_{n,i,*} = \max\{\ t_{n,j} \leq T_{n,i}\ \}, \tag{44}$$

and proceed as if nothing has happened. This is the previous tick scheme, see Zhang (2011) and the references therein.

The rationale for this is the following result, which is shown in Appendix D.2.

Proposition 4. (Previous Tick Sampling.) *Assume that the* $t_{n,i},\ i = 0, \cdots, B'_n$ *is (for each* *n) a non-decreasing sequence of stopping times. Suppose that*

$$\sup_i |T_{n,i,*} - T_{n,i}| \xrightarrow{p} 0 \text{ as } n \to \infty, \tag{45}$$

as well as $T_{n,0,*} = 0$ *and* $T_{n,B'_n,*} = \mathcal{T}$. *In the formal results*[27] *of this paper, the conditions on the microstructure* $\tilde{e}_{n,T_{n,i}}$ *may be replaced by the same conditions on* $\tilde{e}_{n,T_{n,i,*}}$. $\mathcal{F}_{T_{n,i}}$ *may, however, not be replaced by* $\mathcal{F}_{T_{n,i,*}}$.

In practice, this means that the results in Section 3 are unaffected by the previous-tick sampling.

Remark 11. (When there is no Edge Effect.) The condition (45) is required even when the microstructure noise $\tilde{e}_{n,T,*}$ is identically zero. □

Remark 12. (Several Dimensions.) The extension of this theory to several dimensions is straightforward. All our results carry over appropriately for the regular grid $\{T_{n,i},\ i = 0, \cdots, B_n\}$, using the identity $xy = \frac{1}{2}\left((x+y)^2 - x^2 - y^2\right)$.[28] One can then use Proposition 4 in each dimension, since no time change is involved in our proofs. □

5.3 In case 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, $QV_{B,K}(\Theta) = [\mathfrak{T}, \mathfrak{T}]_T - [\mathfrak{T}, \mathfrak{T}]_0 + o_p(1)$. The same is true for $QV_{B,K}(\hat{\Theta})$. 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[29] $(T_{i_j}, T_{i_j+1}]$, $j = 1, \cdots, \nu$. (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

[27] Theorems, propositions, corollaries, and lemmae. We emphasize that (unless the $t_{n,i}$ are nonrandom, or in certain other circumstances), the $T_{n,i,*}$ may not be stopping times. Hence, for example, the argument in Remark 14 (Section 7) may not be valid. Also, $\mathcal{F}_{T_{n,i,*}}$ will not be defined unless $T_{n,i,*}$ is a stopping time. In case of doubt, please make use of the more specific Proposition 5 in Section 6.

[28] See the definition of multivariate quadratic variation in Jacod and Shiryaev (2003, Eq. (I.4.46), p. 52).

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

relevant squares in $QV_{B,K}(\hat{\Theta})$ are computed as $(\hat{\Theta}_{(T_{i+K},T_{i+2K}]} - \hat{\Theta}_{(T_{i-K},T_i]})^2$. Call this quantity $QV_{B,K,\text{modified}}(\hat{\Theta})$. Similarly, for the true process θ, denote the modified averaged quadratic variation by $QV_{B,K,\text{modified}}(\Theta)$. □

The critical piece for analyzing the above construction is then the following, which generalizes Theorem 1 in Section 2.3, by the same methods.

Theorem 6. (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 let $T_i = i\Delta T$. Suppose that $K\Delta T \to 0$, and that $K \to \infty$. Let $\tau_1, \cdots, \tau_v \in (0,T)$ be stopping times. 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*

$$\begin{aligned} \frac{1}{(K\Delta T)^2} QV_{B,K,\mathit{modified}}(\Theta) &= \left(\frac{2}{3}[\theta,\theta]_{\mathcal{T}-} + \frac{2}{3}\sum_{j=1}^{\nu}([\theta,\theta]_{T_{i_j+1}} - [\theta,\theta]_{T_{i_j}})\right)(1+o_p(1)) \\ &\xrightarrow{p} \frac{2}{3}[\theta,\theta]_{\mathcal{T}-} + \frac{2}{3}\sum_{j=1}^{\nu}(\Delta\theta_{\tau_j})^2. \end{aligned} \tag{46}$$

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 4 the TSAVAR (26) remains consistent for $\text{AVAR}_n(\hat{\Theta} - \Theta)$. $QV_{B,K,\text{modified}}(\hat{\Theta})$ 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 the case of many small jumps, it is unlikely that all jumps will be detected. The contiguity results of Zhang (2007), however, may mitigate the problem.

5.4 Block Estimators: the Interface between Block Sizes $\mathcal{M}_n$ and K_n

Estimators are often based on rolling blocks of $\mathcal{M}_n$ observations. See, *e.g.,* Examples 6, 7, 9, and 10 and Remark 14 in Section 7. We thus have two types of block sizes: *(i)* $\mathcal{M}_n$ is used to construct the underlying $\hat{\Theta}$, and *(ii)* K_n (one or more) is used to construct our current $QV_{B,K}(\hat{\Theta})$, and the resulting AVAR and $[\theta,\theta]_{\mathcal{T}-}$ estimators.

The two fundamental comments on this setup are: (a) it is important to not mix up $\mathcal{M}_n$ and K_n, and (b) there is no need for $\mathcal{M}_n$ and K_n to be related.

In the schematic case[30] where observations times are the same as our T_is, this means that the estimator $\hat{\Theta}_{(0,T_i]}$ is not defined for $i < \mathcal{M}_n$. For $i \geq \mathcal{M}$, however, we can seek relief in forward estimators (Section 5.1), so that no matter what value K_n has, we can define $\hat{\Theta}_{(T_{i-K_n},T_i]} = \hat{\Theta}_{(0,T_i]} - \hat{\Theta}_{(0,T_{i-K_n}]}$ from original forward estimators. These will be defined for $K_n + \mathcal{M}_n \leq i \leq B_n$. With this definition, we can marginally alter $QV_{B,K}(\hat{\Theta})$ from (4) (Section 2.1) to

$$QV_{B,K,\mathcal{M}'}(\hat{\Theta}) = \frac{1}{K} \sum_{i=K+\mathcal{M}'}^{B-K} (\hat{\Theta}_{(T_i,T_{i+K}]} - \hat{\Theta}_{(T_{i-K},T_i]})^2, \tag{47}$$

and similarly for $QV_{B,K}(\Theta)$, where $\mathcal{M}'$ is either $\mathcal{M}_n$ or a slightly larger number (in case an estimator based on a single block is undesirable).

All theorems and other formal results go through unaltered if one replaces $QV_{B,K}(\hat{\Theta})$ by $QV_{B,K,\mathcal{M}'}(\hat{\Theta})$, provided $\mathcal{M}'_n \Delta T_n \to 0$ as $n \to \infty$. This is a substantially weaker requirement than the theoretical condition (63) used to analyze edge effects in Remark 14 in Section 7.

6 Guidance: II. Theory: Tools to verify Condition 1

We again emphasize that it is possible to use our methods without first verifying the conditions. This is standard practice in many areas of inference; the observed information, and bootstrapping, are examples where practice is often ahead of theory. We now, however, pass to the question of how to verify conditions. There are three main strategies: discretization, interpolation, and contiguity.

Discretization. For general results, we recommend, in particular, the books by Jacod and Shiryaev (2003), Jacod and Protter (2012), and Aït-Sahalia and Jacod (2014), as well as the many articles cited above, and in these books.

In our context, we assume for greatest generality that data arrives at irregular times, $t_{n,i}$, $i = 0, \cdots, B'_n$. The semimartingale M_n is on the form

$$M_{n,t} = \sum_{j=1}^{i} \chi_j^n, \text{ for } t_{n,i} \leq t < t_{n,i+1}. \tag{48}$$

We are now outside the framework of a fixed filtration used in the rest of the paper, but there is a path. Proposition 5 will be proved in Appendix D.2.

Condition 3. (Alternative Convergence Condition.) *Let θ_t be a semimartingale on the fixed filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq T}, P)$. Let $t_{i,n}, i = 0, \cdots B'_n$ be a nondecreasing sequence*

[30] Otherwise see Section 5.2.

of $(\mathcal{F}_t)$-stopping times so that

$$\sup |t_{i+1,n} - t_{i,n}| \xrightarrow{p} 0 \text{ as } n \to \infty, \tag{49}$$

as well as $t_{n,0} = 0$ and $t_{n,B'_n} = \mathcal{T}$ for each n. Let $M_{n,t}$ be on the form (48) and assume that $M_{n,t}$ is a semimartingale with respect to filtration $\mathcal{F}^n_t = \mathcal{F}_{t_{n,i}}$ for $t_{n,i} \le t < t_{n,i+1}$. Assume the rest of the wording of Condition 1 with the proviso that $\{(e_{n,T_{n,i}}, \tilde{e}_{n,T_{n,i}}) : T_{n,i} \in \mathcal{T}_n\}$ be replaced with the set of random variables $\{(e_{n,T_{n,i,*}}, \tilde{e}_{n,T_{n,i,*}}) : T_{n,i,*} = \max\{ t_{n,j} \le T_{n,i} \}\}$.

Proposition 5. (Satisfying Conditions with a Discrete Time Martingale). *In the formal results[31] of this paper, Condition 1 may be replaced by Condition 3. At the same time, the conditions on the microstructure $(e_{n,T_{n,i}}, \tilde{e}_{n,T_{n,i}})$ should be replaced by the same conditions on $(e_{n,T_{n,i,*}}, \tilde{e}_{n,T_{n,i,*}})$, while $\mathcal{F}_{T_{n,i}}$ may not be replaced by $\mathcal{F}_{T_{n,i,*}}$. With these modifications, all formal results remain valid.*

We can now avail ourselves of the standard Jacod structure. For example, to satisfy Condition 1, we can check the assumptions of Theorem IX.7.19 (p. 589-590), or of Theorem IX.7.28 (p. 590-591) of Jacod and Shiryaev (2003, Chapter IX.7b, p. 589-591)), with $B \equiv Z \equiv G \equiv 0$. To additionally satisfy P-UT, we additionally need, respectively,

$$\sum_{i=1}^{n} |E(h(\chi_i^n)|\mathcal{F}_{(i-1)/n})| = O_p(1) \text{ or } \sum_{i=1}^{n} |E(\chi_i^n|\mathcal{F}_{(i-1)/n})| = O_p(1). \tag{50}$$

Furthermore, If $L_{n,t} = n^{\alpha} M_{n,t}$ can be written as $L_{n,t} = L^{(1)}_{n,t} + L^{(2)}_{n,t}$, Condition 1 is satisfied for $L_{n,t}$ provided it is satisfied for $L^{(1)}_{n,t}$, and provided $L^{(2)}_{n,t} \to 0$ uniformly in probability (ucp), with (for P-UT)

$$\sum_{i=1}^{n} |E(L^{(2)}_{n,t_{n,i}} - L^{(2)}_{n,t_{n,i-1}} |\mathcal{F}_{n,t_{n,i-1}})| = O_p(1), \tag{51}$$

again by Jacod and Shiryaev (2003)[Theorem VI.6.21 (p. 382)]. Both ucp and P-UT are additive (*ibid.,* Remark 6.4, p. 377).

Incidentally, Theorem IX.7.19, or Theorem 7.28, of Jacod and Shiryaev (2003) also guarantee the conditions of Proposition 2 (feasible estimation).

The methodology is illustrated by Example 6, where the paper by Jacod and Rosenbaum (2013) verifies the stable convergence with the help of Jacod and Shiryaev (2003, Theorem IX.7.19 (p. 590)) and where ignorable terms are ucp, and where it remains to show P-UT-ness. The example illustrates that the P-UT property often follows from the same arguments that give rise to stable convergence.

Interpolation. This has to a great extent been the approach of the current authors. Even if the data are discrete, one can create a continuous martingale by interpolation. One can verify

[31] See Footnote 27 in Section 5.2 for caveats.

Condition 1 by checking the assumptions of Zhang (2001, Theorem B.4, pp. 65-67) or Mykland and Zhang (2012, Theorem 2.28, p. 152-153). The P-UT property is here automatic, by Jacod and Shiryaev (2003, Corollary VI.6.30, p. 385). We have included a procedure of this type in Example 1 in Section 7.

The idea of interpolation goes back to Heath (1977), and is related to embedding, cf. the references in Mykland (1995). In our current case, however, one has to be particularly precise, since the process θ_t already lives on the relevant filtration.

CONTIGUITY. The contiguity approach (Mykland and Zhang (2009, 2011, 2012, 2016b,c)) may, when applicable, reduce high frequency martingales to ones that are locally Gaussian. We refer to the cited papers for further discussion.

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 $\widehat{\mathrm{AVAR}}_n$ is much easier to implement when constructing a feasible statistic of the form (2). 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 bi- and multi-power methods (Barndorff-Nielsen and Shephard (2004b, 2006), Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006a,b)), or by truncation[32] (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)), as appropriate. See also Zhang (2007), Christensen, Oomen, and Podolskij (2011), and Bajgrowicz, Scaillet, and Treccani (2015). We emphasize that θ can be a general semimartingale,[33] 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.

In implementation, we assume that $\hat{\Theta}_{(S,T]}$ is the forward estimator from Section 5.1. For examples with irregular observations, we assume the previous-tick scheme from Section 5.2, and in particular that (45) is satisfied. We shall omit the subscript n on t: t_i means $t_{n,i}$.

REMARK 13. (TWO TYPES OF CONDITIONS.) To see how our examples fit into the theory, we need

[32] For the case of removal by truncation, please consult Section 5.3.

[33] In all our examples, the spot values of θ_t exists. See Section 5.3 for further discussion of this.

to check two classes of conditions. One is on the martingale $M_{n,t}$, and they are all in Condition 1 or in the alternative Condition 3. We recall that they are

$$n^\alpha M_n \xrightarrow{\mathcal{L}} L \text{ stably, } n^\alpha M_n \text{ is P-UT, } [L,L]_{\mathcal{T}} \in \mathcal{G}, \text{ and } L \text{ is a martingale conditionally on } [L,L]_{\mathcal{T}}. \tag{52}$$

The edge effects have various conditions attached to them depending on their order of magnitude. They all need to satisfy that $\tilde{e}_{T_i} = o_p(n^{-\alpha})$. The edge conditions are in Sections 3.2-3.3. The easiest condition to satisfy is (29) in Theorem 4, which makes the two scales AVAR and $\widehat{[\theta,\theta]}_{T}$- consistent. This condition also implies (22) in Theorem 3 for the choice of K_n that satisfies the balance condition (30). □

Example 1. (Realized Volatility, No Microstructure Noise.) 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 (2002a)). There is no edge effect, *i.e.*, $\tilde{e}_{T_{n,i,*}} \equiv 0$. By Remark 13, we need to check (52). The stable convergence has been shown by Jacod and Protter (1998) using discretization. We here use interpolation because it gives the P-UT property directly. The interpolated semimartingale has the form $M_{n,t} = \sum_{t_{n,j+1}\le t}(X_{t_{n,j+1}} - X_{t_{n,j}})^2 + (X_t - X_{t_{n,*}})^2 - \int_0^t \sigma_s^2 ds$, where $t_{n,*} = \max_j\{t_{n,j} \le t\}$. See Zhang (2001); Mykland and Zhang (2006, 2012). The requirements $[L,L] \in \mathcal{G}$, and that L be a martingale conditional on $\mathcal{G}$, also follow from the construction in the cited papers, and all theorems in the current paper can be used. □

Example 2. (Bipower Variation, No Microstructure Noise.). The bipower variation $\hat{\Theta}_{(0,T]} = \frac{\pi}{2}\sum_{0<t_i\le T}|\Delta X_{t_{i-1}}||\Delta X_{t_i}|$ (and more generally, Multipower Variation, Barndorff-Nielsen and Shephard (2004b, 2006)) estimates the integrated volatility in a way that is robust to jumps. Since jumps are of the essence in this model, we specify that $dX_t = \mu_t dt + \sigma_t dW_t + dJ_t$, where J_t is a semimartingale for which $[J,J]_t$ is purely discontinuous.

The parameter is $\theta_t = \sigma_t^2$. The convergence rate is $\alpha = 1/2$. We here study the case of equidistant sampling, $t_{n,i} - t_{n,i-1} = \Delta t_n = \mathcal{T}/n$, and for convenience we take $\Delta T_n = \Delta t_n$. Our semimartingale is

$$M_{n,T_i} = \frac{\pi}{2}\sum_{j=2}^{i}|\Delta X_{t_{j-1}}||\Delta X_{t_j}| - \int_0^{T_{i-1}} \sigma_t^2 dt. \tag{53}$$

The papers by Barndorff-Nielsen and Shephard (2004b, 2006), Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006a,b), and Barndorff-Nielsen, Shephard, and Winkel (2006c) have shown stable convergence and the other conditions of (52), with the exception of P-UT property. For details about P-UT property and its proof, see Appendix D.3.

Edge Effect. There is some variability between proofs of whether the integral in (53) has upper limit T_{i-1} or T_i. In the latter case, there is no edge effect. In the former case, by Remark 14 below, $\text{ave}(\tilde{e}^2_{T_i}) = O_p((\Delta T_n)^3)$, by (64).

In conclusion, all theorems and estimators in the current paper can be used to estimate the AVAR of Bipower Variation. □

Example 3. (Classical Two-Scales Realized Volatility.). The parameter remains $\theta_t = \sigma_t^2$. There is now microstructure noise, and observations are of the form

$$Y_{t_i} = X_{t_i} + \epsilon_{t_i} \tag{54}$$

which we here for simplicity take to be iid, or to be stationary with fast mixing dependence. X_t is assumed to be a continuous Itô-semimartingale.

The classical Two-Scales Realized Volatility (TSRV; Zhang, Mykland, and Aït-Sahalia (2005), Aït-Sahalia, Mykland, and Zhang (2011)) has a convergence rate of $\alpha = 1/6$. It is easy to see that Conditon 1 is satisfied. Edge effects, whether alone or by averages, are of order $O_p(n^{-2\alpha})$, *cf.* Zhang, Mykland, and Aït-Sahalia (2005, eq. (A.21), p. 1409), whence Theorem 4 applies. The two-scales AVAR and $\widehat{[\theta,\theta]}_{T-}$ are thus consistent. In fact, Theorem 3 is valid so long as $n^{2\alpha}K_n\Delta T_n \to \infty$. □

Example 4. (Pre-averaging followed by TSRV). The parameter remains $\theta_t = \sigma_t^2$. The observations are as in (54). The convergence rate is $\alpha = 1/4$. The estimator is constructed as follows (Mykland and Zhang (2016a)). One preaverages observations across blocks of size $O(n^{1/2})$ observations, and then calculates a (j,k) TSRV on the basis of the preaveraged observations, where $1 \leq J < K$ are finite. It is shown in (Mykland and Zhang (2016a)) that this estimator of integrated volatility converges stably at rate $\alpha = 1/4$, the semimartingale is P-UT, and the edge effects are benign, of exact order $O_p(n^{-1/2})$. The edge effects are thus small enough to satisfy the edge conditions (22) and (29) in Theorems 3-4 (Section 3.2). We have used this method in Figure 1. Note that in the terms of Section 5.4, and Remark 14 below, $\mathcal{M} = k$.

It is conjectured that the same type of situation pertains to classical pre-averaging (Jacod, Li, Mykland, Podolskij, and Vetter (2009a); Podolskij and Vetter (2009b)), but we have not investigated this. □

Example 5. (Multi-Scale and Kernel Realized Volatility.) The parameter remains $\theta_t = \sigma_t^2$. The observations are as in (54). The convergence rate is $\alpha = 1/4$. We here show that the Multi-Scale Realized Volatility (MSRV, Zhang (2006)) is covered by our current development. Following Bibinger and Mykland (2016), the result also covers Realized Kernel estimators (RK, Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008)).

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),

$$M_{n,t} = M_{n,t}^{(1)} + M_{n,t}^{(2)} + M_{n,t}^{(3)}, \tag{55}$$

where[34]

$$
\begin{aligned}
M_{n,t}^{(1)} &= -2\sum_{i=1}^{\mathcal{M}_n} a_{n,i}\frac{1}{i}\sum_{t_{i+1}\le t}\epsilon_{t_{n,j}}\epsilon_{t_{n,j-i}},\\
M_{n,t}^{(2)} &= \sum_{i=1}^{\mathcal{M}_n} a_{n,i}[X,X]_t^{(n,i)} - \int_0^t \sigma_s^2 ds, \text{ and}\\
M_{n,t}^{(3)} &= 2\sum_{i=1}^{\mathcal{M}_n} a_{n,i}[X,\epsilon]_t^{(i)}.
\end{aligned} \tag{56}
$$

The edge effects, e and $\tilde{e}$, are given by (*Ibid.*, eq. (51), p. 1039, and rewritten form (53), p. 1040)

$$
e_{n,0} = \sum_{j=0}^{\mathcal{M}_n-1}\varpi_{n,j}\epsilon_{t_j}^2 - E\epsilon^2 \text{ and } \tilde{e}_{n,t_k} = \sum_{j=0}^{\mathcal{M}_n-1}\varpi_{n,j}\epsilon_{t_{k-j}}^2 - E\epsilon^2, \text{ where } \varpi_{n,j} = \sum_{i=j+1}^{\mathcal{M}_n}\frac{a_{n,i}}{i}. \tag{57}
$$

With these definitions, and with $\mathcal{M}_n = O(n^{1/2})$, eq. (13) in the current paper is satisfied up to $O_p(n^{-1/2})$ (*ibid.,* Proposition 1, p. 1023).

The terms in (57) are of order $O_p(n^{-1/4})$, and so Condition 1 is violated. Since this magnitude of edge effects is in any case undesirable, we propose to amend the MSRV by estimating the edge effects:

$$
\begin{aligned}
\text{adjusted MSRV}_{n,t_k} &= \text{original MSRV}_{n,t_k} - \hat{e}_{n,t_k} + \hat{e}_{n,0}, \text{ where}\\
\hat{e}_{n,t_k} &= \sum_{j=0}^{\mathcal{M}_n-1}\varpi_{n,j}\left(Y_{t_{k-j}} - \bar{Y}_{t_k}\right)^2 - \frac{1}{2}[X,X]_{\mathcal{T}}^{(n,1)},
\end{aligned} \tag{58}
$$

and similarly for $\hat{e}_{n,0}$, where $\bar{Y}_{t_k}$ is the mean of $Y_{t_{k-\mathcal{M}_n+1}}, \cdots, Y_{t_k}$. From Zhang (2006, Condition 1 (p. 1023) and eq. (54) (p. 1040)),

$$
\sum_{j=0}^{\mathcal{M}_n-1}\varpi_{n,j} = 1 \text{ and } \sum_{j=0}^{\mathcal{M}_n-1}\varpi_{n,j} = O_p(\mathcal{M}_n^{-1}), \tag{59}
$$

we obtain $\hat{e}_{n,t_k} = \tilde{e}_{n,t_k} + O_p(n^{-1/2})$. Hence,

$$
\text{adjusted MSRV}_{n,t_k} = M_{n,t_k} + O_p(n^{-1/2}), \tag{60}
$$

and so the new edge effect is of size $O_p(n^{-1/2})$. Under the conditions of *Ibid.*, Theorem 4 (p. 1031), including $\mathcal{M}_n/n^{1/2} \to c$, it is easy to see that Condition 1 is satisfied.[35]The edge effects are thus

[34]Except that we use $\mathcal{M}_n$ to denote the number of scales (called M_n in Zhang (2006). The square brackets in (56) are discrete sums. The $a_{n,i}$ are given by *Ibid.*, eq. (21)-(22) p. 1026.

[35]The second term in (58) is only available at time $\mathcal{T}$. This means that it can be used to estimate the MSRV at time $\mathcal{T}$. For the intermediate calculations at times $T_{n,i}$ or $T_{n,i,*}$, this is not a concern, however, since the term is

small enough to satisfy the edge conditions (22) and (29) in Theorems 3-4 (Section 3.2).

Similar arguments would extend to the dependent but mixing noise in Aït-Sahalia, Mykland, and Zhang (2011). □

REMARK 14. (EDGE EFFECTS IN BLOCK BASED ESTIMATION.) Estimators are often based on rolling blocks of $\mathcal{M}_n$ observations.[36] This is the case in the following Examples 6, 7, 9, and 10.[37] See also Section 5.4 to the effect that our K_n is unrelated to $\mathcal{M}_n$.

Rolling block estimators frequently have the common feature that the edge effect is (exactly or approximately) on the form $\tilde{e}_{T_i} = -\Theta_{(T_{i-\mathcal{M}_n+1},T_i]}$. We here present a general strategy for dealing with edge effects on this form, and we shall comment on specifics in connection with individual examples. For simplicity, we assume that observations are an equidistant sample every $\Delta t_n = \mathcal{T}/n$ units of time, and we also set $\Delta T_n = \mathcal{T}/n$. (This is the case for all the papers we cite on block estimation.) Assume that the conditions (52) on the martingale $M_{n,t}$ are satisfied.

First of all, use (B.32) in Appendix B to write $\tilde{e}_{T_i} = \Theta''_{(T_{i-\mathcal{M}_n+1},T_i]} - \theta_{T_i}(\mathcal{M}_n - 1)\Delta T_n$, where $\Theta''_{(T_{i-\mathcal{M}_n+1},T_i]}$ is as defined in (11) in Section 2.3. Then absorb $-\theta_{T_i}(\mathcal{M}_n - 1)\Delta T_n$ in the semi-martingale M_n, so that

$$M^{\text{adjusted}}_{n,T_i} = M^{\text{original}}_{n,T_i} - \theta_{T_i}(\mathcal{M}_n - 1)\Delta T_n, \tag{61}$$

and redefine the edge effect as

$$\tilde{e}_{T_i} = \Theta''_{(T_{i-\mathcal{M}_n+1},T_i]}. \tag{62}$$

So long as[38]

$$\mathcal{M}_n \Delta T_n = o(n^{-\alpha}), \tag{63}$$

the limiting martingale and the mode of convergence is unchanged (Jacod and Shiryaev (2003, Lemma VI.3.31, p. 532)). P-UT property is also not affected (*ibid.*, Remark VI.6.4, p. 377). Also, by the same methods as in the Proof of Theorem 1 (see Appendix B), $\tilde{e}_{T_i} = O_p(\mathcal{M}_n \Delta T_n) = o_p(n^{-\alpha})$. Hence, Condition 1, or alternative Condition 3, is satisfied.

As an application of Theorem 7 in Appendix A (the proof is similar to that of Theorem 2 (Appendix B), we obtain that

$$\text{ave}(\tilde{e}^2_{T_i}) = \begin{cases} \frac{1}{3\mathcal{T}}\left((\mathcal{M}_n - 1)\Delta T_n\right)^3 [\theta,\theta]_{\mathcal{T}-}(1 + o_p(1)) \text{ when } \mathcal{M}_n \to \infty \text{ as } n \to \infty, \text{ and} \\ O_p((\Delta T_n)^3) \text{ when } \mathcal{M}_n \text{ remains finite as } n \to \infty, \end{cases} \tag{64}$$

constant in i and thus will cancel when computing $\hat{\Theta}_{(T_i,T_{i+K}]} - \hat{\Theta}_{(T_{i-K},T_i]}$. For purposes of verifying the conditions of our results, we therefore proceed as if $\frac{1}{2}[X,X]^{(n,1)}_{\mathcal{T}}$ is replaced by $E\epsilon^2$.

[36] Many papers use k_n or M_n to denote what we here call $\mathcal{M}_n$. We use the latter symbol to avoid overlap with our own notation.

[37] The block structure is also present in most of our other examples, even if we have not used the structure explicitly. To some extent this is a question of technique of proof.

[38] See Jacod and Rosenbaum (2015) and Theorem 3.1 in Jacod and Rosenbaum (2013) for an important contribution on what can happen otherwise.

whence assumption (29) in Theorem 4 is satisfied. The two scales AVAR and $\widehat{[\theta,\theta]}_{T-}$ are thus consistent. Depending on the size of $\mathcal{M}_n$, further small edge conditions are satisfied. □

Example 6. (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.[39] The estimation of integrals of σ_t^p goes back to Barndorff-Nielsen and Shephard (2002a), who 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 (2006a), 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 $\mathcal{M}_n$ finite. This avoids bias. When using overlapping (rolling) blocks (or moving windows), 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. Conditions (52) are clearly satisfied, by the derivation in the cited papers. Also, by Remark 14, we can use all of the results: Theorem 3-5, and Remark 9.

Another path is to let the block size increase with n, cf. Mykland and Zhang (2011, Section 5, pp. 224-229), and Jacod and Rosenbaum (2013, 2015). As seen in the cited papers, for increasing block size, there is a bias that can be corrected for. In Jacod and Rosenbaum (2013), the corrected estimator is (in their notation) $V'(g)^n$, eq. (3.7), p. 1469, which satisfies assumptions (52). We now discuss how to verify these assumptions. The stable convergence is stated in *ibid.,* Theorem 3.2 (pp. 1469-1470). The P-UT condition is satisfied by noting that in the proof of their Lemma 4.4 (p. 1478-1480), each of the four components obviously also satisfies our equation (51), by being bounded term-wise. In their Lemma 4.5 (pp. 1478, 1480-1481), they proceed by verifying the conditions of Jacod and Shiryaev (2003, Theorem IX.7.28, p. 591), and it is easy to see that the second part of (our) eq. (50) is satisfied, guaranteeing P-UT also for this term in view of Section 6.

The edge effect is part of $V_t^{n,2}$ in Jacod and Rosenbaum (2013, p. 1478). *Ibid.,* assumption (3.6) (p. 1469) yields that condition (63) in Remark 14 is satisfied, whence at least the two scales AVAR and $\widehat{[\theta,\theta]}_{T-}$ are consistent.

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 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

[39] Inference in the presence of noise in considered in Jacod and Protter (2012, Section 16.4-16.5, pp. 512-554).

or without microstructure noise. 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. When there is no microstructure noise, Condition 1 is satisfied by a slight extension of the derivations in the cited papers. Both regression and ANOVA have edge effects due to blocking, as in Example 6. Since $\mathcal{M}_n$ is finite, and according to Remark 14, we can use all results: Theorem 3-5, and Remark 9. □

Example 8. (Estimation of Co-Volatility (*Ex-Post* 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 (2015) for micro-structure, jumps, and asymptotic distributions. Alternatives include the Previous-Tick estimator (Zhang (2011), Bibinger and Mykland (2016)), and Quasi-Likelihood (Shephard and Xiu (2014)). The estimator in Mykland and Zhang (2012, Chapter 2.6.3, p. 172-175) is a hybrid of Hayashi-Yoshida and Quasi-Likelihood. The asymptotic distributions, however, are often quite complex, and the estimation of AVAR is daunting. In comparison, the approach of observed AVAR offers a pleasing alternative to assessing the asymptotic variance of co-volatility. In all these cases, it is quite clear that the stable convergence holds, and that the current paper's Condition 1 is satisfied, including the P-UT property. In terms of edge effects, the Previous-Tick Two-Scales Covariance (TSCV, Zhang (2011)) has exactly the same properties as the classical TSRV (Example 3). This is because of the strong representation property of one in terms of the other (Zhang (2011, eq. (39), p. 41, see also eq. (8), p. 35). The two-scales AVAR and $\widehat{[\theta, \theta]}_{T-}$ based on the Previous-Tick TSCV are thus consistent. Due to the large number of covariance estimators, however, we have not investigated edge effects for the full spectrum of these. □

Example 9. (Continuous Leverage Effect, with or without Microstructure Noise.) 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 Mykland and Zhang (2009, Section 4.3, pp. 1426-1428) and Wang and Mykland (2014) for the case where X_t is continuous, and in Aït-Sahalia, Fan, Laeven, Wang, and Yang (2016) and Kalnina and Xiu (2015) for the case where the process X_t can also have jumps.[40] Wang and Mykland (2014) and Aït-Sahalia, Fan, Laeven, Wang, and Yang (2016) study both the case where there is microstructure noise, and where there is none. All estimators are based on blocks.

We here study the procedure of Aït-Sahalia, Fan, Laeven, Wang, and Yang (2016). Jumps

[40] Aït-Sahalia, Fan, and Li (2013) discusses leverage effect in the parametric framework.

are removed as in Jacod and Todorov (2010). The relevant central limit theorems are Theorem 5.1 (no microstructure noise) and Theorem 7.2 (with microstructure noise). The conditions (52) are satisfied by a slight extension of the proofs of these results. The optimal rates ($\alpha = 1/4$ and $\alpha = 1/8$) are attained in both cases (with choice of parameter $b = 1/2$). The edge effects are essentially on the form described in Remark 14, *cf.* $D(2)_t^n$ (p. 42, for the no-microstructure case, and p. 50 for the case with microstructure noise). In both cases $\mathcal{M}_n$ (called k_n in this paper) is of order $O(n^{2\alpha})$. Thus condition (63) in Remark 14 is satisfied. The two scales AVAR and $\widehat{[\theta,\theta]}_{T-}$ are thus consistent. □

Example 10. (Volatility of Volatility, no Microstructure Noise.) The process X is assumed to be a continuous Itô-semimartingale, with volatility $\sigma_t^2 = d[X,X]_t/dt$ which is itself assumed to be a continuous Itô-semimartingale. The parameter is $\theta_t = d[\sigma^2,\sigma^2]_t/dt$. The convergence rate is $\alpha = 1/4$. The results in the literature on this inference problem are Vetter (2015, Theorems 2.1 and 2.5) and Mykland, Shephard, and Sheppard (2012, Theorem 7 and Corollary 2).

We here focus on the estimator of Vetter (2015). It is on the form (25) in Section 3.2 above, with AVAR_n replaced by the quarticity estimator of Barndorff-Nielsen and Shephard (2002a, 2004a). The estimator is thus a special case of Theorem 3.

Turning to the question of whether the estimator satisfies the conditions of this paper, observe that this is also a rolling block estimator. The conditions (52) are satisfied by a slight extension of the proof of Vetter (2015, Theorems 2.1). $\mathcal{M}_n$ is of order $O(n^{1/2})$, and hence condition (63) in Remark 14 is satisfied. The two scales AVAR and $\widehat{[\theta,\theta]}_{T-}$ (the estimator of the volatility of the volatility of the volatility) are thus consistent, as are the multi-scale estimators.

It should be noted that by computing the (two-scales or multi-scale) estimate $\widehat{[\theta,\theta]}_{\mathcal{T}-}$ for any of the estimators in Examples 3-5, one obtains an estimator of $[\sigma^2,\sigma^2]_{\mathcal{T}-}$ that is consistent in the presence of microstructure. □

8 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 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 (5) in Section 2.2. To operationalize this insight, the two main tools are the Integral-to-Spot Device (Section 2.3), and the mathematical similarity between edge effects and microstructure noise (Section 3.1). The estimation of asymptotic variance (AVAR) is implemented with the help of a two-scales method in Section 3.2, and examples are given in Section 7. Practical and theoretical guidance to how to use the procedure is given in Sections 5-6.

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 (Section 4). 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 only the first order requirement on estimators of AVAR. Further optimization may involve the convergence rate, and the AVAR of $\widehat{AVAR}$. A main question remains of whether there is added benefit in going to a multi-scale procedure. There is also room for a more complete theory of tuning parameter selection, and of multivariate inference. Additional insight may be gained by letting $\Delta T \to 0$ for fixed $\delta = K\Delta T$. 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 5.3).

References

Aït-Sahalia, Y., J. Fan, R. Laeven, C. D. Wang, and X. Yang (2016): "Estimation of continuous and discontinuous leverage effect," *Journal of the American Statistical Association* (in press).

Aït-Sahalia, Y., J. Fan, and Y. Li (2013): "The Leverage Effect Puzzle: Disentangling Sources of Bias at High Frequency," *Journal of Financial Economics*, 109, 224–249.

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

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

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

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

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

——— (2014): *High-Frequency Financial Econometrics*, Princeton, N.J.: Princetion University Press, first ed.

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.

Aït-Sahalia, Y. and D. Xiu (2015): "Principal Component Analysis of High-Frequency Data," Tech. rep., National Bureau of Economic Research.

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

Andersen, T., T. Bollerslev, and N. Meddahi (2005): "Correcting the Errors: Volatility Forecast Evaluation Using High Frequency Data and Realized Volatilities," *Econometrica*, 73, 279–296.

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.

Andreou, E. and E. Ghysels (2002): "Rolling-Sample Volatility Estimators: Some New Theoretical, Simulation and Empirical Results," *Journal of Business and Economic Statistics*, 20, 363–376.

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

Bajgrowicz, P., O. Scaillet, and A. Treccani (2015): "Jumps in high-frequency data: Spurious detections, dynamics, and news," *Management Science*.

Barndorff-Nielsen, O., S. Graversen, J. Jacod, M. Podolskij, and N. Shephard (2006a): "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.

——— (2006b): "Limit theorem for bipower variations in financial econometrics," *Econometric Theory*, 22, 677–719.

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 realised kernels to measure ex-post variation of equity prices in the presence of noise," *Econometrica*, 76(6), 1481–1536.

——— (2011): "Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading," *Journal of Econometrics*, 162, 149–169.

Barndorff-Nielsen, O. E., S. Kinnebrock, 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.

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

——— (2002b): "Estimating quadratic variation using realised variance," *Journal of Applied Econometrics*, 17, 457–477.

——— (2004a): "Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics," *Econometrica*, 72, 885–925.

——— (2004b): "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.

——— (2015): Discussion at Conference in Honor of Ole-Barndorff-Nielsen, Aarhus, 2015.

Barndorff-Nielsen, O. E., N. Shephard, and M. Winkel (2006c): "Limit theorems for multipower variation in the presence of jumps," *Stochastic Processes and Their Applications*, 116, 796–806.

Bibinger, M. and P. A. Mykland (2016): "Inference for Multi-Dimensional High-Frequency Data: Equivalence of Methods, Central Limit Theorems, and an Application to Conditional Independence Testing," *Scandinavian Journal of Statistics* (in press).

Bibinger, M. and M. Vetter (2015): "Estimating the quadratic covariation of an asynchronously observed semimartingale with jumps," *Annals of the Institute of Statistical Mathematics*, 67, 707–743.

Bollerslev, T., A. Patton, and R. Quaedvlieg (2016): “Exploiting the Errors: A Simple Approach for Improved Volatility Forecasting,” *Journal of Econometrics*, 192, 1–18.

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, N. Thamrongrat, and B. Veliyev (2015): “Inference from high-frequency data: A subsampling approach,” *Journal of Econometrics* (in press).

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.

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

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

Ghysels, E., P. A. Mykland, and E. Renault (2012): “In-sample Asymptotics and Across-sample Efficiency Gains for High Frequency Data Statistics,” Tech. rep., University of Chicago.

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., 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 (2015): "Microstructure Noise in the Continuous Case: Approximate Efficiency of the Adaptive Pre-Averaging Method," *Stochastic Processes and Their Applications*, 125, 2910–2936.

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

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

——— (2015): "Estimation of volatility functionals: The case of a $\sqrt{n}$ window," in *Large Deviations and Asymptotic Methods in Finance, Volume 110 of the Springer Proceedings in Mathematics and Statistics*, ed. by P. Friz, J. Gatheral, A. Gulisashvili, A. Jacquier, and J. Teichmann, Berlin: Springer-Verlag, 559–590.

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," *LSE STICERD Research Paper No. EM523*.

Kalnina, I. and D. Xiu (2015): "Nonparametric Estimation of the Leverage Effect using Information from Derivatives Markets," In revision for *Journal of the American Statistical Association*.

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. A. Mykland, E. Renault, L. Zhang, and X. Zheng (2014): "Realized Volatility when Sampling Times are possibly Endogenous," *Econometric Theory*, 30, 580–605.

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.

Meddahi, N. (2002): "A theoretical comparison between integrated and realized volatility," *Journal of Applied Econometrics*, 17, 479–508.

Mykland, P. A. (1995): "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 (2012): "Efficient and feasible inference for the components of financial variation using blocked multipower variation," Technical Report, University of Oxford.

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.

——— (2016a): "The Algebra of Two Scales Estimation," Working Paper, University of Illinois at Chicago and University of Chicago.

——— (2016b): "Between data cleaning and inference: Pre-averaging and other robust estimators of the efficient price," *Journal of Econometrics*, 194, 242–262.

——— (2016c): "Superblocks of Pre-Averaged Data: First Aid for Microsctructure, and a Path to Efficiency," Discussion paper.

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.

——— (2010): "Understanding limit theorems for semimartingales: a short survey," *Statistica Neerlandica*, 64, 329–351.

Pollard, D. (1984): *Convergence of Stochastic Processes*, New York: Springer-Verlag.

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

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.

Robert, C. Y. and M. Rosenbaum (2011): "A New Approach for the Dynamics of Ultra-High-Frequency Data: The Model with Uncertainty Zones," *Journal of Financial Econometrics*, 9, 344–366.

——— (2012): "Volatility and Covariation Estimation when Microstructure Noise and Trading Times are Endogenous," *Mathematical Finance*, 22 (1), 133–164.

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 (2014): "Econometric analysis of multivariate realised QML: estimation of the covariation of equity prices under asynchronous trading," *Chicago Booth Research Paper.*

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.

Stoker, T. M. (1993): "Smoothing Bias in Density Derivative Estimation," *Journal of the American Statistical Association*, 88, 855–863.

Vetter, M. (2015): "Estimation of Integrated Volatility of Volatility with Applications to Goodness-of-fit Testing," *Bernoulli*, 21, 2393–2418.

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.

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 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 General Results on the Triangular Array Convergence of the Quadratic Variation of Semimartingales.

Definition 5. (Orders in Probability.) *For a sequence $\alpha_t^{(n)}$ of semimartingales, we say that $(\alpha_t^{(n)}) = O_p(1)$ if the sequence is tight, with respect to convergence in law relative to the Skorokhod topology on $\mathbb{D}$ (Jacod and Shiryaev (2003, Theorem VI.3.21, p. 350)), and also P-UT (ibid., Chapter VI.3.b, and Definition VI.6.1, p. 377). For scalar random quantities, $O_p(\cdot)$ and $o_p(\cdot)$ are defined as usual, see, e.g., Pollard (1984, Appendix A).*

Condition 4. *Let $\alpha_t^{(n)}$ and $\beta_t^{(n)}$ be sequences (in n) of semimartingales. Each of these sequences are (separately) assumed to be $O_p(1)$.*

Definition 6. (Notation). *The symbol $\mathbb{F}$ will refer to a collection of nonrandom functions $f_\cdot^{(l,n)} \in \mathbb{D}[0,\mathcal{T}]$, $n \in \mathbb{N}$, and $l = 1, ..., 2K_n$ satisfying*

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

Similarly, $\mathbb{G}$ will refer to a collection $g_t^{(l,n)}$ with the same size and properties.

Given $\mathbb{F}$ and $\mathbb{G}$, set

$$\alpha_t^{(l,n)} = \int_0^t f_{s-}^{(l,n)} d\alpha_s^{(n)} \text{ and } \beta_t^{(l,n)} = \int_0^t g_{s-}^{(l,n)} d\beta_s^{(n)} \text{ for } l = 1, ..., 2K_n. \tag{A.2}$$

Also,

$$i \equiv L[2K] \text{ means that } i = 2Kj + L, \text{ where } j \text{ is an integer.} \tag{A.3}$$

Definition 7. (Decomposition of $\mathbb{F}$ and $\mathbb{G}$ by Block.) *Recall that B_n is the set of basic blocks, and that $\Delta T_n = \mathcal{T}/B_n$. With reference to the collection $\mathbb{F}$: For given (l,n), the function $f_t^{(l,n)}$ is allowed to jump at times $T_{K_n j+l}$ but must otherwise satisfy certain compactness properties.*

Specifically, for each $n \in \mathbb{N}$, and $l = 1, ..., 2K_n$, define, for $j \in \mathbb{N} \cap [1, (B_n - l)/(K_n + 1)]$,

$$f_t^{(l,j,n)} = \begin{cases} f_{T_{K_n j+l}}^{(l,n)} \text{ for } t \in [0, T_{K_n j+l}) \\ f_t^{(l,n)} \text{ for } t \in [T_{K_n j+l}, T_{(K_n+1)j+l}) \\ \lim_{t \uparrow T_{(K_n+1)j+l}} f_t^{(l,n)} \text{ for } t \in [T_{(K_n+1)j+l}, \mathcal{T}] \end{cases} \tag{A.4}$$

The set of such $f_t^{(l,j,n)}$ will be denoted $\mathbb{F}'$. $\mathbb{G}'$ is defined similarly.

Theorem 7. Consistency of triangular array rolling quadratic variation.) *Under Condition 4, assume (A.1), and that the sets $\mathbb{F}'$ and $\mathbb{G}'$ (from Definition 7) are relatively compact* *for the Skorokhod topology.*[41] *Also suppose that $K_n \Delta T_n \to 0$ as $n \to \infty$. Then,*

$$\frac{1}{2K_n} \sum_{l=1}^{2K_n} \sum_{K_n \le i \le B_n - K_n, i \equiv l[2K_n]} (\alpha^{(l,n)}_{T_{i+K_n}} - \alpha^{(l,n)}_{T_{i-K_n}})^2 = \frac{1}{2K_n} \sum_{l=1}^{2K_n} [\alpha^{(l,n)}, \alpha^{(l,n)}]_{\mathcal{T}} + o_p(1). \tag{A.5}$$

and similarly for β. Also,

$$\frac{1}{2K_n} \sum_{l=1}^{2K_n} \sum_{K_n \le i \le B_n - K_n, i \equiv l[2K_n]} (\alpha^{(l,n)}_{T_{i+K_n}} - \alpha^{(l,n)}_{T_{i-K_n}})(\beta^{(l,n)}_{T_{i+K_n}} - \beta^{(l,n)}_{T_{i-K_n}}) = \frac{1}{2K_n} \sum_{l=1}^{2K} [\alpha^{(l,n)}, \beta^{(l,n)}]_{\mathcal{T}} + o_p(1). \tag{A.6}$$

Remark 15. (Uniformity in ΔT.) Theorem 7 does not impose any requirement on ΔT_n, except that $\Delta T_n > 0$ and $K_n \Delta T_n \to 0$. See the final comment in the proof of the theorem. □

Before proving our results, we recall the following useful concept.

Definition 8. (The Canonical Decomposition of α.) *We shall be using the canonical decomposition of α_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*

$$\alpha_t = \alpha_0 + \alpha(h)_t + B(h)_t + \breve{\alpha}(h)_t. \tag{A.7}$$

h is called the truncation function. Compared to the notation in our reference work, their X is our α, their $M(h)$ is our $\alpha(h)$, while their $B(h)$ is the same as ours. Also, let $\tilde{C}_t = \langle \alpha(h), \alpha(h) \rangle$. This is the "second modified characteristic" (Ibid., Definition II.2.16, p. 79). For the case of no truncation function, α 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,[42] *and set*

$$D(\alpha)(h)_t = TV(\breve{\alpha})_t + TV(B(h))_t. \tag{A.8}$$

Similar notation applies to $\alpha^{(n)}$, $\beta^{(n)}$, etc. □

Proof of Theorem 7. We prove (A.5). The result (A.6) is obtained similarly but with longer notation. For (A.6), we specifically need that $\alpha^{(n)}$ and $\beta^{(n)}$ be tight, which is assumed, and that $D(\alpha^{(n)})(h)_{\mathcal{T}}$, $D(\beta^{(n)})(h)_{\mathcal{T}}$, $\langle \alpha^{(n)}(h), \alpha^{(n)}(h) \rangle_{\mathcal{T}}$, $\langle \beta^{(n)}(h), \beta^{(n)}(h) \rangle_{\mathcal{T}}$, and $\langle \beta^{(n)}(h), \alpha^{(n)}(h) \rangle_{\mathcal{T}}$ be tight. The first four of these follow from the P-UT property of $\alpha^{(n)}$ and $\beta^{(n)}$ (Jacod and Shiryaev (2003, Theorem VI.6.15)), the final one since $|\langle \beta^{(n)}(h), \alpha^{(n)}(h) \rangle_{\mathcal{T}}| \le (\langle \alpha^{(n)}(h), \alpha^{(n)}(h) \rangle_{\mathcal{T}} + \langle \beta^{(n)}(h), \beta^{(n)}(h) \rangle_{\mathcal{T}})/2$.

[41] A criterion can be found in Jacod and Shiryaev (2003, Theorem VI.1.14(b), p. 328). The condition is satisfied in all our applications (B.24), (C.38), and (C.43).

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

In analogy with (A.2), define $\alpha_t^{(l,,j,n)} = \int_0^t f_{s-}^{(l,j,n)} d\alpha_s^{(n)}$. Also, define

$$Z_{n,l}(t) = \sum_{T_{i+K_n} \leq t, i \equiv l[2K]} (\alpha_{T_{i+K_n}}^{(l,n)} - \alpha_{T_{i-K_n}}^{(l,n)})^2 + (\alpha_t^{(l,n)} - \alpha_{T_{*,L}}^{(l,n)})^2 - [\alpha^{(l,n)}, \alpha^{(l,n)}]_t \tag{A.9}$$

where $T_{*,L} = \max\{T_i \ : \ T_{i+K_n} \leq t, i \equiv L[2K_n]\}$, so that

$$dZ_{n,l}(t) = 2(\alpha_{t-}^{(l,n)} - \alpha_{T_{*,l}}^{(l,n)}) d\alpha_t^{(l,n)}. \tag{A.10}$$

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

$$Z_n(t) = \frac{1}{2K_n} \sum_{l=1}^{2K_n} Z_{l,n}(t) \text{ and } Z_n(h)(t) = \frac{1}{2K_n} \sum_{L=1}^{2K_n} Z_{l,n}(h)(t) \tag{A.11}$$

Observe that $Z_n(\mathcal{T})$ = the difference between the explicit terms on left and right hand sides of (A.5).

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

$$|Z_{l,n}(h)(t) - Z_{l,n}(t)| \ \leq \ 2 \int_0^t |\alpha_{s-}^{(l,n)} - \alpha_{T_{*,l}}^{(l,n)}| dD^{(n)}(h)_t \tag{A.12}$$

where $D^{(n)}(h)$ is defined as in (A.8), and with the original $\alpha^{(n)}$. Also, in the notation of Jacod and Shiryaev (2003, VI.1.8, p. 326), it follows from (A.1) that for all $t \in [0, \mathcal{T}]$ and all $s \in [T_{*,L}, t]$

$$\begin{aligned} |\alpha_{s-}^{(l,n)} - \alpha_{T_{*,l}}^{(l,n)}| \ &\leq \ 2 \max_j w'_{\mathcal{T}}(\alpha^{(l,j,n)}, K_n \Delta T_n) \ + \sup_{T_{*,L} < s < t} |\Delta \alpha_s^{(n)}| \\ &\leq \ 2 \max_j w'_{\mathcal{T}}(\alpha^{(l,j,n)}, K_n \Delta T_n) \ + \ v_n(t-) \end{aligned} \tag{A.13}$$

where $v_n(t-) = \sup_{T_{**} < s < t} |\Delta \alpha_s^{(n)}|$, with $T_{**} = \max\{T_i \ : \ T_{i+2K_n} \leq t, \}$, so that

$$\sup_{0 \leq t \leq \mathcal{T}} |Z_n(h)(t) - Z_n(t)| \ \leq \ 4 \max_{l,j} w'_{\mathcal{T}}(\alpha^{(l,j,n)}(h), K \Delta T) D^{(n)}(h)(\mathcal{T}) \ + \ 2 \int_0^{\mathcal{T}} v_n(t-) dD^{(n)}(h)_t. \tag{A.14}$$

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

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

$$\langle Z_n(h), Z_n(h)\rangle_{\mathcal{T}} \leq 8\left(4\max_{l,j} w'_{\mathcal{T}}(\alpha^{(l,j,n)}, K\Delta T)^2 \tilde{C}_{\mathcal{T}}^{(n)} \; + \; \int_0^{\mathcal{T}} v_n^2(t-) d\tilde{C}_t^{(n)}\right). \tag{A.15}$$

This is because the same bound applies to each $\langle Z_{n,l_1}(h), Z_{n,l_2}(h)\rangle_{\mathcal{T}}$.

We now seek to describe the asymptotic behavior of $\max_{l,j} w'_{\mathcal{T}}(\alpha^{(l,j,n)}(h), K_n\Delta T_n)$ and $v_n(t-)$ so as to control the asymptotic behavior of (A.14)-(A.15).

On the one hand, since $\mathbb{F}'$ from Definition 7 is relatively compact for the Skorokhod topology *(ex. hyp.)*, we obtain from Jacod and Shiryaev (2003, Theorem VI.3.21, p. 350, and Theorem VI.6.22, p. 383) that

$$\max_{l,j} w'_{\mathcal{T}}(\alpha^{(l,j,n)}(h), K_n\Delta T_n) \xrightarrow{p} 0 \quad \text{as } n \to \infty. \tag{A.16}$$

On the other hand, we bound $v_n(t-)$ as follows. Let $\epsilon > 0$ be arbitrary. Since $\alpha^{(n)}$ is tight, we shall without loss of generality be working with a convergent subsequence so that $\alpha^{(n)} \xrightarrow{\mathcal{L}} \alpha$. Redo the canonical decomposition (Definition 8) with a specific truncation function given by $h_\epsilon(x) = x$ if $|x| \leq \epsilon$, and $= \epsilon\ \mathrm{sgn}(x)$ otherwise:

$$\begin{aligned} \alpha_t^{(n)} &= \alpha_0^{(n)} + \alpha^{(n)}(h_\epsilon)_t + B^{(n)}(h_\epsilon)_t + \breve{\alpha}^{(n)}(h_\epsilon)_t \text{ and} \\ \alpha_t &= \alpha_0 + \alpha(h_\epsilon)_t + B(h_\epsilon)_t + \breve{\alpha}(h_\epsilon)_t. \end{aligned} \tag{A.17}$$

Set $v_{n,\epsilon}(t-) = \sup_{T_{**}<s<t} |\Delta\breve{\alpha}^{(n)}(h_\epsilon)_s|$ and observe that

$$v_n(t-) \leq v_{n,\epsilon}(t-) + \epsilon. \tag{A.18}$$

Let $\tau_{n,i}$ be the i^{th} jump time of $\breve{\alpha}^{(n)}(h_\epsilon)_t$, with $\tau_{n,0} = 0$. Similarly, τ_i is the i^{th} jump time of $\breve{\alpha}(h_\epsilon)_t$. We note that, for given $t \in [0, \mathcal{T}]$, and for any $\delta > 0$

$$\begin{aligned} \{v_{n,\epsilon}(t-) = 0\} &\supseteq \cup_i\{\tau_{n,i} \geq t \geq \tau_{n,i-1} + 2K_n\Delta T_n\} \\ &\supseteq \cup_i\{\tau_{n,i} \geq t \geq \tau_{n,i-1} + \delta\} \end{aligned} \tag{A.19}$$

as soon as $\delta \geq 2K_n\Delta T_n$ (and this does happen eventually, by assumption). By invoking Jacod and Shiryaev (2003, Proposition VI.3.15, p. 349) with $\tau_{n,i}$ as $T_i(\breve{\alpha}^{(n)}(h_\epsilon), \frac{\epsilon}{2})$ and τ_i as $T_i(\breve{\alpha}(h_\epsilon), \frac{\epsilon}{2})$, the proposition yields that $(\tau_{n,1}, \cdots, \tau_{n,k}) \xrightarrow{\mathcal{L}} (\tau_1, \cdots, \tau_k)$ as $n \to \infty$ for any k. This is because, the process $\breve{\alpha}^{(n)}(h_\epsilon)$ converges in law to $\breve{\alpha}(h_\epsilon)$ in view of *ibid.*, Proposition VI.3.16, p. 349.

By approximating the indicator of the set $\{\tau_{n,i} \geq t \geq \tau_{n,i-1}\}$ by a continuous function, and then undoing the approximation, we obtain $P\{\tau_{n,i} \geq t \geq \tau_{n,i-1} + \delta\} \to P\{\tau_i \geq t \geq \tau_{i-1} + \delta\}$ as

$n \to \infty$. Since the union (A.19) is disjoint, it follows that

$$\begin{aligned}\liminf_n P\{v_{n,\epsilon}(t-) = 0\} &\geq \sum_{i=1}^{k} P\{\tau_i \geq t \geq \tau_{i-1} + \delta\} \\ &\to P\{\tau_k \geq t\} \text{ as } \delta \downarrow 0 \\ &\to 1 \text{ as } k \to \infty. \end{aligned} \tag{A.20}$$

Hence from $(A.18)$, $P\{v_n(t-) \geq \epsilon\} \to 0$. Since ϵ was arbitrary, we obtain

$$\forall\, t \in [0, \mathcal{T}]:\ v_n(t-) \xrightarrow{p} 0 \text{ and } |v_n(t-)| \leq \sup_{0 \leq s \leq \mathcal{T}} |\Delta \alpha_s^{(n)}|, \tag{A.21}$$

the latter statement assuring dominated convergence.

We can now combine (A.14)-(A.15) with (A.16) and (A.21) to obtain, as $n \to \infty$,

$$\begin{aligned}\sup_{0 \leq t \leq \mathcal{T}} |Z_n(h)(t) - Z_n(t)| &\xrightarrow{p} 0 \text{ and} \\ \langle Z_n(h), Z_n(h) \rangle_{\mathcal{T}} &\xrightarrow{p} 0. \end{aligned} \tag{A.22}$$

The transition to (A.22) did not assume that $D^{(n)}(h)_t$ or $\tilde{C}_{\mathcal{T}}^{(n)}$ have a limit as $n \to \infty$. By the assumption that the $\alpha_t^{(n)}$ is $O_p(1)$ and hence P-UT, however, Jacod and Shiryaev (2003, Theorem VI.6.15, p. 380), yields that $D^{(n)}(h)_{\mathcal{T}}$ and $\tilde{C}_{\mathcal{T}}^{(n)}$ are tight.

From the second line in (A.22), by Lenglart's inequality (Jacod and Shiryaev (2003, Lemma I.3.30, p. 35)),

$$\sup_{0 \leq t \leq \mathcal{T}} |Z_n(h)(t)| \xrightarrow{p} 0. \tag{A.23}$$

Combining (A.23) with the first line of (A.22) yields the result of the Theorem, since $Z_n(\mathcal{T}) =$ the left hand side of (A.5). Since none of the bounds used depend on ΔT_n but only on $K_n \Delta T_n$, the result does not impose any requirement on ΔT_n, except that $\Delta T_n > 0$ and $K_n \Delta T_n \to 0$. *Q.E.D.*

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

Proof of Theorem 1. Because we shall use Theorem 7, we here let all quantities depend on index n. Thus, unlike Definition 2 in Section 2, $K = K_n$, etc, though we shall often omit the subscript when the meaning is obvious. For the purposes of the current proof, one can simply take

$n = B$, but this will no longer be the case in later appendices. Set

$$f_t^{(l,n)} = \frac{1}{K\Delta T} \sum_{K \le i \le B-K; i \equiv l[2K]} \left((T_{i+K} - t) I\{T_{i+K} > t \ge T_i\} + (t - T_{i-K}) I\{T_i > t \ge T_{i-K}\} \right). \tag{B.24}$$

where $i \equiv l[2K]$ means that i is on the form $2Ki + l$. We note that $f_t^{(l)} = f_t^{(l,n)}$ depends on n through ΔT, K, and B. It is easy to see that the family $\mathbb{F} = \{f^{(l,n)}\}$ satisfies (A.1), and that the set $\mathbb{F}'$ (from Definition 7) is indeed relatively compact for the Skorokhod topology.

Define the processes $\theta_t^{(l,n)} = \int_0^t f_{s-}^{(l,n)} d\theta_s$. To motivate the following development, note from Theorem 2 in Section 2.3 that for fixed $i \equiv l[2K]$,

$$\begin{aligned} \frac{1}{K(\Delta T)} (\Theta_{(T_i, T_{i+K}]} - \Theta_{(T_{i-K}, T_i]}) &= \frac{1}{K(\Delta T)} (\Theta'_{(T_i, T_{i+K}]} + \Theta''_{(T_{i-K}, T_i]}) \\ &= \int_{T_{i-K}}^{T_{i+K}} f_t^{(l,n)} d\theta \\ &= \theta_{T_{i+K}}^{(l,n)} - \theta_{T_{i-K}}^{(l,n)}, \end{aligned} \tag{B.25}$$

whence

$$\frac{1}{K^2(\Delta T)^2} \sum_{K \le i \le B-K, i \equiv l[2K]} (\Theta_{(T_i, T_{i+K}]} - \Theta_{(T_{i-K}, T_i]})^2 = \sum_{K \le i \le B-K, i \equiv l[2K]} (\theta_{T_{i+K}}^{(l,n)} - \theta_{T_{i-K}}^{(l,n)})^2 \tag{B.26}$$

and

$$\frac{1}{2} \frac{1}{K^2(\Delta T)^2} QV_{B,K}(\Theta) = \frac{1}{2K_n} \sum_{l=1}^{2K} \sum_{K \le i \le B-K, i \equiv l[2K]} (\theta_{T_{i+K}}^{(l,n)} - \theta_{T_{i-K}}^{(l,n)})^2. \tag{B.27}$$

We now wish to show that

$$\begin{aligned} \frac{1}{2K_n} \sum_{l=1}^{2K} \sum_{K \le i \le B-K, i \equiv l[2K]} (\theta_{T_{i+K}}^{(l)} - \theta_{T_{i-K}}^{(l)})^2 &= \frac{1}{2K} \sum_{l=1}^{2K} [\theta^{(l,n)}, \theta^{(l,n)}]_{\mathcal{T}} + o_p(1) \\ &= \int_0^{\mathcal{T}} f_t^{(n)} d[\theta, \theta]_t + o_p(1) \text{ , where} \end{aligned} \tag{B.28}$$

$$\begin{aligned} f_t^{(n)} &= \frac{1}{2K} \sum_{l=1}^{2K} (f_t^{(l,n)})^2 \\ &= \frac{1}{2K^3(\Delta T)^2} \sum_{K \le i \le B-K} \left((T_{i+K} - t)^2 I\{T_{i+K} \ge t > T_i\} + (t - T_{i-K})^2 I\{T_i \ge t > T_{i-K}\} \right). \end{aligned} \tag{B.29}$$

If K is finite, this is a simple matter of checking that

$$\sum_{K \le i \le B-K, i \equiv l[2K]} (\theta^{(l)}_{T_{i+K}} - \theta^{(l)}_{T_{i-K}})^2 = [\theta^{(l,n)}, \theta^{(l,n)}]_{\mathcal{T}} + o_p(1) \text{ for each } l = 1, \cdots, 2K,$$

where we recall that $i \equiv l[2K]$ means that i is on the form $2Ki + l$. For the general case where K can be finite or infinite, we proceed as follows. The class of functions $f_t^{(l,n)}$ given by (B.24) satisfies the conditions of Theorem 7. So does $\alpha_t^{(n)} = \theta_t$; since the process does not move with n, it is both tight and P-UT. Theorem 7 therefore yields (B.29).

For $t \in (T_{j-1}, T_j] \subseteq (T_K, T_{B-K}]$,

$$\begin{aligned} f_t^{(n)} &= \frac{1}{2K^3(\Delta T)^2} \left(\sum_{j-K \le i \le j-1} (T_{i+K} - t)^2 + \sum_{j \le i \le j+K-1} (t - T_{i-K})^2 \right) \\ &= \frac{1}{3}\left(1 - \frac{1}{K^2}\right) \; + \; \frac{1}{2}\frac{1}{K^2}\left(\left(\frac{T_j - t}{\Delta T}\right)^2 + \left(\frac{t - T_{j-1}}{\Delta T}\right)^2\right) \end{aligned} \tag{B.30}$$

hence, eventually, on all $[\delta, \mathcal{T} - \delta]$, for any $\delta > 0$. Since, for all $t \in [0, \mathcal{T}]$, $0 \le f_t^{(n)} \le 1$, and since $f_{\mathcal{T}}^{(n)} = 0$, Theorem 1 follows. Remark 15 in Appendix A continues to apply, for the same reasons. *Q.E.D.*

Proof of Theorem 2: By Itô's formula, $d(T + \delta - t)(\theta_t - \theta_T) = (T + \delta - t)d\theta_t - (\theta_t - \theta_T)dt$. Integrating from T to $T + \delta$ yields

$$0 = \Theta'_{(T,T+\delta]} - \Theta_{(T,T+\delta]} + \theta_T \delta. \tag{B.31}$$

Similarly $d(t - (T - \delta))(\theta_T - \theta_t) = -(t - (T - \delta))d\theta_t + (\theta_T - \theta_t)dt$. Integrating from $T - \delta$ to T yields

$$0 = -\Theta''_{(T-\delta,T]} - \Theta_{(T-\delta,T]} + \theta_T \delta. \tag{B.32}$$

Combining (B.31)-(B.32) yields the result. *Q.E.D.*

C Proof of Theorem 3, and a more General Result.

We here show a broader result of which Theorem 3 is a corollary. First of all, we replace the "omnibus" Condition 1 by the weaker and more precise Condition 5. Also, it shows what happens when one gives up on forcing negligibility in the form of conditions (22) and $\Delta T = o(n^{-\alpha})$. The former is conceptually important as it separates out what part of Condition 1 is required for the convergence of quadratic variations (as opposed to being a valid asymptotic variance). The latter is useful in case one were tempted to take K fixed in the discontinuous θ_t case. We first state and prove the more general Theorem 8, and then derive Theorem 3.

Condition 5. (Relative size of semi-martingale and edge effect in $\hat{\Theta}$ in (13).) *We assume that $M_{n,t}$ is a sequence of semimartingales. We assume that there is a rate $\alpha > 0$ (which need not be known) so that the sequence of semimartingales $(n^{\alpha} M_{n,t}) = O_p(1)$ in the sense of Definition 5 in Appendix A. We assume that $e_{n,T} = o_p(n^{-\alpha})$ and $\tilde{e}_{n,S} = o_p(n^{-\alpha})$ for any S and T.*

Theorem 8. (More General Expansion of $QV_{B,K}(\widehat{\Theta})$. *Assume that θ_t is a semimartingale on $[0, \mathcal{T}]$, and suppose that Condition 5 holds. Define*

$$\begin{aligned} QV_{B,K}(\Theta, M) &= \frac{1}{K} \sum_{i=K}^{B-K} \left(\Theta_{(T_i, T_{i+K}]} - \Theta_{(T_{i-K}, T_i]}\right) \left((M_{T_{i+K}} - M_{T_i}) - (M_{T_i} - M_{T_{i-K}})\right), \\ QV_{B,K}(M) &= \frac{1}{K} \sum_{i=K}^{B-K} \left((M_{T_{i+K}} - M_{T_i}) - (M_{T_i} - M_{T_{i-K}})\right)^2, \text{ and} \\ R_{n,K} &= \frac{1}{K} \sum_{i=K}^{B-K} (\tilde{e}_{T_{i+K}} - e_{T_i} - \tilde{e}_{T_i} + e_{T_{i-K}})^2, \end{aligned} \tag{C.33}$$

and also

$$\overline{QV}_{B,K}(\hat{\Theta}) = QV_{B,K}(\Theta) + 2QV_{B,K}(\Theta, M) + QV_{B,K}(M) \tag{C.34}$$

Let $K = K_n$ be positive integers, and assume that $K_n \Delta T_n \to 0$. Then, in extension of (23),

$$\frac{1}{2K} \sum_{K \le i \le B-K} (\hat{\Theta}_{(T_{i-K}, T_{i+K}]} - \Theta_{(T_{i-K}, T_{i+K}]})^2 = [M_n, M_n]_T + R_{n,K} + O_p(n^{-\alpha} R_{n,K}^{1/2}). \tag{C.35}$$

Also, in extension of (25),

$$\begin{aligned} \overline{QV}_{B,K}(\hat{\Theta}) \;=\; & 2[M_n, M_n]_T \\ & + (K\Delta T)^2 \frac{2}{3} \left(1 - \frac{1}{K^2}\right) [\theta, \theta]_{\mathcal{T}-} + (\Delta T)^2 \int_0^{\mathcal{T}} \left(\left(\frac{t^* - t}{\Delta T}\right)^2 + \left(\frac{t - t_*}{\Delta T}\right)^2 \right) d[\theta, \theta]_t \\ & + 2\Delta T \int_0^{\mathcal{T}} \left(1 - 2\frac{t - t_*}{\Delta T}\right) d[\theta, M_n]_t \;+\; o_p((K_n \Delta T)^2) + o_p(n^{-2\alpha}) \end{aligned} \tag{C.36}$$

and

$$QV_{B,K}(\hat{\Theta}) = \overline{QV}_{B,K}(\hat{\Theta}) + R_{n,K} + O_p((K\Delta T + n^{-\alpha}) R_{n,K}^{1/2}) \tag{C.37}$$

The convergence in probability is uniform in ΔT_n, so long as $\Delta T_n > 0$ and $K_n \Delta T_n \to 0$.

For the proofs, set $\alpha_t^{(n)} = \theta_t$, $\beta_t^{(n)} = n^{\alpha} M_{n,t}$. Let $f_t^{(l,n)}$ is given by (B.24) above. We shall use two different definitions of $g_t^{(l,n)}$. For both cases, let $\alpha_t^{(l,n)}$ and $\beta_t^{(l,n)}$ be as given by (A.2) .

Proof of (C.35) (Case 1 for $g_t^{(l,n)}$). Set

$$g_t^{(l,n)} = \sum_{K \le i \le B-K; i \equiv l[2K]} I\{T_{i+K} > t \ge T_{i-K}\}. \tag{C.38}$$

From Theorem 7,

$$\begin{aligned}
\frac{1}{2K}\sum_{i=K}^{B-K}(\beta^{(n)}_{T_{i+K}} - \beta^{(n)}_{T_{i-K}})^2 &= \frac{1}{2K_n}\sum_{l=1}^{2K}\sum_{K\le i\le B-K, i\equiv l[2K]}(\beta^{(l,n)}_{T_{i+K}} - \beta^{(l,n)}_{T_{i-K}})(\beta^{(l,n)}_{T_{i+K}} - \beta^{(l,n)}_{T_{i-K}}) \\
&= \frac{1}{2K}\sum_{l=1}^{2K}[\beta^{(l,n)}, \beta^{(l,n)}]_{\mathcal{T}} + o_p(1) \\
&= [\beta^{(n)}, \beta^{(n)}]_{\mathcal{T}} + o_p(1)
\end{aligned} \tag{C.39}$$

Thus, following (13), and using (C.39), write

$$\begin{aligned}
&\frac{1}{2K}\sum_{i=K}^{B-K}\left(\hat{\Theta}_{(T_{i-K},T_{i+K}]} - \Theta_{(T_{i-k},T_{i+K}]})\right)^2 \\
&= \frac{1}{2K}\sum_{i=K}^{B-K}\left(n^{-\alpha}(\beta^{(n)}_{T_{i+K}} - \beta^{(n)}_{T_{i-K}}) + (\tilde{e}_{T_{i+K}} - e_{T_i})\right)^2 \\
&= n^{-2\alpha}[\beta^{(n)}, \beta^{(n)}]_{\mathcal{T}} + R_{n,K} + O_p((K\Delta T + n^{-\alpha})R_{n,K}^{1/2}).
\end{aligned} \tag{C.40}$$

by Cauchy-Schwarz. Since $n^{-2\alpha}[\beta^{(n)}, \beta^{(n)}]_{\mathcal{T}} = [M_n, M_n]_{\mathcal{T}}$, (C.35) is proved. Remark 15 in Appendix A remains valid for the same reasons, and also in view of Proof of Theorem 1. *Q.E.D.*

Proof of the rest of Theorem 8 (Case 2 for $g_t^{(l,n)}$). Recall that

$$\begin{aligned}
\hat{\Theta}_{(T_i,T_{i+K}]} - \hat{\Theta}_{(T_{i-K},T_i]} &= \Theta_{(T_i,T_{i+K}]} - \Theta_{(T_{i-K},T_i]} \\
&\quad + (M_{T_{i+K}} - M_{T_i}) - (M_{T_i} - M_{T_{i-K}}) + (\tilde{e}_{T_{i+K}} - e_{T_i} - \tilde{e}_{T_i} + e_{T_{i-K}})
\end{aligned} \tag{C.41}$$

We obtain from Cauchy-Schwarz that

$$QV_{B,K}(\hat{\Theta}) = \overline{QV}_{B,K}(\hat{\Theta}) + R_{n,K} + O_p(\overline{QV}_{B,K}(\hat{\Theta})^{1/2}R_{n,K}^{1/2}) \tag{C.42}$$

whence (C.37) follows from (C.36)

It remains to show (C.36). The first term in (C.34) is covered by Theorem 1 in Section 2.3. To handle the two remaining terms, we redefine

$$g_t^{(l,n)} = \sum_{K \le i \le B-K; i \equiv l[2K]} \left(I\{T_{i+K} > t \ge T_i\} - I\{T_i > t \ge T_{i-K}\}\right), \tag{C.43}$$

but keep the rest of the notation from the beginning of this section (Appendix C). Note that $f_t^{(l,n)}$ is absolutely continuous, and that $g_t^{(l,n)} = -(K\Delta T)df_t^{(l,n)}/dt$ (except at discontinuities), whence by Fubini's Theorem, where $f_t^{(n)}$ is given in equation (B.29),

$$\begin{aligned}\sum_{l=1}^{2K} g_t^{(l,n)} f_t^{(l,n)} &= -\frac{1}{2}(K\Delta T)\frac{d}{dt}\sum_{l=1}^{2K}(f_t^{(l,n)})^2 \\ &= -(K^2\Delta T)\frac{d}{dt}f_t^{(n)} \\ &= 1 - 2\frac{t-t_*}{\Delta T}\end{aligned} \tag{C.44}$$

eventually for all $t \in [\delta, \mathcal{T} - \delta]$, by (B.30). One can alternatively verify (C.44) directly.

From Theorem 7,

$$\begin{aligned}\frac{1}{2}n^{2\alpha}QV_{B,K}(M) &= \frac{1}{2K_n}\sum_{l=1}^{2K}\sum_{K\le i\le B-K, i\equiv l[2K]}(\beta_{T_{i+K}}^{(l,n)} - \beta_{T_{i-K}}^{(l,n)})(\beta_{T_{i+K}}^{(l,n)} - \beta_{T_{i-K}}^{(l,n)}) \\ &= \frac{1}{2K}\sum_{l=1}^{2K}[\beta^{(l,n)}, \beta^{(l,n)}]_{\mathcal{T}} + o_p(1) \\ &= [\beta^{(n)}, \beta^{(n)}]_{\mathcal{T}} + o_p(1)\end{aligned} \tag{C.45}$$

and

$$\begin{aligned}\frac{1}{2K\Delta T}n^{\alpha}QV_{B,K}(\Theta, M) &= \frac{1}{2K_n}\sum_{l=1}^{2K}\sum_{K\le i\le B-K, i\equiv l[2K]}(\alpha_{T_{i+K}}^{(l,n)} - \alpha_{T_{i-K}}^{(l,n)})(\beta_{T_{i+K}}^{(l,n)} - \beta_{T_{i-K}}^{(l,n)}) \\ &= \frac{1}{2K}\sum_{l=1}^{2K}[\alpha^{(l,n)}, \beta^{(l,n)}]_{\mathcal{T}} + o_p(1) \\ &= \frac{1}{2K}\sum_{l=1}^{2K}\int_0^{\mathcal{T}} g_t^{(l,n)} f_t^{(l,n)} d[\theta, \beta^{(n)}]_t + o_p(1) \\ &= \frac{1}{2K}\int_0^{\mathcal{T}}\left(1 - 2\frac{t-t_*}{\Delta T}\right)d[\theta, \beta^{(n)}]_t + o_p(1)\end{aligned} \tag{C.46}$$

by (C.44). *Q.E.D.*

Remaining Proof of Theorem 3. Condition 1 implies Condition 5. Eq. (22) is the same as requiring that $\sum_i e_{T_i}^2 = o_p(K_n n^{-2\alpha})$ and $\sum_i \tilde{e}_{T_i}^2 = o_p(K_n n^{-2\alpha})$, whence $R_{n,K} = o_p(n^{-2\alpha})$. Expressions (23) and (25) them follow directly from Theorem 8 when assuming Condition 1. This is because of (18) in Proposition 1. For expression (18), we also have invoked the assumption (24). *Q.E.D.*

REMARK 16. (AVAR *vs.* AMSE.) There are situations of interest when Condition 5 is satisfied, but the additional assumptions of Condition 1 are not. Most notably, consider the situation where $[L, L]_{\mathcal{T}}$ is not $\mathcal{G}$-measurable but instead just integrable. For simplicity, assume that $L_{n,t} = n^{-\alpha} M_{n,t}$ converges in law to L_t relative to the Skorokhod metric on $\mathbb{D}$ (as oppsed to just being tight). In this case, (15) needs to be replaced by

$$\mathrm{AMSE}(\hat{\Theta} - \Theta) = n^{-2\alpha}[L, L]_{\mathcal{T}} + o_p(n^{-2\alpha}), \tag{C.47}$$

where AMSE is the asymptotic mean squared error. This situation arises, for example, in the case of endogenous sampling times for realized volatility (Li, Mykland, Renault, Zhang, and Zheng (2014)). The same phenomenon occurs under direct estimation of skewness (Kinnebrock and Podolskij (2008, Example 6), Mykland and Zhang (2009, Example 3, p. 1414-1416)). □

D Stable Convergence and of the P-UT Condition

D.1 Concepts

STABLE CONVERGENCE (Definition 3 in Section 3.1) allows you to take the information from the data (represented by sigma-field $\mathcal{G}$) into the asymptotic distribution. Most commonly, this information is the quadratic variation $[L, L]_{\mathcal{T}}$, which plays the rôle of variance in the asymptotic distribution, but which can be consistently estimated from the data by any consistent estimator of $n^{2\alpha}[M_n, M_n]_{\mathcal{T}}$. This is the contents of Proposition 1. .

General conditions for stable convergence to hold can be found in Hall and Heyde (1980), and has a quite general formulation in Jacod and Shiryaev (2003, Theorem VI.6.26 (p. 384)). Stable convergence of estimators has also been found in countless articles in specific situations, including in high frequency data. See also the book by Jacod and Protter (2012) and the review paper by Podolskij and Vetter (2010).

The amount of data $\mathcal{G}$ that one wishes to carry to asymptopia may vary. The theory described in this paper will work for any $\mathcal{G} \subseteq \mathcal{F}$, so long as $[L, L]_{\mathcal{T}}$ is $\mathcal{G}$-measurable. (This is true under minimal conditions, see Proposition 6 at the end of this section.) One may, however, wish to carry other information. First, for suitably chosen $\mathcal{G}$, stable convergence commutes with measure change (Mykland and Zhang (2009, Proposition 1, p. 1408)), and this can simplify analysis. Second, stable convergence can help weaken conditions with the assistance of *localization*, see, *e.g.,* Jacod and Protter (2012, Lemma 4.4.9, p. 118-121), and Mykland and Zhang (2012, Section 2.4.5, pp. 160-161). In common practice, the information in $\mathcal{G}$ will include latent efficient prices X_t and parameter processes θ_t, but typically not information from the microstructure noise, if present in the model (Zhang, Mykland, and Aït-Sahalia (2005), Zhang (2006), Jacod, Li, Mykland, Podolskij, and Vetter (2009a), Podolskij and Vetter (2009b), Jacod and Protter (2012), and many others).

Thus, $L_{n,t} = n^{\alpha} M_{n,t}$ may in some circumstances not be $\mathcal{G}$ measurable.

For general discussions of stable convergence, see Jacod and Protter (1998, Section 2, pp. 169-170), Jacod and Shiryaev (2003, Chapter VIII.5c-d, pp. 512-519), Jacod and Protter (2012, Chapter 2.2.1, pp. 46-50), and Mykland and Zhang (2012, Section 2.4, pp. 150-161). For further background on stable convergence, see Rényi (1963), Aldous and Eagleson (1978), , Rootzén (1980), and Zhang (2001). Stable convergence was originally thought of as a form of conditional convergence (Jacod and Shiryaev (2003, top of p. 513)).

REMARK 17. In this paper, convergence in law for processes is relative to the Skorokhod topology on the space $\mathbb{D} = \mathbb{D}[0, \mathcal{T}]$ of *càdlàg* functions $[0, \mathcal{T}] \to \mathbb{R}$. In Definition 3, the pair (L_n, Y) converges in the product topology. In other words, $(L_n, Y) \xrightarrow{\mathcal{L}} (L, Y)$ means that $Ef(L_n)g(Y) \to Ef(L)g(Y)$, for all bounded continuous $f : \mathbb{D} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$. For more on the Skorokhod topology, see Jacod and Shiryaev (2003, Chapter VI.1-2, pp. 325-346). Note that $\mathcal{F}_t$ can depend on n, *cf.* the discretization discussion in the next section. □

THE PREDICTABLY UNIFORMLY TIGHT (P-UT) CONDITION is described and studied in Jacod and Shiryaev (2003, Chapter VI.6, p. 377-388). It is an additional regularity condition which avoids certain idiosyncrasies associated with regular process convergence. If the sequence of semimartingales L_n is tight in the Skorokhod topology, one can take as definition of P-UT that if H_n is a bounded family of predictable processes, then $\int_0^T H_{n,t} dL_{n,t}$ is tight for each T (*ibid,* Definition 6.1, p. 377, and Corollary 6.20, p. 381). Also, by *ibid.,* Theorem VI.6.22 (p. 383), if $(H_{n,+}, L_n) \xrightarrow{\mathcal{L}} (H, L)$ (and subject to regularity conditions), then $\int H_{n,t} dL_{n,t} \xrightarrow{\mathcal{L}} \int H_t dL_t$. Also, and this is important for the current paper, $[L_n, L_n] \xrightarrow{\mathcal{L}} [L, L]$ (*ibid,* Theorem VI.6.26, p. 384). Finally, P-UT prevents the predictable finite variation part of L_n from turning into a different type of process (*ibid,* Theorem 6.15 (iii), p. 380, and Theorem VI.6.21, p. 382).

We have seen in Sections 6 and 7 that there is little additional burden in verifying the P-UT condition once one proves stable convergence. Also, a sufficient condition for a sequence of local martingales $L_{n,t}$ to be P-UT is that (Jacod and Shiryaev (2003, Corollary VI.6.30, p. 385))

$$\sup_n E \sup_{0 \le t \le \mathcal{T}} |\Delta L_{n,t}| < \infty. \tag{D.48}$$

The condition (D.48) is weaker than what is usually required for a central limit theorem,[43], and it does in particular not impose asymptotic negligibility. If (D.48) still seems too strong, the requirement can be localized using stable convergence, as described above in this section.

As an illustration of how stable convergence blends with P-UT:

PROOF OF PROPOSITION 1. Let $(\mathcal{F}_t^L)$ be the filtration generated by the process L_t, on the extension $(\tilde{\Omega}, \tilde{\mathcal{G}}, \tilde{P})$. Since, by assumption, L_t is a local martingale with respect to filtration $(\mathcal{G} \vee \mathcal{F}_t^L)$,

[43]See, for example, Hall and Heyde (1980, conditions (3.18) and (3.20), p. 58).

then it follows that $L_t^2 - [L,L]_t$ is also a local martingale w.r.t. $\mathcal{G} \vee \mathcal{F}_t^L$, and hence $E(L_\mathcal{T} \mid \mathcal{G}) = 0$ and $E(L_\mathcal{T}^2 - [L,L]_\mathcal{T} \mid \mathcal{G}) = 0$. Hence, $\mathrm{Var}(L_\mathcal{T}|\mathcal{G}) = [L,L]_\mathcal{T}$. Set $L_{n,t} = n^\alpha M_{n,t}$. Since $L_{n,t}$ is P-UT, Jacod and Shiryaev (2003, Proposition VI.2.1, p. 377 and Theorem VI.6.26, p. 384)) yields that $[L_n,L_n]_\mathcal{T} \xrightarrow{\mathcal{L}} [L,L]_\mathcal{T}$ stably in law as $n \to \infty$. However, since $[L,L]_\mathcal{T}$ is $\mathcal{G}$ measurable and hence defined on the original space, $[L_n,L_n] \xrightarrow{p} [L,L]_\mathcal{T}$ by Jacod and Protter (2012, eq. (2.2.7), p 47). (It is enough for the "<=" part of the cited result that the limiting random variable be $\mathcal{G}$ measurable.) *Q.E.D.*

We finish with a result on minimal stable convergence.[44]

Proposition 6. (Automatic Minimal Stable Convergence.) *Assume that the sequence of semimartingales $L_n = n^\alpha M_n$ converges in law to L, and is P-UT. Also assume that $[L_n,L_n]_\mathcal{T}$ converges in probability. Call this limit V (so $[L_n,L_n]_\mathcal{T} \xrightarrow{p} V$). Let $\mathcal{G}$ be the sigma-field generated by V. Then there is an extension $(\tilde{\Omega},\tilde{\mathcal{G}},\tilde{P})$ of $(\Omega,\mathcal{G},P)$ so that L_n converges stably in law with respect to $\mathcal{G}$ as $n \to \infty$. Also, on this extension, $[L,L]_\mathcal{T} = V$, and $\mathcal{F}_\mathcal{T}^L$ is conditionally independent of $\mathcal{F}$ given $\mathcal{G}$.*

D.2 Proofs of Propositions 4, 5, and 6

Proof of Proposition 5 in Section 6. The only modification that is required in our proofs is to replace our parameter process by $\theta_{n,t} = \theta_{t_{n,i}}$ for $t_{n,i} \le t < t_{n,i+1}$. Since (the original $(\mathcal{F}_t)$ adapted) θ_t is a semimartingale, then so is $\theta_{n,t}$. Also, $\theta_{n,t}$ converges in probability to θ_t in the Skorokhod topology (Jacod and Shiryaev (2003, Proposition VI.6.37, p 387)) (and hence also in law). Also, $\theta_{n,t}$ is P-UT (*ibid,* Definition VI.6.1, p.377) since the relevant predictable functions on filtration $\mathcal{F}_{t_{n,i}}$ is a subset of the corresponding predictable functions on filtration $\mathcal{F}_t$.

For example, the proof of Theorem 1 in Appendix B goes through with $\theta_{n,t}$ in lieu of θ_t, because Theorem 7 in Appendix A allows time varying $\theta_{n,t}$. The times $T_{n,i}$ are not changed in derivations that do not involve microstructure noise.

Arguments involving only $(e_{n,T_{n,i}}, \tilde{e}_{n,T_{n,i}})$ are directly converted to $(e_{n,T_{n,i,*}}, \tilde{e}_{n,T_{n,i,*}})$. *Q.E.D.*

Proof of Proposition 4 in Section 5.2. This is a corollary to Proposition 5. If Condition 1 is valid (in its original form) for $M_{n,t}$, it certainly also holds when discretized as in Condition 3, again using Jacod and Shiryaev (2003, Proposition VI.6.37, p 387). This shows the result. *Q.E.D.*

Proof of Proposition 6. Let $f : \mathbb{D} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be bounded and continuous. Since $L_{n,t}$ is P-UT, Jacod and Shiryaev (2003, Proposition VI.2.1, p. 377 and Theorem VI.6.26, p. 384)) yields that $(L_n,[L_n,L_n]_\mathcal{T}) \xrightarrow{\mathcal{L}} (L,[L,L]_\mathcal{T})$ (in the non-stable sense), *i.e.,* $Ef(L_n)g([L_n,L_n]_\mathcal{T}) =$

[44]The following proposition is conceptually related to Hall and Heyde (1980, condition (3.19), p. 58).

$Ef(L)g([L,L]_{\mathcal{T}}) + o(1)$. On the other hand, by the assumed convergence in probability, $|Ef(L_n)g([L_n,L_n]_{\mathcal{T}}) - Ef(L_n)g(V)| \leq \sup_x |f(x)|E|g([L_n,L_n]_{\mathcal{T}}) - g(V)| \to 0$.

We now construct our extension as in Jacod and Protter (2012, p. 36): $\tilde{\Omega} = \Omega \times \mathbb{D}[0,T]$ with product sigma-field, where the sigma-field on $\mathbb{D}[0,T]$ is derived from the Skorokhod topology (Jacod and Shiryaev (2003, Theorem VI.1.14c, p. 328)). The transition probability is given as the regular conditional probability $Q(L|V)$ (Ash (1972, Theorem 6.6.5, p. 265)), where Q is defined as the joint distribution of $(L,[L,L]_{\mathcal{T}})$ on $\mathbb{D}[0,T] \times \mathbb{R}$ (with corresponding product sigma-field).

With these definitions, $[L,L]_{\mathcal{T}} = V$, and hence, from the above,

$$Ef(L)g(V) = Ef(L)g([L,L]_{\mathcal{T}}) = Ef(L_n)g([L_n,L_n]_{\mathcal{T}}) + o(1) = Ef(L_n)g(V) + o(1) \text{ as } n \to \infty.$$

Hence, the stable convergence follows. The remaining statements of the proposition hold by construction. *Q.E.D.*

D.3 P-UT property in Example 2 in Section 7

P-UT PROPERTY FOR $n^{1/2}M_n$. We make the following assumptions: (i) μ_t is locally integrable and σ_t^2 is continuous,[45] and (ii)

$$\sum_{j=1}^{n} |\Delta J_{t_{j-1}}||\Delta J_{t_j}| = O_p(n^{-1/2}). \tag{D.49}$$

In particular, equation (D.49) is satisfied when $J_t = J_t^{(1)} + J_t^{(2)}$, where $J^{(1)}$ has finitely many jumps and $J^{(2)}$ is a purely discontinuous Itô-semimartingale (see, for example, Jacod and Protter (2012, Definition 2.1.1, p. 35, see also Theorem 2.1.2, p. 37)).

PROOF OF P-UT PROPERTY. Without changing either assumptions or conclusions, we absorb the $\mu_t dt$ term into dJ_t, so that $dX_t = \sigma_t dW_t + dJ_t$. $[J,J]_t$ is unchanged, and so is the statement (D.49). From (D.49) as well as Jacod and Shiryaev (2003, Definition VI.6.1 and the additivity VI.6.4, both p. 377), it follows that to verify P-UT of the original M_n, it is enough that the P-UT property holds on a modified $\tilde{M}_n$ which has the same form as (53) but with X replaced by X^c, where $dX_t^c = \sigma_t dW_t$. For this process, it is easy to verify P-UT under the contiguous sequence of measures Q_n from Mykland and Zhang (2009, Section 3, pp. 1416-1421). and using the big block-small block device (Mykland, Shephard, and Sheppard (2012, Appendix A.5, 32-33)), again using Definition VI.6.1 from Jacod and Shiryaev (2003). But this definition is invariant to contiguous change of measure, and hence $\tilde{M}_n$ is P-UT under the original measure P. It follows that the original $n^{1/2}M_n$ is P-UT. *Q.E.D.*

[45] The spot volatility is also a semimartingale since $\theta_t = \sigma_t^2$. The continuity assumption is merely for convenience and can be reduced to an assumption that σ_t^2 be locally bounded.