

Realised quantile-based estimation of the integrated variance*

Kim Christensen[†]

Roel Oomen[‡]

Mark Podolskij[§]

January, 2010

Abstract

In this paper, we propose a new jump robust quantile-based realised variance measure of ex-post return variation that can be computed using potentially noisy data. The estimator is consistent for the integrated variance and we present feasible central limit theorems which show that it converges at the best attainable rate and has excellent efficiency. Asymptotically, the quantile-based realised variance is immune to finite activity jumps and outliers in the price series, while in modified form the estimator is applicable with market microstructure noise and therefore operational on high-frequency data. Simulations show that it has superior robustness properties in finite sample, while an empirical application illustrates its use on equity data.

Keywords: Finite activity jumps; Market microstructure noise; Order statistics; Outliers; Realised variance.

JEL Classification: C10; C80.

*Podolskij gratefully acknowledges financial support from CREATES funded by the Danish National Research Foundation. We would like to thank Peter Bank, Álvaro Cartea, Fulvio Corsi, Dick van Dijk, Dobrislav Dobrev, Bruce Lehman, Haikady Nagaraja, Kevin Sheppard (discussant), Neil Shephard and the seminar participants at the London-Oxford Financial Econometrics Study Group at Imperial College, Humboldt-Universität zu Berlin, ECARES of Université Libre de Bruxelles, LSE Department of Statistics, Federal Reserve Board, Washington, Amsterdam Business School, Tinbergen Institute, Rotterdam, the Workshop on Mathematical Finance for Young Researchers, Quantitative Products Laboratory, Berlin, and the SITE Summer Workshop 2008, Stanford for valuable comments and discussions.

[†]CREATES, School of Economics and Management, Aarhus University, Building 1322, Bartholins Allé 10, 8000 Aarhus, Denmark. E-mail: kchristensen@creates.au.dk.

[‡]Deutsche Bank AG, Winchester House, 1 Great Winchester Street, London EC2N 2DB, UK and affiliated with the Department of Quantitative Economics, the University of Amsterdam, The Netherlands. E-mail: roel.ca.oomen@gmail.com. Phone: +44 (0) 207 54 56989.

[§]Podolskij is with ETH Zürich, Department of Mathematics, Rämistrasse 101, CH-8092 Zürich, Switzerland and affiliated with CREATES, University of Aarhus, Denmark. E-mail: mark.podolskij@math.ethz.ch.

1 Introduction

In recent years, our understanding of asset price dynamics has been significantly enhanced by the increasing availability of intra-day financial tick data in conjunction with the rapid development and harnessing of the necessary econometric tools. Realised variance, defined as the sum of squared intra-period returns, has been a key driver in this literature (e.g. Andersen, Bollerslev, Diebold, and Labys, 2001; Barndorff-Nielsen and Shephard, 2002) as it provides a simple yet highly efficient way to consistently estimate the quadratic variation of a price process. However, when faced with the realities of high-frequency data, the use of realised variance is limited in two important ways. First, realised variance is overly sensitive to an inherent feature of high-frequency data, namely market microstructure noise. In fact, it destroys the consistency of the estimator. Second, realised variance is an estimator of the total variation – the sum of diffusive and jump variation – and consequently cannot distinguish between these two fundamentally different sources of risk. As emphasized by Aït-Sahalia (2004), the ability to disentangle jumps from volatility is key for the valuation of derivatives (e.g. Merton, 1976; Duffie, Pan, and Singleton, 2000), risk measurement and management (e.g. Duffie and Pan, 2001), as well as asset allocation (e.g. Jarrow and Rosenfeld, 1984; Liu, Longstaff, and Pan, 2003). Andersen, Bollerslev, and Diebold (2007) and Bollerslev, Kretschmer, Pigorsch, and Tauchen (2009) also point to its importance for the empirical modeling of asset price dynamics and forecasting of volatility.

In response, two largely separate strands of literature have emerged. One on jump-robust realised variance measures, most notably the influential bi-power variation of Barndorff-Nielsen and Shephard (2004) and the threshold estimators of Jacod (2008); Mancini (2004a, 2009). The other on noise-robust realised variance measures, where the main approaches are based on subsampling (see Zhang, 2006; Zhang, Mykland, and Aït-Sahalia, 2005), kernel-based autocovariance adjustments (see Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2008; Zhou, 1996), and pre-averaging methods (see Jacod, Li, Mykland, Podolskij, and Vetter, 2009; Podolskij and Vetter, 2009).¹ Yet, the jump-robust realised variance measures are typically not robust to noise and the noise-robust realised variance measures are typically not robust to jumps. The contribution of this paper is to integrate both issues: we develop a new quantile-based realised variance (QRV) measure, which constitutes the first estimator of integrated variance that, at the same time, is highly efficient and simultaneously robust to jumps and microstructure noise. We present a complete asymptotic theory for this class of estimators, with central limit theorems that hold up in the presence of finite activity jumps.

¹Other methods include sparse sampling (e.g. Andersen, Bollerslev, Diebold, and Labys, 2000; Bandi and Russell, 2008), pre-filtering (Andersen, Bollerslev, Diebold, and Ebens, 2001; Bollen and Inder, 2002; Hansen, Large, and Lunde, 2008), model-based corrections (e.g. Corsi, Zumbach, Müller, and Dacorogna, 2001), and wavelet-based methods of Fan and Wang (2007). Variations and extensions of the realised kernel and subsampling approach can be found in Sun (2006) and Nolte and Voev (2007), respectively.

In addition to noise and jump robustness, another appealing feature of our estimator is that it is robust to outliers in the price process. Outliers are unavoidable in high-frequency data (due to, for instance, delayed trade reporting, recording errors, decimal misplacement, etc.) and are often hard to filter out systematically due to their random and irregular nature while the vast quantities of data make “visual inspection” impractical. In the paper, we illustrate that this feature of our estimator can be crucial in practical applications.

The construction of the QRV is based on the fundamental relationship between quantiles and the spread of the normal distribution. A stylized example provides the key intuition. For i.i.d. Gaussian data with mean zero and variance σ^2 , we know that the 95% quantile equals 1.645σ . Thus, with a measurement of a sample quantile, volatility can be estimated by inverting this relationship. Of course, other quantiles can be added to improve the efficiency of the estimator and as long as the selected quantiles are sufficiently far away from the extreme tails we have robustness to outliers. Estimators of this kind have a long history in the statistics literature: they can be traced back to Pearson (1920) and are further studied by Mosteller (1946), Eisenberger and Posner (1965), and David (1970).

The scope of this paper is, however, more ambitious than the above illustration in that we aim to develop an estimator of the integrated variance under very weak conditions on the underlying process, need to deal with microstructure noise, and consider limits where data are sampled at an increasing rate over a fixed time window. More formally, let $\{X_{i/N}\}_{i=0}^N$ denote a time-series of logarithmic asset price observations and define returns as $\Delta_i^N X = X_{i/N} - X_{(i-1)/N}$. The basic idea is to split the sample into n smaller blocks each containing m returns, $D_i^m X = (\Delta_k^N X)_{(i-1)m+1 \leq k \leq im}$ for $i = 1, \dots, n$ with $N = mn$, and then study the behavior of the return quantiles on these blocks as n grows large. In particular, we consider a $(k \times 1)$ vector of quantiles $\bar{\lambda} = (\lambda_1, \dots, \lambda_k)'$ and define the QRV as:

$$QRV_N(m, \bar{\lambda}, \alpha) \equiv \alpha' QRV_N(m, \bar{\lambda}),$$

where α is a $(k \times 1)$ vector of quantile weights, and $QRV_N(m, \bar{\lambda})$ is a $(k \times 1)$ vector with j th entry equal to

$$QRV_N(m, \lambda_j) = \frac{m}{N} \sum_{i=1}^{N/m} \frac{q_i(m, \lambda_j)}{\nu_1(m, \lambda_j)}, \quad \text{for } \lambda_j \in (1/2, 1).$$

Here

$$q_i(m, \lambda) = g_{\lambda m}^2 \left(\sqrt{N} D_i^m X \right) + g_{m-\lambda m+1}^2 \left(\sqrt{N} D_i^m X \right),$$

is the realised symmetric squared λ -quantile, $\nu_1(m, \lambda_j)$ is a normalizing constant that measures the expectation of $q_i(m, \lambda)$ under a standard normal, and the function $g_k(x) = x_{(k)}$ extracts the k th order statistic of $x = (x_1, \dots, x_m)$.

It is quite intuitive that QRV as defined above provides consistent and jump-robust estimates of the integrated variance as $n \rightarrow \infty$. As the number of blocks grows, they span an increasingly

short interval so that in the limit and under weak assumptions on the price process, each block contains at most one jump and volatility within the block is locally constant. In this scenario, the term $q_i(m, \lambda_j)/\nu_1(m, \lambda_j)$ constitutes an estimator of the (scaled) return variance over the i^{th} block, which is robust to jumps by the assumption that $\lambda_{max} = \max\{\bar{\lambda}\} < 1$. Summing across blocks then naturally yields a consistent estimator for the integrated variance. From the above, it is also clear that the QRV estimator can be formulated based on overlapping blocks. In the paper, we show that such a subsampled version of QRV further improves the efficiency of the estimator. In terms of asymptotic theory, we derive feasible central limit theorems for QRV, which show that for fixed m and in the absence of noise, our estimator converges to the integrated variance at rate $N^{-1/2}$. With microstructure noise, our (modified) estimator converges at rate $N^{-1/4}$. In both cases, this is known to be the fastest possible rate. We carry on to show that for suitable choice of parameters the asymptotic variance of the QRV is close to the maximum likelihood bound in either situation. Moreover, all our consistency and central limit theorems are shown to be robust to the presence of finite activity jumps.

Implementation of the QRV requires the choice of some “tuning” parameters, such as the number of blocks n or block length m , the quantiles $\bar{\lambda}$, and the quantile weights α . For given block length and quantiles, we show how to select the quantile weights optimally to minimize the asymptotic variance of QRV. In the special case where $m \rightarrow \infty$, we can express these optimal weights in a simple way. This proves convenient for the implementation of the QRV, as it turns out we can use these asymptotic weights even for finite m with hardly any loss of efficiency. Similarly, the choice of m and $\bar{\lambda}$ can be based on asymptotic efficiency considerations, but we also emphasize that the robustness of QRV in finite sample is controlled by the joint choice of these parameters as the $(1 - \lambda_{max})m - 1$ largest negative and positive returns are discarded in each block of data. In the paper, we provide detailed guidance on how to choose m and $\bar{\lambda}$ based on both asymptotic theory and practical considerations.

There are a number of recent papers related to our work. Podolskij and Vetter (2009) combine pre-averaging with the bi-power variation measure to also obtain a jump- and noise-robust volatility estimator. In comparison to QRV, their estimator is inefficient and because it is based on bi-power variation it is not robust to outliers and the corresponding central limit theorems are not robust to jumps. Andersen, Dobrev, and Schaumburg (2008) develop two jump-robust measures named MinRV and MedRV, but they rule out microstructure noise. Interestingly, we show that their estimators can be nested in our class of estimators and by extension we provide an asymptotic theory for the MinRV and MedRV in the presence of microstructure noise. Further, the idea to get rid of jumps through quantiles is somewhat similar in spirit to threshold estimators of the integrated variance (see, e.g., Mancini, 2004b; Jacod, 2008), where return observations larger than a pre-determined threshold are removed before computing realised variance. Because these estimators

use a global threshold to pre-truncate the data there is a risk of retaining jumps when the threshold is set too high or removing an excessive amount of observations when the threshold does not fully encapsulate the returns in high volatility episodes. In both cases the estimator will be biased. In contrast, QRV sets a local threshold that adapts naturally to the magnitude of the returns observations in each block through the choice of quantiles. As such, it is able to be robust to both small and large jumps even in the presence of time-varying volatility. Finally, we note that QRV is also related to the literature on L-statistics (e.g. van der Vaart, 1998), and it bears some resemblance to the classical idea of trimmed mean estimation.

The remainder of this paper is organized as follows. In Section 2, we introduce the QRV estimator and present the associated asymptotic theory. We also conduct an extensive simulation study to gauge its finite sample performance and discuss how this compares to other related estimators. In Section 3, we develop a modified version of QRV that remains consistent and asymptotically efficient with microstructure noise. Section 4 contains an empirical application and Section 5 concludes. Proofs and some additional discussion of alternative formulations of QRV can be found in the appendix.

2 Quantile-based realised variance measurement

Let $X = (X_t)_{t \geq 0}$ denote the log-price process, defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ and adapted to the filtration $(\mathcal{F}_t)_{t \geq 0}$. The theory of financial economics states that if X evolves in a frictionless market, then it has to be of semimartingale form (see Back, 1991). In this paper we start from the assumption that X is a pure Brownian semimartingale, i.e. a continuous sample path process of the form:

$$X_t = X_0 + \int_0^t a_u du + \int_0^t \sigma_u dW_u, \quad t \geq 0, \quad (1)$$

where $a = (a_t)_{t \geq 0}$ is a predictable locally bounded drift function, $\sigma = (\sigma_t)_{t \geq 0}$ is an adapted càdlàg volatility process and $W = (W_t)_{t \geq 0}$ a standard Brownian motion.

To prove our CLTs, we will work under some stronger assumptions on σ .

Assumption (V) σ does not vanish (V_1) and it satisfies the equation:

$$\sigma_t = \sigma_0 + \int_0^t a'_u du + \int_0^t \sigma'_u dW_u + \int_0^t v'_u dB'_u, \quad t \geq 0, \quad (V_2)$$

where $a' = (a'_t)_{t \geq 0}$, $\sigma' = (\sigma'_t)_{t \geq 0}$ and $v' = (v'_t)_{t \geq 0}$ are adapted càdlàg, $B' = (B'_t)_{t \geq 0}$ is a Brownian motion, and $W \perp\!\!\!\perp B'$ (here $A \perp\!\!\!\perp B$ means that A and B are stochastically independent).

This means that σ has its own Brownian semimartingale structure. Note the appearance of W in σ , which allows for leverage effects. If X is a unique strong solution of a stochastic differential equation

then, under some smoothness assumptions on the volatility function $\sigma = \sigma(t, X_t)$, assumption (V₂) (with $v'_s = 0$ for all s) is a consequence of Ito's formula. Thus, assumption (V₂) is fulfilled for many financial models and, even though it is not a necessary condition, it simplifies the proofs considerably. A more general treatment, including the case where σ jumps, can be found in Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006). We rule out these technical details here, as they are not important to our exposition.

In what follows, we also make use of the concept of stable convergence in law.

Definition 1 *A sequence of random variables, $(Z_n)_{n \in \mathbb{N}}$, converges stably in law with limit Z , defined on an appropriate extension of $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, if and only if for every \mathcal{F} -measurable, bounded random variable Y and any bounded, continuous function g , the convergence $\lim_{n \rightarrow \infty} \mathbb{E}[Yg(Z_n)] = \mathbb{E}[Yg(Z)]$ holds. We write $Z_n \xrightarrow{d_s} Z$, if $(Z_n)_{n \in \mathbb{N}}$ converges stably in law to Z .*

Stable convergence implies weak convergence, or convergence in law, which can be defined as above by taking $Y = 1$, see Rényi (1963) or Aldous and Eagleson (1978) for more details about the properties of stably converging sequences. The extension of this concept to stable convergence of processes is discussed in Jacod and Shiryaev (2003, pp. 512–518). In our context, we need the stable convergence to transform the infeasible mixed Gaussian central limit theorems (CLT) proved below into feasible ones that can be implemented in practice.

Central to the theory of semimartingales is the quadratic variation process, defined as:

$$[X]_t = \operatorname{plim}_{n \rightarrow \infty} \sum_{i=1}^n (X_{t_i} - X_{t_{i-1}})^2,$$

for any sequence of partitions $0 = t_0 < t_1 < \dots < t_n = t$ such that $\sup_i \{t_i - t_{i-1}\} \rightarrow 0$ as $n \rightarrow \infty$ (see, e.g. Protter, 2004, pp. 66). For the process in Eq. (1), the quadratic variation is equal to integrated variance (IV hereafter), i.e.

$$[X]_t = \int_0^t \sigma_u^2 du.$$

As in Andersen and Bollerslev (1998), and many of the subsequent papers in this area, here the object of econometric interest is the IV.

2.1 The estimator and its properties

From now on, we will work on the unit time interval without loss of generality, i.e. $t \in [0, 1]$. We assume that X is observed at equidistant points $t_i = i/N$, for $i = 0, \dots, N$. The increments of X – the continuously compounded returns – are denoted as:

$$\Delta_i^N X = X_{i/N} - X_{(i-1)/N}, \quad (2)$$

for $i = 1, \dots, N$. We further assume that $N = nm$, where m, n are natural numbers. Specifically, we consider n contiguous subintervals or blocks $[(i-1)/n, i/n]$, each containing m returns, i.e.

$$D_i^m X = (\Delta_k^N X)_{(i-1)m+1 \leq k \leq im}, \quad (3)$$

for $i = 1, \dots, n$. The idea of building subintervals is quite natural in the current setting and the mathematical intuition for this approach is discussed in Mykland (2010). In the asymptotic analysis below, we concentrate on the case where m is fixed and $n \rightarrow \infty$, but also briefly comment on the case where n is fixed and $m \rightarrow \infty$, and $n, m \rightarrow \infty$.

Define the function $g_k : \mathbb{R}^m \rightarrow \mathbb{R}$ such that

$$g_k(x) = x_{(k)}, \quad (4)$$

where $x_{(k)}$ is the k th order statistic of $x = (x_1, \dots, x_m)$. Also define the realised (symmetric) squared λ -quantile on the i^{th} block as

$$q_i(m, \lambda) = g_{\lambda m}^2(\sqrt{N} D_i^m X) + g_{m-\lambda m+1}^2(\sqrt{N} D_i^m X), \quad (5)$$

where λm is a natural number. Note that the function $q_i(m, \lambda)$ is even in X , so its value does not change if we replace X by $-X$. Also note that $\sqrt{N} D_i^m X$ has been normalized to be $O_p(1)$.

We are now in a position to introduce the quantile-based realised variance (QRV, hereafter):

$$QRV_N(m, \bar{\lambda}, \alpha) \equiv \alpha' QRV_N(m, \bar{\lambda}), \quad (6)$$

where $\bar{\lambda} = (\lambda_1, \dots, \lambda_k)'$ is a $(k \times 1)$ vector of quantiles with $\lambda_j \in (1/2, 1)$ for $j = 1, \dots, k$, α is a $(k \times 1)$ non-negative weighting vector with $|\alpha|_1 = 1$, and $QRV_N(m, \bar{\lambda})$ is a $(k \times 1)$ vector with j th entry equal to

$$QRV_N(m, \lambda_j) = \frac{m}{N} \sum_{i=1}^{N/m} \frac{q_i(m, \lambda_j)}{\nu_1(m, \lambda_j)}. \quad (7)$$

The scaling factor ν in Eq. (7) is given by:

$$\nu_r(m, \lambda) = \mathbb{E} \left[\left(|U_{(\lambda m)}|^2 + |U_{(m-\lambda m+1)}|^2 \right)^r \right], \quad (8)$$

for $r > 0$, where $U_{(\lambda m)}$ is the (λm) -th order statistic of an independent standard normal sample $\{U_i\}_{i=1}^m$.

We show below that QRV constitutes a consistent and highly efficient estimator of the IV under very weak conditions on X . Moreover, due to its reliance on return quantiles (with $\lambda < 1$), the QRV is asymptotically robust to finite activity jumps. This holds even when $k = 1$ and the estimator is constructed using only a single pair of quantiles. Still, combining multiple pairs of quantiles as in Eq. (6) improves the efficiency of the QRV, and we can explicitly characterize the optimal quantile weights that minimize its asymptotic variance.

Two further remarks are in order. First, while it is possible to use asymmetric quantiles in Eq. (5), this is suboptimal in the current setting due to the symmetry of the normal distribution, and we therefore do not consider this case. Moreover, when $q_i(m, \lambda)$ is based on asymmetric quantiles, the resulting CLT becomes infeasible, because the function $q_i(m, \lambda)$ is not even in X anymore (see Kinnebrock and Podolskij, 2008, for further discussion). Second, to compute QRV the scaling factor $\nu_r(m, \lambda)$ is needed. These values can be obtained to any desired degree of accuracy by straightforward simulation or numerical integration using the joint density function of the order statistics:

$$f_{U_{(m-\lambda m+1)}, U_{(\lambda m)}}(x, y) = 1_{\{x < y\}} \frac{m!}{(m - \lambda m)!} \frac{(\Phi(y) - \Phi(x))^{2\lambda m - m - 2} ((1 - \Phi(y))\Phi(x))^{m - \lambda m}}{(2\lambda m - m - 2)!(m - \lambda m)!} \phi(x)\phi(y),$$

for $\lambda \in (1/2, 1)$, where ϕ and Φ are the standard normal density and distribution functions.

We now present the main limit results of the QRV.

Theorem 1 *For the process X in Eq. (1), and $N = mn$ with m fixed, as $N \rightarrow \infty$*

$$QRV_N(m, \bar{\lambda}, \alpha) \xrightarrow{p} IV,$$

where $IV = \int_0^1 \sigma_u^2 du$.

Proof see Appendix C ■

Theorem 2 *For the process X in Eq. (1), with condition (V) satisfied and $N = mn$ with m fixed, as $N \rightarrow \infty$*

$$\sqrt{N} (QRV_N(m, \bar{\lambda}, \alpha) - IV) \xrightarrow{d_s} \sqrt{\theta(m, \bar{\lambda}, \alpha)} \int_0^1 \sigma_u^2 dW'_u,$$

where

$$\theta(m, \bar{\lambda}, \alpha) = \alpha' \Theta(m, \bar{\lambda}) \alpha, \quad (9)$$

and the $k \times k$ matrix $\Theta(m, \bar{\lambda}) = (\Theta(m, \bar{\lambda})_{sl})_{1 \leq s, l \leq k}$ is given by

$$\Theta(m, \bar{\lambda})_{ij} = m \frac{\nu_1(m, \lambda_i, \lambda_j) - \nu_1(m, \lambda_i)\nu_1(m, \lambda_j)}{\nu_1(m, \lambda_i)\nu_1(m, \lambda_j)},$$

with

$$\nu_1(m, \lambda_i, \lambda_j) = \mathbb{E}[(|U_{(m\lambda_i)}|^2 + |U_{(m-m\lambda_i+1)}|^2)(|U_{(m\lambda_j)}|^2 + |U_{(m-m\lambda_j+1)}|^2)], \quad (10)$$

and where W' is another Brownian motion defined on an extension of $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ with $W' \perp\!\!\!\perp \mathcal{F}$. Because σ is independent of W' , this implies stable convergence to a mixed normal distribution:

$$\sqrt{N} (QRV_N(m, \bar{\lambda}, \alpha) - IV) \xrightarrow{d_s} MN(0, \theta(m, \bar{\lambda}, \alpha)IQ),$$

where $IQ = \int_0^1 \sigma_u^4 du$ is the integrated quarticity.

Proof see Appendix C ■

Theorem 1 and 2 show that QRV is a consistent estimator of the IV under very weak conditions on the process X and that it converges at rate $N^{-1/2}$, the best attainable in this setting. For a given block size m and quantiles $\bar{\lambda}$, the weighting vector α can be chosen optimally to minimize the asymptotic variance of QRV, i.e.

$$\alpha^* = \frac{\Theta^{-1}(m, \bar{\lambda})\iota}{\iota'\Theta^{-1}(m, \bar{\lambda})\iota}, \quad (11)$$

where ι is a $(k \times 1)$ vector of ones. The asymptotic efficiency of QRV is characterized by the constant θ in Eq. (9), which is equal to $\theta(m, \bar{\lambda}, \alpha^*) = (\iota'\Theta^{-1}(m, \bar{\lambda})\iota)^{-1}$ when optimal weights are used. In Section 2.4, we shall see that θ takes on values between 3 and 4 when using a single (pair of) quantile around 0.9 and rapidly approaches the parametric lower bound of 2 as multiple quantiles are used. Thus, QRV can attain the efficiency of the maximum likelihood estimator in the parametric no-jump version of this problem, while still retaining robustness to jumps, when they do appear in the price process. In fact, the consistency and CLT of the QRV are unaffected by the presence of finite activity jumps, as the following proposition highlights.

Proposition 1 *Theorem 1 and 2 remain valid for an extension of the process X in Eq. (1) that incorporates finite activity jumps, i.e.*

$$X_t = X_0 + \int_0^t a_u du + \int_0^t \sigma_u dW_u + \sum_{i=1}^{q(t)} J_i, \quad t \geq 0, \quad (12)$$

where $q = (q(t))_{t \geq 0}$ is a finite activity counting process and $J = (J_i)_{i=1}^{q(t)}$ are non-zero random variables representing the jumps in X .

The intuition for this result is clear: with m fixed and $N \rightarrow \infty$, the number of observations in each block remains constant but the time interval it spans shrinks so that, in the limit, it contains at most one jump with probability one. Combined with the restriction $\lambda < 1$, which ensures we leave out at least the largest negative and positive return when constructing the QRV, this naturally implies that the estimator is asymptotically immune to finite activity jumps.

To conclude this section, we point out that our quantile-based approach can also be used to estimate the integrated quarticity in a similar fashion. In particular, define

$$QRQ_N(m, \bar{\lambda}, \alpha) \equiv \alpha' QRQ_N(m, \bar{\lambda}), \quad (13)$$

where $\bar{\lambda} = (\lambda_1, \dots, \lambda_k)'$ is a $(k \times 1)$ vector of quantiles with $\lambda_j \in (1/2, 1)$ for $j = 1, \dots, k$, α is a $(k \times 1)$ non-negative weighting vector with $|\alpha|_1 = 1$, and $QRQ_N(m, \bar{\lambda})$ is a $(k \times 1)$ vector with j th entry equal to:

$$QRQ_N(m, \lambda_j) = \frac{1}{\nu^{iq}(m, \lambda_j)} \frac{m}{N} \sum_{i=1}^{N/m} \left(g_{\lambda_j m}^4 \left(\sqrt{N} D_i^m X \right) + g_{m-\lambda_j m+1}^4 \left(\sqrt{N} D_i^m X \right) \right), \quad (14)$$

with

$$\nu^{iq}(m, \lambda) = \mathbb{E} [|U_{(\lambda m)}|^4 + |U_{(m-\lambda m+1)}|^4].$$

As in Theorem 1 we have consistency for m fixed and $N \rightarrow \infty$, i.e. $QRQ_N(m, \bar{\lambda}, \alpha) \xrightarrow{p} IQ$. This estimator can now be used to formulate a feasible CLT for QRV as follows:

$$\sqrt{N} \frac{QRV_N(m, \bar{\lambda}, \alpha_{iv}) - IV}{\sqrt{\theta(m, \bar{\lambda}, \alpha_{iv}) QRQ_N(m, \bar{\lambda}, \alpha_{iq})}} \xrightarrow{d} N(0, 1).$$

2.2 A subsampling implementation of QRV

The QRV estimator developed above is constructed using empirical return quantiles computed over non-overlapping contiguous intervals of data. In this section, we discuss a subsampled version of the QRV, which is still more efficient than its blocked counterpart. In particular, we define:

$$QRV_N^{sub}(m, \bar{\lambda}, \alpha) \equiv \alpha' QRV_N^{sub}(m, \bar{\lambda}), \quad (15)$$

where α and $\bar{\lambda}$ are as above, and $QRV_N^{sub}(m, \bar{\lambda})$ is a $(k \times 1)$ vector with j th entry equal to:

$$QRV_N^{sub}(m, \lambda_j) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \frac{q_i^{sub}(m, \lambda_j)}{\nu_1(m, \lambda_j)}, \quad (16)$$

and

$$q_i^{sub}(m, \lambda) = g_{\lambda m}^2 \left(\sqrt{N} D_{i,m} X \right) + g_{m-\lambda m+1}^2 \left(\sqrt{N} D_{i,m} X \right),$$

where $D_{i,m} X = (\Delta_k^N X)_{i \leq k \leq i+m-1}$ for $i \geq 1$.

Because the estimator in Eq. (15) is basically a subsampled version of the blocked QRV in Eq. (6), it is also consistent for the IV by Theorem 1. Moreover, by a triangular inequality argument it is clear that subsampling improves the asymptotic efficiency.

Theorem 3 *For the process X in Eq. (1) with condition (V) satisfied and $N = mn$ with m fixed, as $N \rightarrow \infty$*

$$\sqrt{N} (QRV_N^{sub}(m, \bar{\lambda}, \alpha) - IV) \xrightarrow{d_s} MN(0, \theta^{sub}(m, \bar{\lambda}, \alpha) IQ), \quad (17)$$

where

$$\theta^{sub}(m, \bar{\lambda}, \alpha) = \alpha' \Theta^{sub}(m, \bar{\lambda}) \alpha, \quad (18)$$

and the $k \times k$ matrix $\Theta^{sub}(m, \bar{\lambda}) = (\Theta^{sub}(m, \bar{\lambda})_{sl})_{1 \leq s, l \leq k}$ is given by

$$\Theta^{sub}(m, \bar{\lambda})_{ij} = \frac{1}{m} \Theta(m, \bar{\lambda})_{ij} + \frac{2}{\nu_1(m, \lambda_i) \nu_1(m, \lambda_j)} \sum_{k=1}^m \text{cov} \left(|U_{(m\lambda_i)}^{(0)}|^2 + |U_{(m-m\lambda_i+1)}^{(0)}|^2, |U_{(m\lambda_j)}^{(k)}|^2 + |U_{(m-m\lambda_j+1)}^{(k)}|^2 \right),$$

where $U^{(0)} = \{U_i\}_{i=1}^m$, $U^{(k)} = \{U_i\}_{i=1+k}^{m+k}$ and $\{U_i\}_{i=1}^{m+k}$ is an independent standard normal sample. Furthermore, the convergence in Eq. (17) is robust to finite activity jumps, i.e. it also holds for the processes defined in Eq. (12).

Proof see Appendix C ■

As before, asymptotically optimal weights can be assigned to the quantiles so as to minimize $\theta^{sub}(m, \bar{\lambda}, \alpha)$. In Section 2.4, we illustrate the efficiency improvement that results from subsampling.

2.3 QRV with $m \rightarrow \infty$

Up to this point, we have considered the case where m is fixed and $n \rightarrow \infty$. In this limit, QRV is consistent under very weak assumptions on the log-price X . We now briefly discuss the case where $m \rightarrow \infty$. To get the corresponding asymptotic results in this limit, much stronger assumptions need to be imposed on X . In particular, when n is fixed a sufficient condition for Theorem 1 and 2 to hold is that σ is constant (this follows directly from classical order statistic results, see for instance David, 1970). In this case, the asymptotic constants are as given in the following proposition.

Proposition 2 *We have*

$$\begin{aligned}\nu_1(\lambda) &\equiv \lim_{m \rightarrow \infty} \nu_1(m, \lambda) = 2c_\lambda^2, \\ \nu_1(\lambda_i, \lambda_j) &\equiv \lim_{m \rightarrow \infty} \nu_1(m, \lambda_i, \lambda_j) = 4c_{\lambda_i}^2 c_{\lambda_j}^2, \\ \Theta(\bar{\lambda})_{ij} &\equiv \lim_{m \rightarrow \infty} \Theta(m, \bar{\lambda})_{ij} = \lim_{m \rightarrow \infty} \Theta^{sub}(m, \bar{\lambda})_{ij} = 2 \frac{(1 - \lambda_j)(2\lambda_i - 1)}{\phi(c_{\lambda_i})\phi(c_{\lambda_j})c_{\lambda_i}c_{\lambda_j}}, \\ \theta(\bar{\lambda}, \alpha) &\equiv \lim_{m \rightarrow \infty} \theta(m, \bar{\lambda}, \alpha) = \lim_{m \rightarrow \infty} \theta^{sub}(m, \bar{\lambda}, \alpha) = \alpha' \Theta(\bar{\lambda}) \alpha,\end{aligned}$$

with $\lambda_i \leq \lambda_j$, where c_α and ϕ denote the α -quantile and density function of the standard normal distribution.

Proof see Appendix C ■

These results are interesting for a number of reasons. Firstly, when $m \rightarrow \infty$ the asymptotic covariance matrix Θ can be expressed in closed form, and it is identical for the blocking and subsampling version of the QRV. This, in turn, allows for fast and easy calculation of optimal quantile weights α^* . In Section 2.4, we demonstrate that the use of these weights leads to only a very limited efficiency loss – even for small m – which makes it attractive from a practical point of view. Secondly, in certain applications the constant volatility assumption can sometimes be justified (e.g. when sampling over a short horizon or on a suitably deformed time scale), and a single block implementation of QRV (i.e. $n = 1$ and m large) may be preferred purely for the sake of computational simplicity. Interestingly, we will see in Section 2.4 that the efficiency loss associated with using only a single block is small, particularly when combining multiple quantiles. Finally, the above limit allows us to clarify the relation between the QRV and some related estimators that

have appeared in the literature. In particular, with $m \rightarrow \infty$, $n = 1$, and constant volatility, QRV corresponds to the estimator of David (1970), the only difference being that the latter estimates the standard deviation instead of the variance. Also, in Appendix B we discuss an alternative formulation of QRV based on absolute returns. This estimator nests the MinRV and MedRV of Andersen, Dobrev, and Schaumburg (2008), and we use the $m \rightarrow \infty$ limit to show its equivalence to the QRV based on signed returns introduced above.

As an aside, if we want to relax the constant volatility assumption above, it is still possible to let $m \rightarrow \infty$, but then we also need $n \rightarrow \infty$. Moreover, stronger assumptions on the dynamics of X are needed: for consistency we require condition (V) and $m/n \rightarrow 0$, whereas to obtain a CLT we additionally need a , σ' and v' to satisfy condition (V) and $m^3/n \rightarrow 0$. We then have the following result:

$$QRV_N(m, \bar{\lambda}, \alpha) \xrightarrow{P} IV,$$

and

$$\sqrt{N} (QRV_N(m, \bar{\lambda}, \alpha) - IV) \xrightarrow{d_s} MN(0, \theta(\bar{\lambda}, \alpha)IQ).$$

A formal proof of these results can be found in a separate appendix, i.e. see Christensen, Oomen, and Podolskij (2010).

2.4 Implementation and asymptotic efficiency of QRV

Implementing the QRV requires selection of quantiles $\bar{\lambda}$, quantile weights α , and block-length m . The asymptotic theory developed above can be used to determine the optimal quantile weights but, as we will now discuss, it can also be used to guide the choice of $\bar{\lambda}$ and m . We will also use the theory to compare the efficiency of QRV to the leading alternative realised variance measures. Section 2.5 will follow up this discussion with simulations investigating QRV's finite sample performance and robustness to jumps and outliers.

Table 1 reports the asymptotic efficiency constant θ of the blocked and subsampled QRV estimator for different m and $\bar{\lambda}$. A number of important observations regarding the efficiency and the preferred implementation of the estimator can be made.

First, considering Panels A and B, we see that QRV is a highly efficient estimator particularly when using multiple quantiles. For instance, when combining the five quantiles listed in Panel A, QRV has an asymptotic efficiency of around 2.2. This compares favorably to the leading jump-robust bi-power variation measure of Barndorff-Nielsen and Shephard (2004) for which the corresponding figure is $\pi^2/4 + \pi - 3 \simeq 2.61$. Moreover, by including additional quantiles in the construction of QRV we can push its efficiency arbitrarily close to 2, so that in the limit it attains the ML efficiency of realised variance (Jacod, 1994; Jacod and Protter, 1998; Barndorff-Nielsen and Shephard, 2002) while still retaining robustness to jumps. Comparing the blocked QRV with the subsampled QRV

Table 1: Asymptotic efficiency of QRV with single and multiple quantiles

$\bar{\lambda}$	\	m	blocked QRV			subsampled QRV			∞
			20	40	100	20	40	100	
<i>Panel A: single quantile</i>									
0.80		4.24	4.29	4.31		3.54	3.73	3.92	4.32
0.85		3.56	3.58	3.59		3.02	3.14	3.27	3.60
0.90		3.10	3.14	3.15		2.67	2.75	2.86	3.16
0.95		2.88	2.99	3.07		2.52	2.62	2.75	3.13
0.98		—	—	3.58		—	—	3.16	3.88
<i>Panel B: multiple quantiles with optimal weights $\alpha_{(m)}^*$</i>									
0.80 – 0.95		2.40	2.41	2.42		2.27	2.29	2.32	2.42
0.80 – 0.98		—	—	2.19		—	—	2.13	2.19
<i>Panel C: multiple quantiles with asymptotically optimal weights $\alpha_{(\infty)}^*$</i>									
0.80 – 0.95		2.41	2.41	2.42		2.31	2.32	2.33	2.42
0.80 – 0.98		—	—	2.19		—	—	2.14	2.19

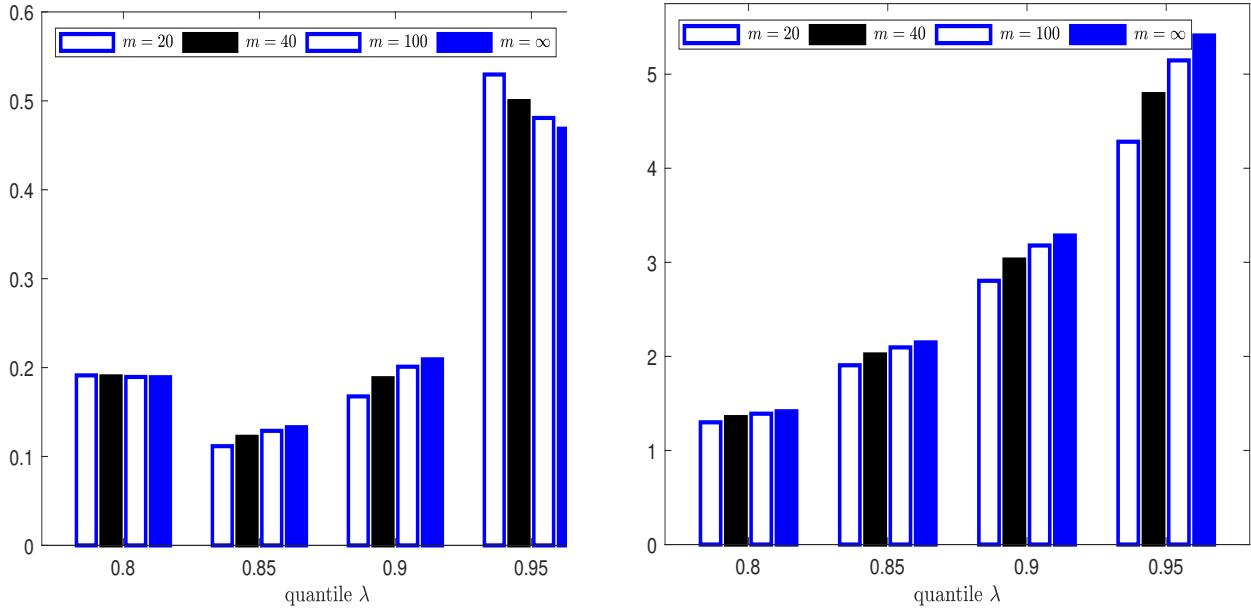
Note. This table reports the asymptotic efficiency constants $\theta(m, \bar{\lambda}, \alpha)$ for the blocked QRV and $\theta^{sub}(m, \bar{\lambda}, \alpha)$ for the subsampled QRV as given in Eqs. (9) and (18) for different values of m . The last column reports the limiting value $\theta(\bar{\lambda}, \alpha)$ as in Proposition 2 (which is the same for the blocked and subsampled QRV). Panel A reports the results for a selection of single quantiles. Panels B and C combine these quantiles using exact “finite m ” optimal weights following Theorems 2 and 3 and asymptotically optimal weights following Proposition 2 respectively.

we confirm an efficiency gain associated with subsampling, albeit the benefit is modest particularly when using multiple quantiles.²

Second, comparing Panels B and C, we see that the use of limiting “ $m \rightarrow \infty$ ” optimal quantile weights following Proposition 2 instead of exact “finite- m ” optimal weights following Theorem 2 and 3 leads to only a marginal deterioration in efficiency. This is very attractive from a practical viewpoint, for it means that we can simply use the limiting closed-form expression for Θ to obtain reliable and near-optimal quantile weights instead of calculating its exact finite-sample counterpart for each and every m and $\bar{\lambda}$ we may consider. As an aside, we point out that the use of asymptotic scaling factors ν_1 should be avoided as it will induce potentially serious biases in the estimator even

²The efficiency gain associated with subsampling can be very substantial when we consider higher powers of quantiles, for instance when estimating integrated quarticity. Also with microstructure noise, the efficiency gain is of a factor 2 – 3. Because we start Section 3 immediately with a subsampled version of our noise robust-estimator this is not explicitly discussed.

Figure 1: Optimal quantile weights and scaling factors for varying block size m .



Note. This figure reports the optimal quantile weights α (Panel A) as in Eq. (11) and scaling factors ν_1 (Panel B) as in Eq. (8) for QRV using blocking with $\bar{\lambda} = \{0.80, 0.85, 0.90, 0.95\}$ and varying m .

for moderately large m . Figure 1 further illustrates these effects by plotting the optimal quantile weights and scaling factors for different values of m .

Third, from Panel A we observe that the choice of quantile can be informed by efficiency considerations. It is quite intuitive that quantiles close to the mode of the return distribution are not very informative about the variance of the process. At the same time, quantiles far into the tail region tend to be erratic. The optimal choice of quantile balances this tradeoff to extract the maximum amount of information regarding the spread of the distribution. From Panel A we see that, depending on the choice of m , the optimal quantile lies in the region $0.90 - 0.95$. Of course, with $k > 1$, quantiles outside this region may be added to exploit the covariance structure of the order statistics. On the choice of block length, we see that there are modest efficiency gains to be had by choosing it small (it can be shown that θ is monotonically increasing in m). However, when multiple quantiles are used, the gain is negligible.

From the above discussion it is clear that the asymptotic theory is very helpful in guiding the choice of quantiles and block length. In particular, we can determine the optimal (mix of) quantiles and their weights by maximizing asymptotic efficiency of the resulting estimator. In addition, the theory indicates a weak preference for choosing a small m and the use of subsampling. In practice, there are two other important considerations in addition to asymptotic efficiency, namely (i) QRV's ability to estimate the IV in finite sample and (ii) QRV's robustness to jumps and outliers. For QRV to accurately measure the IV, σ is required to be locally constant on each subinterval. Thus,

also from this viewpoint, a small m or equivalently a large n is desirable. On the other hand, jump robustness is controlled by the *joint* choice of m and $\lambda_{max} = \max\{\bar{\lambda}\}$: over a block of length m , QRV is robust to $(1 - \lambda_{max})m - 1$ positive and negative jumps. These observations suggest that the optimal choice of $\bar{\lambda}$ and m should satisfy the following three conditions (i) m is sufficiently small to ensure good “locality” of the estimator, (ii) λ includes a quantile in the highly informative tail region, and (iii) $(1 - \lambda_{max})m$ is sufficiently large to ensure robustness against jumps and outliers. Sections 2.5 and 4 will further illustrate the above using simulations and empirical analysis.

2.5 Finite sample performance and jump robustness

The results so far indicate that the asymptotic efficiency of QRV is excellent. The simulations below are designed to gauge the finite sample performance of the estimator. We pay particular attention to bias, efficiency, and robustness to jumps and outliers. We also compare the performance of the QRV to recently developed alternative estimators.

To simulate the log-price X , we adopt the following model:

$$dX_t = \sigma_t dW_t, \quad t \in [0, 1], \quad (19)$$

where W is a standard Brownian motion and the dynamics of σ_t are as specified below. The baseline scenario is a constant volatility Brownian motion (“BM”), i.e.

$$\sigma_t^2 = 0.0391. \quad (20)$$

To assess QRV’s ability to handle time-varying volatility, we use a Heston-type stochastic volatility (“SV”) model

$$d\sigma_t^2 = (0.3141 - 8.0369\sigma_t^2)dt + \sigma_t \sqrt{0.1827} dB_t, \quad (21)$$

where B is another Brownian motion with $B \perp\!\!\!\perp W$. To gauge the impact of leverage, we also simulate from Eq. (21) with $dW_t dB_t = -0.75dt$ (“SV-LEV”). Finally, we consider two more variance specifications that are both capable of generating erratic and highly volatile sample paths. The first is a model proposed by Aït-Sahalia (1996) that incorporates stochastic elasticity of variance and non-linear drift (“SEV-ND”), i.e.

$$d\sigma_t^2 = (-0.554 + 21.32\sigma_t^2 - 209.3\sigma_t^4 + 0.005\sigma_t^{-2})dt + \sqrt{0.017\sigma_t^2 + 53.97\sigma_t^{5.76}} dB_t. \quad (22)$$

with $B \perp\!\!\!\perp W$. The second is a two-factor stochastic volatility model (“SV2F-LEV”) analyzed in Chernov, Gallant, Ghysels, and Tauchen (2003), i.e.

$$\begin{aligned} \sigma_t^2 &= s \cdot \exp(-1.2 + 0.04f_t^{(1)} + 1.5f_t^{(2)}), \\ df_t^{(1)} &= -0.000137f_t^{(1)}dt + dB_t^{(1)}, \\ df_t^{(2)} &= -1.386f_t^{(2)}dt + (1 + 0.25f_t^{(2)})dB_t^{(2)}, \end{aligned} \quad (23)$$

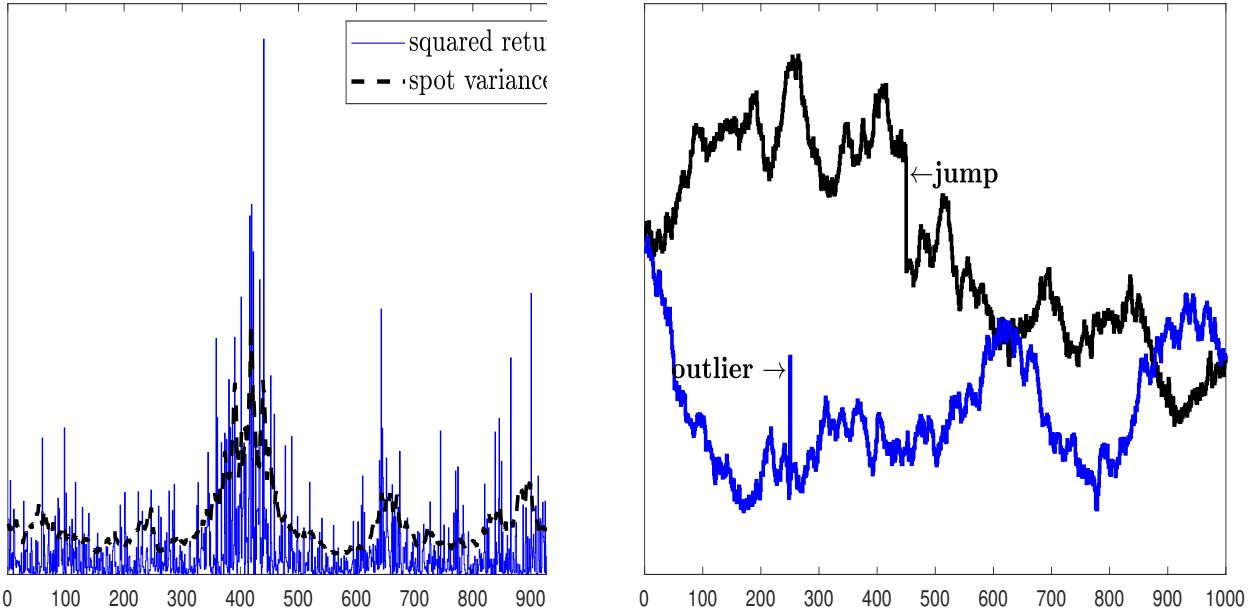
where $dW_t dB_t^{(1)} = dW_t dB_t^{(2)} = -0.3dt$ and $s\text{-exp}$ denotes a “spliced” exponential function as specified and discussed in Chernov, Gallant, Ghysels, and Tauchen (2003).

The above stochastic volatility models cover a wide range of dynamic specifications and thus provide a good testing ground for QRV. The parameter values for the BM, SV, and SEV-ND models in Eqs. (20–22) are taken from the empirical study by Bakshi, Ju, and Ou-Yang (2006) whereas the parameters for the SV2F-LEV model in Eq. (23) are taken from Huang and Tauchen (2005). It should be noted that while these studies typically calibrate the models from daily data to an annual horizon, here we simulate the processes over the unit interval so that, effectively, we compress a year’s worth of variation into a single day. As a result, we end up simulating highly erratic volatility paths, which serves to challenge the QRV estimator to the extreme. As an illustration, see Panel A of Figure 2 for a simulated return and variance series.

To study the robustness of QRV, we also simulate from the BM model and add jumps. In particular, we add a fixed number of n_J Gaussian jumps at random points in the sample with a combined jump variation v_J measured as a fraction of the IV. We consider four scenarios: $\{n_J, v_J\} = \{1, \frac{1}{4}\}$, i.e. one large jump accounting for 20% of total variation, $\{n_J, v_J\} = \{5, \frac{1}{4}\}$, i.e. five medium jumps accounting for 20% of total variation, $\{n_J, v_J\} = \{10, \frac{1}{4}\}$, i.e. ten small jumps accounting for 20% of total variation, and $\{n_J, v_J\} = \{5, \frac{1}{2}\}$, i.e. five large jumps accounting for a third of total variation. Additionally, we consider a scenario where the price series is contaminated by “outliers”. Such spurious and deviant price observations are commonly encountered in high-frequency data due to, for instance, delayed trade reporting, misplaced decimal points, data errors, etc. (see Section 4 for some examples). In our simulations, we position a single outlier of random size at a random point in the series, ensuring that its variation accounts for 20% of total variation. As an illustration, see Panel B of Figure 2 for a simulated price series with a jump and outlier added.

To simulate the process in Eq. (19), we use an Euler discretization scheme and set $N = 1,000$. QRV is computed as in Eq. (6) and (15) using four pairs of quantiles $\bar{\lambda} = \{0.80, 0.85; 0.90; 0.95\}$, asymptotic weights derived from Proposition 2, and three different choices of block length, namely $m = \{20, 40, 100\}$ or equivalently $n = \{50, 25, 10\}$. For comparison, we also compute realised variance (RV) in addition to three recently proposed jump-robust estimators, i.e. bi-power variation (BPV) of Barndorff-Nielsen and Shephard (2004), threshold realised variance (TRV) of Jacod (2008);

Figure 2: SV and jump simulation.



Note. Panel A plots a time-series of $N = 1,000$ squared returns simulated from the SEV-ND model over the unit interval. Panel B plots a time-series of $N = 1,000$ prices simulated from the BM+jump and BM+outlier model over the unit time interval.

Mancini (2004a, 2009), and MedRV of Andersen, Dobrev, and Schaumburg (2008):

$$\begin{aligned}
 RV_N &= \sum_{i=1}^N |\Delta_i^N X|^2, \\
 BPV_N &= \frac{\pi}{2} \sum_{i=2}^N |\Delta_i^N X| |\Delta_{i-1}^N X|, \\
 TRV_N &= \sum_{i=1}^N |\Delta_i^N X|^2 \mathbb{1}_{\{|\Delta_i^N X| < cN^{-\bar{\omega}}\}}, \quad \text{for } \bar{\omega} \in (0, 1/2), \\
 MedRV_N &= \frac{\pi}{6 - 4\sqrt{3} + \pi} \frac{N}{N-2} \sum_{i=2}^{N-1} \text{median}(|\Delta_{i-1}^N X|, |\Delta_i^N X|, |\Delta_{i+1}^N X|)^2.
 \end{aligned}$$

To implement TRV, we set $\bar{\omega} = 0.47$ and $c = 6\sqrt{IV}$, where IV is estimated using BPV. This parameter choice is in line with Aït-Sahalia and Jacod (2009a) and ensures that, in our simulation setup, TRV is unbiased in the absence of jumps (alternatively, we could have lowered c , gaining robustness to jumps but introducing a downward bias under SV).

Over 100,000 independent simulation runs, we compute a “bias” measure $\mathbb{E}(\widehat{IV}/IV)$ and an “efficiency” measure $\text{var}(\sqrt{N}(\widehat{IV} - IV)/\sqrt{IQ})$, where $\widehat{IV} = \{QRV_N, RV_N, BPV_N, TRV_N, MedRV_N\}$. If the estimator is unbiased we expect the bias statistic to be one. Moreover, from the relevant asymptotic results we know the efficiency statistic should be 2 for RV and TRV, 2.6 for BPV, 3.0

for MedRV, and around 2.4 and 2.3 for our implementation of the blocked and subsampled QRV, respectively.

From the results in Table 2 several interesting patterns emerge. First consider the scenarios without jumps. With model BM, all estimators perform as expected. They are unbiased and their efficiency measure is close to what the asymptotic distribution theory predicts, indicating that it affords a good approximation to finite sample performance. When introducing stochastic volatility through model SV, we find that QRV is biased downwards when few blocks are selected. However, this bias is small for $m = 100 / n = 10$ and negligible for $m = 20 / n = 50$. Leverage (SV-LEV) does not have a noticeable impact on any of the results. Under the SEV-ND and SV2F-LEV variance specifications, both generating high volatility-of-volatility, the QRV estimator still performs well provided that a sufficient number of subintervals are selected. With a highly erratic volatility path, we need to use large n or small m to ensure good locality of the estimator in line with the discussion above. Finally, comparing the blocked QRV to the subsampled QRV, we see that they perform similarly in terms of bias but that the efficiency of the subsampled QRV deteriorates when m increases. The intuition for this is that the subsampling procedure places less weight on the observations in the first and last block than it places on all other observations. In the asymptotic analysis this effect disappears as $n \rightarrow \infty$, but in finite sample it can adversely affect the efficiency of the estimator, particularly in the presence of stochastic volatility. Importantly, the blocking implementation of QRV does not suffer from this and may thus be preferred in situations where n is relatively small.

Next, consider the scenarios with jumps. As expected, QRV enjoys superior robustness to jumps. The small bias we observe (not exceeding 4% across all scenarios considered) can be explained by noting that the jumps added to the process distort the original ordering of the diffusive returns. This in turn biases the empirical return quantiles used to construct QRV.³ Importantly, however, this effect is largely independent of the jump size so that QRV maintains excellent robustness in finite sample. Also, with outliers simulated as described above, we see that QRV is virtually unaffected both in terms of bias and in terms of efficiency. Turning to the benchmark estimators, it is well known that in the current setting RV estimates total variation, i.e. $(1 + v_J)IV$, explaining the bias and low efficiency when evaluated against IV. BPV is asymptotically immune to jumps, but biased in finite sample: for the BMJ model considered here $\mathbb{E}(BPV_N/IV) \simeq 1 + 2\sqrt{n_J v_J/N}$ which can be substantial when jumps become more frequent or volatile. Also, with an outlier in the price series –

³To further clarify intuition for this, consider the following example. Suppose we have a ranked sequence of diffusive returns $\{r_{(1)}, r_{(2)}, \dots, r_{(m)}\}$ from the BM model. With $\lambda = 0.95$ we can estimate IV unbiasedly using $r_{(5)}$ and $r_{(95)}$ as described above. Now suppose a positive jump J is added to, say, $r_{(60)}$. If the jump is sufficiently large, the ordered returns sequence becomes $\{r_{(1)}, r_{(2)}, \dots, r_{(59)}, r_{(61)}, \dots, r_{(m)}, r_{(60)} + J\}$ and the “realised” $\lambda = 0.95$ quantile is now $r_{(96)}$. Thus, as jumps are added to the price process, the original ordering of returns can be disrupted, leading to a small upward bias in QRV. This bias, however, is only weakly related to the size of the jump.

effectively constituting two consecutive jumps of opposite sign – the key assumption underlying the robustness of BPV and MedRV is violated leaving both estimators severely biased. Finally, TRV’s performance with large jumps and outliers is comparable to that of QRV, but deteriorates when jumps become smaller and more frequent. TRV requires one to specify a uniform threshold for the sample which makes it sensitive to small jumps if it is set too high and sensitive to stochastic volatility if it is set too low. In this respect, a nice feature of QRV is that it effectively sets a block specific “threshold” which naturally adapts to the magnitude of returns through the choice of quantiles.

3 QRV with market microstructure noise

It has long been recognized that market microstructure effects in high-frequency data – such a bid-ask bounce and non-synchronous trading – distort the statistical properties of returns (e.g. Epps, 1979; Fisher, 1966; Niederhoffer and Osborne, 1966) and are detrimental to RV as an estimator of the IV, see, e.g., Zhou (1996). In this section, we develop a modified version of the QRV that is robust to noise and delivers consistent estimates of the IV.

On a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, we consider the noisy diffusion model

$$Y_{i/N} = X_{i/N} + u_{i/N}, \quad (24)$$

for $i = 0, 1, \dots, N$. Here, the “efficient” price X is a Brownian semimartingale as in Eq. (1). The microstructure noise u is an i.i.d. process, independent of X , with

$$\mathbb{E}(u_{i/N}) = 0, \quad \mathbb{E}(u_{i/N}^2) = \omega^2.$$

The process Y in Eq. (24) is constructed as follows. Suppose that X is defined on a filtered probability space $(\Omega^0, \mathcal{F}^0, (\mathcal{F}_t^0)_{t \geq 0}, \mathbb{P}^0)$. We define a second probability space $(\Omega^1, \mathcal{F}^1, (\mathcal{F}_t^1)_{t \geq 0}, \mathbb{P}^1)$, where Ω^1 denotes $\mathbb{R}^{[0,1]}$ and \mathcal{F}^1 the product Borel- σ -field on Ω^1 . Next, let \mathbb{Q} be a probability measure on \mathbb{R} (the marginal law of u). For any $t \geq 0$, $\mathbb{P}_t^1 = \mathbb{Q}$ and \mathbb{P}^1 denotes the product $\otimes_{t \in [0,1]} \mathbb{P}_t^1$. The filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, on which we define the process Y , is given as

$$\left. \begin{aligned} \Omega &= \Omega^0 \times \Omega^1, & \mathcal{F} &= \mathcal{F}^0 \times \mathcal{F}^1, & \mathcal{F}_t &= \bigcap_{s > t} \mathcal{F}_s^0 \times \mathcal{F}_s^1, \\ \mathbb{P} &= \mathbb{P}^0 \otimes \mathbb{P}^1. \end{aligned} \right\}$$

The i.i.d. assumption on u is a natural starting point to analyze the noise case and is widely used in the literature. Moreover, it has some empirical support at moderate sampling frequencies (see, e.g., Hansen and Lunde, 2006; Diebold and Strasser, 2013, for a further discussion of this assumption). As before, the object of econometric interest is the IV of X , with the additional challenge that inference is now based on noisy high-frequency data.

The modified QRV measure we develop below uses pre-averaged data, which gives it robustness to noise. A similar approach to noise reduction is studied in Podolskij and Vetter (2009) and applied to the BPV estimator. Although the main contribution of this paper is the development of QRV as a highly efficient jump- and outlier-robust measure of the IV, we point out that the treatment of our estimator in the presence of noise goes well beyond Podolskij and Vetter (2009). First, the procedure developed here makes much more efficient use of the data. Specifically, in a constant volatility setting, the asymptotic variance of QRV can be as low as $8.5\sigma^3\omega$, which is very close to the lower bound of the ML estimator under parametric assumptions and substantially more efficient than the estimator used by Podolskij and Vetter (2009), which has an asymptotic variance of around $26\sigma^3\omega$. Second, our asymptotic theory holds under much weaker assumptions on the noise distribution. Third, we prove that the CLT of the noise-corrected QRV is robust to finite activity jumps, which is something that is not possible for the BPV. Fourth, even though there is no explicit formula available for the conditional variance in the CLT, we provide a consistent estimator of this quantity, which permits a feasible CLT.

3.1 Construction of the estimator

Choose a natural number $K = K(N)$ with

$$K = cN^{1/2} + o(N^{1/4}), \quad (25)$$

for some constant $c > 0$, and consider a weight function h on $[0, 1]$, which is continuous, piecewise continuously differentiable having a piecewise Lipschitz derivative h' with $h(0) = h(1) = 0$ and that satisfies $\int_0^1 h^2(s)ds > 0$. A typical example, that is used in our simulations in Section 3.3, is $h(x) = x \wedge (1 - x)$.

Define the return-like statistic

$$\bar{Y}_j^N = \sum_{i=1}^{K-1} h\left(\frac{i}{K}\right) \Delta_{j+i}^N Y, \quad (26)$$

and also set $\psi_1 = \int_0^1 (h'(x))^2 dx$ and $\psi_2 = \int_0^1 h^2(x)dx$.

Remark 1 In practice, it is better to use the Riemann approximations

$$\psi_1^n = K \sum_{j=1}^K \left(h\left(\frac{j}{K}\right) - h\left(\frac{j-1}{K}\right) \right)^2, \quad \psi_2^n = \frac{1}{K} \sum_{j=1}^{K-1} h^2\left(\frac{j}{K}\right)$$

of ψ_1 and ψ_2 to improve the finite sample properties, because ψ_1^n and ψ_2^n are the "true" constants that appear in the computations.

Next, select a sub-sequence using data observed in the interval $[i/N, (i + m(K - 1))/N]$:

$$\overline{D}_i^N Y = \{\overline{Y}_{i+(j-1)(K-1)}^N\}_{j=1}^m, \quad \text{for } i = 0, 1, \dots, N - m(K - 1)$$

and compute

$$q_i^*(m, \lambda) = g_{\lambda m}^2 \left(N^{1/4} \overline{D}_i^N Y \right) + g_{m-\lambda m+1}^2 \left(N^{1/4} \overline{D}_i^N Y \right).$$

The noise-corrected QRV measure (QRV* hereafter) is now defined as:

$$QRV_N^*(m, \bar{\lambda}, \alpha) \equiv \alpha' QRV_N^*(m, \bar{\lambda}), \quad (27)$$

where $\bar{\lambda}$ and α are as above, and the j th element of $QRV_N^*(m, \bar{\lambda})$ is given by:

$$QRV_N^*(m, \lambda_j) = \frac{1}{c\psi_2(N - m(K - 1) + 1)} \sum_{i=0}^{N-m(K-1)} \frac{q_i^*(m, \lambda_j)}{\nu_1(m, \lambda_j)}. \quad (28)$$

Note that the constant K controls the stochastic order of the term \overline{Y}_j^N , since

$$\overline{u}_j^N = O_p \left(\sqrt{\frac{1}{K}} \right), \quad \overline{X}_j^N = O_p \left(\sqrt{\frac{K}{N}} \right). \quad (29)$$

Thus, when K is chosen as in Eq. (25) the stochastic orders of the quantities in Eq. (29) are balanced (this implies the best rate of convergence), and under mild conditions we have that

$$N^{1/4} \overline{Y}_j^N | \mathcal{F}_{j/N} \stackrel{a}{\sim} N \left(0, c\psi_2 \sigma_{j/N}^2 + \frac{\psi_1}{c} \omega^2 \right).$$

This demonstrates the rationale of the filtering procedure underlying the construction of QRV*, namely while $N^{1/4} \overline{Y}_j^N$ is affected by the noise through ω^2 , it behaves like $\sqrt{N}(X_{i/N} - X_{(i-1)/N})$.

3.2 Asymptotic properties

Our first result shows the consistency of QRV* (after a proper bias correction).

Theorem 4 *Assume that m is a fixed number and $\mathbb{E}(u_i^4) < \infty$. As $N \rightarrow \infty$, it holds that*

$$QRV_N^*(m, \bar{\lambda}, \alpha) - \frac{\psi_1}{c^2 \psi_2} \omega^2 \xrightarrow{p} IV.$$

Proof see Appendix C ■

In practice, we can form consistent estimates of ω^2 , e.g. $\widehat{\omega}^2 = \frac{1}{2N} \sum_{i=1}^N |Y_{i/N} - Y_{(i-1)/N}|^2$ as in Bandi and Russell (2006), $\widehat{\omega}^2 = -\frac{1}{N-1} \sum_{i=1}^{N-1} (Y_{(i+1)/N} - Y_{i/N})(Y_{i/N} - Y_{(i-1)/N})$ as in Oomen (2006), or with the parametric MA(1)-based maximum likelihood estimator of Aït-Sahalia, Mykland, and Zhang (2005). As a consequence, we have the convergence

$$QRV_N^*(m, \bar{\lambda}, \alpha) - \frac{\psi_1}{c^2 \psi_2} \widehat{\omega}^2 \xrightarrow{p} IV.$$

This result is robust to the presence of finite activity jumps. Also note that because $\widehat{\omega}^2$ is a \sqrt{N} -estimator of ω^2 , it will not influence the CLT of the slower converging QRV*.

To prove the CLT, it is useful to introduce some further notation.

Definition 2 For $x \in \mathbb{R}$, $u \in [0, 1]$, $l = 1, \dots, m$ and λ_1, λ_2 we define the quantity

$$f_{m,l,x,u}(\lambda_1, \lambda_2) = \text{cov}\left(g_{\lambda_1 m}^2(S) + g_{m-\lambda_1 m+1}^2(S), g_{\lambda_2 m}^2(T) + g_{m-\lambda_2 m+1}^2(T)\right), \quad (30)$$

where $S = (S_1, \dots, S_m)^T$, $T = (T_1, \dots, T_m)^T$ are centered and jointly normal with

- (i) $S_i \perp S_j$, $T_i \perp T_j$ for all $i \neq j$.
- (ii) $\text{var}(S_i) = \text{var}(T_i) = c\psi_2 x^2 + \frac{\psi_1}{c}\omega^2$ for all i .
- (iii) $\text{cov}(S_{i+l-1}, T_i) = cw_h(u)x^2 + \frac{1}{c}w_h'(u)\omega^2$ for all i .
- (iv) $\text{cov}(S_{i+l}, T_i) = cw_h(1-u)x^2 + \frac{1}{c}w_h'(1-u)\omega^2$ for all i .
- (v) $\text{cov}(S_i, T_j) = 0$ for all $|i+l-j-1| > 1$.

Here the function $w_h(u)$ is defined by

$$w_h(u) = \int_0^{1-u} h(y)h(y+u)dy.$$

When $\lambda = \lambda_1 = \lambda_2$ we use the notation $f_{m,l,x,u}(\lambda) = f_{m,l,x,u}(\lambda_1, \lambda_2)$.

Notice that h' (the derivative of h) exists almost everywhere, so the quantity w_h' makes sense.

Theorem 5 Assume that m is a fixed number, $\mathbb{E}(u_i^8) < \infty$, the marginal distribution \mathbb{Q} of u is symmetric around 0 and that condition (V) is satisfied. As $N \rightarrow \infty$

$$N^{1/4} \left(QRV_N^*(m, \bar{\lambda}, \alpha) - \frac{\psi_1}{c^2\psi_2} \widehat{\omega}^2 - IV \right) \xrightarrow{d} MN\left(0, \frac{2}{c\psi_2^2} \Sigma_m(\lambda_1, \dots, \lambda_k)\right),$$

where $\Sigma_m(\lambda_1, \dots, \lambda_k) = (\Sigma_m(\lambda_1, \dots, \lambda_k)_{sl})_{1 \leq s, l \leq k}$ is given by

$$\Sigma_m(\lambda_1, \dots, \lambda_k)_{sl} = \frac{1}{\nu_{1,m}(\lambda_s)\nu_{1,m}(\lambda_l)} \sum_{l=1}^m \int_0^1 \int_0^1 f_{m,l,\sigma_t,u}(\lambda_s, \lambda_l) dt du.$$

Furthermore, this convergence is robust to the presence of finite activity jumps. ■

Proof see Appendix C

A couple of points are worth highlighting. First, the rate of convergence in Theorem 5 is $N^{-1/4}$, which is known to be optimal in the noisy diffusion model (Gloter and Jacod, 2001a,b). Second, Jacod, Li, Mykland, Podolskij, and Vetter (2009) show that when the IV is estimated using a “sum-of-squares” estimator based on filtered data:

$$\frac{1}{c\psi_2(N-K)} \sum_{j=1}^{N-K} |\bar{Y}_j^N|^2,$$

the lowest attainable variance for the choice $h(x) = x \wedge (1-x)$ is roughly $8.5\sigma^3\omega$ assuming σ is constant (the variance of the ML estimator is $8\sigma^3\omega$). Consequently, the lower bound for the variance of QRV* is also $8.5\sigma^3\omega$ (note that for a suitable choice of parameters the realised kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) can attain a variance of $8.002\sigma^3\omega$). Finally, even though there is no explicit expression for the conditional variance in the CLT, it is nonetheless possible to estimate it from the data.

Proposition 3 *Assume that m is fixed and $\mathbb{E}(u_i^8) < \infty$. As $N \rightarrow \infty$, it holds that*

$$\begin{aligned} \frac{1}{c\psi_2^2(K-1)(N-3m(K-1)+3)} \sum_{i=m(K-1)-1}^{N-2m(K-1)+1} \frac{q_i^*(m, \lambda_s)}{\nu_1(m, \lambda_s)} \sum_{j=i-m(K-1)+1}^{i+m(K-1)-1} \frac{q_j^*(m, \lambda_l) - q_{i+m(K-1)}^*(m, \lambda_l)}{\nu_1(m, \lambda_l)} \\ \xrightarrow{P} \frac{2}{c\psi_2^2} \Sigma_m(\lambda_1, \dots, \lambda_k)_{sl}. \end{aligned}$$

Proof see Appendix C ■

Using the estimator from Proposition 3, we obtain a feasible CLT for QRV* in the exact same manner as discussed in Section 2.1 for QRV.

3.3 Finite sample performance and noise robustness

The simulations below are designed to illustrate the performance of QRV* in the presence of market microstructure noise, comment on reasonable choice of the pre-averaging window width K , and make a comparison to alternative estimators. For ease of exposition, we mean QRV* to include the bias correction term $\psi_1 \widehat{\omega}^2 / (\psi_2 c^2)$ throughout the remainder of this paper. To simulate the “efficient” price process, we use the BM model as in Eqs. (19 – 20) and add i.i.d. noise as in Eq. (24). To ensure our simulation setup is realistic, we base our choice of parameters on a comprehensive set of summary statistics of global equity trade data as reported in Appendix A. We set the number of high-frequency return observations $N = \{1,000; 10,000\}$ representing typical small-to-mid and large-cap stocks. The level of microstructure noise is set to $\omega^2 = \gamma^2 IV/N$, where $\gamma^2 = \{0.25; 2.50; 10\}$. From Table 3, we see that this covers average, high, and extreme levels of noise. Note that the noise is normalized with respect to the IV, and that the so-called noise ratio γ (see

Oomen, 2006) has a natural interpretation in relation to the bias of the RV, as $\mathbb{E}(RV) = IV(1+2\gamma^2)$. To implement QRV*, we use the quantiles as before, i.e. $\bar{\lambda} = \{0.80; 0.85; 0.90; 0.95\}$, set $m = 40$, estimate ω^2 as in Oomen (2006), and vary K between 1 and 25. To provide a benchmark for our results, we compute the multi-scale RV (MSRV) of Zhang (2006):

$$MSRV_N(q) = \sum_{j=1}^q \frac{a_j}{j} \sum_{h=0}^{j-1} \gamma_{h,j}(0), \quad (31)$$

where q denotes the number of subsamples and

$$\gamma_{h,q} = \sum_{i=1}^N (Y_{iq+h} - Y_{(i-1)q+h})^2, \quad \text{and} \quad a_j = (1 - 1/q^2)^{-1} \left(\frac{j}{q^2} h(j/q) - \frac{j}{2q^3} h'(j/q) \right),$$

for $h(x) = 12(x - 1/2)$. In the simulations, we use the optimal number of subsamples, which can be chosen as $q_Z^* = c^* \sqrt{N}$, where

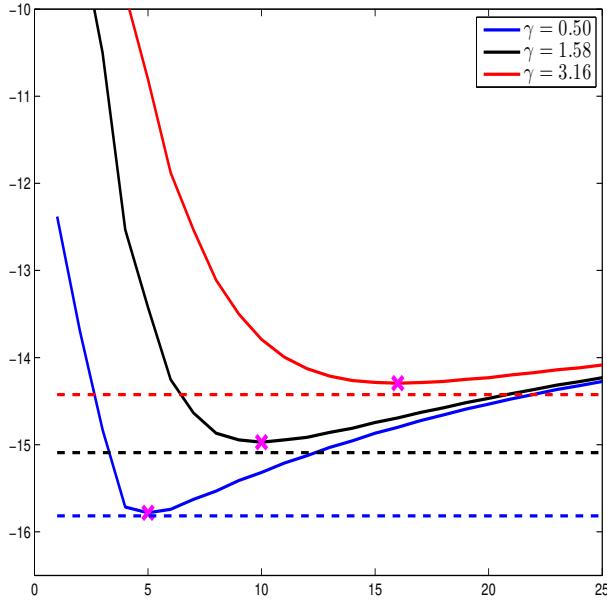
$$c^* = \arg \min_c \left\{ 2 \frac{52}{35} c IQ + \frac{48}{5} c^{-1} \omega^2 (IV + \omega^2/2) + 48 c^{-3} \omega^4 \right\}. \quad (32)$$

Both MSRV and QRV* are consistent for the IV and converge at rate $N^{-1/4}$, but the MSRV is not robust to jumps or outliers. The same is true for the two-scale RV of Zhang, Mykland, and Aït-Sahalia (2005) and the realised kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008). Because the performance of these estimators is very similar in the current setting, we concentrate on MSRV to conserve space.

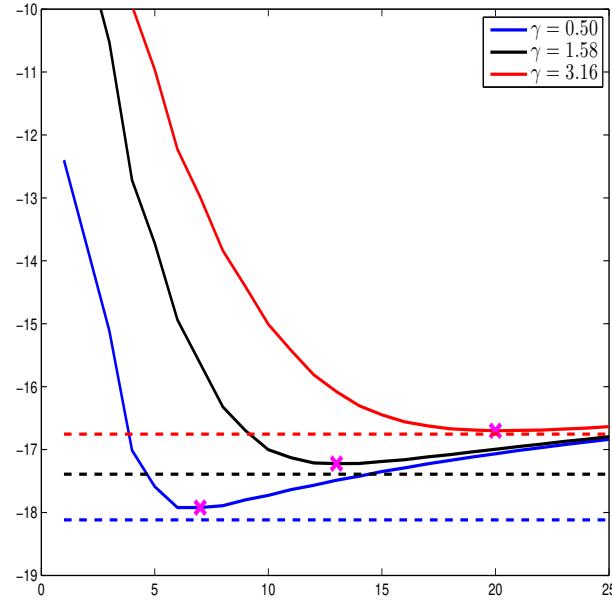
Figure 3 plots the log MSE of QRV* as a function of K for the simulations described above. The dashed horizontal lines indicate the performance of MSRV. As expected, the MSE minimizing choice of K increases in γ and N : the optimal choice of pre-averaging window width balances the noise reduction it achieves at the cost of efficiency loss. Perhaps most importantly, we see that the performance of QRV* is comparable to that of MSRV across all scenarios considered. While QRV* is only slightly inferior to MSRV in terms of efficiency, it comes of course with the benefit of being robust to jumps and outliers. The empirical application below will further illustrate this point. Regarding the optimal choice of K we can make the following observations. Although we have no explicit asymptotic guidance available, it is clear that the choice of window width can be informed by simulations as conducted here. In particular, for given N and noise ratio γ – the first quantity is readily available and the second can be estimated straightforwardly from the data – the optimal value for K can be read off a simulated MSE curve like the one in Figure 3. Also note that the MSE loss function is highly asymmetric in K , and so a conservative choice of pre-averaging window is generally preferred. This also helps to reduce the effects of price discreteness often encountered in high-frequency data (see Section 4) and makes the estimator less sensitive to potential violations of the i.i.d. noise assumption.

Figure 3: Performance of QRV* in the presence of noise.

Panel A: log MSE of QRV* with $N = 1,000$



Panel B: log MSE of QRV* with $N = 10,000$



Note. This figure plots the (log) MSE of the bias-corrected QRV* for sample size $N = 1,000$ (Panel A) and $N = 10,000$ (Panel B) and various levels of microstructure noise γ . The crosses indicate the minimum MSE, and thus identify the optimal choice of K . The dashed horizontal lines indicate the log MSE of MSRV using optimal number of subsamples.

4 Empirical illustration

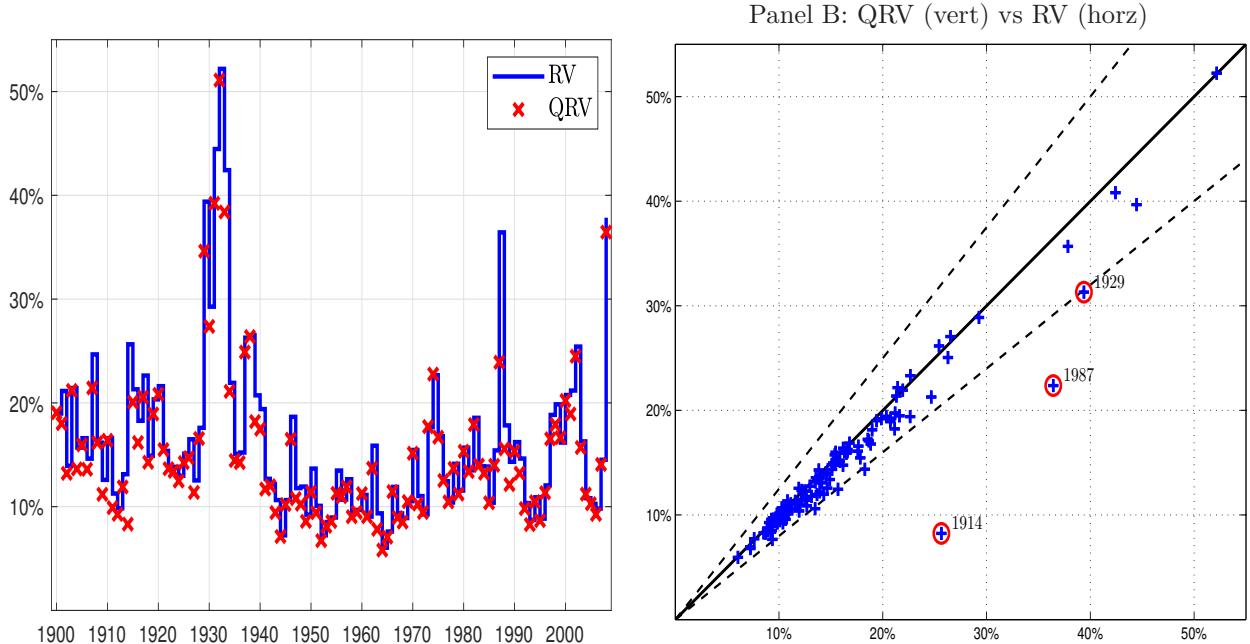
In this section, we apply the QRV estimator to a variety of equity data. The aim here is to illustrate the practical implementation of QRV and highlight some of its empirical properties. We use clean low-frequency data over long horizons as well as noisy high-frequency data over short horizons and find that in both cases the performance of the QRV is good compared to RV and its microstructure noise robust counterparts. To facilitate the discussion and interpretation of our results, we express all estimates as annualized standard deviations throughout this section.

4.1 QRV with “clean” low-frequency data

The use of QRV, like any other RV measure, is not merely limited to high-frequency data over short horizons, but can also be applied to low-frequency data over longer horizons. In the latter case the impact of market microstructure noise is benign and can be ignored for all practical purposes. See, for instance, Schwert (1989) who calculates monthly RVs using daily data or, more recently, Andersen, Bollerslev, Diebold, and Wu (2006) who study quarterly RVs and realised betas calculated from daily data.

As our first illustration, we look at daily data for the Dow Jones Industrial Average (DJIA) stock

Figure 4: QRV with daily Dow Jones Industrial Average index data.



Note. Year-by-year QRV and RV estimates from daily DJIA index data over the period January 2, 1900 through December 30, 2008. QRV is calculated using $\bar{\lambda} = \{0.80; 0.85; 0.90; 0.95\}$ and $m = 60$. The estimates are reported as annualized standard deviation. The dashed lines in Panel B mark the region where RV and QRV differ by more than 25%.

index over the sample period January 1900 through December 2008, i.e. 27,306 daily observations spanning more than a century.⁴ For each year in the sample, we estimate the ex-post return variation using QRV and RV from these daily data (i.e. $N \simeq 250$). To implement QRV, we use $\bar{\lambda} = \{0.80, 0.85, 0.90, 0.95\}$ with $m = 60$ and use the subsampling implementation. As volatility is widely documented to be very persistent, a quarterly block length provides sufficient locality. At the same time, with $(1 - \lambda_{\max})m - 1 = 2$, we are robust to up to four jumps or two outliers per quarter. We found, however, that our results were insensitive to reasonable alternative choices of quantiles and block length. Also, the subsampled and blocked implementation of QRV yield very similar estimates.

In Figure 4, we plot the time series of variance estimates in Panel A and a cross plot of RV (on the horizontal axis) versus QRV (on the vertical axis) in Panel B. We can see that the instances where QRV deviates substantially from RV all correspond to years with extreme market movements. For example, in 1914, the DJIA closed on July 30, 1914 at 71.42 and reopened more than 4 months later on December 14, 1914 at 56.76, reflecting a 20% drop in value. In 1929, the start of the great depression, the DJIA index fell 13.5% on October 28, another 11.7% the next day, only to rebound by 12.3% on October 30. Similarly, in 1987 the stock market crashed again, experiencing

⁴Source: Dow Jones Indexes, <http://www.djindexes.com/>

a daily return of -22.6% on October 19 which, even with the RV estimate of 38% for that year, constitutes a nine-standard deviation event. All this illustrates the robustness of QRV to jumps. In the remaining years, the QRV estimates are close to those of RV (with a sample correlation exceeding 0.99) indicating good efficiency and absence of any systematic biases. As an aside, note that BPV and MedRV are not robust to the jump scenario experienced in 1929 with three consecutive large returns.

Over the full 108 year sample period, we calculate an average annualized volatility estimate of 18.05% for RV and 16.75% for QRV, which suggests that roughly 14% of total variation can be attributed to jumps. If we leave out the three years discussed above, these figures drop to 17.4% and 16.5% , respectively, indicating that about 9% of total variation is due to jumps. Interestingly, using BPV Andersen, Bollerslev, and Diebold (2007) estimate the jump contribution to total variation at 14.4% using 5 minute S&P500 futures data over the period 1990 – 2002, Huang and Tauchen (2005) estimate the contribution at 7% for 5 minute S&P500 cash data from 1997 – 2002 and 4.5% for 5 minute S&P500 futures data from 1982 – 2002, while Corsi and Renò (2012) estimate the contribution at around 10% using 5 minute S&P500 futures data from 1990 – 2007. In a related study, Eraker, Johannes, and Polson (2003) estimate continuous time jump diffusion models using daily S&P500 index returns over the period 1980 – 1999 and measure the jump contribution between 8.2% and 14.7% depending on the model considered. While a direct comparison of these results is difficult (as different data, sampling frequency, horizon, and econometric techniques are used), it does illustrate that our estimates are plausible and in line with the extant literature.

4.2 QRV with “noisy” high-frequency data

Our second illustration is based on noisy high-frequency data with volatility computed over short daily horizons. We study Apple Inc. (AAPL) trade data over the period May 1, 2006 through December 30, 2008, which were extracted from the NYSE TAQ database. We only include trades from the primary exchange (i.e. NASDAQ) and aggregate records with the same millisecond precision time-stamp into one observation using the volume-weighted average trade price. To illustrate the robustness of QRV to jumps and outliers, we do *not* filter the data on qualifiers. The final dataset contains a record of 35,419,565 observations over 672 trading days, an average of about 52,708 trades per day.

Despite the deep liquidity of AAPL, its trade data are inherently noisy due to presence of bid-ask spread bounce. This is confirmed in Panel A of Figure 5, where we find substantial autocorrelation in returns. Also, with trade data at this frequency, price discreteness is a concern: from Panel B we see that the vast majority of return observation are either zero or plus/minus one tick, and virtually all observations are less than 8 ticks in magnitude. For these reasons, it is clearly inappropriate to apply the standard QRV. Instead, we use our noise-robust QRV*. To implement

this estimator we use, as before, four pairs of quantiles $\bar{\lambda} = \{0.80, 0.85, 0.90, 0.95\}$, set $K = 15$, and $m = 240$. This choice of parameters ensures substantial robustness to jumps and outliers with $(1 - \lambda_{\max})m - 1 = 11$. Moreover, with an effective block length of about 30 minutes for an average day (i.e. $15 \times 240 / 52708 \times 390$), the estimator is sufficiently “local” for it to pick up time variation in volatility. Also, because the data is sampled in event time, return volatility is homogenized to a large extent and this makes the results very insensitive to the choice of m . Finally, the choice of K is guided by simulations as in Section 3.3: unreported results show that the optimal K is around 10 for representative values of AAPL sample size and noise level (i.e. $N \approx 50,000$ and $\gamma \approx 0.5$). As mentioned above, a conservative choice of K is advised to account for price discreteness and other features of the data not captured by the BM plus i.i.d. noise model. Besides, given the large amount of data available here, efficiency is less of a concern. For these reasons we set $K = 15$, but it turns out that virtually identical results are obtained with $K = 10$. The effect of pre-averaging is nicely illustrated in Panels C and D of Figure 5. The pronounced serial correlation observed in raw returns is virtually eliminated for the pre-averaged data. At the same time, price discreteness is heavily reduced and the return distribution is now much closer to Gaussian.

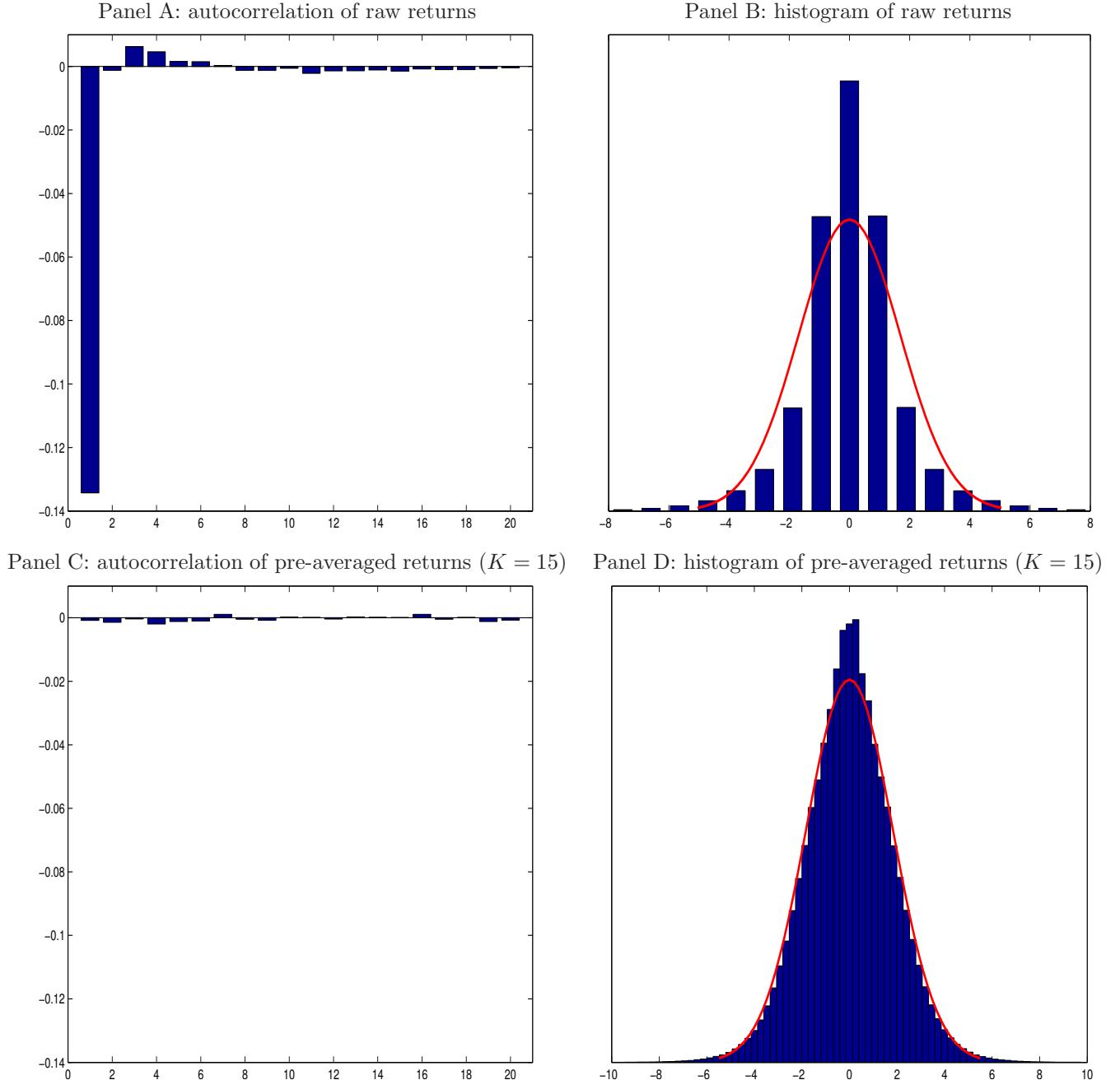
Figure 6 draws a time series (in Panel A) and scatter plot (in Panel B) of QRV* compared to MSRV of Zhang (2006).⁵ As in the previous illustration, we find a close alignment between the two estimates with a few noticeable exceptions. We highlight two here. On May 11, 2007 our sample of unfiltered trade data contains numerous “out-of-sync” records as indicated by the qualifier “Z” in the TAQ data.⁶ The time series of highly erratic transaction prices is displayed in Panel C of Figure 6, from which it is evident that standard volatility estimators will be heavily distorted. On the raw data, MSRV estimates daily volatility at 39.3%, which more than halves to 19.2% after removing the spurious observations. In sharp contrast, QRV gives reliable estimates both on the raw and cleaned data, i.e. 20.0% and 18.0% respectively. The second example is October 11, 2007. See Panel D of Figure 6 for the price path: the unfiltered data contains a number of trades that are seemingly executed at prices well below fair market value. Closer inspection of the data reveals that several – but not all – of these observations were flagged by the NASDAQ as suspicious (using a “yellow flag” indicated by qualifier “Y”⁷) or re-opening prints (as indicated by qualifier “O” or “5”).

⁵The MSRV is implemented with an optimal bandwidth as in Eq. (32), estimated for each day in the sample separately. ω^2 is estimated from the first-order autocovariance of the trade data, whereas IV and IQ estimates are calculated by subsampling the equivalent of 5 minute data in trade time (i.e. each subsample consists of 79 price observations per day). The average optimal bandwidth is 3.28, with a minimum of 2, a maximum of 9, and $q^* = 3$ for more than three out of every four days.

⁶NASDAQ Market Data Distribution (2008) describes this qualifier as “Sold Out of Sequence is used when a trade is printed (reported) out of sequence and at a time different from the actual transaction time.”

⁷NASDAQ Market Data Distribution (2008) describes this qualifier as “Market Centers will have the ability to identify regular trades being reported during specific events as out of the ordinary by appending a new sale condition code Yellow Flag (“Y”) on each transaction reported.”

Figure 5: Summary statistics of “noisy” AAPL trade data.



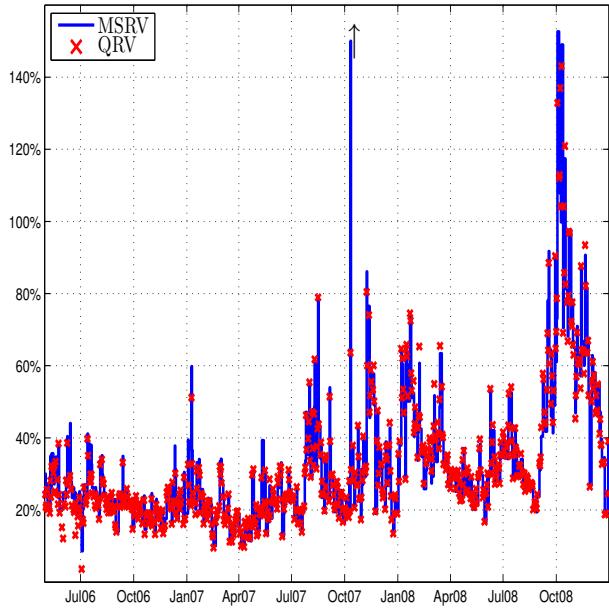
Note. Summary statistics of raw and pre-averaged AAPL trade data over period May 1, 2006 through December 30, 2008. Returns are expressed in basis points.

On the raw data, MSRV estimates volatility at 4,183.29% which drops to 67.88% after removing the spurious data records. QRV, on the other hand, again enjoys remarkable robustness to these outliers and estimates volatility at 63.59% on the raw data and at 62.90% on the cleaned data.

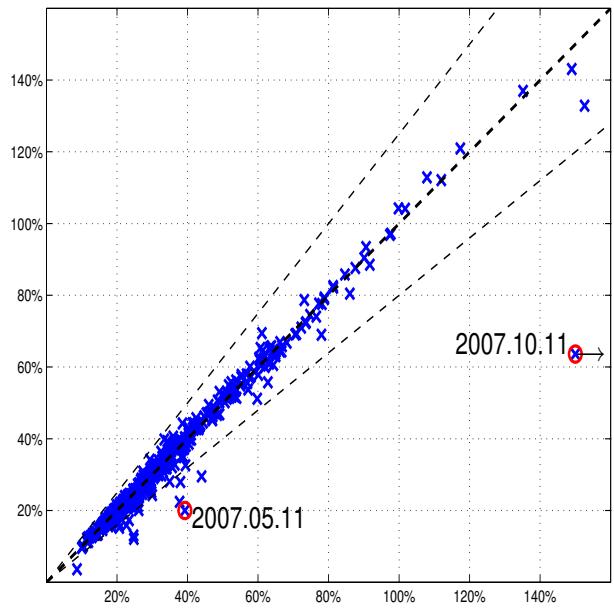
Averaging over all days in the sample (excluding October 11, 2007), we compute a return

Figure 6: QRV* with “noisy” AAPL trade data.

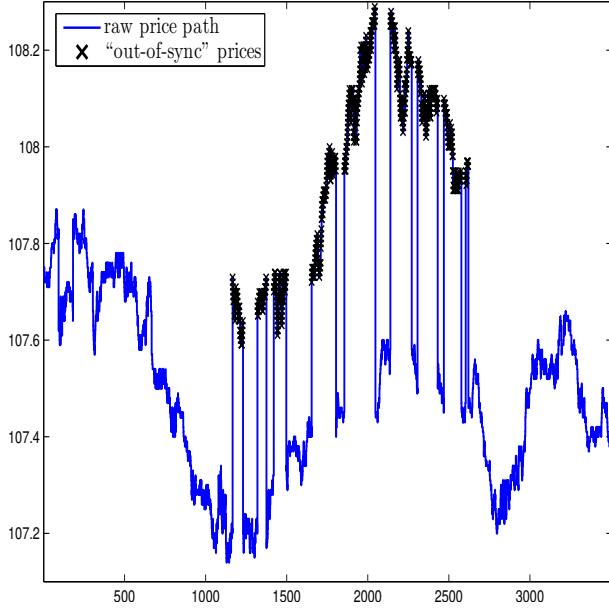
Panel A: ORV and MSRV against time



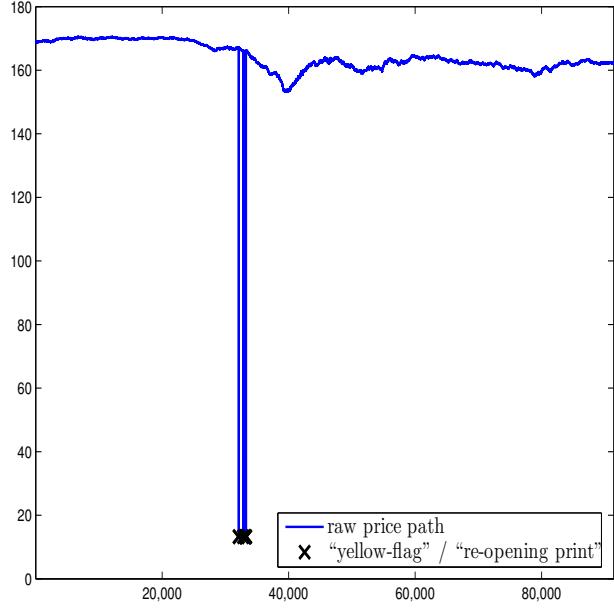
Panel B: QRV vs MSRV



Panel C: trade prices 2007.05.11 (9:50 – 10:05)



Panel D: trade prices 2007.10.11 (9:45 – 16:00)



Note. Day-by-day QRV and MSRV estimates from unfiltered tick-by-tick trade prices of AAPL over the period May 1, 2006 through December 30, 2008. QRV* is calculated using $\bar{\lambda} = \{0.80; 0.85; 0.90; 0.95\}$, $m = 240$, and $K = 15$. MSRV is calculated using the optimal number of subsamples as in Eq. (32). Estimates are expressed as annualized standard deviation. The thin dashed lines in Panel B mark the region where MSRV and QRV differ by more than 25%. The MSRV estimate for 2007.10.11 is 4,183.29%, but to preserve the scaling of the graph it is capped at 150% (as indicated by the arrow).

volatility of 37.32% with MSRV and 37.15% with QRV*. In stark contrast to our previous illustration with daily data and the studies by Andersen, Bollerslev, and Diebold (2007); Corsi and Renò (2012);

Eraker, Johannes, and Polson (2003); Huang and Tauchen (2005) amongst others, we now find that less than 1% of total variation can be attributed to jumps! This is rather surprising particularly because we are considering here a single stock instead of a well diversified index like the S&P500 and cover a period that includes the exceptionally turbulent 2008. Moreover, because some of the larger deviations between QRV and MSRV in our sample are due to further spurious data points, we fail to identify a single instance, where a true jump in the price is observed at this frequency. Hence, this raises the question whether previously identified jumps in the literature are in fact actual jumps in the price, which become difficult to identify at ultra-high frequency due to for instance market microstructure noise or, on the other hand, whether they are simply a consequence of (sparse) sampling and vanish when moving to tick-by-tick event time. We argue the latter and provide anecdotal evidence, which suggests that price jumps identified at a low sampling frequency may in fact be bursts of volatility. This logic is also what underlies the family of jump tests developed by Aït-Sahalia and Jacod (2009a), namely, true jumps can only be identified by increasing the sampling frequency to the limit.

Consider April 24, 2007: an important day for AAPL with the SEC dropping tax fraud charges against CEO Steve Jobs⁸, and a quarterly earnings announcement to be released on the following day. Panel A of Figure 7 plots the intra-day price path at a conventional 5 minute frequency, and we observe a temporary 2% drop in share price with an instant recovery around 13:30. The widely used BPV jump test statistic is highly significant at -6.8 (RV is 41.2% and BPV 32.4%) and unequivocally identifies this day to contain at least one jump. Contrast this with Panel B of Figure 7, where we plot the AAPL share price over the one hour window from 12:45 to 13:45 at ultra-high trade frequency. Now, the largest trade-by-trade move is only \$0.13 at a price of around \$92 or merely 14 basis points. At the same time, more than one-third of the total daily volume was traded in this one hour window. On these data the MSRV and QRV* estimates closely agree at 30.4% and 29.8% respectively. Taken together, the hypothesis that there was a burst of volatility around 13:30, which is mistakenly identified as a jump at lower frequency, seems the more plausible one. This argument can be supported further by noting that the New York Stock Exchange, which actively makes markets in AAPL, contractually obliges their designated market makers to maintain a fair and orderly market which, as highlighted in NYSE Rule 104, implies “*the maintenance of price continuity with reasonable depth*”.⁹ See Hasbrouck (2007) for further discussion on this.

⁸See <http://www.sec.gov/news/press/2007/2007-70.htm>.

⁹NYSE Rule 104 on the dealings and responsibilities of a designated market marker (DMM) states “*The function of a member acting as a DMM on the Floor of the Exchange includes the maintenance, in so far as reasonably practicable, of a fair and orderly market on the Exchange in the stocks in which he or she is so acting. The maintenance of a fair and orderly market implies the maintenance of price continuity with reasonable depth, to the extent possible consistent with the ability of participants to use reserve orders, and the minimizing of the effects of temporary disparity between supply and demand. In connection with the maintenance of a fair and orderly market,*

The suggestion that price jumps are less common than previously thought does not in any way negate the importance of QRV or other jump-robust RV measures. Firstly, QRV is robust not only to jumps but also to outliers. As we have seen above, outliers do occur in high-frequency data and cannot always be filtered out perfectly based on trade qualifiers (this is particularly true for less recent non-US data). They are therefore a challenge to the econometrician and here the value of QRV is indisputable. Secondly, numerous scenarios remain where true price jumps can be observed, for instance over intra-day auctions, lunch breaks or when circuit breakers or volatility auctions are activated.¹⁰ Also, the market maker’s precise obligation changes by exchange and regardless of this, its influence is not without limit so that with exceptional news releases or large market orders in an illiquid market, jumps can of course still occur. Finally, jumps can be commonplace when moving beyond equity markets governed by a market maker charged with the obligation to maintain price continuity. Consider, for example, electricity markets, where storage is costly, power plants can fail, and sudden unpredictable swings in demand for energy – often for exogenous reasons – can lead to substantial temporary price spikes (see, e.g., Bessembinder and Lemmon, 2002; Longstaff and Wang, 2004).

5 Concluding remarks

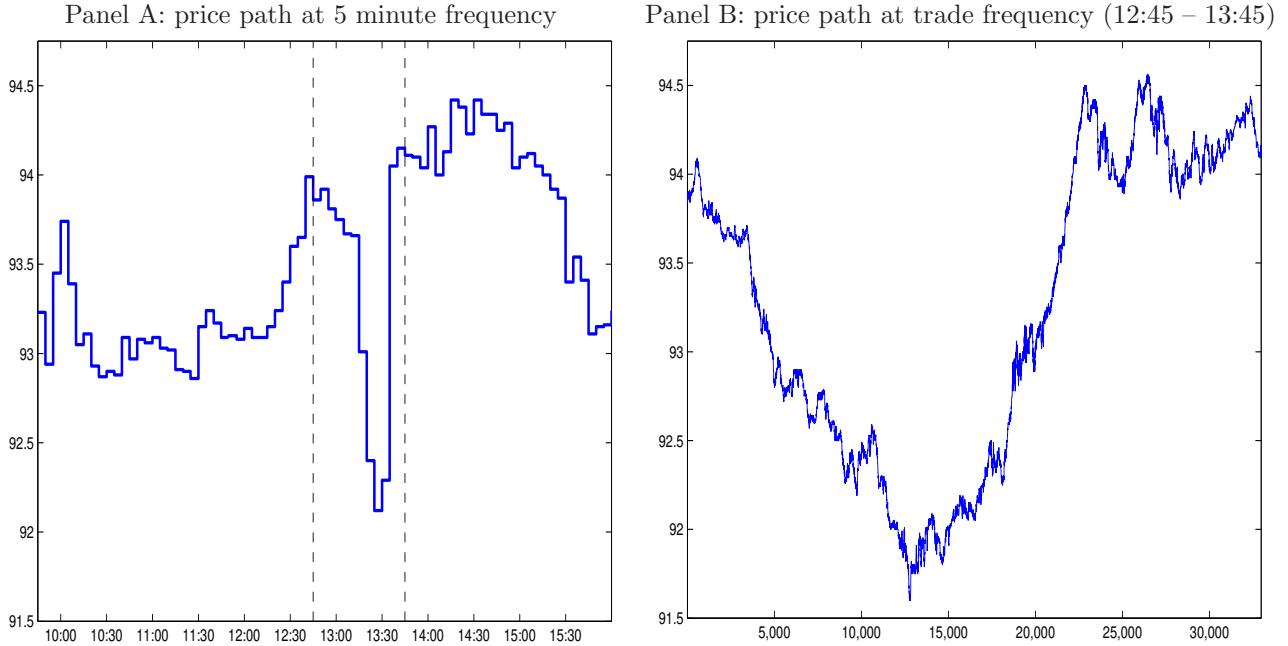
In this paper we develop a new quantile-based realised variance measure that is consistent for the integrated variance and robust to jumps and outliers. A modified version, based on pre-averaged data, is also introduced and we show that in the presence of microstructure noise it retains consistency and attains the best possible convergence rate of $N^{-1/4}$. Importantly, our estimator is highly efficient making it the first estimator of integrated variance in the literature that is, at the same time, efficient, and robust to both jumps, outliers and microstructure noise. From a practical viewpoint, the estimator is easy to implement and is relatively insensitive to the particular choice of tuning parameters. Extensive simulations and empirical applications illustrate the excellent performance of our estimator.

The methodology outlined in this paper can be extended into various directions. For instance, it is possible to develop a joint distribution theory for RV and QRV allowing the construction of a formal jump test in the spirit of Barndorff-Nielsen and Shephard (2006). Also, it is possible

it is commonly desirable that a member acting as DMM engage to a reasonable degree under existing circumstances in dealings for the DMM’s own account when lack of price continuity, lack of depth, or disparity between supply and demand exists or is reasonably to be anticipated.” (source: <http://rules.nyse.com>)

¹⁰For example, the German trading platform Xetra has daily intra-day auction between 13:00 – 13:17 for the various segments. The Tokyo (Hong Kong) stock exchange shuts down from 11:00 – 12:30 (12:30 – 14:30). The NYSE uses a circuit breaker where a 10% intra-day move in the DJIA triggers a market wide trading halt of up to one hour.

Figure 7: Jumps or burst of volatility?



Note. Panel A plots the price of AAPL on April 24, 2007 at a 5 minute frequency (using last tick interpolation on trade data). Panel B plots the price at trade frequency from 12:45 to 13:45.

to modify QRV to produce jump and noise robust estimates of the integrated quarticity, a key quantity when making inference about integrated variance and testing for jumps. With these tools available, it may then be interesting to revisit some of the empirical work on non-parametric jump tests (e.g. Aït-Sahalia and Jacod, 2009a,b; Barndorff-Nielsen and Shephard, 2006; Christensen and Podolskij, 2009; Fan and Wang, 2007; Jiang and Oomen, 2008; Lee and Mykland, 2008) and, inspired by our empirical findings, reevaluate the role that jumps play in financial equity price dynamics (e.g. Andersen, Bollerslev, and Diebold, 2007; Eraker, Johannes, and Polson, 2003; Huang and Tauchen, 2005). Recent work by Aït-Sahalia and Jacod (2009a), Barndorff-Nielsen, Shephard, and Winkel (2006) and Woerner (2006) has shown that bi-power variation is not only robust to finite activity jumps (as considered in this paper) but also to certain infinite activity jump specifications. An investigation of the properties of QRV in such a scenario might be of interest and allow for further comparison to alternative jump robust estimators. Finally, in this paper we maintained the assumption of i.i.d. noise but this may be relaxed to allow for dependent noise (as studied by, for instance, Aït-Sahalia, Mykland, and Zhang, 2011; Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2008; Jacod, Li, Mykland, Podolskij, and Vetter, 2009). All the above is well beyond the scope of the current paper and will be left for future research.

Table 2: Performance of QRV with stochastic volatility, jumps, and outliers

model	blocked QRV (m, n)			subsampled QRV (m, n)			benchmarks			
	(20, 50)	(40, 25)	(100, 10)	(20, 50)	(40, 25)	(100, 10)	RV	BPV	TRV	MedRV
<i>Panel A: "bias" measure $\mathbb{E}(\widehat{IV}/IV)$</i>										
BM	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SV	1.00	0.99	0.98	1.00	0.99	0.98	1.00	1.00	1.00	1.00
SV-LEV	1.00	0.99	0.98	1.00	0.99	0.98	1.00	1.00	1.00	1.00
SEV-ND	1.00	0.99	0.99	1.00	0.99	0.98	1.00	1.00	1.00	1.00
SV2F-LEV	1.00	1.00	0.99	1.00	0.99	0.98	1.00	1.00	1.00	1.00
BMJ($n_J = 1, v_J = \frac{1}{4}$)	1.00	1.00	1.00	1.00	1.00	1.00	1.25	1.03	1.01	1.00
BMJ($n_J = 5, v_J = \frac{1}{4}$)	1.02	1.02	1.02	1.02	1.02	1.02	1.25	1.06	1.05	1.02
BMJ($n_J = 10, v_J = \frac{1}{4}$)	1.04	1.04	1.03	1.04	1.04	1.03	1.25	1.08	1.12	1.03
BMJ($n_J = 5, v_J = \frac{1}{2}$)	1.03	1.02	1.02	1.03	1.02	1.02	1.50	1.09	1.05	1.02
BM-outlier	1.01	1.01	1.01	1.01	1.01	1.01	1.25	1.21	1.02	1.33
<i>Panel B: "efficiency" measure $\text{var}(\sqrt{N}(\widehat{IV} - IV)/\sqrt{IQ})$</i>										
BM	2.41	2.42	2.42	2.33	2.38	2.49	2.00	2.60	2.00	2.96
SV	2.42	2.41	2.37	2.37	2.51	3.28	2.01	2.61	2.01	2.96
SV-LEV	2.40	2.39	2.33	2.36	2.49	3.27	2.02	2.60	2.02	2.94
SEV-ND	2.37	2.35	2.29	2.33	2.46	3.29	2.00	2.61	2.00	2.95
SV2F-LEV	2.38	2.36	2.32	2.34	2.51	3.52	1.99	2.59	1.99	2.93
BMJ($n_J = 1, v_J = \frac{1}{4}$)	2.44	2.44	2.44	2.36	2.40	2.51	127.74	3.66	2.20	2.99
BMJ($n_J = 5, v_J = \frac{1}{4}$)	3.02	2.54	2.52	2.77	2.49	2.59	27.87	3.80	3.51	3.29
BMJ($n_J = 10, v_J = \frac{1}{4}$)	3.16	2.68	2.61	2.90	2.61	2.69	15.53	3.84	4.88	3.41
BMJ($n_J = 5, v_J = \frac{1}{2}$)	4.63	2.60	2.52	3.81	2.52	2.59	104.66	5.24	3.48	4.06
BM-outlier	2.46	2.47	2.46	2.38	2.42	2.53	127.22	89.24	2.89	237.02

Note. This table reports the bias and efficiency measure for QRV (and RV, BPV, TRV, and MedRV for comparison) under various model specifications for $N = 1,000$. The bias measure in Panel A is equal to 1 for an unbiased IV estimator. The efficiency measure in Panel B takes on a minimum attainable value of 2 for the MLE.

A Summary statistics of trade data

Table 3: Summary statistics of trade data (2008)

universe	# of observations N			noise ratio γ		
	Q5	Q50	Q95	Q5	Q50	Q95
<i>Panel A: US</i>						
S&P600	157	751	2,417	0.10	0.34	0.73
S&P400	604	1,749	4,710	0.12	0.36	0.76
S&P500	1,477	4,174	12,355	0.14	0.37	0.93
S&P100	2,945	7,338	20,707	0.17	0.40	1.06
DJ30	4,701	9,562	23,686	0.22	0.45	0.97
<i>Panel B: Europe</i>						
DJ Stoxx Small 200	158	772	2,225	0.25	0.59	1.15
DJ Stoxx Mid 200	352	1,419	3,689	0.30	0.63	1.16
DJ Stoxx Large 200	999	3,634	11,169	0.34	0.66	1.28
DJ Stoxx50	3,161	6,975	15,860	0.40	0.71	1.40
<i>Panel C: Asia-pacific</i>						
S&P ASX200	199	744	2,957	0.30	0.74	1.75
S&P Topix 150	370	1,070	2,639	0.34	1.03	3.59
Hang Seng	465	1,260	4,090	0.39	0.88	2.26
<i>Panel D: Emerging markets (BRIC)</i>						
Ibovespa (Brazil)	261	1,130	5,617	0.32	0.66	1.21
DJ Titans 10 (Russia)	543	6,066	22,230	0.58	1.03	1.29
DJ BRIC 50 (India)	726	2,098	4,987	0.16	0.49	0.90
DJ BRIC 50 (China)	1,177	2,328	5,197	0.60	1.17	2.96

Note. This table reports the 5th, 50th, and 95th percentile of the number of observations (i.e. number of trades per day) N and the noise ratio $\gamma^2 = \omega^2/(IV/N)$ computed across all names in each universe and all days over the period Jan 2, 2008 through Dec 31, 2008. The index constituents of January 2009 are used.

To motivate the choice of parameters used for the simulations in Section 3.3, Table 3 reports summary statistics of the number of intra-day trade price observations N and the level of microstructure noise as measured by the noise ratio $\gamma^2 = N\omega^2/IV$ (see Oomen, 2006) for various stock universes. The data is taken from Reuters DataScope Tick History and covers the period January 2, 2008

through December 30, 2008. For the US and Europe, the selection covers small-caps (S&P600, DJ Stoxx Small 200), mid-caps (S&P400, DJ Stoxx Mid 200), large-caps (S&P500, S&P100, DJ Stoxx Large 200) and blue-chips (DJ30, DJ Stoxx50). For the Asia-pacific region and emerging markets, the universes cover large-caps only. The tick data are filtered on trade conditions, only trades from the primary exchange are included¹¹, and trades with identical millisecond precision time-stamp are aggregated into one observation with a volume-weighted average trade price. All this is done to maximize the cleanliness of the data. To compute the noise ratio (for each stock and day in the sample), the microstructure noise variance ω^2 is estimated as the negative of the first-order autocovariance of trade returns following Oomen (2006). The IV is proxied by an ad-hoc implementation of the realised (Bartlett) kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) with a bandwidth parameter equal to five as suggested by Gatheral and Oomen (2010). The summary statistics in Table 3 are computed across all names in the universe and days in the sample. This is justified as N and γ are reasonable stable over the sample period.

¹¹For the US in particular, and Europe to lesser extent, this depresses the average number of trades per day as large fractions of volume are executed on competing exchanges and multi-lateral trading platforms.

B The QRV with absolute returns: an alternative formulation

In the main text, we defined QRV based on signed returns using symmetrized quantiles. Here we give an alternative formulation of the QRV based on the quantiles of absolute returns and show how these specifications are related. We base our exposition on the subsampled version of the QRV, as it can be linked to some other estimators proposed in the literature, but it should be clear that we could equally well work with a blocking QRV.

Define

$$q_i^{abs,sub}(m, \lambda) = g_{\lambda m}^2 \left(\sqrt{N} |D_{i,m} X| \right), \quad \text{for } \lambda \in [0, 1).$$

As we are now dealing with absolute returns, there is no need to symmetrize $q_i^{abs,sub}(m, \lambda)$. We write

$$QRV_N^{abs,sub}(m, \bar{\lambda}, \alpha) \equiv \alpha' QRV_N^{abs,sub}(m, \bar{\lambda}), \quad (33)$$

with the j th element of $QRV_N^{abs,sub}(m, \bar{\lambda})$ given by:

$$QRV_N^{abs,sub}(m, \lambda_j) = \frac{1}{N-m} \sum_{i=1}^{N-m} \frac{q_i^{abs,sub}(m, \lambda_j)}{\nu_1^{abs}(m, \lambda_j)}, \quad \text{for } \lambda_j \in [0, 1), \quad (34)$$

and $j = 1, \dots, k$, where $\nu_r^{abs}(m, \lambda)$ is defined as:

$$\nu_r^{abs}(m, \lambda) = \mathbb{E} \left[\left(|U|_{(\lambda m)} \right)^{2r} \right]. \quad (35)$$

The consistency and central limit theorems derived for QRV based on signed returns, extend directly to the case where we use absolute returns by replacing $\nu_r(m, \lambda)$ in Eq. (8) with $\nu_r^{abs}(m, \lambda)$ above and $\nu_1(m, \lambda_i, \lambda_j)$ in Eq. (10) with

$$\nu_1^{abs}(m, \lambda_i, \lambda_j) = \mathbb{E} \left[\left(|U|_{(\lambda_i m)} \right)^2 \left(|U|_{(\lambda_j m)} \right)^2 \right]. \quad (36)$$

The corresponding asymptotic constants for the $m \rightarrow \infty$ limit of Proposition 2 are now:

$$\begin{aligned} \nu_1^{abs}(\lambda) &\equiv \lim_{m \rightarrow \infty} \nu_1^{abs}(m, \lambda) = d_\lambda, \\ \nu_1^{abs}(\lambda_i, \lambda_j) &\equiv \lim_{m \rightarrow \infty} \nu_1^{abs}(m, \lambda_i, \lambda_j) = d_{\lambda_i} d_{\lambda_j}, \\ \Theta^{abs}(\bar{\lambda})_{ij} &\equiv \lim_{m \rightarrow \infty} \Theta^{abs}(m, \bar{\lambda})_{ij} = \frac{\lambda_i(1-\lambda_j)}{p(d_{\lambda_i})p(d_{\lambda_j})d_{\lambda_i}d_{\lambda_j}}, \end{aligned}$$

with $\lambda_i < \lambda_j$, where d_α and p denote the α -quantile and density function of the χ_1^2 -distribution. With noise, the results presented for signed returns also extend to the formulation based absolute returns. The only substantive change would be the corresponding redefinition of Eq. (30), i.e.

$$f_{m,l,x,u}^{abs}(\lambda_1, \lambda_2) = \text{cov} \left(g_{\lambda_1 m}^2(|S|), g_{\lambda_2 m}^2(|T|) \right).$$

It is quite intuitive that QRV_N^{sub} and $QRV_N^{abs,sub}$ are closely related, when the quantiles are chosen as $\lambda_i^{abs} = 2\lambda_i - 1$. While it is hard to formalize this intuition, in unreported simulations we find that the performance of these estimators is indistinguishable. In the $m \rightarrow \infty$ limit, however, we can be more precise and prove their equivalence. Consider $\bar{\lambda} = \lambda$ for $\lambda \in (1/2, 1)$. Then it follows that $U_{(\lambda m)} \xrightarrow{p} c_\lambda$ as $m \rightarrow \infty$, where c_λ is the λ -quantile of $N(0, 1)$. On the other hand, $|U|_{((2\lambda-1)m)} \xrightarrow{p} c_\lambda$, and since $c_\lambda > 0$ for $\lambda \in (1/2, 1)$, $|U_{(\lambda m)}|$ and $|U|_{((2\lambda-1)m)}$ have the same asymptotic behavior. By symmetry of the normal distribution, it also holds that $|U_{(m-\lambda m+1)}| \xrightarrow{p} c_\lambda$, and thus $(|U_{(\lambda m)}| + |U_{(m-\lambda m+1)}|)/2$ and $|U|_{((2\lambda-1)m)}$ also have the same asymptotic behaviour. It now follows that the expressions for the asymptotic variance of the two estimators are identical for $m \rightarrow \infty$, when the quantiles are chosen this way.

There are, however, two important ways in which the QRV estimator based on absolute returns is distinguished from its counterpart on signed returns. First, by a suitable choice of λ^{abs} it is possible to discard only a single observation per block (e.g. $m = 100$ with $\lambda^{abs} = 0.99$). This allows for the use of observations in the extreme tail, at the cost of losing robustness to outliers. Second, the estimator based on absolute returns nests the MinRV and MedRV proposed by Andersen, Dobrev, and Schaumburg (2008) as special cases. Specifically, with $m = 2$ and $\lambda^{abs} = \frac{1}{2}$ we have MinRV and with $m = 3$ and $\lambda^{abs} = \frac{2}{3}$ we have MedRV.

C Proofs

In this part of the paper, we state the proofs of the theorems given in the main text. Throughout, we use the approximation

$$\Delta_i^N X \approx \sigma_{\frac{i-1}{N}} \Delta_i^N W.$$

Thus, to prove our asymptotic results we first replace $\Delta_i^N X$ with $\sigma_{\frac{i-1}{N}} \Delta_i^N W$ and then show that the error caused by this approximation is asymptotically negligible.

Let us fix some notations. We set

$$\beta_i^n = \sqrt{N} \left(\sigma_{\frac{i-1}{n}} \Delta_k^N W \right)_{(i-1)m+1 \leq k \leq im},$$

and define

$$w_i^{(n,m)}(\lambda) = g_{\lambda m}^2(\beta_i^n) + g_{m-\lambda m+1}^2(\beta_i^n).$$

Before we start to prove the main results, we state a simple Lemma.

Lemma 1 *The function g_k defined in (4) has the following properties:*

1. g_k is continuous.
2. g_k is differentiable on the set $\{x \in \mathbb{R}^m \mid x_i \neq x_j, 1 \leq i < j \leq m\}$, that is

$$\frac{1}{\epsilon} [g_k(x + \epsilon y) - g_k(x)] \rightarrow y_{k*} \quad \epsilon \searrow 0,$$

where $y \in \mathbb{R}^m$ and

$$k* = i \text{ with } x_i = x_{(k)}.$$

In the following we assume without loss of generality that a , σ , a' , σ' and v' are bounded (for details see e.g. Section 3 in Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard, 2006). Moreover, the constants used in the proofs will all be denoted by C .

Proof of Theorem 1 Here we show the consistency of $QRV_N(m, \bar{\lambda}, \alpha)$. Since $\sum_{j=1}^k \alpha_j = 1$ it is sufficient to prove Theorem 1 for $\bar{\lambda} = \lambda \in (1/2, 1)$ and $\alpha = 1$ (i.e. $k = 1$). In this case we use the notation $QRV_N(m, \lambda) = QRV_N(m, \bar{\lambda}, \alpha)$. First, we define:

$$\xi_i^n = \nu_{1,m}^{-1}(\lambda) w_i^{(n,m)}(\lambda),$$

$$U_n = \frac{1}{n} \sum_{i=1}^n \xi_i^n.$$

Note that:

$$\mathbb{E} \left[\xi_i^n \mid \mathcal{F}_{\frac{i-1}{n}} \right] = \sigma_{\frac{i-1}{n}}^2,$$

so

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\xi_i^n \mid \mathcal{F}_{\frac{i-1}{n}} \right] \xrightarrow{p} \int_0^1 \sigma_u^2 du. \quad (37)$$

Now, by setting

$$\eta_i^n = \xi_i^n - \mathbb{E} \left[\xi_i^n \mid \mathcal{F}_{\frac{i-1}{n}} \right], \quad (38)$$

we get:

$$\mathbb{E} \left[|\eta_i^n|^2 \mid \mathcal{F}_{\frac{i-1}{n}} \right] = \frac{\nu_{2,m}(\lambda) - \nu_{1,m}^2(\lambda)}{\nu_{1,m}^2(\lambda)} \sigma_{\frac{i-1}{n}}^4.$$

Therefore,

$$\frac{1}{n^2} \sum_{i=1}^n \mathbb{E} \left[|\eta_i^n|^2 \mid \mathcal{F}_{\frac{i-1}{n}} \right] \xrightarrow{p} 0.$$

Hence, the assertion $U_n \xrightarrow{p} \int_0^1 \sigma_u^2 du$ follows directly from (37). Now, we are left to prove that

$$QRV_N(m, \lambda) - U_n \xrightarrow{p} 0. \quad (39)$$

Note that

$$QRV_N(m, \lambda) - U_n = \frac{\nu_{1,m}^{-1}(\lambda)}{n} \sum_{i=1}^n \zeta_i^n,$$

where

$$\zeta_i^n = q_i^{(n,m)}(\lambda) - w_i^{(n,m)}(\lambda).$$

Now, we use the decomposition

$$\zeta_i^n = \zeta_i^n(1) + \zeta_i^n(2),$$

where $\zeta_i^n(k)$, $k = 1, 2$, are given by

$$\zeta_i^n(1) = g_{\lambda m}^2 \left(\sqrt{N} D_i^m X \right) - g_{\lambda m}^2(\beta_i^n), \quad (40)$$

$$\zeta_i^n(2) = g_{m-\lambda m+1}^2 \left(\sqrt{N} D_i^m X \right) - g_{m-\lambda m+1}^2(\beta_i^n). \quad (41)$$

In the following, we show that

$$\frac{\nu_{1,m}^{-1}(\lambda)}{n} \sum_{i=1}^n \zeta_i^n(1) \xrightarrow{p} 0. \quad (42)$$

The corresponding result for $\zeta_i^n(2)$ can be proven similarly. We begin with the following Lemma.

Lemma 2 For $x \in \mathbb{R}^m$, we define a norm $\|x\| = \sum_{k=1}^m |x_k|$. Then we have

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\|\sqrt{N} D_i^m X - \beta_i^n\|^2 \right] \rightarrow 0. \quad (43)$$

Proof of Lemma 2 The boundedness of the drift function a and $\|x\|^2 \leq m \sum_{k=1}^m |x_k|^2$ yield

$$\begin{aligned}\mathbb{E} \left[\|\sqrt{N} D_i^m X - \beta_i^n\|^2 \right] &\leq mC \left(\frac{1}{n} + N \sum_{k=1}^m \mathbb{E} \left[\int_{\frac{i-1}{n} + \frac{k-1}{N}}^{\frac{i-1}{n} + \frac{k}{N}} |\sigma_u - \sigma_{\frac{i-1}{n}}|^2 du \right] \right) \\ &= mC \left(\frac{1}{n} + N \mathbb{E} \left[\int_{\frac{i-1}{n}}^{\frac{i}{n}} |\sigma_u - \sigma_{\frac{i-1}{n}}|^2 du \right] \right).\end{aligned}$$

Hence, the left side of (43) is smaller than

$$mC \left(\frac{1}{n} + m \mathbb{E} \int_0^1 |\sigma_u - \sigma_{[nu]/n}|^2 du \right).$$

Because σ is bounded and càdlàg (and so $\sigma_u - \sigma_{[nu]/n} \xrightarrow{p} 0$ with an exception of countably many $u \in [0, 1]$), the assertion of Lemma 2 is proved by Lebesgue's theorem. \blacksquare

Next, we set

$$m_A(\epsilon) = \sup \{ |g_{\lambda m}^2(x) - g_{\lambda m}^2(y)| : \|x - y\| \leq \epsilon, \|x\| \leq A \}.$$

For all $\epsilon \in (0, 1]$ and $A > 1$, we obtain the estimate

$$\begin{aligned}\zeta_i^n(1) &\leq C \left(m_A(\epsilon) + A^2 \mathbb{1}_{\{\|\sqrt{N} D_i^m X - \beta_i^n\| > \epsilon\}} \right. \\ &\quad \left. + \left(g_{\lambda m}^2(\sqrt{N} D_i^m X) + g_{\lambda m}^2(\beta_i^n) \right) \left(\mathbb{1}_{\{\|\sqrt{N} D_i^m X\| > A\}} + \mathbb{1}_{\{\|\beta_i^n\| > A\}} \right) \right) \\ &\leq C \left(m_A(\epsilon) + \frac{A^2 \|\sqrt{N} D_i^m X - \beta_i^n\|^2}{\epsilon^2} + \frac{(\|\sqrt{N} D_i^m X\| + \|\beta_i^n\|)^3}{A} \right).\end{aligned}$$

The boundedness of a and σ and Burkholder's inequality imply that

$$\mathbb{E} \left[\|\sqrt{N} D_i^m X\|^p \right] + \mathbb{E} [\|\beta_i^n\|^p] \leq C_p, \tag{44}$$

for all $p \geq 0$. This means that

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} [\zeta_i^n(1)] \leq C \left(m_A(\epsilon) + \frac{1}{A} + \frac{A^2}{n\epsilon^2} \sum_{i=1}^n \mathbb{E} [\|\sqrt{N} D_i^m X - \beta_i^n\|^2] \right).$$

Because $m_A(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$ for every A , we obtain by Lemma 2

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} [\zeta_i^n(1)] \rightarrow 0, \tag{45}$$

by first choosing A large, then ϵ small and finally n large. Then (42) holds and the proof of Theorem 1 is complete. \blacksquare

Proof of Theorem 2 We proceed with a three-stage proof of Theorem 2. First, we prove a CLT for the sequence $\bar{U}_n = (\bar{U}_n^1, \dots, \bar{U}_n^k)$ with

$$\bar{U}_n^j = \sqrt{\frac{m}{n}} \sum_{i=1}^n \eta_i^n(j), \quad \eta_i^n(j) = \nu_{1,m}^{-1}(\lambda_j) \left(w_i^{(n,m)}(\lambda_j) - \mathbb{E} \left[w_i^{(n,m)}(\lambda_j) \mid \mathcal{F}_{\frac{i-1}{n}} \right] \right).$$

More precisely, we show that $\bar{U}_n \xrightarrow{d_s} MN(0, \Theta(m, \bar{\lambda})IQ)$. Next, we again consider the case $k = 1$, $\bar{\lambda} = \lambda$. The second step is to define a new sequence:

$$U'_n = \nu_{1,m}^{-1}(\lambda) \sqrt{\frac{m}{n}} \sum_{i=1}^n \left(q_i(m, \lambda) - \mathbb{E} \left[q_i(m, \lambda) \mid \mathcal{F}_{\frac{i-1}{n}} \right] \right),$$

and show the result

$$U'_n - \bar{U}_n \xrightarrow{p} 0.$$

Finally, in part III, we prove the convergences:

$$\begin{aligned} & \sqrt{\frac{m}{n}} \sum_{i=1}^n \left(\nu_{1,m}^{-1}(\lambda) \mathbb{E} \left[q_i(m, \lambda) \mid \mathcal{F}_{\frac{i-1}{n}} \right] - \mathbb{E} \left[\xi_i^n \mid \mathcal{F}_{\frac{i-1}{n}} \right] \right) \xrightarrow{p} 0, \\ & \sqrt{\frac{m}{n}} \left(\sum_{i=1}^n \mathbb{E} \left[\xi_i^n \mid \mathcal{F}_{\frac{i-1}{n}} \right] - \int_0^1 \sigma_u^2 du \right) \xrightarrow{p} 0. \end{aligned}$$

Clearly, the afore-mentioned steps imply the assertion of Theorem 2.

Proof of part I Notice that:

$$\frac{m}{n} \sum_{i=1}^n \mathbb{E} \left[\eta_i^n(j) \eta_i^n(l) \mid \mathcal{F}_{\frac{i-1}{n}} \right] \xrightarrow{p} m \frac{\nu_1(m, \lambda_j, \lambda_l) - \nu_1(m, \lambda_j) \nu_1(m, \lambda_l)}{\nu_1(m, \lambda_j) \nu_1(m, \lambda_l)} \int_0^1 \sigma_u^4 du.$$

for any $1 \leq j, l \leq k$. Moreover, since $W \stackrel{d}{=} -W$ and $w_i^{(n,m)}(\lambda_j)$ is an even functional in W , we have

$$\mathbb{E} \left[\eta_i^n(j) \Delta_i^n W \mid \mathcal{F}_{\frac{i-1}{n}} \right] = 0, \quad 1 \leq j \leq k.$$

Next, let $H = (H_t)_{t \in [0,1]}$ be a bounded martingale on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, which is orthogonal to W (i.e., with quadratic covariation $[W, H] = 0$, almost surely). By the Clark's representation theorem (see, e.g., Karatzas and Shreve, 1998, Appendix E) we obtain

$$\eta_i^n(j) = \sigma_{\frac{i-1}{n}}^2 \int_{\frac{i-1}{n}}^{\frac{j}{n}} F_s^n(j) dW_s$$

for some predictable process $F_s^n(j)$. Then,

$$\sqrt{\frac{m}{n}} \sum_{i=1}^n \mathbb{E} \left[\eta_i^n \Delta_i^n H \mid \mathcal{F}_{\frac{i-1}{n}} \right] = 0,$$

because $[W, H] = 0$. Finally, stable convergence in law follows by Theorem IX 7.28 in Jacod and Shiryaev (2003):

$$\bar{U}_n \xrightarrow{d_s} MN(0, \Theta(m, \bar{\lambda})IQ),$$

which completes the proof of part I. □

Proof of part II We begin by setting

$$\delta_i^n = \nu_{1,m}^{-1}(\lambda) \sqrt{\frac{m}{n}} \left(q_i(m, \lambda) - w_i^{(n,m)}(\lambda) \right),$$

and obtain the identity:

$$U'_n - \bar{U}_n = \sum_{i=1}^n \left(\delta_i^n - \mathbb{E} \left[\delta_i^n \mid \mathcal{F}_{\frac{i-1}{n}} \right] \right).$$

To complete the second step, it suffices that

$$\sum_{i=1}^n \mathbb{E} [|\delta_i^n|^2] \rightarrow 0.$$

We omit the proof of this result, as it be shown by using exactly the same methods behind the proof of the convergence in (45) in Theorem 1. \square

Proof of part III It holds that:

$$\sqrt{\frac{m}{n}} \left(\sum_{i=1}^n \mathbb{E} \left[\xi_i^n \mid \mathcal{F}_{\frac{i-1}{n}} \right] - \int_0^1 \sigma_u^2 du \right) = \sqrt{mn} \sum_{i=1}^n \int_{\frac{i-1}{n}}^{\frac{i}{n}} \left(\sigma_{\frac{i-1}{n}}^2 - \sigma_u^2 \right) du.$$

Exploiting the results of Section 8 (Part 2) in Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006) (recall that m is a fixed number), we find that, under condition (V), the convergence

$$\sqrt{\frac{m}{n}} \left(\sum_{i=1}^n \mathbb{E} \left[\xi_i^n \mid \mathcal{F}_{\frac{i-1}{n}} \right] - \int_0^1 \sigma_u^2 du \right) \xrightarrow{p} 0, \quad (46)$$

holds. Now, we prove the first convergence of part III stated above. Under condition (V), we introduce the decomposition

$$\sqrt{N} D_i^m X - \beta_i^n = \mu_i^n(1) + \mu_i^n(2),$$

where $\mu_i^n(1)$ and $\mu_i^n(2)$ are m -dimensional vectors with components defined by

$$\begin{aligned} (\mu_i^n(1))_k &= \sqrt{N} \int_{\frac{i-1}{n} + \frac{k-1}{N}}^{\frac{i-1}{n} + \frac{k}{N}} \left(a_u - a_{\frac{i-1}{n}} \right) du + \sqrt{N} \int_{\frac{i-1}{n} + \frac{k-1}{N}}^{\frac{i-1}{n} + \frac{k}{N}} \left(\int_{\frac{i-1}{n} + \frac{k-1}{N}}^u a'_s ds \right. \\ &\quad \left. + \int_{\frac{i-1}{n} + \frac{k-1}{N}}^u \left(\sigma'_s - \sigma'_{\frac{i-1}{n}} \right) dW_s + \int_{\frac{i-1}{n} + \frac{k-1}{N}}^u \left(v'_s - v'_{\frac{i-1}{n}} \right) dB'_s \right) dW_u, \\ (\mu_i^n(2))_k &= \sqrt{N} \left(\frac{1}{N} a_{\frac{i-1}{n}} + \sigma'_{\frac{i-1}{n}} \int_{\frac{i-1}{n} + \frac{k-1}{N}}^{\frac{i-1}{n} + \frac{k}{N}} \left(W_u - W_{\frac{i-1}{n}} \right) dW_u \right. \\ &\quad \left. + v'_{\frac{i-1}{n}} \int_{\frac{i-1}{n} + \frac{k-1}{N}}^{\frac{i-1}{n} + \frac{k}{N}} \left(B'_u - B'_{\frac{i-1}{n}} \right) dW_u \right), \end{aligned} \quad (47)$$

for $k = 1, \dots, m$. Moreover, we decompose

$$\nu_{1,m}^{-1}(\lambda)q_i(m, \lambda) - \xi_i^n = \theta_i^n(1) + \theta_i^n(2),$$

where

$$\begin{aligned} \theta_i^n(1) &= \nu_{1,m}^{-1}(\lambda) \left(g_{\lambda m}^2 \left(\sqrt{N} D_i^m X \right) - g_{\lambda m}^2 (\beta_i^n + \mu_i^n(2)) \right. \\ &\quad \left. + g_{m-\lambda m+1}^2 \left(\sqrt{N} D_i^m X \right) - g_{m-\lambda m+1}^2 (\beta_i^n + \mu_i^n(2)) \right), \\ \theta_i^n(2) &= \nu_{1,m}^{-1}(\lambda) \left(g_{\lambda m}^2 (\beta_i^n + \mu_i^n(2)) - g_{\lambda m}^2 (\beta_i^n) \right. \\ &\quad \left. + g_{m-\lambda m+1}^2 (\beta_i^n + \mu_i^n(2)) - g_{m-\lambda m+1}^2 (\beta_i^n) \right). \end{aligned} \tag{48}$$

Using the same methods as for the proof of (45) in Theorem 1, we obtain

$$\sqrt{\frac{m}{n}} \sum_{i=1}^n \mathbb{E} [|\theta_i^n(1)|] \rightarrow 0,$$

which implies

$$\sqrt{\frac{m}{n}} \sum_{i=1}^n \mathbb{E} [\theta_i^n(1) | \mathcal{F}_{\frac{i-1}{n}}] \xrightarrow{p} 0.$$

Thus, we are left to prove that

$$\sqrt{\frac{m}{n}} \sum_{i=1}^n \mathbb{E} [\theta_i^n(2) | \mathcal{F}_{\frac{i-1}{n}}] \xrightarrow{p} 0.$$

Now we apply Lemma 1 to the term $\theta_i^n(2)$ with $\epsilon = N^{-1/2}$ and

$$x = \beta_i^n,$$

$$y = \sqrt{N} \mu_i^n(2).$$

Notice that as σ does not vanish, we have $(\beta_i^n)_k \neq (\beta_i^n)_l$ for all $1 \leq k < l \leq m$ almost surely, and consequently the assumptions of Lemma 1 are satisfied. Finally, we obtain the stochastic expansion

$$\begin{aligned} \sqrt{\frac{m}{n}} \sum_{i=1}^n \mathbb{E} [\theta_i^n(2) | \mathcal{F}_{\frac{i-1}{n}}] &= 2\nu_{1,m}^{-1}(\lambda) \sqrt{\frac{m}{n}} \sum_{i=1}^n \mathbb{E} \left[g_{\lambda m} (\beta_i^n) \left(\mu_i^n(2) \right)_{(\lambda m)*} \right. \\ &\quad \left. + g_{m-\lambda m+1} (\beta_i^n) \left(\mu_i^n(2) \right)_{(m-\lambda m+1)*} \mid \mathcal{F}_{\frac{i-1}{n}} \right] + o_p(1), \end{aligned}$$

where we recall that $(\lambda m)*$ is defined by

$$(\lambda m)* = k \quad \text{with} \quad (\beta_i^n)_k = (\beta_i^n)_{(\lambda m)}.$$

Now $(W, V) \stackrel{d}{=} -(W, V)$ and

$$g_{\lambda m}(\beta_i^n) \left(\mu_i^n(2) \right)_{(\lambda m)*} + g_{m-\lambda m+1}(\beta_i^n) \left(\mu_i^n(2) \right)_{(m-\lambda m+1)*}$$

is odd in (W, V) , which implies that

$$\mathbb{E} \left[g_{\lambda m}(\beta_i^n) \left(\mu_i^n(2) \right)_{(\lambda m)*} + g_{m-\lambda m+1}(\beta_i^n) \left(\mu_i^n(2) \right)_{(m-\lambda m+1)*} \mid \mathcal{F}_{\frac{i-1}{n}} \right] = 0. \quad (49)$$

Consequently,

$$\sqrt{\frac{m}{n}} \sum_{i=1}^n \mathbb{E} \left[\theta_i^n(2) \mid \mathcal{F}_{\frac{i-1}{n}} \right] \xrightarrow{p} 0,$$

which completes the proof of part III and, hence, Theorem 2 holds. \blacksquare

Proof of Proposition 1 Here we show that the CLT in Theorem 2 is robust to the presence of finite activity jumps. We consider a process X of the type (12). Let X^c denote the continuous part of X and set $J_n = \{1 \leq i \leq n \mid X \text{ jumps on the interval } [\frac{i-1}{n}, \frac{i}{n}]\}$. Again it is sufficient to assume that $k = 1$ and $\bar{\lambda} = \lambda \in (1/2, 1)$. Now we use the decomposition

$$\sqrt{N}(QRV_N(m, \lambda) - IV) = \sqrt{N} \left(\frac{1}{\nu_1(m, \lambda)} \frac{m}{N} \sum_{i \in J_n^c} q_i(m, \lambda) - IV \right) + \frac{1}{\nu_1(m, \lambda)} \frac{m}{\sqrt{N}} \sum_{i \in J_n} q_i(m, \lambda).$$

Recall that X has only finitely many jumps on compact intervals (a.s.), so the second sum on the right-hand side is finite. Therefore, the first term on the right-hand side converges stably to the limit described in Theorem 2 (since this is a statistic based on X^c). We need to show that

$$\frac{1}{\sqrt{N}} q_i(m, \lambda) \xrightarrow{p} 0$$

for any $i \in J_n$. It is well-known that the probability of having two jumps within the interval $[\frac{i-1}{n}, \frac{i}{n}]$ is negligible, so we assume that X jumps one time at $s \in [\frac{i-1}{n}, \frac{i}{n}]$ for some $i \in J_n$. Recall that $q_i(m, \lambda) = g_{\lambda m}^2(\sqrt{N}D_i^m X) + g_{m-\lambda m+1}^2(\sqrt{N}D_i^m X)$. Due to (44) we have that $(\Delta X_s = X_s - X_{s-})$

$$\frac{1}{\sqrt{N}} q_i(m, \lambda) 1_{\{|\Delta X_s| \geq 2||\sqrt{N}D_i^m X^c||\}} \leq \frac{C}{\sqrt{N}} ||\sqrt{N}D_i^m X^c||^2 \xrightarrow{p} 0,$$

since on $\{|\Delta X_s| \geq 2||\sqrt{N}D_i^m X^c||\}$ the jump is contained in $(D_i^m X)_{(0)}$ or $(D_i^m X)_{(m)}$, which both do not appear in $q_i(m, \lambda)$ because $\lambda \in (1/2, 1)$. On the other hand, we have that

$$\frac{1}{\sqrt{N}} q_i(m, \lambda) 1_{\{|\Delta X_s| < 2||\sqrt{N}D_i^m X^c||\}} \leq \frac{C \left(||\sqrt{N}D_i^m X^c||^2 + |\Delta X_s|^2 \right) ||D_i^m X^c||^2}{\sqrt{N} |\Delta X_s|^2} \xrightarrow{p} 0$$

again due to (44). This completes the proof of Proposition 1. \blacksquare

Proof of Theorem 3 In the following we assume that the process X is continuous (the robustness to finite activity jumps is shown as in Theorem 2). Here we show the CLT for the subsampled statistic $QRV_N^{sub}(m, \bar{\lambda}, \alpha)$. Note that the summands $q_i^{sub}(m, \lambda_j)$ ($1 \leq j \leq k$) in the definition of $QRV_N^{sub}(m, \bar{\lambda}, \alpha)$ are m-dependent (i.e. $q_i^{sub}(m, \lambda_j)$ and $q_l^{sub}(m, \lambda_j)$ are correlated for $|i - l| < m$), which makes the proof more complicated.

We will apply “big blocks & small blocks”-technique to break this dependence. More precisely, we will build big blocks of size pm , which will be separated by a small block of size m . This procedure ensures the (conditional) independence of big blocks, whereas the small blocks become asymptotically negligible when we later let p converge to infinity.

For this purpose we require some additional notations. First, set

$$a_i(p) = i(p+1)m, \quad b_i(p) = i(p+1)m + pm,$$

and let $A_i(p)$ denote the set of integers l with $a_i(p) \leq l < b_i(p)$ and $B_i(p)$ the set of integers l with $b_i(p) \leq l < a_{i+1}(p)$. Furthermore, let $j_N(p)$ denote the largest integer j with $b_j(p) \leq N$. Notice that $j_N(p) = O(N/p)$. Finally, we set $i_N(p) = (j_N(p) + 1)(p+1)m$.

Next, as in the proof of Theorem 2, we define an approximation of $D_{i,m}X$ by

$$D_{i,m}^l = \left(\sigma_{\frac{l}{N}} \Delta_j^N W \right)_{i \leq j \leq i+m-1}$$

with $l \leq i$, and we set

$$q_{i,l}^{sub}(m, \lambda) = g_{\lambda m}^2(\sqrt{N} D_{i,m}^l) + g_{m-\lambda m+1}^2(\sqrt{N} D_{i,m}^l).$$

We further set

$$\Upsilon_{i,l}^N = q_{i,l}^{sub}(m, \lambda) - \mathbb{E}[q_{i,l}^{sub}(m, \lambda) | \mathcal{F}_{\frac{l}{N}}],$$

and

$$\tilde{\Upsilon}_j^N = \begin{cases} \frac{1}{\nu_1(m, \lambda)} \frac{1}{N} \Upsilon_{j,a_i(p)}^N, & j \in A_i(p) \\ \frac{1}{\nu_1(m, \lambda)} \frac{1}{N} \Upsilon_{j,b_i(p)}^N, & j \in B_i(p) \\ \frac{1}{\nu_1(m, \lambda)} \frac{1}{N} \Upsilon_{j,i_N(p)}^N, & j \geq i_N(p) \end{cases}$$

Finally, we define

$$\zeta(p, 1)_j^N = \sum_{l=a_j(p)}^{b_j(p)-1} \tilde{\Upsilon}_l^N, \quad \zeta(p, 2)_j^N = \sum_{l=b_j(p)}^{a_{j+1}(p)-1} \tilde{\Upsilon}_l^N,$$

and

$$M(p)^N = \sum_{j=0}^{j_N(p)} \zeta(p, 1)_j^N, \quad N(p)^N = \sum_{j=0}^{j_N(p)} \zeta(p, 2)_j^N, \quad C(p)^N = \sum_{j=i_N(p)}^N \tilde{\Upsilon}_j^N$$

Notice that the big blocks are collected in $M(p)^N$, the small blocks are contained in $N(p)^N$ and $C(p)^N$ is the sum of the border terms.

Recall that the quantities $M(p)^N, N(p)^N, C(p)^N$ are constructed from the approximations of the true returns. Consequently, we obtain the decomposition

$$\sqrt{N} (QRV_N^{sub}(m, \lambda) - IV) = \sqrt{N} (M(p)^N + N(p)^N + C(p)^N) + \gamma_N(p) , \quad (50)$$

where $\gamma_N(p)$ stands for the approximation error when $D_{i,m}X$ is replaced by $D_{i,m}^l$ (plus the error that is due to the replacement of $1/(N-m+1)$ by $1/N$ in the definition of $\tilde{\Upsilon}_j^N$, which is obviously asymptotically negligible). Exactly as in Part II and III of the proof of Theorem 2 we deduce that

$$\lim_{p \rightarrow \infty} \limsup_{N \rightarrow \infty} P(|\gamma_N(p)| > \epsilon) = 0$$

for any $\epsilon > 0$. Since $C(p)^N$ contains a fixed number of summands (for fixed p and m) and each summand is of order $1/N$, we also obtain

$$\lim_{p \rightarrow \infty} \limsup_{N \rightarrow \infty} P(\sqrt{N}|C(p)^N| > \epsilon) = 0.$$

Now notice that the summands in the definition of $N(p)^N$ are uncorrelated. Consequently, we obtain

$$\mathbb{E}[|N(p)^N|^2] \leq \frac{C}{pN},$$

which implies that

$$\lim_{p \rightarrow \infty} \limsup_{N \rightarrow \infty} P(\sqrt{N}|N(p)^N| > \epsilon) = 0.$$

We are left to prove the CLT for $\sqrt{NM}(p)^N$, where $M(p)^N = (M(p)_1^N, \dots, M(p)_k^N)$ and each $M(p)_l^N$ is associated with $\lambda_l \in (1/2, 1)$. Set

$$M(p)_l^N = \sum_{j=0}^{j_N(p)} \zeta(p, 1)_{j,l}^N$$

to emphasize the dependence of $\zeta(p, 1)_{j,l}^N$ on λ_l . As in Part I of the proof of Theorem 2 we obtain

$$N^{1/2} \sum_{j=0}^{j_N(p)} \mathbb{E}[\zeta(p, 1)_{j,l}^N \Delta W(p)_j^N | \mathcal{F}_{\frac{a_j(p)}{N}}] = 0$$

$$N^{1/2} \sum_{j=0}^{j_N(p)} \mathbb{E}[\zeta(p, 1)_{j,l}^N \Delta H(p)_j^N | \mathcal{F}_{\frac{a_j(p)}{N}}] = 0 ,$$

where $\Delta Y(p)_j^N = Y_{\frac{b_j(p)}{N}} - Y_{\frac{a_j(p)}{N}}$ for any process Y , and H is a bounded martingale that is orthogonal to W . A straightforward computation shows that, for any fixed p ,

$$N \sum_{j=0}^{j_N(p)} \mathbb{E}[\zeta(p, 1)_{j,l}^N \zeta(p, 1)_{j,i}^N | \mathcal{F}_{\frac{a_j(p)}{N}}] \xrightarrow{p} \Theta^{sub}(m, \bar{\lambda})_{il}(p) IQ$$

with

$$\Theta^{sub}(m, \bar{\lambda})_{il}(p) = \frac{p}{p+1} \left(\frac{1}{m} \Theta(m, \bar{\lambda})_{il} + \frac{2}{\nu_1(m, \lambda_i)\nu_1(m, \lambda_l)} \sum_{k=1}^{m-1} \left(1 - \frac{k}{pm} \right) \text{cov} \left(|U_{(m\lambda_i)}^{(0)}|^2 + |U_{(m-m\lambda_i+1)}^{(0)}|^2, |U_{(m\lambda_l)}^{(k)}|^2 + |U_{(m-m\lambda_l+1)}^{(k)}|^2 \right) \right),$$

where $U^{(0)}$ and $U^{(k)}$ are defined in Theorem 3. Now, we deduce by Theorem IX 7.28 in Jacod and Shiryaev (2003):

$$\sqrt{N}M(p)^N \xrightarrow{d_s} Y_p = MN(0, \Theta^{sub}(m, \bar{\lambda})(p)IQ)$$

for any fixed p . On the other hand, we have that $Y_p \xrightarrow{p} Y = MN(0, \Theta^{sub}(m, \bar{\lambda})IQ)$ when $p \rightarrow \infty$. This completes the proof of Theorem 3. \blacksquare

Proof of Proposition 2 First, recall that $U_{(\lambda m)} \xrightarrow{p} c_\lambda$ as $m \rightarrow \infty$. Thus, we immediately obtain the identities

$$\nu_1(\lambda) = 2c_\lambda^2, \quad \nu_1(\lambda_i, \lambda_j) = 4c_{\lambda_i}^2 c_{\lambda_j}^2.$$

To prove the other identities we derive a joint CLT for $(U_{(\lambda_1 m)}^{(0)}, U_{(\lambda_2 m)}^{(k)})$ for $\lambda_1 \geq \lambda_2$ (and both are in the interval $(1/2, 1)$). Let Φ denote the $N(0, 1)$ -distribution and ϕ its density. The crucial step is the so-called Bahadur representation that says:

$$U_{(\lambda_1 m)}^{(0)} - c_{\lambda_1} = \frac{\lambda_1 - \hat{F}_n(c_{\lambda_1})}{\phi(c_{\lambda_1})} + O(m^{-3/4} \log m) \text{ a.s.,}$$

where

$$\hat{F}_n(c_{\lambda_1}) = \frac{1}{m} \sum_{i=1}^m 1_{\{U_i \leq c_{\lambda_1}\}}$$

is the empirical distribution function. Clearly, we also have

$$U_{(\lambda_2 m)}^{(k)} - c_{\lambda_2} = \frac{\lambda_2 - \hat{F}_n(c_{\lambda_2})^{(k)}}{\phi(c_{\lambda_2})} + O(m^{-3/4} \log m) \text{ a.s.,}$$

where

$$\hat{F}_n(c_{\lambda_2})^{(k)} = \frac{1}{m} \sum_{i=1+k}^{m+k} 1_{\{U_i \leq c_{\lambda_2}\}}.$$

Now we apply the CLT for empirical distribution functions. Assume that $k/m = \mu + o(m^{-1/2})$ ($\mu \in [0, 1]$). We obtain the following CLT:

$$\sqrt{m} \begin{pmatrix} \frac{1}{m} \sum_{i=1}^k 1_{\{U_i \leq c_{\lambda_1}\}} - \mu \lambda_1 \\ \frac{1}{m} \sum_{i=k+1}^m 1_{\{U_i \leq c_{\lambda_1}\}} - (1-\mu) \lambda_1 \\ \frac{1}{m} \sum_{i=k+1}^m 1_{\{U_i \leq c_{\lambda_2}\}} - (1-\mu) \lambda_2 \\ \frac{1}{m} \sum_{i=m}^{m+k} 1_{\{U_i \leq c_{\lambda_2}\}} - \mu \lambda_2 \end{pmatrix} \xrightarrow{d} N(0, \Sigma)$$

with

$$\Sigma = \begin{pmatrix} \mu\lambda_1(1-\lambda_1) & 0 & 0 & 0 \\ 0 & (1-\mu)\lambda_1(1-\lambda_1) & (1-\mu)\lambda_2(1-\lambda_1) & 0 \\ 0 & (1-\mu)\lambda_2(1-\lambda_1) & (1-\mu)\lambda_2(1-\lambda_2) & 0 \\ 0 & 0 & 0 & \mu\lambda_2(1-\lambda_2) \end{pmatrix}$$

Next, we apply the Δ -method for the function:

$$f(x, y, z, w) = \left(\frac{1}{\phi(c_{\lambda_1})}(x+y), \frac{1}{\phi(c_{\lambda_2})}(z+w) \right).$$

Clearly,

$$Df(x, y, z, w) = \begin{pmatrix} \frac{1}{\phi(c_{\lambda_1})} & \frac{1}{\phi(c_{\lambda_1})} & 0 & 0 \\ 0 & 0 & \frac{1}{\phi(c_{\lambda_2})} & \frac{1}{\phi(c_{\lambda_2})} \end{pmatrix}$$

and we obtain the CLT (set $Df = Df(\mu\lambda_1, (1-\mu)\lambda_1, \mu\lambda_2, (1-\mu)\lambda_2)$):

$$\sqrt{m} \begin{pmatrix} U_{(\lambda_1 m)}^{(0)} - c_{\lambda_1} \\ U_{(\lambda_2 m)}^{(k)} - c_{\lambda_2} \end{pmatrix} \xrightarrow{d} N(0, Df \Sigma (Df)')$$

with

$$Df \Sigma (Df)' = \begin{pmatrix} \frac{\lambda_1(1-\lambda_1)}{\phi^2(c_{\lambda_1})} & \frac{(1-\mu)\lambda_2(1-\lambda_1)}{\phi(c_{\lambda_1})\phi(c_{\lambda_2})} \\ \frac{(1-\mu)\lambda_2(1-\lambda_1)}{\phi(c_{\lambda_1})\phi(c_{\lambda_2})} & \frac{\lambda_2(1-\lambda_2)}{\phi^2(c_{\lambda_2})} \end{pmatrix}$$

Now apply again the Δ -method for the function $f(x, y) = (x^2, y^2)$:

$$\sqrt{m} \begin{pmatrix} |U_{(\lambda_1 m)}^{(0)}|^2 - c_{\lambda_1}^2 \\ |U_{(\lambda_2 m)}^{(k)}|^2 - c_{\lambda_2}^2 \end{pmatrix} \xrightarrow{d} N \left(0, 4 \begin{pmatrix} \frac{c_{\lambda_1}^2 \lambda_1(1-\lambda_1)}{\phi^2(c_{\lambda_1})} & \frac{c_{\lambda_1} c_{\lambda_2} (1-\mu)\lambda_2(1-\lambda_1)}{\phi(c_{\lambda_1})\phi(c_{\lambda_2})} \\ \frac{c_{\lambda_1} c_{\lambda_2} (1-\mu)\lambda_2(1-\lambda_1)}{\phi(c_{\lambda_1})\phi(c_{\lambda_2})} & \frac{c_{\lambda_2}^2 \lambda_2(1-\lambda_2)}{\phi^2(c_{\lambda_2})} \end{pmatrix} \right)$$

From the latter CLT we deduce that

$$\Theta(\bar{\lambda})_{12} = 2 \frac{(1-\lambda_1)(2\lambda_2-1)}{\phi(c_{\lambda_1})\phi(c_{\lambda_2})c_{\lambda_1}c_{\lambda_2}}.$$

Finally, recall that

$$\begin{aligned} \Theta^{sub}(m, \bar{\lambda})_{12} &= \frac{1}{m} \Theta(m, \bar{\lambda})_{12} \\ &+ \frac{2}{\nu_1(m, \lambda_1)\nu_1(m, \lambda_2)} \sum_{k=1}^{m-1} \text{cov} \left(|U_{(m\lambda_1)}^{(0)}|^2 + |U_{(m-m\lambda_1+1)}^{(0)}|^2, |U_{(m\lambda_2)}^{(k)}|^2 + |U_{(m-m\lambda_2+1)}^{(k)}|^2 \right). \end{aligned}$$

Clearly

$$\frac{1}{m} \Theta(m, \bar{\lambda})_{12} \rightarrow 0, \quad \nu_1(m, \lambda_i) \rightarrow 2c_{\lambda_i}^2, \quad \frac{1}{m} \sum_{k=1}^{m-1} \left(1 - \frac{k}{m} \right) \rightarrow 1/2$$

as $m \rightarrow \infty$. Hence, for the term $\Theta^{sub}(m, \bar{\lambda})_{12}$ we asymptotically obtain (by replacing again μ by k/m):

$$\Theta^{sub}(m, \bar{\lambda})_{12} \rightarrow 2 \frac{(1-\lambda_1)(2\lambda_2-1)}{\phi(c_{\lambda_1})\phi(c_{\lambda_2})c_{\lambda_1}c_{\lambda_2}}.$$

■

Proof of Theorem 4 As in Theorem 1 it is sufficient to consider the 1-dimensional case $k = 1$, $\bar{\lambda} = \lambda \in (1/2, 1)$. For $l \leq i$ we set

$$\beta_i^{*N} = N^{1/4} \{ \sigma_{\frac{i}{N}} \bar{W}_{i+(j-1)K}^N + \bar{u}_{i+(j-1)K}^N \}_{j=1}^m,$$

and define

$$w_i^{*(n,m)} = g_{\lambda m}^2(\beta_i^{*N}) + g_{m-\lambda m+1}^2(\beta_i^{*N}). \quad (51)$$

First, we state the following Lemma, which is crucial for our proofs (especially for the CLT).

Lemma 3 Assume that $\mathbb{E}(u_i^4) < \infty$. Then it holds that

$$\mathbb{E}[w_i^{*(n,m)} | \mathcal{F}_{\frac{i}{N}}] = \nu_1(m, \lambda)(c\psi_2\sigma_{\frac{i}{N}}^2 + \frac{c}{\psi_1}\omega^2) + o_p(N^{-1/4})$$

uniformly in i .

Proof of Lemma 3 By representation (26) we obtain the (conditional) convergence in distribution

$$\beta_i^{*N} | \mathcal{F}_{\frac{i}{N}} \xrightarrow{d} N_m \left(0, \text{diag} \left(c\psi_2\sigma_{\frac{i}{N}}^2 + \frac{c}{\psi_1}\omega^2, \dots, c\psi_2\sigma_{\frac{i}{N}}^2 + \frac{c}{\psi_1}\omega^2 \right) \right).$$

Denote by P_m^N the distribution of the left-hand side and by Φ_m the distribution of the right-hand side (by ϕ_m we denote the Lebesgue density of Φ_m). Now we apply the Edgeworth-expansion result presented in Lahiri (see Theorem 6.1 and 6.2 therein):

$$\left| \int (g_{\lambda m}^2 + g_{m-\lambda m+1}^2) d(P_m^N - \Phi_m - N^{-1/4} P_{m,1}) \right| = o_p(N^{-1/4}),$$

which holds under the condition $\mathbb{E}(u_i^4) < \infty$. The quantity $P_{m,1}$ is the second order term of the Edgeworth-expansion that has an odd density $p_{m,1}$ that is given by

$$p_{m,1}(x) = - \sum_{|\nu|=3} \frac{\chi_\nu}{\nu!} D^\nu \phi_m(x),$$

where $\nu = (\nu_1, \dots, \nu_m)$ is a nonnegative integer vector ($|\nu| = \nu_1 + \dots + \nu_m$, $\nu! = \nu_1! \cdots \nu_m!$), χ_ν are constants that depend on the cumulants of the marginal distribution of the process u and $D^\nu = \partial/\partial x_1^{\nu_1} \cdots \partial/\partial x_m^{\nu_m}$. Since $g_{\lambda m}^2 + g_{m-\lambda m+1}^2$ is an even function, we deduce that

$$\int (g_{\lambda m}^2 + g_{m-\lambda m+1}^2) dP_m(1) = 0$$

(because $p_{m,1}$ is odd). Thus

$$\mathbb{E}[w_i^{*(n,m)} | \mathcal{F}_{\frac{i}{N}}] = \nu_1(m, \lambda)(c\psi_2\sigma_{\frac{i}{N}}^2 + \frac{c}{\psi_1}\omega^2) + o_p(N^{-1/4}),$$

which holds uniformly in i because σ is bounded. This completes the proof of Lemma 3. ■

By the same methods as presented in the proof of Theorem 1 (see e.g. (42)) we conclude that

$$QRV_N^*(m, \lambda) - \frac{1}{\nu_1(m, \lambda)} \frac{1}{c\psi_2(N - m(K - 1) + 1)} \sum_{i=0}^{N-m(K-1)} w_i^{*(n,m)} \xrightarrow{p} 0. \quad (52)$$

Moreover, by Lemma 3, the convergence

$$\frac{1}{\nu_1(m, \lambda)} \frac{1}{c\psi_2(N - m(K - 1) + 1)} \sum_{i=0}^{N-m(K-1)} \mathbb{E}[w_i^{*(n,m)} | \mathcal{F}_{\frac{i}{N}}] \xrightarrow{p} IV + \frac{\psi_1 \omega^2}{c^2 \psi_2} \quad (53)$$

holds. In view of (52) and (53) we are left to proving

$$\frac{1}{\nu_1(m, \lambda)} \frac{1}{c\psi_2(N - m(K - 1) + 1)} \sum_{i=0}^{N-m(K-1)} \eta_i^{*N} \xrightarrow{p} 0, \quad \eta_i^{*N} = w_i^{*(n,m)} - \mathbb{E}[w_i^{*(n,m)} | \mathcal{F}_{\frac{i}{N}}]. \quad (54)$$

Observe that due to the construction of β_i^{*N} , the boundedness of σ and $\mathbb{E}(u_i^4) < \infty$ we obtain the estimate

$$\mathbb{E}[\eta_i^{*N} \eta_j^{*N}] \leq C, \quad |i - j| < m(K - 1), \quad (55)$$

whereas $\mathbb{E}[\eta_i^{*N} \eta_j^{*N}] = 0$ for $|i - j| \geq m(K - 1)$. Since m is fixed, we deduce the estimate

$$\mathbb{E}\left[\left| \frac{1}{\nu_1(m, \lambda)} \frac{1}{c\psi_2(N - m(K - 1) + 1)} \sum_{i=0}^{N-m(K-1)} \eta_i^{*N} \right|^2 \right] \leq \frac{C}{K}, \quad (56)$$

which completes the proof of Theorem 4. ■

Proof of Theorem 5 We assume that the process X is continuous (the robustness to finite activity jumps is shown as in Theorem 2) and show the CLT for the noise robust statistic $QRV_N^*(m, \bar{\lambda}, \alpha)$. The summands $q_i^*(m, \lambda_j)$ ($1 \leq j \leq k$) in the definition of $QRV_N^{sub}(m, \bar{\lambda}, \alpha)$ are now $m(K - 1)$ -dependent, so we have to apply “big blocks & small blocks”-technique once again to break this dependence. More precisely, we will build big blocks of size $pm(K - 1)$, which will be separated by a small block of size $m(K - 1)$ (again this procedure ensures the (conditional) independence of big blocks, whereas the small blocks become asymptotically negligible when we later let p converge to infinity). Quite often we will use the same notations as in the proof of Theorem 3 to emphasize the strong parallels between these proofs.

First, set

$$a_i(p) = i(p + 1)m(K - 1), \quad b_i(p) = i(p + 1)m(K - 1) + pm(K - 1),$$

and let $A_i(p)$ denote the set of integers l with $a_i(p) \leq l < b_i(p)$ and $B_i(p)$ the set of integers l with $b_i(p) \leq l < a_{i+1}(p)$. Furthermore, let $j_N(p)$ denote the largest integer j with $b_j(p) \leq N$. Notice that $j_N(p) = O(\sqrt{N}/p)$. Finally, we set $i_N(p) = (j_N(p) + 1)(p + 1)m(K - 1)$.

Next, we define an approximation of $\overline{D}_i^N Y$ by

$$\overline{D}_{i,l}^N = \{\sigma_{\frac{l}{N}} \overline{W}_{i+(j-1)(K-1)}^N + \overline{u}_{i+(j-1)(K-1)}^N\}_{j=1}^m$$

with $l \leq i$, and we set

$$q_{i,l}^*(m, \lambda) = g_{\lambda m}^2(N^{1/4} \overline{D}_{i,l}^N) + g_{m-\lambda m+1}^2(N^{1/4} \overline{D}_{i,l}^N).$$

We further set

$$\Upsilon_{i,l}^N = q_{i,l}^*(m, \lambda) - \mathbb{E}[q_{i,l}^*(m, \lambda) | \mathcal{F}_{\frac{i}{N}}],$$

and

$$\tilde{\Upsilon}_j^N = \begin{cases} \frac{1}{\nu_1(m, \lambda)} \frac{1}{c\psi_2 N} \Upsilon_{j, a_i(p)}^N, & j \in A_i(p) \\ \frac{1}{\nu_1(m, \lambda)} \frac{1}{c\psi_2 N} \Upsilon_{j, b_i(p)}^N, & j \in B_i(p) \\ \frac{1}{\nu_1(m, \lambda)} \frac{1}{c\psi_2 N} \Upsilon_{j, i_N(p)}^N, & j \geq i_N(p) \end{cases}$$

Finally, we define

$$\zeta(p, 1)_j^N = \sum_{l=a_j(p)}^{b_j(p)-1} \tilde{\Upsilon}_l^N, \quad \zeta(p, 2)_j^N = \sum_{l=b_j(p)}^{a_{j+1}(p)-1} \tilde{\Upsilon}_l^N,$$

and

$$M(p)^N = \sum_{j=0}^{j_N(p)} \zeta(p, 1)_j^N, \quad N(p)^N = \sum_{j=0}^{j_N(p)} \zeta(p, 2)_j^N, \quad C(p)^N = \sum_{j=i_N(p)}^N \tilde{\Upsilon}_j^N$$

Notice that the big blocks are collected in $M(p)^N$, the small blocks are contained in $N(p)^N$ and $C(p)^N$ is the sum of the border terms.

Recall that the terms $M(p)^N, N(p)^N, C(p)^N$ are constructed from the approximations $\overline{D}_{i,l}^N$. Thus

$$N^{1/4} \left(QRV_N^*(m, \bar{\lambda}, \alpha) - \frac{\psi_1}{c^2 \psi_2} \hat{\omega}^2 - IV \right) = N^{1/4} (M(p)^N + N(p)^N + C(p)^N) + \gamma_N(p), \quad (57)$$

where $\gamma_N(p)$ stands for the approximation error when $\overline{D}_i^N Y$ is replaced by $\overline{D}_{i,l}^N$ (plus the error that is due to the replacement of $1/(N-m(K-1)+1)$ by $1/N$ in the definition of $\tilde{\Upsilon}_j^N$, which is obviously asymptotically negligible). Now, the convergence

$$\lim_{p \rightarrow \infty} \limsup_{N \rightarrow \infty} P(|\gamma_N(p)| > \epsilon) = 0,$$

for any $\epsilon > 0$, follows through the lines of Part II and III of the proof of Theorem 2. In fact, there are two additional difficulties that has to be shown in a different way. First of all we have to prove (under assumption (V)) that

$$\frac{1}{\nu_1(m, \lambda)} \frac{1}{c\psi_2(N-m(K-1)+1)} \sum_{i=0}^{N-m(K-1)} \mathbb{E}[w_i^{*(n,m)} | \mathcal{F}_{\frac{i}{N}}] - \left(IV + \frac{\psi_1 \omega^2}{c^2 \psi_2} \right) = o_p(N^{-1/4})$$

(which is the counterpart of (46)). But the afore-mentioned estimate follows immediately from Lemma 3 and (46). The other difficulty arises when we have to deal with the counterpart of the identity (49), which in this case would involve a noise term (and have a slightly different form). However, this type of identity remains true, because $(W, V, u) \stackrel{d}{=} -(W, V, u)$ since the marginal distribution of u is assumed to be symmetric around 0.

Since $C(p)^N$ contains at most $(p+1)m(K-1)$ summands and each summand is of order $1/N$, we also obtain

$$\lim_{p \rightarrow \infty} \limsup_{N \rightarrow \infty} P(N^{1/4}|C(p)^N| > \epsilon) = 0.$$

Now notice that the summands in the definition of $N(p)^N$ are uncorrelated and each summand is of order K/N . Consequently, we obtain

$$\mathbb{E}\left[|N(p)^N|^2\right] \leq \frac{C}{p\sqrt{N}},$$

which implies that

$$\lim_{p \rightarrow \infty} \limsup_{N \rightarrow \infty} P(N^{1/4}|N(p)^N| > \epsilon) = 0.$$

We are left to proving the CLT for $N^{1/4}M(p)^N$, where $M(p)^N = (M(p)_1^N, \dots, M(p)_k^N)$ and each $M(p)_l^N$ is associated with $\lambda_l \in (1/2, 1)$. Set

$$M(p)_l^N = \sum_{j=0}^{j_N(p)} \zeta(p, 1)_{j,l}^N$$

to emphasize the dependence of $\zeta(p, 1)_{j,l}^N$ on λ_l (we also associate $\Upsilon_{j,i}^N(l)$ with λ_l).

As in Part I of the proof of Theorem 2 we obtain

$$N^{1/4} \sum_{j=0}^{j_N(p)} \mathbb{E}[\zeta(p, 1)_{j,l}^N \Delta W(p)_j^N | \mathcal{F}_{\frac{a_j(p)}{N}}] = 0,$$

where $\Delta Y(p)_j^N = Y_{\frac{b_j(p)}{N}} - Y_{\frac{a_j(p)}{N}}$ for any process Y . As in Jacod, Li, Mykland, Podolskij, and Vetter (2009) (see (5.58)) we have that

$$N^{1/4} \sum_{j=0}^{j_N(p)} \mathbb{E}[\zeta(p, 1)_{j,l}^N \Delta H(p)_j^N | \mathcal{F}_{\frac{a_j(p)}{N}}] \xrightarrow{p} 0$$

for any bounded martingale H that is orthogonal to W .

Now notice that (26) implies the identities

$$N^{1/2} \mathbb{E}[\bar{W}_j^N \bar{W}_i^N] = cw_h \left(\frac{|j-i|}{K} \right) + O(K^{-1}), \quad N^{1/2} \mathbb{E}[\bar{u}_j^N \bar{u}_i^N] = \frac{1}{c} w_{h'} \left(\frac{|j-i|}{K} \right) \omega^2 + O(K^{-1}), \quad |j-i| < K-1, \quad (58)$$

where the functional $w_f(u)$ is given in Definition 2. When $|j - i| \geq K - 1$ the above covariances are 0. Assume now that $j > i$ with $j = i + (l - 1)(K - 1) + d$ for some $1 \leq l \leq m$ and $0 \leq d \leq K - 2$, and $a_z(p) \leq i, j \leq b_z(p) - 1$. Due to the identities in (58) we obtain

$$\mathbb{E}[\Upsilon_{j,a_z(p)}^N(l_1)\Upsilon_{i,a_z(p)}^N(l_2)|\mathcal{F}_{\frac{a_z(p)}{N}}] = f_{m,l,\sigma_{a_z(p)},\frac{d}{K}}(\lambda_{l_1}, \lambda_{l_2}) + O_p(K^{-1}),$$

where $f_{m,l,x,u}(\lambda_{l_1}, \lambda_{l_2})$ is given in Definition 2 ($1 \leq l_1, l_2 \leq k$). This implies that

$$\begin{aligned} & N^{1/2} \sum_{j=0}^{j_N(p)} \mathbb{E}[\zeta(p, 1)_{j,l_1}^N \zeta(p, 1)_{j,l_2}^N | \mathcal{F}_{\frac{a_j(p)}{N}}] \\ & \xrightarrow{p} \frac{p}{p+1} \frac{2\nu_{1,m}^{-1}(\lambda_{l_1}) \nu_{1,m}^{-1}(\lambda_{l_2})}{c\psi_2^2} \sum_{l=1}^m \left(1 - \frac{l-1}{pm}\right) \int_0^1 \int_0^1 f_{m,l,\sigma_t,u}(\lambda_{l_1}, \lambda_{l_2}) dt du = \Gamma(p)_{l_1, l_2}. \end{aligned}$$

Now, we deduce by Theorem IX 7.28 in Jacod and Shiryaev (2003):

$$\sqrt{N}M(p)^N \xrightarrow{d_s} Z_p = MN(0, \Gamma(p))$$

for any fixed p . On the other hand, we have that $Z_p \xrightarrow{p} Z = MN\left(0, \frac{2}{c\psi_2^2} \Sigma_m(\lambda_1, \dots, \lambda_k)\right)$ when $p \rightarrow \infty$. This completes the proof of Theorem 5. \blacksquare

Proof of Proposition 3 Due to the polarization identity it is sufficient to prove Proposition 3 for $k = 1$, $\bar{\lambda} = \lambda \in (1/2, 1)$. Recall the definition of $w_i^{*(n,m)}$ in (51). As in the proof of Theorem 4 we have that

$$\begin{aligned} & \frac{\nu_1^{-2}(m, \lambda)}{c\psi_2^2(K-1)(N-3m(K-1)+3)} \sum_{i=m(K-1)-1}^{N-2m(K-1)+1} \left(q_i^*(m, \lambda) \left(\sum_{j=i-m(K-1)+1}^{i+m(K-1)-1} \{q_j^*(m, \lambda) - q_{i+m(K-1)}^*(m, \lambda)\} \right) \right. \\ & \left. - w_i^{*(n,m)} \left(\sum_{j=i-m(K-1)+1}^{i+m(K-1)-1} \{w_j^{*(n,m)} - w_{i+m(K-1)}^{*(n,m)}\} \right) \right) \xrightarrow{p} 0. \end{aligned}$$

Next, a straightforward (but somewhat tedious) calculation shows that

$$\begin{aligned} & \frac{\nu_1^{-2}(m, \lambda)}{c\psi_2^2(K-1)(N-3m(K-1)+3)} \times \\ & \sum_{i=m(K-1)-1}^{N-2m(K-1)+1} \mathbb{E} \left[w_i^{*(n,m)} \left(\sum_{j=i-m(K-1)+1}^{i+m(K-1)-1} \{w_j^{*(n,m)} - w_{i+m(K-1)}^{*(n,m)}\} \right) \middle| \mathcal{F}_{\frac{i-m(K-1)+1}{N}} \right] \\ & \xrightarrow{p} \frac{2\nu_{1,m}^{-2}(\lambda)}{c\psi_2^2} \sum_{l=1}^m \int_0^1 \int_0^1 f_{m,l,\sigma_t,u}(\lambda) dt du. \end{aligned}$$

On the other hand, we deduce that

$$\begin{aligned} & \frac{\nu_1^{-2}(m, \lambda)}{c\psi_2^2(K-1)(N-3m(K-1)+3)} \sum_{i=m(K-1)-1}^{N-2m(K-1)+1} \left(w_i^{*(n,m)} \left(\sum_{j=i-m(K-1)+1}^{i+m(K-1)-1} \{w_j^{*(n,m)} - w_{i+m(K-1)}^{*(n,m)}\} \right) \right. \\ & \left. - \mathbb{E} \left[w_i^{*(n,m)} \left(\sum_{j=i-m(K-1)+1}^{i+m(K-1)-1} \{w_j^{*(n,m)} - w_{i+m(K-1)}^{*(n,m)}\} \right) \middle| \mathcal{F}_{\frac{i-m(K-1)+1}{N}} \right] \right) \xrightarrow{p} 0 \end{aligned}$$

as in (54). This completes the proof of Proposition 3. ■

References

- Aït-Sahalia, Y., 1996, “Testing continuous-time models of the spot interest rate,” *Review of Financial Studies*, 9(2), 385–426.
- , 2004, “Disentangling diffusion from jumps,” *Journal of Financial Economics*, 74(3), 487–528.
- Aït-Sahalia, Y., and J. Jacod, 2009a, “Estimating the degree of activity of jumps in high frequency data,” *Annals of Statistics*, 37(5), 2202–2244.
- , 2009b, “Testing for jumps in a discretely observed process,” *Annals of Statistics*, 37(1), 184–222.
- Aït-Sahalia, Y., P. A. Mykland, and L. Zhang, 2005, “How often to sample a continuous-time process in the presence of market microstructure noise,” *Review of Financial Studies*, 18(2), 351–416.
- , 2011, “Ultra high frequency volatility estimation with dependent microstructure noise,” *Journal of Econometrics*, 160(1), 160–175.
- Aldous, D. J., and G. K. Eagleson, 1978, “On mixing and stability of limit theorems,” *Annals of Probability*, 6(2), 325–331.
- Andersen, T. G., and T. Bollerslev, 1998, “Answering the skeptics: Yes, standard volatility models do provide accurate forecasts,” *International Economic Review*, 39(4), 885–905.
- Andersen, T. G., T. Bollerslev, and F. X. Diebold, 2007, “Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility,” *Review of Economics and Statistics*, 89(4), 701–720.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens, 2001, “The distribution of realized stock return volatility,” *Journal of Financial Economics*, 61(1), 43–76.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys, 2000, “Great realizations,” *Risk*, 13(3), 105–108.
- , 2001, “The distribution of realized exchange rate volatility,” *Journal of the American Statistical Association*, 96(453), 42–55.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and J. Wu, 2006, “Realized beta: Persistence and predictability,” in *Advances in Econometrics*, ed. by T. B. Fomby, and D. Terrell. Cambridge University Press, Cambridge, pp. 1–40.
- Andersen, T. G., D. Dobrev, and E. Schaumburg, 2008, “Duration-based volatility estimation,” Working paper, Northwestern University.
- Back, K., 1991, “Asset prices for general processes,” *Journal of Mathematical Economics*, 20(4), 371–395.
- Bakshi, G. S., N. Ju, and H. Ou-Yang, 2006, “Estimation of continuous-time models with an application to equity volatility,” *Journal of Financial Economics*, 82(1), 227–249.
- Bandi, F. M., and J. R. Russell, 2006, “Separating microstructure noise from volatility,” *Journal of Financial Economics*, 79(3), 655–692.
- , 2008, “Microstructure noise, realized variance, and optimal sampling,” *Review of Economic Studies*, 75(2), 339–369.

- Barndorff-Nielsen, O. E., S. E. Graversen, J. Jacod, M. Podolskij, and N. Shephard, 2006, “A central limit theorem for realized power and bipower variations of continuous semimartingales,” in *From Stochastic Calculus to Mathematical Finance: The Shiryaev Festschrift*, ed. by Y. Kabanov, R. Lipster, and J. Stoyanov. Springer, Heidelberg, pp. 33–68.
- Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard, 2008, “Designing realized kernels to measure the ex post variation of equity prices in the presence of noise,” *Econometrica*, 76(6), 1481–1536.
- Barndorff-Nielsen, O. E., and N. Shephard, 2002, “Econometric analysis of realized volatility and its use in estimating stochastic volatility models,” *Journal of the Royal Statistical Society: Series B*, 64(2), 253–280.
- , 2004, “Power and bipower variation with stochastic volatility and jumps,” *Journal of Financial Econometrics*, 2(1), 1–48.
- , 2006, “Econometrics of testing for jumps in financial economics using bipower variation,” *Journal of Financial Econometrics*, 4(1), 1–30.
- Barndorff-Nielsen, O. E., N. Shephard, and M. Winkel, 2006, “Limit theorems for multipower variation in the presence of jumps,” *Stochastic Processes and their Applications*, 116(5), 796–806.
- Bessembinder, H., and M. L. Lemmon, 2002, “Equilibrium pricing and optimal hedging in electricity forward markets,” *Journal of Finance*, 57(3), 1347–1382.
- Bollen, B., and B. Inder, 2002, “Estimating daily volatility in financial markets utilizing intraday data,” *Journal of Empirical Finance*, 9(5), 551–562.
- Bollerslev, T., U. Kretschmer, C. Pigorsch, and G. Tauchen, 2009, “A discrete-time model for daily S&P500 returns and realized variations: jumps and leverage effects,” *Journal of Econometrics*, 150(2), 151–166.
- Chernov, M., A. R. Gallant, E. Ghysels, and G. Tauchen, 2003, “Alternative models for stock price dynamics,” *Journal of Econometrics*, 116(1–2), 225–257.
- Christensen, K., R. C. A. Oomen, and M. Podolskij, 2010, “Appendix to Realised quantile-based estimation of the integrated variance,” Working paper.
- Christensen, K., and M. Podolskij, 2009, “Range-based estimation of quadratic variation,” Working paper.
- Corsi, F., and R. Renò, 2012, “Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling,” *Journal of Business and Economic Statistics*, 30(3), 368–380.
- Corsi, F., G. Zumbach, U. A. Müller, and M. M. Dacorogna, 2001, “Consistent high-precision volatility from high-frequency data,” *Economic Notes*, 30(2), 183–204.
- David, H. A., 1970, *Order Statistics*. John Wiley and Sons, Chichester, 1st edn.
- Diebold, F. X., and G. H. Strasser, 2013, “On the correlation structure of microstructure noise: A financial economic approach,” *Review of Economic Studies*, 80(4), 1304–1337.
- Duffie, D., and J. Pan, 2001, “Analytical value-at-risk with jumps and credit risk,” *Finance and Stochastics*, 5(2), 155–180.
- Duffie, D., J. Pan, and K. J. Singleton, 2000, “Transform analysis and asset pricing for affine jump-diffusions,” *Econometrica*, 68(6), 1343–1376.

- Eisenberger, I., and E. C. Posner, 1965, "Systematic statistics used for data compression in space telemetry," *Journal of the American Statistical Association*, 60(309), 97–133.
- Epps, T. W., 1979, "Comovements in stock prices in the very short run," *Journal of the American Statistical Association*, 74(366), 291–298.
- Eraker, B., M. Johannes, and N. Polson, 2003, "The impact of jumps in volatility and returns," *Journal of Finance*, 58(3), 1269–1300.
- Fan, J., and Y. Wang, 2007, "Multi-scale jump and volatility analysis for high-frequency financial data," *Journal of the American Statistical Association*, 102(480), 1349–1362.
- Fisher, L., 1966, "Some new stock-market indexes," *Journal of Business*, 39(1–2), 191–225.
- Gatheral, J., and R. C. A. Oomen, 2010, "Zero-intelligence realized variance estimation," *Finance and Stochastics*, 14(2), 249–283.
- Gloster, A., and J. Jacod, 2001a, "Diffusions with measurement errors. I - local asymptotic normality," *ESAIM: Probability and Statistics*, 5, 225–242.
- , 2001b, "Diffusions with measurement errors. II - measurement errors," *ESAIM: Probability and Statistics*, 5, 243–260.
- Hansen, P. R., J. Large, and A. Lunde, 2008, "Moving average-based estimators of integrated variance," *Econometric Reviews*, 27(1–3), 79–111.
- Hansen, P. R., and A. Lunde, 2006, "Consistent ranking of volatility models," *Journal of Econometrics*, 131(1–2), 97–121.
- Hasbrouck, J., 2007, *Empirical Market Microstructure*. Oxford University Press, Oxford, 1st edn.
- Huang, X., and G. Tauchen, 2005, "The relative contribution of jumps to total price variance," *Journal of Financial Econometrics*, 3(4), 456–499.
- Jacod, J., 1994, "Limit of random measures associated with the increments of a Brownian semimartingale," Preprint number 120, Laboratoire de Probabilités, Université Pierre et Marie Curie, Paris.
- , 2008, "Asymptotic properties of realized power variations and related functionals of semimartingales," *Stochastic Processes and their Applications*, 118(4), 517–559.
- Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter, 2009, "Microstructure noise in the continuous case: The pre-averaging approach," *Stochastic Processes and their Applications*, 119(7), 2249–2276.
- Jacod, J., and P. E. Protter, 1998, "Asymptotic error distributions for the Euler method for stochastic differential equations," *Annals of Probability*, 26(1), 267–307.
- Jacod, J., and A. N. Shiryaev, 2003, *Limit Theorems for Stochastic Processes*. Springer, Berlin, 2nd edn.
- Jarrow, R. A., and E. R. Rosenfeld, 1984, "Jump risks and the intertemporal capital asset pricing model," *Journal of Business*, 57(3), 337–351.
- Jiang, G. J., and R. C. A. Oomen, 2008, "Testing for jumps when asset prices are observed with noise – A "swap variance" approach," *Journal of Econometrics*, 144(2), 352–370.
- 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(6), 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(6), 2535–2563.
- Liu, J., F. Longstaff, and J. Pan, 2003, "Dynamic asset allocation with event risk," *Journal of Finance*, 58(1), 231–259.
- Longstaff, F. A., and A. W. Wang, 2004, "Electricity forward prices: A high-frequency empirical analysis," *Journal of Finance*, 59(4), 1877–1900.
- Mancini, C., 2004a, "Estimating the integrated volatility in stochastic volatility models with Lévy type jumps," working paper, manuscript Università di Firenze.
- , 2004b, "Estimation of the characteristics of jump of a general Poisson-diffusion model," *Scandinavian Actuarial Journal*, 2004(1), 42–52.
- , 2009, "Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps," *Scandinavian Journal of Statistics*, 36(2), 270–296.
- Merton, R. C., 1976, "Option pricing when underlying stock returns are discontinuous," *Journal of Financial Economics*, 3(1–2), 125–144.
- Mosteller, F., 1946, "On some useful "inefficient" statistics," *Annals of Mathematical Statistics*, 17(4), 377–408.
- Mykland, P. A., 2010, "A Gaussian calculus for inference from high frequency data," *Annals of Finance*, 8(2), 235–258.
- NASDAQ Market Data Distribution, 2008, "UTP trade data feed (UTDF) specifications," Version 10.4d, Available at: <http://www.nasdaqtrader.com/>.
- Niederhoffer, V., and M. F. M. Osborne, 1966, "Market making and reversal on the stock exchange," *Journal of the American Statistical Association*, 61(316), 897–916.
- Nolte, I., and V. Voev, 2007, "Estimating high-frequency based (co-) variances: A unified approach," Working paper, University of Konstanz.
- Oomen, R. C. A., 2006, "Comment on 2005 JBES invited address "Realized variance and market microstructure noise" by Peter R. Hansen and Asger Lunde," *Journal of Business and Economic Statistics*, 24(2), 195–202.
- Pearson, K., 1920, "On the probable errors of frequency constants. Part III," *Biometrika*, 13(1), 113–132.
- Podolskij, M., and M. Vetter, 2009, "Bipower-type estimation in a noisy diffusion setting," *Stochastic Processes and their Applications*, 119(9), 2803–2831.
- Protter, P. E., 2004, *Stochastic Integration and Differential Equations*. Springer, Berlin, 1st edn.
- Rényi, A., 1963, "On stable sequences of events," *Sankhya: The Indian Journal of Statistics; Series A*, 25(3), 293–302.
- Schwert, G. W., 1989, "Why does stock market volatility change over time," *Journal of Finance*, 44(5), 1115–1153.
- Sun, Y., 2006, "Best quadratic unbiased estimators of integrated variance in the presence of market microstructure noise," Working paper, University of California, San Diego.

- van der Vaart, A. W., 1998, *Asymptotic Statistics*. Cambridge University Press, Cambridge, 1st edn.
- Woerner, J. H. C., 2006, “Power and multipower variation: Inference for high-frequency data,” in *Proceedings of the International Conference on Stochastic Finance 2004*, ed. by A. N. Shiryaev, M. do R. Grossinho, P. Oliveira, and M. Esquivel, pp. 343–364, Berlin. Springer.
- Zhang, L., 2006, “Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach,” *Bernoulli*, 12(6), 1019–1043.
- Zhang, L., P. A. Mykland, and Y. Aït-Sahalia, 2005, “A tale of two time scales: determining integrated volatility with noisy high-frequency data,” *Journal of the American Statistical Association*, 100(472), 1394–1411.
- Zhou, B., 1996, “High-frequency data and volatility in foreign-exchange rates,” *Journal of Business and Economic Statistics*, 14(1), 45–52.