AARHUS
UNIVERSITY
BUSINESS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS AND BUSINESS

CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2011-47

Asymptotic theory of range-based multipower variation

Kim Christensen and Mark Podolskij

Department of Economics and Business Aarhus University Bartholins Allé 10 DK-8000 Aarhus C Denmark	Email: oekonomi@au.dk Tel: +45 8942 1610

Asymptotic theory of range-based multipower variation

Kim Christensen[*] **Mark Podolskij**[†]

This version: October, 2011

Forthcoming in the Journal of Financial Econometrics

Abstract

In this paper, we present a realised range-based multipower variation theory, which can be used to estimate return variation and draw jump-robust inference about the diffusive volatility component, when a high-frequency record of asset prices is available. The standard range-statistic – routinely used in financial economics to estimate the variance of securities prices – is shown to be biased when the price process contains jumps. We outline how the new theory can be applied to remove this bias by constructing a hybrid range-based estimator. Our asymptotic theory also reveals that when high-frequency data are sparsely sampled, as is often done in practice due to the presence of microstructure noise, the range-based multipower variations can produce significant efficiency gains over comparable subsampled return-based estimators. The analysis is supported by a simulation study and we illustrate the practical use of our framework on some recent TAQ equity data.

JEL Classification: C10; C80.

Keywords: High-frequency data; Integrated variance; Realised multipower variation; Realised range-based multipower variation; Quadratic variation.

[*]Corresponding author: Aarhus University, CREATES, Department of Economics and Business, University Park, Building 1322, 8000 Aarhus C, Denmark. Phone: (+45) 87 16 55 02, e-mail: kchristensen@creates.au.dk.

[†]University of Heidelberg, Institute of Applied Mathematics, Im Neuenheimer Feld 294, Room 215, 69120 Heidelberg, Germany. E-mail: m.podolskij@uni-heidelberg.de.

1 Introduction

The standard, arbitrage-free continuous time setting for securities prices in financial economics shows that, in frictionless markets, return variation admits a general decomposition into a continuous, diffusive volatility component and discontinuous jumps (e.g. Andersen, Bollerslev, Diebold, and Labys, 2003; Back, 1991; Delbaen and Schachermayer, 1994). In the past few years, our ability to assess the relative significance of these two fundamentally distinct sources of risk has taken a major step forward with the increasing availability and use of high-frequency data. This has opened the way for non-parametric estimation of the quadratic return variation and, for example, allows us to split this composite measure of risk into the integrated variance and the sum of the squared jumps. It follows a long-standing tradition within asset- and derivatives pricing, portfolio allocation and risk management of using low-frequency data (e.g., daily or weekly) to fit parametric stochastic volatility or jump-diffusion models (e.g. Alizadeh, Brandt, and Diebold, 2002; Andersen, Benzoni, and Lund, 2002; Bates, 1996, 2011; Chernov, Gallant, Ghysels, and Tauchen, 2003; Heston, 1993; Hull and White, 1987; Gallant, Hsu, and Tauchen, 1999).

A large body of work in the high-frequency space is based on a framework called realised (return-based) multipower variation, which relies on statistics constructed from intraday returns (e.g. Aït-Sahalia and Jacod, 2009a,b, 2011; Andersen, Bollerslev, and Diebold, 2007; Barndorff-Nielsen and Shephard, 2004, 2006; Barndorff-Nielsen, Shephard, and Winkel, 2006; Corsi, Pirino, and Renò, 2010; Dobrev and Szerszen, 2010; Huang and Tauchen, 2005; Mancini, 2004, 2009; Todorov, 2009). In practice, the presence of market imperfections (such as price discreteness and bid-ask spreads) means that standard return-based multipower variation estimators often use sparse sampling (e.g. Bollerslev and Todorov, 2011; Corsi and Renò, 2010; Tauchen and Zhou, 2011). The idea being that at a moderate frequency, for example the 5-minute frequency, the impact of microstructure noise is small enough to be ignored. Of course, for many liquid assets the data are much more abundant, so this principle often entails a significant loss of information, and much recent research has focused on developing estimators that are more resistant to the noise (e.g. Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2008; Christensen, Oomen, and Podolskij, 2010; Fan and Wang, 2007; Jacod, Li, Mykland, Podolskij, and

Vetter, 2009; Large, 2010; Podolskij and Vetter, 2009a; Zhang, Mykland, and Aït-Sahalia, 2005, among others). A branch of related work can be found in, e.g., Andersen, Dobrev, and Schaumburg (2008); Lee and Mykland (2008a,b); Mykland, Shephard, and Sheppard (2010).

In this paper, we formulate a complete realised *range-based* multipower variation theory, which builds directly on the return-based multipower variation by simply replacing absolute returns with ranges, suitably scaled. We outline how it allows us to estimate aggregate return variation and form jump-robust estimates of integrated variance using the range. The analysis shows that if high-frequency data are being sparsely sampled, using a realised range-statistic can produce considerable efficiency gains relative to a standard return-based estimator, even when the latter employs subsampling to exhaust the entire database (e.g. Zhang, Mykland, and Aït-Sahalia, 2005; Zhou, 1996). Intuitively, the range partially distills some of the information contained in intermediate data not used by a sparsely sampled return-based estimator, and this turns out to be a more effective way of doing it compared to subsampling of low-frequency returns. Another appealing key feature of this theory is that realised range-based multipower variation estimators can be made increasingly robust to jumps without losing asymptotic efficiency. We use both simulations and empirical data to illustrate how these findings manifest, for example in order to construct feasible jump-robust confidence intervals for the integrated variance. Moreover, we show that this advantage largely prevails also in the presence of a realistic level of market microstructure noise. At low-frequency sampling, the range therefore offers a parsimonious, yet highly efficient, framework, which avoids the need for doing complicated corrections in order to combat the noise.

Indeed, the main motivation for using the range is that realised range-based estimation of the integrated variance is known to be very efficient in pure diffusion models (e.g. Christensen and Podolskij, 2007; Martens and van Dijk, 2007; Parkinson, 1980). Interestingly, however, the properties of the high-low remain unchartered territory in jump-diffusion models and, as we show here, the standard range-estimator suffers from systematic biases, when jumps are added to the price equation. We propose to rectify this bias using a hybrid range-estimator, which has the form of a linear combination of the original range-statistic and a jump-robust measure of integrated variance.

It should be noted that the way we retrieve jump-robust measures of integrated multipower variation follows exactly the procedure of the return-based framework, which depends on terms in the proximity of a jump to get small. As such, it inherits some of the weaknesses associated with this approach, although our simulations do demonstrate that the range has superior finite sample robustness. In this respect, an interesting route is adopted by Dobrev (2007), who extends the standard range-statistic (used here) – based on the single largest price move – into a generalized range theory, which maximizes the sum of multiple price moves. The generalized range is also jump-robust (potentially to some forms of infinite activity jump processes), but this feature descends from scaling constants, which may entail some advantages in finite samples. This topic is related to a multitude of recent alternative jump-robust estimators based on truncation (e.g. Aït-Sahalia and Jacod, 2009a,b; Andersen, Dobrev, and Schaumburg, 2008; Christensen, Oomen, and Podolskij, 2010, 2011; Corsi, Pirino, and Renò, 2010; Mancini, 2004, 2009).

The paper proceeds as follows. In section 2, we set notation and invoke a standard arbitrage-free continuous time jump-diffusion semimartingale model. We also briefly review some aspects of the theory of return-based multipower variation, before we switch to studying realised range-based multipower estimation. The key theoretical results are presented in Theorem 2 and a novel combination estimator is proposed in Eq. (23). In section 3, we conduct a Monte Carlo study to investigate the finite sample properties of our new range-based multipower variations. We also inspect the asymptotic approximation of the jump-robust range-based tripower variance. In section 4, we progress with some empirical results using high-frequency data from the TAQ database. In section 5, we conclude and offer directions for future research. An appendix contains the derivations of our theoretical results.

2 Theoretical framework

In this section, we derive new non-parametric theory, based on the price range, for consistently estimating return variation, and we show how it can be applied to filter out the continuous variation part from the squared jumps.

2.1 The model

The theory is developed for a univariate log-price, say $p = (p_t)_{t\geq 0}$, which is defined on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$. p evolves in continuous time and is adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$, which holds all relevant information released with the passing of time.

As standard in asset pricing theory, we assume that p is a member of the class of jump-diffusion semimartingales that satisfy the generic representation:[1]

$$p_t = p_0 + \int_0^t \mu_u \mathrm{d}u + \int_0^t \sigma_u \mathrm{d}W_u + \sum_{i=1}^{N_t} J_i, \tag{1}$$

where $\mu = (\mu_t)_{t\geq 0}$ is a locally bounded and predictable drift term, $\sigma = (\sigma_t)_{t\geq 0}$ is a càdlàg volatility process, $W = (W_t)_{t\geq 0}$ a standard Brownian motion, $N = (N_t)_{t\geq 0}$ a finite activity simple counting process, and $J = \{J_i\}_{i=1,\ldots,N_t}$ is a sequence of non-zero random variables.[2] Here, N represents the total number of jumps in p that has occurred up to and including time t, while J are the corresponding jump sizes. Note that the drift term is of order $\mathrm{d}t$ and is therefore negligible over short intervals of time, as typically considered in the high-frequency literature. As such, the model induces two main sources of risk, namely diffusive volatility and jumps.

The quadratic variation of the cumulative return process is then given by

$$[\,p\,]_t = \int_0^t \sigma_u^2 \mathrm{d}u + \sum_{i=1}^{N_t} J_i^2, \tag{2}$$

i.e. the integrated diffusive variance coefficient and the sum of the squared jumps. The quadratic variation plays a key role in high-frequency volatility estimation due to the following definition from stochastic integration theory:

$$[\,p\,]_t = \underset{n\to\infty}{\text{p-lim}} \sum_{i=1}^{n} (p_{t_i} - p_{t_{i-1}})^2, \tag{3}$$

[1] Asset prices must be semimartingales under rather weak conditions (e.g. Back, 1991; Delbaen and Schachermayer, 1994).

[2] A simple counting process, N, is of finite activity provided that $N_t < \infty$ for $t \geq 0$, almost surely. In this paper, we do not explore infinite activity jump processes, although these models have been studied in the context of realised multipower variation (e.g. Aït-Sahalia and Jacod, 2009a,b; Barndorff-Nielsen, Shephard, and Winkel, 2006; Todorov and Tauchen, 2010; Woerner, 2005, 2006).

for any sequence of partitions $0 = t_0 < t_1 < \ldots < t_n = t$ such that $\max_{1 \leq i \leq n} \{t_i - t_{i-1}\} \to 0$ as $n \to \infty$ (e.g. Protter, 2004). This result is important, because it shows that we can infer the latent quadratic return variation from high-frequency observations of p and, in the limit, consistently estimate it as more and more data are filled in the interval $[0, t]$. The practical relevance of the quadratic variation is stressed in several papers (e.g. Andersen, Bollerslev, and Diebold, 2010; Andersen, Bollerslev, Diebold, and Labys, 2003; Barndorff-Nielsen and Shephard, 2007). It is, for example, closely related to the conditional variance, which features prominently in many pillars of financial economics.

In the special case where there are no jumps in p (i.e. $N \equiv 0$), it holds that

$$p_t = p_0 + \int_0^t \mu_u \mathrm{d}u + \int_0^t \sigma_u \mathrm{d}W_u, \tag{4}$$

and the jump-diffusion process in Eq. (1) narrows down to a stochastic volatility model with continuous sample paths, for which the quadratic variation equals the integrated variance, $[\,p\,]_t = \int_0^t \sigma_u^2 \mathrm{d}u$. We should note that some of our results are derived under this assumption.

Below, in the CLTs only, we are also going to impose some regularity conditions on σ:

Assumption (V): *σ does not vanish* $\mathbf{(V_1)}$ *and it satisfies the equation:*

$$\sigma_t = \sigma_0 + \int_0^t \mu'_u \mathrm{d}u + \int_0^t \sigma'_u \mathrm{d}W_u + \int_0^t v'_u \mathrm{d}B'_u, \quad \text{for } t \geq 0, \tag{$\mathbf{V_2}$}$$

where $\mu' = (\mu'_t)_{t \geq 0}$, $\sigma' = (\sigma'_t)_{t \geq 0}$ and $v' = (v'_t)_{t \geq 0}$ are càdlàg, with μ' also being locally bounded and predictable, and $B' = (B'_t)_{t \geq 0}$ is a Brownian motion independent of W.

Assumption $\mathbf{(V_1)}$ is a weak regularity condition, which is fulfilled for many financial models. Assumption $\mathbf{(V_2)}$ amounts to saying that σ is of continuous semimartingale form.[3] It appears restrictive, because it rules out jumps in σ, which is at odds with empirical evidence (e.g. Eraker, Johannes, and Polson, 2003; Todorov, 2010). But, it should be pointed out that the assumption is not a necessary condition and, hence, it is not required for our results to go through. It

[3]Note the appearance of W in σ, which allows for leverage effects (e.g. Christie, 1982).

can be dispensed with in favor of a more flexible specification, allowing σ to jump, at the cost of substantial extra rigor in the proofs (along the lines of Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard, 2006). Here, we rule out such technical details to preserve a leaner and more clear-cut exposition.

2.2 The data

Throughout the remainder of the text, we will work on the unit interval, $t \in [0, 1]$. We think of this as representing the part of the day, for which high-frequency data are at our disposal. In our empirical application, where we are going to apply the estimators introduced in this section on a day-by-day basis to high-frequency data from NYSE- and NASDAQ-listed stocks, it is thus natural to let the unit interval be the hours spanned by the regular trading session.

The foundation for our econometric analysis is a high-frequency record of p, supposed to be available at equidistant times $t_i = i/N$, $i = 0, 1, \ldots, N$.[4] Throughout, we assume $N = nm$, for $n, m \in \mathbb{N}$. Here, n is the number of subintervals of the form $[(i-1)/n, i/n]$, for $i = 1, \ldots, n$, on which we shall compute various statistics, while m is the number of price changes within such a time interval.[5] Note that in the asymptotics we let $n \to \infty$, while, for simplicity, we keep m fixed throughout.

Thus, the challenge is to infer the quadratic variation, and its split into the continuous and discontinuous part, from a set of discrete high-frequency data.

We define intraday returns at sampling frequency n as follows

$$r_{i\Delta,\Delta} = p_{i/n} - p_{(i-1)/n}, \quad \text{for } i = 1, \ldots, n, \tag{5}$$

where $\Delta = 1/n$ is the time distance between price observations.

[4] In practice, high-frequency data are irregularly spaced and equidistant prices are imputed from the observed ones. Two approaches are linear interpolation (e.g. Andersen and Bollerslev, 1997) or the previous-tick method suggested by Wasserfallen and Zimmermann (1985). The former method has an unfortunate property in connection with estimating quadratic return variation, see Hansen and Lunde (2006, Lemma 1).

[5] This way of blocking high-frequency data into smaller pieces is natural in our setup, see, for example, Mykland (2010).

Note that for $n < N$, the sequence $r_{i\Delta,\Delta}$ is not exhausting all the available high-frequency data. In effect, we are assuming that returns are constructed at a coarser sampling frequency n, relative to the total amount of N "ultra" high-frequency returns that can potentially be made. In practice, high-frequency data are polluted with measurement error due to market microstructure frictions, such as price discreteness, bid-ask spreads etc. (e.g. Hansen and Lunde, 2006). The presence of such noise can be detrimental to standard estimators of return variation, in particular at very high sampling frequencies. Although our ability to cope with and account for the impact of noise in the estimation of return variation has significantly improved in recent years[6], it still remains common in applied work to use low-frequency tick data, for example 5-minute data are often used (e.g. Bollerslev and Todorov, 2011; Corsi and Renò, 2010; Tauchen and Zhou, 2011). It is precisely this type of sparse sampling, which motivates us to suggest using the range-statistic below.

In this paper, we do not explicitly control for the noise.[7] Instead, we are going to assume that the econometrician has selected n low enough such that potential biases from the noise can be ignored.

2.3 Return-based estimation of quadratic variation

With the setup in place, we can proceed by estimating quadratic variation and its two components using standard model-free return-based measures. The realised variance, proposed in Andersen and Bollerslev (1998) and Barndorff-Nielsen and Shephard (2002), is defined as a sum of squared intraday returns and is at sampling frequency n given by:

$$RV^n = \sum_{i=1}^{n} r_{i\Delta,\Delta}^2 \xrightarrow{p} \int_0^1 \sigma_u^2 \mathrm{d}u + \sum_{i=1}^{N_1} J_i^2. \tag{6}$$

The consistency of RV^n for the quadratic return variation is immediate in light of Eq. (3). However, use of the realised variance by itself is not sufficient to learn about the composition of quadratic variation, so in order to isolate the integrated variance and squared jumps, we are going to require more material.

[6] A representative, but necessarily incomplete, list of papers in this field, include Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008); Christensen, Oomen, and Podolskij (2010); Jacod, Li, Mykland, Podolskij, and Vetter (2009); Podolskij and Vetter (2009a,b); Zhang, Mykland, and Aït-Sahalia (2005); Zhang (2006).

[7] Christensen, Podolskij, and Vetter (2009) analyze the impact of noise on the range-statistic and propose a bias-correction to it.

To accomplish this, consider the bipower and tripower variance:

$$BV^n = \frac{n}{n-1}\sum_{i=1}^{n-1}\prod_{j=1}^{2}\frac{|r_{(i+j-1)\Delta,\Delta}|}{\mu_1}, \tag{7}$$

$$TV^n = \frac{n}{n-2}\sum_{i=1}^{n-2}\prod_{j=1}^{3}\frac{|r_{(i+j-1)\Delta,\Delta}|^{2/3}}{\mu_{2/3}}, \tag{8}$$

where $\mu_r = E(|Z|^r)$ and $Z \sim N(0,1)$. The theoretical underpinnings of these estimators were derived in previous work by Barndorff-Nielsen and Shephard (2004) and Barndorff-Nielsen, Shephard, and Winkel (2006), see also Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006).

In particular, they are consistent estimators of the integrated variance under both the jump-diffusion and pure diffusion model given by Eq. (1) and (4), i.e. as $n \to \infty$

$$BV^n \xrightarrow{p} \int_0^1 \sigma_u^2 \mathrm{d}u \qquad \text{and} \qquad TV^n \xrightarrow{p} \int_0^1 \sigma_u^2 \mathrm{d}u. \tag{9}$$

The intuition for the jump-robustness is that, with only a finite number of jumps in the log-price, all "jump" returns are eventually (i.e., for n large enough) paired with a "continuous" return. The latter has order $O_p(\sqrt{\Delta})$, and this way jumps get knocked out of the probability limit.[8]

We can subsequently retrieve the sum of the squared jumps, e.g. by using TV^n

$$RV^n - TV^n \xrightarrow{p} \sum_{i=1}^{N_1} J_i^2. \tag{10}$$

Note that, in absence of price jumps, all the above three estimators are targeting the integrated variance. A natural way of comparing them in this special case is then done via their limiting distributions.

[8]There are a number of alternative ways of estimating the integrated variance robustly in the presence of jumps. Aït-Sahalia and Jacod (2009a,b) and Mancini (2004, 2009), for example, use threshold elimination of "large" returns before computing the realised variance, while Andersen, Dobrev, and Schaumburg (2008) and Christensen, Oomen, and Podolskij (2010) propose to infer diffusive volatility from the quantiles of high-frequency returns. A couple of recent papers also show how to improve finite sample jump robustness of the bipower variance and how to make it more efficient, see Corsi, Pirino, and Renò (2010) and Mykland, Shephard, and Sheppard (2010).

As an example, if p follows the diffusion process defined by Eq. (4), the CLT of RV^n has the form

$$\sqrt{n}\left(RV^n - \int_0^1 \sigma_u^2 \mathrm{d}u\right) \xrightarrow{d_s} MN\left(0, 2\int_0^1 \sigma_u^4 \mathrm{d}u\right), \tag{11}$$

where $MN(0,V)$ stands for a centered mixed normal distribution with conditional variance V, and $\int_0^1 \sigma_u^4 \mathrm{d}u$ is the integrated quarticity.[9] The only way this result is altered for the BV^n and TV^n is that the factor multiplying the integrated quarticity increases to 2.6 and 3.1.

The key insight to draw from this discussion is that, in general, RV^n estimates the quadratic return variation and, in the absence of jumps, it is most efficient.[10] But RV^n is not jump-robust, whereas both BV^n and TV^n are robust in the probability limit. Moreover, only the CLT of the tripower variance also holds true in the presence of jumps (e.g. Barndorff-Nielsen, Shephard, and Winkel, 2006).[11] Hence, using products of lagged returns (while keeping the sum of their powers equal to two) produces jump-robust estimates of integrated variance. The degree of robustness increases with the addition of extra lags, but we pay a price for this by losing some efficiency, if in fact there are no jumps. It is interesting to note this trade-off for later comparison, because the properties of the range-statistic turn out to be different.

2.3.1 Subsampling

In the above, the sparsely sampled return-based estimators potentially ignore a lot of information about the true return variation, because a large portion of the total amount of available high-frequency data is effectively discarded upfront. But, even if we do not want push the sampling frequency beyond n, it is easy to soak up more efficiency by subsampling

[9] Throughout the paper, the symbol "$\xrightarrow{d_s}$" is used to denote convergence in law stably. We refer to Rényi (1963) for a formal definition of stable convergence in law. Moreover, the motivation for using this type of convergence in the high-frequency volatility setting is explained in great detail by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008).

[10] In theory, the realised variance at the highest sampling frequency N has the efficiency of the ML estimator in parametric versions of this problem. This shows that, in principle, we should construct the realised variance based on all data, but of course this result ignores problems associated with microstructure noise.

[11] This means that we can use the asymptotic distribution of TV^n to construct confidence intervals for the integrated variance in both the jump and no-jump scenario, something we consider below.

the data, as suggested by Zhou (1996) and Zhang, Mykland, and Aït-Sahalia (2005). This can be accomplished by simply shifting the starting point from which low-frequency returns are computed.

Let

$$r_{i\Delta,\Delta,j} = p_{i/n+j/N} - p_{(i-1)/n+j/N}, \quad \text{for } i = 1, \ldots, n \text{ and } j = 0, 1 \ldots, m-1. \tag{12}$$

Then, we can compute and average m realised variance estimates:

$$SRV^{n,m} = \frac{1}{m} \sum_{j=0}^{m-1} RV^{n,j}, \quad \text{where} \quad RV^{n,j} = \sum_{i=1}^{n} r^2_{i\Delta,\Delta,j}.^{12} \tag{13}$$

The bipower and tripower variance can also be modified in this way and the subsampled version of these estimators will be called $SBV^{n,m}$ and $STV^{n,m}$ in the following.

The averaging performed by Eq. (13) results in further efficiency gains, and it is known that the asymptotic variance factor of the subsampled realised variance can be brought down from 2 to 1.33 as the number of subsamples $m \to \infty$ (e.g. Zhang, Mykland, and Aït-Sahalia, 2005). To our knowledge, the reduction in variance associated with the subsampled bipower and tripower variance is not known in closed-form, but it will be accessed with Monte Carlo simulations below, where we compare it with the realised range-based estimators introduced next.

2.4 Range-based estimation of quadratic variation

The range provides an alternative way of learning about the quadratic return variation (e.g. Parkinson, 1980). Its use in the high-frequency context was initiated by Christensen and Podolskij (2007) and Martens and van Dijk (2007), who developed the so-called realised range-based variance.

Write

$$s_{p_{i\Delta,\Delta},m} = \max_{0 \le s,t \le m} \left(p_{\frac{i-1}{n}+\frac{t}{N}} - p_{\frac{i-1}{n}+\frac{s}{N}} \right), \qquad \text{for } i = 1, \ldots, n, \tag{14}$$

[12] Notation is a bit loose here. We can only compute the nth return for the initial subsample estimate, while doing so for the rest requires a log-price observation outside of the unit interval. Thus, the last $m-1$ subsample estimates are based on only $n-1$ high-frequency returns and we apply an additional small sample correction to compensate for the missing summand.

as the range over the interval $[(i-1)/n, i/n]$.[13] Then,

$$RRV_b^{n,m} = \frac{1}{\lambda_{2,m}} \sum_{i=1}^{n} s^2_{p_{i\Delta,\Delta},m}, \tag{15}$$

is the realised range-based variance at sampling frequency n. Here,

$$\lambda_{r,m} = \mathbb{E}\big(s^r_{W,m}\big) \quad \text{with} \quad s_{W,m} = \max_{s,t=0,1,\ldots,m} \big\{W_{t/m} - W_{s/m}\big\} \tag{16}$$

is the rth moment of the range of a standard Brownian motion on the unit interval $[0,1]$, where this expectation is based only on observations of the process at equidistant times $t_j = j/m$, for $j = 0, 1, \ldots, m$.[14] The subscript b appearing in the definition of $RRV_b^{n,m}$ indicates that it will be biased in the presence of jumps, as we detail below.

Assuming p is a pure diffusion model as in Eq. (4), and under condition (V), the main theoretical findings of Christensen and Podolskij (2007) can be summarized as saying that

$$\sqrt{n}\left(RRV_b^{n,m} - \int_0^1 \sigma_u^2 \mathrm{d}u\right) \xrightarrow{d_s} MN\left(0, \Lambda_m \int_0^1 \sigma_u^4 \mathrm{d}u\right), \tag{17}$$

where $\Lambda_m = \left(\lambda_{4,m} - \lambda^2_{2,m}\right)/\lambda^2_{2,m}$.

The Λ_m factor appearing in the CLT of $RRV_b^{n,m}$ depends on m, the total number of price changes available in each interval of the form $[(i-1)/n, i/n]$, $i = 1, \ldots, n$. We add that, for $m \geq 2$ as considered here, Λ_m is always strictly smaller than the asymptotic variance coefficient of two for realised variance (see Figure 1).[15] This comparison also extends to the subsampled version of realised variance. Hence, if data are recorded at a finer resolution than the sampling frequency, and so long as the influence of noise is not too severe, it is always better to construct a realised range-statistic than to

[13] Below, in the appendix of proofs, we also make use of the range of a standard Brownian motion over the interval $[(i-1)/n, i/n]$, which is denoted by $s_{W_{i\Delta,\Delta},m}$, simply replacing p with W in the definition of Eq. (14).

[14] A couple of points are worth highlighting here. First, the constants $\lambda_{r,m}$ used to rescale the range-statistic hinges on the assumption that data be equidistant, which is typically not the case in practice. We shall return to this below. Second, and in contrast to the return-based estimators, the range-based scalings are not available in closed-form. Although this is a drawback, it is relatively easy to estimate them by simulation. To facilitate this step, tables of $\lambda_{r,m}$, as a function of r and m, can be obtained from the authors upon request.

[15] As m grows large, Λ_m converges to a value of about 0.4, which is a restatement of the original result by Parkinson (1980).

compute realised variance. In effect, we are better able to recoup information about the integrated variance contained in intermediate data by computing a price range on that interval rather than subsampling low-frequency returns. But, it requires that p follows the stochastic volatility model in Eq. (4) and it is therefore not valid in general. Still, it highlights the potential of the range and motivates us to analyze its properties in the jump-diffusion context.

2.4.1 Extension to jump-diffusion processes

To the best of our knowledge, little is known about the high-low estimator in models with jumps, as defined by Eq. (1). It turns out that in its raw form, the realised range-based variance is inconsistent for the quadratic return variation, if there are jumps in the price process, as highlighted in Theorem 1.

Theorem 1. *Assume that p follows the jump-diffusion process defined in Eq.* (1)*. As $n \to \infty$, it holds that:*

$$RRV_b^{n,m} \xrightarrow{p} \int_0^1 \sigma_u^2 \mathrm{d}u + \frac{1}{\lambda_{2,m}} \sum_{i=1}^{N_1} J_i^2, \tag{18}$$

Proof. See appendix. □

As the theorem shows, $RRV_b^{n,m}$ is downward biased in the jump-diffusion framework, because it incorrectly scales down the squared jumps by $\lambda_{2,m}$. To understand this, consider a model that consists solely of jumps (i.e., with drift and diffusion coefficient set to zero). Then, as $n \to \infty$, the sum of squared ranges converges to the sum of squared jumps, so the re-scaling is not required.

Although the conclusion of Theorem 1 is disappointing, the structure of the inconsistency unveiled by it does suggest a quick fix for constructing a hybrid range-statistic that can estimate the whole quadratic return variation.

We could, for example, do the following:

$$\lambda_{2,m} RRV_b^{n,m} + (1 - \lambda_{2,m})\, STV^{n,m} \xrightarrow{p} \int_0^1 \sigma_u^2 \mathrm{d}u + \sum_{i=1}^{N_1} J_i^2, \tag{19}$$

i.e., we take a linear combination of $RRV_b^{n,m}$ and $STV^{n,m}$ using the weights $(\lambda_{2,m}, 1 - \lambda_{2,m})$. This effectively amounts

to undoing the re-scaling of the squared ranges in Eq. (15) and then using a jump-robust estimator to subtract the excess portion of the integrated variance that emanates from this operation.

2.4.2 Realised range-based multipower variation

The main problem with this approach is that in doing the bias-correction to the realised range-based variance, we are relying on a return-based estimator as the robust measure of integrated variance, which conflicts with our intention of using the range-statistic. As such, a jump-robust range-based estimator of the integrated variance is required, which, again to best of our knowledge, has not been proposed in the literature. To fill this hole, we are therefore going to introduce and study a complete range-based multipower variation theory, as laid out next.

Definition. *The realised range-based multipower variation with parameter* $(q_1, \ldots, q_k) \in \mathbb{R}^k_+$ *is defined as:*

$$RMV^{n,m}_{(q_1,\ldots,q_k)} = \frac{n}{n-k+1} n^{q_+/2-1} \sum_{i=1}^{n-k+1} \prod_{j=1}^{k} \frac{s^{q_j}_{p_{(i+j-1)\Delta,\Delta},m}}{\lambda_{q_j,m}}, \tag{20}$$

where $q_+ = \sum_{j=1}^{k} q_j$.

$RMV^{n,m}_{(q_1,\ldots,q_k)}$ is composed of suitably scaled range-based cross-products raised to the powers $(q_1, \ldots, q_k)$ and it constitutes a direct analogue to the general definition of realised multipower variation (e.g. Barndorff-Nielsen, Shephard, and Winkel, 2006).

Remark. *Note that* $s_{p_{i\Delta,\Delta},m} \geq 0$ *and so* $RMV^{n,m}_{(q_1,\ldots,q_k)}$ *is non-negative by construction.*

We should point out that, with proper modifications throughout, $(i + j - 1)$ may be replaced by $(i + Kj - 1)$ for any finite positive integer K in the definition of $RMV^{n,m}_{(q_1,\ldots,q_k)}$. Such "staggering" of the data has been suggested in Andersen, Bollerslev, and Diebold (2007) and Barndorff-Nielsen and Shephard (2006). Moreover, Huang and Tauchen (2005) show how extra lagging can help to reduce the impact of microstructure noise in this type of estimators by effectively breaking the serial correlation in returns induced by the noise.

The next result states the theoretical properties of the realised range-based multipower variation.

Theorem 2. *Assume that p follows the diffusion process defined in Eq.* (4)*. As $n \to \infty$, it holds that:*

$$RMV^{n,m}_{(q_1,...,q_k)} \xrightarrow{p} \int_0^1 |\sigma_u|^{q+} \mathrm{d}u. \tag{21}$$

Moreover, if condition (V) is satisfied, it additionally holds that

$$\sqrt{n}\left(RMV^{n,m}_{(q_1,...,q_k)} - \int_0^1 |\sigma_u|^{q+} \mathrm{d}u\right) \xrightarrow{d_s} MN\left(0, \Lambda^m_{(q_1,...,q_k)} \int_0^1 |\sigma_u|^{2q+} \mathrm{d}u\right), \tag{22}$$

where

$$\Lambda^m_{(q_1,...,q_k)} = \frac{\prod_{j=1}^{k} \lambda_{2q_j,m} - (2k-1)\prod_{j=1}^{k} \lambda^2_{q_j,m} + 2\sum_{h=1}^{k-1}\prod_{j=1}^{h} \lambda_{q_j,m} \prod_{j=k-h+1}^{k} \lambda_{q_j,m} \prod_{j=1}^{k-h} \lambda_{q_j+q_{j+h},m}}{\prod_{j=1}^{k} \lambda^2_{q_j,m}}.$$

Furthermore, the consistency result in Eq. (21) *is robust to jumps if* $\max_{1\leq j\leq k}(q_j) < 2$*, while the CLT in Eq.* (22) *is robust to jumps under the stronger condition* $\max_{1\leq j\leq k}(q_j) < 1$*.*

Proof. See appendix. □

Remark. *Note that the rate of convergence is not influenced by m and no assumptions on the ratio n/m are required.*

Theorem 2 lays the foundation for producing range-based estimates of integrated power variation of various orders. It also shows what is required for such estimates to be robust against jumps in their probability limit and asymptotic distribution.

In light of the return-based multipower variation theory, there is nothing too surprising about the conclusions of the theorem. As such, the interested reader should note, while going through the appendix, that the recipe used to prove the results is to some extent "standard" by now. This means that several of the steps taken to deduce the properties of $RMV^{n,m}_{(q_1,...,q_k)}$ borrow directly from extant literature (e.g. Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard, 2006; Christensen and Podolskij, 2007). However, the range is a complicated functional, which has a number of subtle, technical implications in the analysis. In the proofs, we highlight where these complications arise and we also try to pinpoint what is "new" relative to the existing theory.

In this paper, the full force of Theorem 2, which can of course be used to estimate many interesting objects, is not required. We will mainly focus on defining jump-robust realised range-based estimates of the integrated variance by cloning the return-based bipower and tripower variance.[16] Thus, we define

$$RBV^{n,m} = \frac{n}{n-1}\sum_{i=1}^{n-1}\prod_{j=1}^{2}\frac{s_{p_{(i+j-1)\Delta,\Delta},m}}{\lambda_{1,m}}, \text{ i.e. } q_1 = q_2 = 1, \tag{23}$$

$$RTV^{n,m} = \frac{n}{n-2}\sum_{i=1}^{n-2}\prod_{j=1}^{3}\frac{s^{2/3}_{p_{(i+j-1)\Delta,\Delta},m}}{\lambda_{2/3,m}}, \text{ i.e. } q_1 = q_2 = q_3 = 2/3, \tag{24}$$

for which it holds that

$$RBV^{n,m} \xrightarrow{p} \int_0^1 \sigma_u^2 \mathrm{d}u \quad \text{and} \quad RTV^{n,m} \xrightarrow{p} \int_0^1 \sigma_u^2 \mathrm{d}u, \tag{25}$$

while only the limiting distribution of the $RTV^{n,m}$ remains unchanged in the jump setting, because the maximum of its powers is strictly smaller than unity.

Figure 1 plots the asymptotic variance coefficient $\Lambda^m_{(q_1,\ldots,q_k)}$, as a function of m, for the three range-based estimators considered here.[17] As evident, the variance decreases monotonically with increasing m, and by the time m reaches ten, a large majority of the potential efficiency gain has been attained. To put this in perspective, consider using the popular 5-minute sampling frequency. Then, $m = 10$ is equivalent to actually observing the price every 30 seconds, which is not unrealistic for many liquid series.

– Insert Figure 1 about here –

How much efficiency is being sacrificed to obtain jump robustness in the range-based setting? The answer, which can be gauged from the figure, is quite surprising. Note that, for $m \leq 3$, the ordering of the estimators is as expected, with

[16] Below, a jump-robust estimator of the integrated quarticity is also used.

[17] Note that the special case $m = 1$ means that we are using only $p_{(i-1)/n}$ and $p_{i/n}$ to compute the range-statistic on $[(i-1)/n, i/n]$. With no interior data available, the absolute return and the range are identical. This explains why the variance of the realised range-based estimators, for $m = 1$, coincides with those known from the return-based multipower variation theory, see, e.g., Eq. (11) and the following discussion.

$RRV_b^{n,m}$ being the most efficient. However, and very intriguingly, the rankings of the variances are swapped for $m \geq 4$, rendering the realised range-based tripower variance not only the more robust estimator of the integrated variance but also the most efficient![18]

While the message conveyed by Figure 1 is compelling, it also clashes with intuition, because the range-based multipower estimators are designed to be increasingly robust to jumps. We would expect this feature to come at a cost, if there are no jumps in the data, vis-à-vis the trade-off embedded in the return-based estimators. In general, the efficiency of a multipower variation statistic is a property of the underlying moments of Brownian motion of the given functional, i.e. in our setting the range. So, there is nothing to stop a multipower variation statistic with extra lags from being more efficient. But, apart from saying that in the proofs the constants, which pop up here and there, make it a fact, we lack a convincing, intuitive explanation about, why it is true. The marginal reduction in variance from adding extra lags disappears fast, though, and from a practical perspective the range-based tripower variance appears sufficient to capture almost all incremental efficiency gain.

Using these results, we will close this subsection by introducing a new, purely range-based estimator that is consistent for the quadratic variation of the jump-diffusion semimartingale defined by Eq. (1):

$$RRV^{n,m} \equiv \lambda_{2,m} RRV_b^{n,m} + (1 - \lambda_{2,m})\, RTV^{n,m} \xrightarrow{p} \int_0^1 \sigma_u^2 \mathrm{d}u + \sum_{i=1}^{N_1} J_i^2, \tag{26}$$

which is our preferred approach of estimating total return variation with the range.

2.4.3 The joint distribution of $RRV_b^{n,m}$ and $RTV^{n,m}$

The univariate convergence in law of $RRV_b^{n,m}$ and $RTV^{n,m}$, which is available from Theorem 2, can be expanded to cover their joint asymptotic distribution, which is required as a basic ingredient for conducting range-based non-parametric

[18] Taken together, the above theory implies that in many practical cases, the $RRV_b^{n,m}$ of Christensen and Podolskij (2007) and Martens and van Dijk (2007) should not be applied as a standalone estimator. In the presence of jumps, it is a biased estimate of quadratic variation, and here we should subsume $RRV_b^{n,m}$ into the combined estimator given by Eq. (26) below, while, in the absence of jumps, the variance of $RRV_b^{n,m}$ is inferior to the range-based multipower variation alternatives, if $m \geq 4$.

tests for the presence of jumps, a topic that has attracted considerable attention in past research (e.g. Aït-Sahalia and Jacod, 2009a,b, 2011; Barndorff-Nielsen and Shephard, 2004, 2006; Christensen, Oomen, and Podolskij, 2011; Huang and Tauchen, 2005; Jiang and Oomen, 2008; Lee and Mykland, 2008b; Li, 2011).

Proposition 1. *Assume that p follows the diffusion process defined in Eq.* (4)*, where the conditions on σ given by (V) are satisfied. As $n \to \infty$, it holds that:*

$$\sqrt{n}\begin{pmatrix} RRV_b^{n,m} - \int_0^1 \sigma_u^2 \mathrm{d}u \\ RTV^{n,m} - \int_0^1 \sigma_u^2 \mathrm{d}u \end{pmatrix} \xrightarrow{d_s} MN\left(\mathbf{0}, \int_0^1 \sigma_u^4 \mathrm{d}u \begin{bmatrix} \Lambda^m_{RRV_b^{n,m}} & \Lambda^m_{RRV_b^{n,m},RTV^{n,m}} \\ \Lambda^m_{RRV_b^{n,m},RTV^{n,m}} & \Lambda^m_{RTV^{n,m}} \end{bmatrix}\right), \tag{27}$$

with

$$\Lambda^m_{RRV_b^{n,m},RTV^{n,m}} = \frac{2}{\lambda^2_{1,m}\lambda^3_{2/3,m}}\left(\lambda_{1,m}\lambda^2_{5/3,m}\lambda^2_{2/3,m} + \lambda^4_{5/3,m}\lambda_{2/3,m} - 2\lambda^2_{1,m}\lambda^3_{2/3,m}\right). \tag{28}$$

Because the jump detection analysis is not implemented in this paper, we exclude a formal verification of the proposition (the proof can be forwarded upon request).[19] But, it should be noted that the bivariate extension is relatively simple to derive.

3 Simulation study

In this section, we document some aspects of the above asymptotic analysis by means of Monte Carlo experiments. The purpose of the study is to understand whether the theoretical, large sample properties of the realised range-based estimators are preserved in smaller, but more realistic, sample sizes.

For the continuous piece of the model, we adopt a dynamic two-factor stochastic volatility process, which can generate highly erratic sample paths for the log-price and volatility series. It is based on previous empirical work carried out by Chernov, Gallant, Ghysels, and Tauchen (2003), so that our setup ensures that the simulation design captures many salient

[19] Previous drafts of this article dealt more formally with the range-based jump detection analysis. An electronic copy of this material can be found on the web or acquired by emailing the authors.

features of real equity data (e.g., leverage correlation and volatility feedback); a market considered in our empirical application below.

In particular, the first building block is as follows:

$$\mathrm{d}p_t = \mu \mathrm{d}t + \text{s-exp}\left[\beta_0 + \beta_1 \sigma_t^{(1)} + \beta_2 \sigma_t^{(2)}\right] \mathrm{d}W_t, \tag{29}$$

with

$$\begin{aligned} \mathrm{d}\sigma_t^{(1)} &= \alpha_1 \sigma_t^{(1)} \mathrm{d}t + \mathrm{d}B_t^{(1)}, \\ \mathrm{d}\sigma_t^{(2)} &= \alpha_2 \sigma_t^{(2)} \mathrm{d}t + \left[1 + \alpha_3 \sigma_t^{(2)}\right] \mathrm{d}B_t^{(2)}. \end{aligned} \tag{30}$$

Here, s-exp(x) is the so-called "spliced" exponential function.[20] The parameter values for the entire system are taken from Huang and Tauchen (2005), i.e. $(\mu, \beta_0, \beta_1, \beta_2, \alpha_1, \alpha_2, \alpha_3) = (0.03, -1.2, 0.04, 1.5, -0.000137, -1.386, 0.25)$ and corr$(\mathrm{d}W_t, \mathrm{d}B_t^{(1)}) =$ corr$(\mathrm{d}W_t, \mathrm{d}B_t^{(2)}) = -0.3$.

To specify the discontinuous piece of p, a compound Poisson process is used. We fix a constant intensity parameter $\kappa = 0.4$ per time unit. Thus, a jump in p is experienced every 2.5 replication, on average, and we draw the corresponding jump sizes from a normal distribution, $J_i \sim N\left(0, p_{jmp} \int_0^1 \sigma_u^2 \mathrm{d}u\right)$. In our simulations, we use $p_{jmp} = 0.25/n_J$, where n_J is the total number of jumps in a given simulation run. Quadratic jump variation is thus taken to be proportional to the integrated variance, with a typical squared jump being larger in size on high volatility days. Together with our selection of κ, this choice implies that the unconditional jump proportion is about 8% of total return variation; a figure that broadly agrees with a consensus measure from the extant, recent literature (e.g. Andersen, Bollerslev, and Huang, 2011; Bollerslev, Law, and Tauchen, 2008; Corsi and Renò, 2010; Tauchen and Zhou, 2011; Todorov, 2009).

A standard Euler approximation scheme is applied to the set of stochastic differential equations given by Eq. (29) – (30). Jumps are scattered randomly throughout the day. We process a total of 10,000 simulations and assume that $N = 2,340$. The latter choice is calibrated to match our empirical work in Section 4, where we study data from the

[20] We refer to Chernov, Gallant, Ghysels, and Tauchen (2003) for more information about how the s-exp(x) function operates. In short, s-exp(x) slows down the growth rate of the exponential function at high values of the input x.

NYSE TAQ database. The regular trading session at NYSE spans 6.5 hours, or 23,400 seconds, and for the sample period covered, we refresh the price every 10 seconds, which motivates our selection of N here.[21] We use the sampling frequencies $n = 26, 39, 78$, which translates into 15-, 10- and 5-minute sampling. Also, we construct return-based estimators by using subsampling, as explained above.

Finally, and although we do not model microstructure noise explicitly in this paper, our approach is motivated by the existence of such frictions. Therefore, we also gauge the performance of our estimators in the presence of noise. In particular, we add to p an i.i.d. noise process u – independent of p – such that $\mathbb{E}(u) = 0$ and $\mathbb{E}\left(u^2\right) = \varpi^2$. We let u have a two-point distribution: $\Pr(u = \pm\varpi) = 1/2$. This choice has been further analyzed in Christensen, Podolskij, and Vetter (2009) and can loosely be thought of as representing a form of bid-ask spread. The magnitude of the noise is controlled by $\gamma = \sqrt{\dfrac{\varpi^2 \int_0^1 \sigma_u^2 \mathrm{d}u}{N}}$, where γ is the noise ratio parameter (see, e.g., Oomen, 2006). Christensen, Oomen, and Podolskij (2010) report a comprehensive set of empirical γ estimates. In accordance with their results, we set $\gamma = 0.50$, which reflects the typical amount of noise found in high-frequency data from the U.S. stock market.

3.1 Simulation results

We start by looking closer at the ability of the range-based estimators to provide unbiased and efficient measures of quadratic variation and integrated variance. In Table 1, we report the relative bias and root mean squared error (rmse) of the various statistics for the three sampling frequencies $n = 26, 39, 78$. The relative bias is computed as a ratio of the estimate to its population target, averaged across simulations, and should equal 1 for an unbiased statistic. The number in parenthesis below the relative bias is the rmse, which has been multiplied by a factor 1,000.

– Insert Table 1 about here –

[21] In order to minimize discretization bias, we first create a compact realization of Eq. (29) – (30) based on simulating a total of 23,400 "second-by-second" log-price updates and from this we extract every 10th data point.

We first navigate through the left-hand portion of the table, which covers the results in the absence of microstructure noise. The presence of noise case is summarized towards the end of the section. First, as noticeable from the table, the realised range-based estimators are mildly biased. This was to be expected for the jump-robust measures, because n is not sufficiently high in these simulations to fully eradicate the impact of jumps. Interestingly, though, the relative bias of $RBV^{n,m}$ and $RTV^{n,m}$ is a bit smaller that what we compute for the return-based competitors, illustrating that the jump-robust range-statistics appear less sensitive to this effect in small samples. Second, and by contrast to the subsampled realised variance, there is also a slight bias in the combined estimator $RRV^{n,m}$, but it is only about one percent at the 5-minute sampling frequency. In any instance, the rmse of the range-based estimators is much smaller compared to the equivalent return-based estimators, which reinforces the statements from the previous section, based on large sample theory. Thus, the range-statistic maintains also a comfortable lead in sampling stability in finite samples. Lastly, the table reaffirms that the range-based tripower extension is more efficient than its bipower companion. As readily seen, it is also less biased. Hence, going forward we restrict attention to $RTV^{n,m}$.

– Insert Figure 2 about here –

To corroborate the analysis, we turn to Figure 2, which provides a reality check on the accuracy of the limiting normal distribution derived for $RTV^{n,m}$. Note that Panel A is based on the standardized version of the CLT from Theorem 2, while Panel B uses the delta method to conclude that

$$\frac{\sqrt{n}\left(\ln RTV^{n,m} - \ln \int_0^1 \sigma_u^2 \mathrm{d}u\right)}{\sqrt{\Lambda^m_{RTV^{n,m}} \int_0^1 \sigma_u^4 \mathrm{d}u / \left(\int_0^1 \sigma_u^2 \mathrm{d}u\right)^2}} \xrightarrow{d} N(0,1). \tag{31}$$

As apparent from Panel A, the sampling distribution of $RTV^{n,m}$ deviates from the standard normal with some distortions both in the tails and center area of the density, although the fit does gradually improve as n rises. On the other hand, the log-based distribution theory in Panel B tracks the finite sample distribution of $\ln RTV^{n,m}$ somewhat better at all

sampling frequencies. Thus, as a practical recommendation, we advocate using the log-based approximation, which has the added virtue of enforcing non-negativity on confidence bands for the integrated variance.

– Insert Figure 3 about here –

Next, we consider the problem of making inference about diffusive return variation, as it would be done in practice, where the integrated variance is unknown. Figure 3 shows 95% confidence intervals for the integrated variance, using the log-based distribution theory and $n = 78$.[22] As an illustration, we took a sequence of ten simulations, some of which include both small and large jumps. The x-axis labels report, in trials with jumps, how large is the jump as a ratio of total quadratic variation. In order to present a feasible theory, we replaced the unobserved integrated quarticity by a consistent jump-robust realised range-based multipower variation estimator. Among several candidates, we chose a tripower estimator for the job (with parameter $q_j = 4/3$, for $j = 1, \ldots, 3$). Also, as a benchmark in the chart, we compare with the return-based confidence intervals using $STV^{n,m}$, where the feasible bands are computed using a subsampled return-based tripower quarticity estimator.

Across all 10,000 simulations, the coverage rates are almost equal, whether using the return- or range-based approach. For example, at the 95% level displayed in the figure, the range-based intervals include the integrated variance 93.38% of the times, while this number is marginally better at 93.81% for the return-based interval. Meanwhile, at the 99% confidence level, the rates change to 98.56% and 98.46%, respectively, yielding almost identical performance. As evident, however, the range-based confidence intervals typically deliver sharper inference with much tighter bands, which is a reflection of the smaller asymptotic variance embedded in such estimators.

Finally, consider the right-hand part of Table 1, which discloses how the results are altered, when the efficient log-price data are concealed behind a realistic level of microstructure noise. As apparent, the range by and large preserves its relative efficiency advantage also under noise, at least for the moderate sampling frequencies considered in this paper. It

[22] Note that because the log-based approximation is used, the confidence intervals are not completely symmetric around the point estimate.

is precisely in this setting, so often used in practice, we believe the range-based estimators hold some potential compared to the mainstream estimators. They are easy to implement, very efficient and fairly robust to noise at low-frequency. But, as consistent with the analysis of Christensen, Podolskij, and Vetter (2009), we also observe that the noise has a larger impact on the range vis-à-vis the return for a fixed value of n. Thus, if higher sampling frequencies are required, the noise will eventually swamp the advantages of the range and a noise-robust estimator should be adopted instead.

4 Empirical application

We illustrate some features of the range-based multipower variation theory for a few members of the Dow Jones Industrial Average index. The exposition is based on transaction data for the following three constituents: American Express (AXP), Merck (MRK) and Exxon (XOM). We also include data for the S&P 500 Depository Receipts (SPY); an exchange-traded fund tracking the S&P 500. We retrieved high-frequency data for these tickers from the TAQ database via the WRDS interface. The sample period spans the whole of 2007 – 2009 and, as such, includes part of the ongoing financial crisis. Prior to analysis, we filtered the raw data for outliers, applying a set of rules that follow the guide proposed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009).[23] Moreover, we restrict attention to the regular trading hours and so remove updates with a timestamp outside 9:30AM to 4:00PM Eastern Standard Time.

Before we commence with the empirical analysis, it is worth to elaborate on a couple of practical points. Firstly, in real markets an important source of noise is price discreteness. Because of this feature – but also other aspects of the market microstructure (e.g., the practice of splitting large block trades into smaller slices) – it is not unusual to find long stretches in the data, where the price either repeats or only alternates between the bid and ask quotation (the so-called bid-ask bounce). Here, we recall that the range-based theory requests the number of "price changes" m contained in each sampling interval, on which the range is being computed, as this serves as a prerequisite for returning the appropriate

[23] In our application, we maintain the complete set of transaction data from every exchange, but we synchronize with quotes originating from the primary exchange only.

scaling factor $\lambda_{r,m}$. Hence, if we were to count every observation in the data as an increment to the price process, such practical features would tend to unduly propagate m. The outcome is an inflation of $\lambda_{r,m}$, which transforms into a downward bias in the range-based estimators.[24] Meanwhile, designing a good algorithm to tally the "true" number of price changes requires us to deliver a formal definition of what we perceive to constitute an increment to the price process; a very subjective and challenging task.

Secondly, and equally important, the theory also calls upon an equidistant grid of log-price observations, while real trade arrivals are, of course, random. With irregularly spaced observations, the scaling factors $\lambda_{r,m}$, which are built from equidistant data, are no longer correct and using these can have a profound effect on the estimation. Consider, for example, a uniform sampling scheme from a Brownian motion, in which 2,341 log-price observations are drawn without replacement from the entire set of 23,401 possible one-second time stamps available in the 6.5 hours trading day, yielding $N = 2,340$ irregularly spaced returns. Suppose also, as in the simulation section, that we split the data into $n = 78$ intervals containing $m = 30$ returns each. Then, Table 2 illustrates the resulting relative bias and mse:

– Insert Table 2 about here –

As apparent, with irregular spacings the use of $\lambda_{r,m}$ leads to non-trivial downward biases in the range-based estimators, which reduces their advantage from an mse point of view. The solution to this problem is to tailor $\lambda_{r,m}$ to the observation times. This implies simulating scaling factors on the fly, which is less appealing. But, it can be done (e.g. Rossi and Spazzini, 2009). We feel this issue is not too worrying for the data series under investigation here, which are comprised of deeply liquid securities.[25] However, when working with older data sets or illiquid series, this issue has

[24] The problem is in some sense akin to the no-trade bias of bipower variation, see Corsi, Pirino, and Renò (2010).

[25] For example, if we modify the above simulation exercise to draw 11,700 irregular observations of Brownian motion (matching closer with our empirical data) and then construct an artificial 10-second record of log-prices using previous-tick interpolation (as done below), the downward bias is reduced to a mere 0.2% in both $RRV^{n,m}$ and $RTV^{n,m}$ across 10,000 trials.

more substance, and here it would be an advantage, and probably necessary, to suitably account for the irregular nature of high-frequency data by simulating grid-specific scalings.

As a consequence of the above features, i.e. price discreteness and irregular spacing, a further modification of the theory is necessary for the empirical use of the realised range-based multipower variation framework. Therefore, we settled on the following compromise. Throughout the trading day, we collect a new observation of p every 10th second, using previous-tick interpolation to replace missing values by the most recent transaction price. We then construct the realised range- and return-based estimators using 5-minute sampling (i.e., setting $n = 78$ and $m = 30$). Table 3 holds descriptive statistics of the data and resulting series.

– Insert Table 3 about here –

As can be gleaned from the table, the range-based multipower variations deliver estimates of quadratic return variation and integrated variance, which are in line or marginally above the corresponding return-based estimates. This holds across the entire sample or by yearly subsample periods, as reported in Table 3. Browsing through the four equities, the range suggests that jump variation, i.e. the proportion of total variation produced by jumps, is in the order of 4.6% – 11.7%, which broadly agrees with the return-based estimates and also the previous literature. This is comforting, as we would expect the unconditional sample averages of both return-based and range-based estimation to be broadly in line. It is also interesting to note that SPY, representing the S&P 500 index, has the lowest estimated jump proportion.

– Insert Figure 4 about here –

In Panel A of Figure 4, we apply the range-statistic to estimate the overall level of return variation and its composition using the high-frequency data from MRK as an illustration. In the graph, all series are converted to an annualized standard deviation term. As readily observed, in general the $RRV^{n,m}$ and $RTV^{n,m}$ series tend to swing in parallel, although there

a few notable departures, as revealed by the jump variation (JV) figure also reported in the chart. Panel B investigates such an instance by plotting the high-frequency data for MRK on January 25, 2008, where the JV measure is about 75%. On this day, Merck published a press release regarding its anti-cholesterol product Mevacor, which received a "not approvable" letter from the FDA, while government regulators also said they were analyzing recent results from the clinical trials of the company's Vytorin drug (the so-called ENCHANCE study). The news were largely perceived as bad by the market, moving the equity into deep negative territory, although most of the losses were recovered before the end of trading. Of course, to reach statistical conclusions about the presence of jumps in the sample path, we would need to do a formal hypothesis test, but the plot and point estimates do indicative this.

– Insert Figure 5 about here –

Finally, in Figure 5, we compare the range-based and subsampled return-based tripower estimators of the integrated variance. In the left panel, we plot the autocorrelation functions of the two estimators up to 100 lags. As evident, they are virtually identical and both display the typical long-range dependence that has been observed in volatility many times before (e.g. Andersen, Bollerslev, Diebold, and Labys, 2003). The range-statistic is slightly more persistent than the return-based companion, which is in fact true across all equities considered here (not reported). Although the differences between the two series are larger in magnitude for some of the other symbols, it is not clear if this can be exploited to produce superior forecasting accuracy. We leave this task for future work to decide. In the right panel, we mimic the approach taken in the simulation section to construct feasible confidence intervals for the integrated variance in the highly volatile month of September, 2008, which featured the demise of Lehman Brothers, the 4th largest investment bank in the U.S. at that time. Notice the widening of the error bands (via the estimate of integrated quarticity) as σ increases, which reflects the surge in volatility in the aftermath of the bankruptcy. Again, and according to theory, the range-based intervals are smaller than the return-based ones, thus delivering sharper inference about the integrated variance.

5 Conclusions and directions for future research

This paper has presented the concept of realised range-based multipower variation and shown how it can be used to estimate the ex-post quadratic return variation and conduct jump-robust inference about integrated variance. The large sample asymptotic theory was backed by both a simulation study and an empirical application, illustrating the potential of the range-statistic. The range was motivated by the typical sparse sampling of return-based estimators caused by the presence of microstructure noise in financial high-frequency data. Of course, the range itself is affected by the noise component, as we demonstrated with numerical simulations. In a companion paper (Christensen, Podolskij, and Vetter, 2009), we study more formally the impact of noise on the standard realised range-based variance, and the interested reader is encouraged to read this material.

The theory developed here casts new light on the range and displays its properties in a general semimartingale model with stochastic volatility and finite activity jumps. But there are still several interesting issues left untouched. First, covariance risk is important in financial economics and range-based measures are notoriously difficult to extend to the multivariate setting. Range-based covariance estimation has been studied in Brandt and Diebold (2006) and Bannouh, van Dijk, and Martens (2009), although they make rather strong assumptions about the underlying process driving the evolution of asset prices over time. Moreover, those papers are based on a polarization identity, which in general does not guarantee a positive semi-definite covariance matrix estimate. In an ongoing paper, we are working on an extension of some of the concepts discussed here to multivariate processes, and we hope to be able to publish some results soon. Second, it could be worth to consider alternative available tools, which may refine the asymptotic approximations presented here, for example bootstrapping, Edgeworth expansions or Box-Cox transformations, as has been suggested by Aït-Sahalia, Mykland, and Zhang (2011); Gonçalves and Meddahi (2009, 2011) in the context of the realised variance. Finally, as the realised range-based variance of Christensen and Podolskij (2007); Martens and van Dijk (2007), which in this paper was shown to be biased, has been found to produce good forecasts of future return variation (e.g. Patton and Sheppard, 2009), we look forward to a more thorough analysis of the forecasting capabilities of realised range-based

multipower variations introduced here, for example following the lines of Andersen, Bollerslev, and Diebold (2007).

6 Acknowledgements

This work was supported by CREATES, which is funded by the Danish National Research Foundation, and the Deutsche Forschungsgemeinschaft through grant SFB 475 ”Reduction of complexity in multivariate data structures”. Mark Podolskij also received funding from the Microstructure of Financial Markets in Europe (MicFinMa) network. Previous drafts of this article were circulated under the titles “Asymptotic theory for range-based estimation of quadratic variation of discontinuous semimartingales” and “Range-based estimation of quadratic variation.” Parts of the paper were prepared, while Kim Christensen was visiting University of California, San Diego (UCSD), Rady School of Management, whose hospitality is gratefully acknowledged. We are indebted to George Tauchen (co-editor) and an anonymous referee for their valuable and insightful feedback. In addition, we thank Allan Timmermann, Asger Lunde, Holger Dette, Morten Nielsen, Neil Shephard, Per Mykland, Roel Oomen, Rossen Valkanov as well as conference and seminar participants at the 2006 CIREQ conference on “Realized Volatility” in Montréal, Canada, the “61st European Meeting of the Econometric Society” (ESEM) 2006, Austria, the ESF workshop on “High Frequency Econometrics and the Analysis of Foreign Exchange Markets” at Warwick Business School, United Kingdom, the “International Conference on High Frequency Finance” in Konstanz, Germany, the “Statistical Methods for Dynamical Stochastic Models” conference in Mainz, Germany, the “Stochastics in Science, in Honor of Ole E. Barndorff-Nielsen” conference in Guanajuato, Mexico, the Rady School of Management, UCSD, and University of Florence for comments and suggestions. The usual disclaimer applies.

Appendix of proofs

Proof of Theorem 1

First, we decompose

$$RRV_b^{n,m} = \frac{1}{\lambda_{2,m}} \left(\sum_{i \in \Gamma_n} s^2_{p_{i\Delta,\Delta,m}} + \sum_{i \in \Gamma_n^c} s^2_{p_{i\Delta,\Delta,m}} \right),$$

where $\Gamma_n = \{1 \leq i \leq n \mid \text{the process } p \text{ jumps on } [(i-1)/n, i/n]\}$. Because the jump part of p has finite activity, there are only finitely many jumps on $[0,1]$, so the first sum in the above decomposition is finite (almost surely). In addition, with a probability converging towards 1, there is at most one jump per interval $[(i-1)/n, i/n]$. Combined with the results of Christensen and Podolskij (2007) (or Theorem 2 in this paper), this implies that

$$\frac{1}{\lambda_{2,m}} \sum_{i \in \Gamma_n} s^2_{p_{i\Delta,\Delta,m}} \xrightarrow{p} \frac{1}{\lambda_{2,m}} \sum_{i=1}^{N_1} J_i^2 \qquad \text{and} \qquad \frac{1}{\lambda_{2,m}} \sum_{i \in \Gamma_n^c} s^2_{p_{i\Delta,\Delta,m}} \xrightarrow{p} \int_0^1 \sigma_u^2 \mathrm{d}u,$$

i.e.

$$RRV_b^{n,m} \xrightarrow{p} \int_0^1 \sigma_u^2 \mathrm{d}u + \frac{1}{\lambda_{2,m}} \sum_{i=1}^{N_1} J_i^2,$$

as asserted. ■

Proof of Theorem 2

Preliminaries and some notation

First, we note that as $t \mapsto \sigma_t$ is càdlàg, all powers of σ are locally integrable with respect to the Lebesgue measure, so that for any t and $s > 0$, $\int_0^t |\sigma_u|^s \mathrm{d}u < \infty$. Moreover, and without loss of generality, we will restrict the functions μ, σ, μ', σ', v' and σ^{-1} to be bounded (e.g., Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard, 2006, Section 3).

Next, if a process X^n is of the form:

$$X^n = \sum_{i=1}^{n} \zeta_i^n,$$

for an array (ζ_i^n) and $X^n \xrightarrow{p} 0$, we say that (ζ_i^n) is asymptotically negligible (AN).

Also, in the below we employ generic constants, which are denoted by C or C_p (the latter notation is applied, when the constant depends on some external parameter p).

Moreover, we define

$$\beta_{i,j}^n = \sqrt{n}|\sigma_{\frac{i-1}{n}}|s_{W_{(i+j-1)\Delta,\Delta},m}, \qquad j = 1, \ldots, k. \tag{32}$$

This construct is used to locally approximate the true (rescaled) range $\sqrt{n}s_{p_{(i+j-1)\Delta,\Delta},m}$. We suppress the dependence of $\beta_{i,j}^n$ on m for notational convenience. Finally, we also set

$$g_j(x) = \frac{1}{\lambda_{q_j,m}} x^{q_j}, \qquad j = 1, \ldots, k,$$

and

$$\rho_x(f) = \mathbb{E}[f(|x|s_{W,m})].$$

It should be noted that

$$\rho_x(g_j) = |x|^{q_j}, \qquad j = 1, \ldots, k.$$

Structure of the multipower variation proof

Before we write down the proof of the realised range-based multipower variation estimator, we will briefly sketch the main steps and ideas behind it.

(1) First, we prove consistency of the estimator

$$\widetilde{RMV}_{(q_1,\ldots,q_k)}^{n,m} = \frac{1}{n} \sum_{i=1}^{n-k+1} \prod_{j=1}^{k} \frac{\left(\beta_{i,j}^n\right)^{q_j}}{\lambda_{q_j,m}}.$$

which is an approximation of $RMV_{(q_1,\ldots,q_k)}^{n,m}$ based on the representation in Eq. (32). The trick is then to show that the error committed by using this approximation goes to 0. That this is true follows exactly as in the return-based setting (e.g., Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard, 2006, Section 6).

(2) The proof of the central limit theorem also starts by figuring out the corresponding result for the approximating sequence, i.e. we first deduce that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n-k+1}\left\{\prod_{j=1}^{k}\frac{\left(\beta_{i,j}^{n}\right)^{q_j}}{\lambda_{q_j,m}}-\prod_{j=1}^{k}\rho_{\sigma_{\frac{i-1}{n}}}(g_j)\right\}\xrightarrow{d_s}\sqrt{\Lambda_{(q_1,\ldots,q_k)}^{m}}\int_0^1|\sigma_u|^{q_+}\mathrm{d}B_u.$$

The basic tool used to work out this assertion is Theorem IX 7.28 in Jacod and Shiryaev (2003). We note that the computation of the asymptotic conditional variance is quite tedious and requires some very lengthy calculations.

(3) In the next step, we justify the approximation made in step (2). Here, we assume without loss of generality that $k=2$ and prove a CLT for the canonical process:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{g_1\left(\sqrt{n}s_{p_{i\Delta,\Delta},m}\right)g_2\left(\sqrt{n}s_{p_{(i+1)\Delta,\Delta},m}\right)-\mathbb{E}\left[g_1\left(\sqrt{n}s_{p_{i\Delta,\Delta},m}\right)g_2\left(\sqrt{n}s_{p_{(i+1)\Delta,\Delta},m}\right)\mid\mathcal{F}_{\frac{i-1}{n}}\right]\right\}$$

This part is in fact also shown as in the return-based case (e.g., Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard, 2006, Section 5).

(4) And, finally, the last step is to show that the process from step (3) and the original normalized statistic in Theorem 2 are asymptotically equivalent. This is the most complicated part. It is much more involved than with return-based estimation, because the supremum is not a smooth functional. On the other hand, as with returns, additional problems arise when powers $q_j\leq 1$ appear in the multipower variation, because $f(x)=|x|^p$ is non-differentiable for $p\in(0,1]$.

A Proof of consistency

Write

$$\widetilde{RMV}_{(q_1,\ldots,q_k)}^{n,m}=\sum_{i=1}^{n-k+1}\chi_i^n$$

with

$$\chi_i^n=\frac{1}{n}\prod_{j=1}^{k}\frac{\left(\beta_{i,j}^{n}\right)^{q_j}}{\lambda_{q_j,m}}.$$

As explained above, this statistic serves to approximate the true estimator $RMV^{n,m}_{(q_1,\ldots,q_k)}$. Now, as in Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006, Section 6), it holds that

$$RMV^{n,m}_{(q_1,\ldots,q_k)} - \widetilde{RMV}^{n,m}_{(q_1,\ldots,q_k)} \xrightarrow{p} 0. \tag{33}$$

On the other hand:

$$\sum_{i=1}^{n-k+1} \mathbb{E}\Big[\chi_i^n \mid \mathcal{F}_{\frac{i-1}{n}}\Big] = \frac{1}{n} \sum_{i=1}^{n-k+1} |\sigma_{\frac{i-1}{n}}|^{q_+} \xrightarrow{p} \int_0^1 |\sigma_u|^{q_+} \mathrm{d}u, \tag{34}$$

and

$$\sum_{i=1}^{n-k+1} \Big(\chi_i^n - \mathbb{E}\Big[\chi_i^n \mid \mathcal{F}_{\frac{i-1}{n}}\Big]\Big) \xrightarrow{p} 0, \tag{35}$$

because the above summands are conditionally k-independent. Putting Eq. (33) together with Eq. (34), and using that the sequence is Eq. (35) is AN, we are able to conclude that

$$RMV^{n,m}_{(q_1,\ldots,q_k)} \xrightarrow{p} \int_0^1 |\sigma_u|^{q_+} \mathrm{d}u.$$

This completes the proof of the range-based multipower variation consistency property. ■

B Proof of the central limit theorem

We divide the proof into several steps, as detailed in the overview above.

B.1 A central limit theorem for the approximation

First, we introduce the quantity

$$U'^{n,m} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n-k+1} \left\{ \prod_{j=1}^{k} \frac{\big(\beta^n_{i,j}\big)^{q_j}}{\lambda_{q_j,m}} - \prod_{j=1}^{k} \rho_{\sigma_{\frac{i-1}{n}}}(g_j) \right\},$$

which is an approximation of $\sqrt{n}\left(RMV^{n,m}_{(q_1,\ldots,q_k)} - \int_0^1 |\sigma_u|^{q_+} \mathrm{d}u\right)$.

Lemma 1. *Assume that p follows the diffusion process defined in Eq.* (4)*. As $n \to \infty$, it holds that*

$$U'^{n,m} \xrightarrow{d_s} \sqrt{\Lambda^m_{(q_1,\ldots,q_k)}} \int_0^1 |\sigma_u|^{q_+} \mathrm{d}B_u.$$

Before giving the proof of the Lemma, note that the following estimates are true, when p is a diffusion:

$$\mathbb{E}\left[|\beta^n_{i,j}|^q\right] \leq C_q, \qquad j = 1, \ldots, k,$$

for all $q > 0$ and a fixed $m \in \mathbb{N}$.

Proof of Lemma 1: By shifting the indices, we form a decomposition

$$U'^{n,m} = \sum_{i=j}^{n} \zeta^n_i + o_p(1)$$

with

$$\zeta^n_i = \frac{1}{\sqrt{n}} \sum_{j=1}^{k} \left\{ g_j(\beta^n_{i-j+1,j}) - \rho_{\sigma_{\frac{i-j}{n}}}(g_j) \right\} \prod_{l=1}^{j-1} g_l(\beta^n_{i-j+1,l}) \prod_{l=j+1}^{k} \rho_{\sigma_{\frac{i-j}{n}}}(g_l).$$

We remark that the term $g_j(\beta^n_{i-j+1,j}) - \rho_{\sigma_{\frac{i-j}{n}}}(g_j)$ is measurable with respect to $\mathcal{F}_{\frac{i}{n}}$, while the other factors in the definition of ζ^n_i are $\mathcal{F}_{\frac{i-1}{n}}$-measurable.

By Theorem IX 7.28 in Jacod and Shiryaev (2003), the proof of Lemma 1 will follow, if we can verify the following five conditions on the sequence (ζ^n_i):

(A) $\displaystyle\sum_{i=k}^{n} \mathbb{E}\left[\zeta^n_i \mid \mathcal{F}_{\frac{i-1}{n}}\right] \xrightarrow{p} 0,$

(B) $\displaystyle\sum_{i=k}^{n} \mathbb{E}\left[|\zeta^n_i|^2 \mid \mathcal{F}_{\frac{i-1}{n}}\right] \xrightarrow{p} \Lambda^m_{(q_1,\ldots,q_k)} \int_0^1 |\sigma_u|^{2q_+} \mathrm{d}u,$

(C) $\displaystyle\sum_{i=k}^{n} \mathbb{E}\left[\zeta^n_i \left(W_{\frac{i}{n}} - W_{\frac{i-1}{n}}\right) \mid \mathcal{F}_{\frac{i-1}{n}}\right] \xrightarrow{p} 0,$

(D) $\displaystyle\sum_{i=k}^{n} \mathbb{E}\left[|\zeta^n_i|^4 \mid \mathcal{F}_{\frac{i-1}{n}}\right] \xrightarrow{p} 0,$

(E) $\displaystyle\sum_{i=k}^{n} \mathbb{E}\left[\zeta^n_i \left(N_{\frac{i}{n}} - N_{\frac{i-1}{n}}\right) \mid \mathcal{F}_{\frac{i-1}{n}}\right] \xrightarrow{p} 0,$

where the last condition must hold for any bounded martingale N that is orthogonal to W (i.e. with quadratic covariation $[N, W] = 0$).

As ζ_i^n is a martingale difference, it follows that

$$\mathbb{E}\left[\zeta_i^n \mid \mathcal{F}_{\frac{i-1}{n}}\right] = 0.$$

This verifies condition **(A)**. Next, because ζ_i^n is an even functional in W and $W \stackrel{d}{=} -W$, we get that

$$\mathbb{E}\left[\zeta_i^n \left(W_{\frac{i}{n}} - W_{\frac{i-1}{n}}\right) \mid \mathcal{F}_{\frac{i-1}{n}}\right] = 0,$$

which implies condition **(C)**. Moreover, we also deduce that

$$\mathbb{E}\left[|\zeta_i^n|^4 \mid \mathcal{F}_{\frac{i-1}{n}}\right] \leq \frac{C}{n^2},$$

and this proves **(D)**. That condition **(E)** is true follows from the work of Christensen and Podolskij (2007).

So, we are left with the task of proving condition **(B)**, which requires a straightforward but somewhat tedious calculation. We start out by defining

$$\rho_{i-j,i-h}(f,g) = \int f\left(|\sigma_{\frac{i-j-1}{n}}|x\right) g\left(|\sigma_{\frac{i-h-1}{n}}|x\right) \delta_m(x)\mathrm{d}x, \qquad 1 \leq j,h \leq k,$$

where δ_m denotes the density of $s_{W,m}$. We note that due to the continuity of σ

$$\sup_{i \leq n} \sup_{1 \leq j,h \leq k} |\rho_{i-j,i-h}(f,g) - \rho_{\sigma_{\frac{i-k}{n}}}(fg)| \xrightarrow{p} 0, \tag{36}$$

a result that is used in the computations below. By also setting

$$\mu_{i,j}^n = \left\{g_j\left(\beta_{i-j+1,j}^n\right) - \rho_{\sigma_{\frac{i-j}{n}}}(g_j)\right\} \prod_{l=1}^{j-1} g_l\left(\beta_{i-j+1,l}^n\right) \prod_{l=j+1}^{k} \rho_{\sigma_{\frac{i-j}{n}}}(g_l)$$

we find the identity

$$\mathbb{E}\left[|\zeta_i^n|^2 \mid \mathcal{F}_{\frac{i-1}{n}}\right] = \sum_{j=1}^{k} \mathbb{E}\left[|\mu_{i,j}^n|^2 \mid \mathcal{F}_{\frac{i-1}{n}}\right] + 2\sum_{j=1}^{k-1}\sum_{h=1}^{k-j} \mathbb{E}\left[\mu_{i,h}^n \mu_{i,h+j}^n \mid \mathcal{F}_{\frac{i-1}{n}}\right].$$

From this, we deduce that

$$\mathbb{E}\left[|\mu_{i,j}^n|^2 \mid \mathcal{F}_{\frac{i-1}{n}}\right] = \left\{\rho_{\sigma_{\frac{i-j}{n}}}(g_j^2) - \rho_{\sigma_{\frac{i-j}{n}}}^2(g_j)\right\} \prod_{l=1}^{j-1} g_l^2\left(\beta_{i-j+1,l}^n\right) \prod_{l=j+1}^{k} \rho_{\sigma_{\frac{i-j}{n}}}^2(g_l)$$

and

$$\mathbb{E}\left[\mu^n_{i,h}\mu^n_{i,h+j} \mid \mathcal{F}_{\frac{i-1}{n}}\right] = \left\{\rho_{i-h,i-h-j}(g_{h+j}g_h) - \rho_{\sigma_{\frac{i-h-j}{n}}}(g_{h+j})\rho_{\sigma_{\frac{i-h}{n}}}(g_h)\right\}\prod_{l=1}^{h-1} g_l\left(\beta^n_{i-h+1,l}\right)$$
$$\times \prod_{l=h+1}^{k}\rho_{\sigma_{\frac{i-j}{n}}}(g_l)\prod_{l=1}^{h+j-1} g_l\left(\beta^n_{i-h-j+1,l}\right)\prod_{l=h+j+1}^{k}\rho_{\sigma_{\frac{i-h-j}{n}}}(g_l).$$

Note that

$$\prod_{l=1}^{h-1} g_l\left(\beta^n_{i-h+1,l}\right)\prod_{l=1}^{h+j-1} g_l\left(\beta^n_{i-h-j+1,l}\right) = \prod_{l=1}^{h-1} g_l\left(\beta^n_{i-h+1,l}\right) g_{l+j}\left(\beta^n_{i-h-j+1,l+j}\right)\times\prod_{l=1}^{j} g_l\left(\beta^n_{i-h-j+1,l}\right).$$

We remark that $g_l\left(\beta^n_{i-h+1,l}\right)$ and $g_{l+j}\left(\beta^n_{i-h-j+1,l+j}\right)$ include the same increment of the Brownian motion W. Putting the pieces together and then calling upon Eq. (36), we conclude that

$$\sum_{i=j}^{n}\mathbb{E}\left[|\zeta^n_i|^2 \mid \mathcal{F}_{\frac{i-1}{n}}\right] \xrightarrow{p} \left\{\left(\prod_{j=1}^{k}\lambda_{q_j,m}\right)^{-2}\left(\sum_{j=1}^{k}\left\{\lambda_{2q_j,m}-\lambda^2_{q_j,m}\right\}\prod_{l=1}^{j-1}\lambda_{2q_l,m}\prod_{l=j+1}^{k}\lambda^2_{q_l,m}\right.\right.$$
$$\left.\left. + 2\sum_{j=1}^{k-1}\sum_{h=1}^{k-j}\left\{\lambda_{q_{h+j}+q_h,m}-\lambda_{q_{h+j},m}\lambda_{q_h,m}\right\}\prod_{l=h+j+1}^{k}\lambda_{q_l,m}\prod_{l=h+1}^{k}\lambda_{q_l,m}\prod_{l=1}^{h-1}\lambda_{q_l+q_{l+j},m}\prod_{l=1}^{j}\lambda_{q_l,m}\right)\right\}\int_0^1|\sigma_u|^{2q_+}\mathrm{d}u.$$

What is left is to show that the constant in front of $\int_0^1|\sigma_u|^{2q_+}\mathrm{d}u$ equals $\Lambda^m_{(q_1,\ldots,q_k)}$. First, make the observation that

$$\sum_{j=1}^{k}\left\{\lambda_{2q_j,m}-\lambda^2_{q_j,m}\right\}\prod_{l=1}^{j-1}\lambda_{2q_l,m}\prod_{l=j+1}^{k}\lambda^2_{q_l,m} = \sum_{j=1}^{k}\left(\prod_{l=1}^{j}\lambda_{2q_l,m}\prod_{l=j+1}^{k}\lambda^2_{q_l,m} - \prod_{l=1}^{j-1}\lambda_{2q_l,m}\prod_{l=j}^{k}\lambda^2_{q_l,m}\right)$$
$$= \prod_{l=1}^{k}\lambda_{2q_l,m} - \prod_{l=1}^{k}\lambda^2_{q_l,m},$$

where we applied a telescopic sum argument. Using the notation

$$a_{h,j} = \prod_{l=h+j+1}^{k}\lambda_{q_l,m}\prod_{l=h+1}^{k}\lambda_{q_l,m}\prod_{l=1}^{h}\lambda_{q_l+q_{l+j},m}\prod_{l=1}^{j}\lambda_{q_l,m}$$

allows us to reduce this further

$$2\sum_{j=1}^{k-1}\sum_{h=1}^{k-j}\left\{\lambda_{q_{h+j}+q_h,m}-\lambda_{q_{h+j},m}\lambda_{q_h,m}\right\}\prod_{l=h+j+1}^{k}\lambda_{q_l,m}\prod_{l=h+1}^{k}\lambda_{q_l,m}\prod_{l=1}^{h-1}\lambda_{q_l+q_{l+j},m}\prod_{l=1}^{j}\lambda_{q_l,m}$$

$$=2\sum_{j=1}^{k-1}\sum_{h=1}^{k-j}(a_{h,j}-a_{h-1,j})=2\sum_{j=1}^{k-1}(a_{k-j,j}-a_{0,j})$$

$$=-2(k-1)\prod_{l=1}^{k}\lambda_{q_l,m}^2+2\sum_{j=1}^{k-1}\prod_{l=k-j+1}^{k}\lambda_{q_l,m}\prod_{l=1}^{k-j}\lambda_{q_l+q_{l+j},m}\prod_{l=1}^{j}\lambda_{q_l,m},$$

which, upon collecting pieces, completes the proof. ■

B.2 Justification of the approximation: I

To keep ideas fixed, we concentrate here on the $k = 2$ setting (i.e., range-based bipower variation). The scheme of the proof does not change for a general k, but of course it does get slightly more complicated, although not by much.

First, we define the process

$$U(g_1,g_2)^{n,m}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{g_1\left(\sqrt{n}s_{p_{i\Delta,\Delta},m}\right)g_2\left(\sqrt{n}s_{p_{(i+1)\Delta,\Delta},m}\right)$$

$$-\mathbb{E}\left[g_1\left(\sqrt{n}s_{p_{i\Delta,\Delta},m}\right)g_2\left(\sqrt{n}s_{p_{(i+1)\Delta,\Delta},m}\right)\mid\mathcal{F}_{\frac{i-1}{n}}\right]\Big\},$$

Then, the convergence

$$U(g_1,g_2)^{n,m}-U'^{n,m}\xrightarrow{p}0$$

follows as in the return-based world, see Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006, Section 5). As a consequence

$$U(g_1,g_2)^{n,m}\xrightarrow{d_s}\sqrt{\Lambda^m_{(q_1,q_2)}}\int_0^1|\sigma_u|^{q_+}\mathrm{d}B_u,$$

which completes this part. ■

B.3 Justification of the approximation: II

Again, and without loss of generality (but a pleasant loss of complexity), we assume that $k = 2$. In light of the previous steps, the CLT will follow, if we can prove the convergence

$$\sqrt{n}\left(RMV^{n,m}_{(q_1,q_2)} - \int_0^1 |\sigma_u|^{q_+}\mathrm{d}u\right) - U(g,h)^{n,m} \xrightarrow{p} 0,$$

for a fixed $m \in \mathbb{N}$. We do this by proving that

$$\zeta_i^{n,m} = \frac{1}{\sqrt{n}}\mathbb{E}\left[g_1(\sqrt{n}s_{p_{i\Delta,\Delta,m}})g_2(\sqrt{n}s_{p_{(i+1)\Delta,\Delta,m}}) \mid \mathcal{F}_{\frac{i-1}{n}}\right] - \sqrt{n}\int_{\frac{i-1}{n}}^{\frac{i}{n}} \rho_{\sigma_u}(g_1)\,\rho_{\sigma_u}(g_2)\,\mathrm{d}u,$$

is AN. To accomplish this, we split $\zeta_i^{n,m}$ into:

$$\zeta_i^{n,m} = \zeta_i^{\prime n,m} + \zeta_i^{\prime\prime n},$$

where

$$\zeta_i^{\prime n,m} = \frac{1}{\sqrt{n}}\left(\mathbb{E}\left[g_1\left(\sqrt{n}s_{p_{i\Delta,\Delta,m}}\right)g_2\left(\sqrt{n}s_{p_{(i+1)\Delta,\Delta,m}}\right) \mid \mathcal{F}_{\frac{i-1}{n}}\right] - \mathbb{E}\left[g_1\left(\beta^n_{i,1}\right) \mid \mathcal{F}_{\frac{i-1}{n}}\right]\mathbb{E}\left[g_2\left(\beta^n_{i,2}\right) \mid \mathcal{F}_{\frac{i-1}{n}}\right]\right),$$

$$\zeta_i^{\prime\prime n} = \sqrt{n}\int_{\frac{i-1}{n}}^{\frac{i}{n}}\left(\rho_{\sigma_u}(g_1)\,\rho_{\sigma_u}(g_2) - \rho_{\sigma_{\frac{i-1}{n}}}(g_1)\,\rho_{\sigma_{\frac{i-1}{n}}}(g_2)\right)\mathrm{d}u.$$

It follows from Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006, Section 8) that $\zeta_i^{\prime\prime n}$ is AN. So, the only missing piece is to show the sequence $\zeta_i^{\prime n,m}$ is also AN and we are done. We set

$$\xi_i^{n,m} = \sqrt{n}s_{p_{i\Delta,\Delta,m}} - \beta^m_{i,1}, \qquad \xi_i^{\prime n,m} = \sqrt{n}s_{p_{(i+1)\Delta,\Delta,m}} - \beta^m_{i,2}.$$

Using assumption (V_2), we introduce the random variables:

$$\zeta(1)_i^{n,m} = \sqrt{n}\max\Bigg(\int_{s,t\in I_{i,n}^m}^t \sigma_{\frac{i-1}{n}}\mathrm{d}W_u + \int_s^t \mu_{\frac{i-1}{n}}\mathrm{d}u$$

$$+\int_s^t \left\{\sigma'_{\frac{i-1}{n}}\left(W_u - W_{\frac{i-1}{n}}\right) + v'_{\frac{i-1}{n}}\left(B'_u - B'_{\frac{i-1}{n}}\right)\right\}\mathrm{d}W_u\Bigg) - \beta_{i,1}^n, \tag{37}$$

$$\zeta(2)_i^{n,m} = \sqrt{n}\Bigg\{\max\Big(\int_{s,t\in I_{i,n}^m}^t \mu_u\mathrm{d}u + \int_s^t \sigma_u\mathrm{d}W_u\Big) - \max\Bigg(\int_{s,t\in I_{i,n}^m}^t \sigma_{\frac{i-1}{n}}\mathrm{d}W_u + \int_s^t \mu_{\frac{i-1}{n}}\mathrm{d}u$$

$$+\int_s^t \left\{\sigma'_{\frac{i-1}{n}}\left(W_u - W_{\frac{i-1}{n}}\right) + v'_{\frac{i-1}{n}}\left(B'_u - B'_{\frac{i-1}{n}}\right)\right\}\mathrm{d}W_u\Bigg)\Bigg\}, \tag{38}$$

where $I_{i,n}^m = \left\{t \mid t = \frac{i-1}{n} + \frac{j}{N} \text{ for some } 0 \le j \le m\right\}$. We get that

$$\xi_i^{n,m} = \zeta(1)_i^{n,m} + \zeta(2)_i^{n,m},$$

with an identical decomposition holding for $\xi_i'^{n,m}$. As in Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006, Section 7.1) we obtain, under assumption (V_2), the estimate

$$\mathbb{E}\left[|\xi_i^{n,m}|^q\right] \le Cn^{-\frac{q}{2}},$$

for any $q > 0$ and uniformly in i. We rewrite $\zeta_i'^{n,m}$ as follows

$$\zeta_i'^{n,m} = \mathbb{E}\left[\delta_i^{n,m} \mid \mathcal{F}_{\frac{i-1}{n}}\right],$$

with $\delta_i^{n,m}$ defined by:

$$\delta_i^{n,m} = \frac{1}{\sqrt{n}}\left(g_1\left(\sqrt{n}s_{p_{i\Delta,\Delta},m}\right)g_2\left(\sqrt{n}s_{p_{(i+1)\Delta,\Delta},m}\right) - g_1\left(\beta_{i,1}^n\right)g_2\left(\beta_{i,2}^n\right)\right).$$

Next, we observe that

$$\delta_i^{n,m} = \frac{1}{\sqrt{n}}g_1\left(\sqrt{n}s_{p_{i\Delta,\Delta},m}\right)\left(g_2\left(\sqrt{n}s_{p_{(i+1)\Delta,\Delta},m}\right) - g_2\left(\beta_{i,2}^n\right)\right)$$

$$+\frac{1}{\sqrt{n}}\left(g_1\left(\sqrt{n}s_{p_{i\Delta,\Delta},m}\right) - g\left(\beta_{i,1}^n\right)\right)g_2\left(\beta_{i,2}^n\right) \equiv \delta_i'^{n,m} + \delta_i''^{n,m}.$$

We show that

$$\mathbb{E}\left[\delta_i''^{n,m} \mid \mathcal{F}_{\frac{i-1}{n}}\right]$$

is AN, but omit the proof for $\mathbb{E}\left[\delta_i'^{n,m} \mid \mathcal{F}_{\frac{i-1}{n}}\right]$ to save space. Note that

$$\sum_{i=1}^{n} \mathbb{E}\left[\delta_i''^{n,m} \mid \mathcal{F}_{\frac{i-1}{n}}\right] = \sum_{i=1}^{n} \mathbb{E}\left[\vartheta_i^{n,m} \mid \mathcal{F}_{\frac{i-1}{n}}\right] + o_p(1)$$

with

$$\vartheta_i^{n,m} = \frac{1}{\sqrt{n}} g_2\left(\beta_{i,2}^n\right) \nabla g_1\left(\beta_{i,1}^n\right) \xi_i^{n,m}.$$

While the above approximation appears to be a simple application of the mean value theorem, this result is actually highly non-trivial, because the function g_1 is not differentiable at 0, when $q_1 < 1$. Nonetheless, $\nabla g_1\left(\beta_{i,1}^n\right)$ is well-defined a.s. and $\mathbb{E}\left[\nabla g_1\left(\beta_{i,1}^n\right)\right] < \infty$ under assumption (V_1). The approximation can be shown as in Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006, Section 8).

Recall that

$$\xi_i^{n,m} = \zeta\left(1\right)_i^{n,m} + \zeta\left(2\right)_i^{n,m},$$

with $\zeta\left(1\right)_i^{n,m}$ and $\zeta\left(2\right)_i^{n,m}$ defined by (37) and (38), respectively. Set

$$\begin{aligned}
f_{in}\left(s,t\right) &= \sqrt{n}\sigma_{\frac{i-1}{n}}\left(W_t - W_s\right), \\
g_{in}\left(s,t\right) &= n\int_s^t \mu_{\frac{i-1}{n}} \mathrm{d}u + n\int_s^t \left\{\sigma'_{\frac{i-1}{n}}\left(W_u - W_{\frac{i-1}{n}}\right) + v'_{\frac{i-1}{n}}\left(B'_u - B'_{\frac{i-1}{n}}\right)\right\} \mathrm{d}W_u \\
&= \mu_{\frac{i-1}{n}} g_{in}^1\left(s,t\right) + \sigma'_{\frac{i-1}{n}} g_{in}^2\left(s,t\right) + v'_{\frac{i-1}{n}} g_{in}^3\left(s,t\right),
\end{aligned}$$

to achieve the identity:

$$\zeta\left(1\right)_i^{n,m} = \max_{s,t\in I_{i,n}^m}\left(f_{in}\left(t,s\right) + \frac{1}{\sqrt{n}} g_{in}\left(t,s\right)\right) - \max_{s,t\in I_{i,n}^m} f_{in}\left(t,s\right).$$

Imposing assumption (V_1):

$$\begin{aligned}(t_{in}^{*m}(W), s_{in}^{*m}(W)) &= \arg\max_{s,t\in I_{i,n}^m} f_{in}(s,t) \\ &= \arg\max_{s,t\in I_{i,n}^m} \sqrt{n}(W_t - W_s) \\ &\stackrel{d}{=} \arg\max_{s,t=0,1,\ldots,m} (W_{t/m} - W_{s/m}).\end{aligned}$$

A standard result then states that the pair $(t_{in}^{*m}(W), s_{in}^{*m}(W))$ is unique, almost surely (e.g. Karatzas and Shreve, 1991, p. 107). Next, the following results, which are proven in Christensen and Podolskij (2007), present a useful stochastic expansion for $\zeta(1)_i^{n,m}$ and a result on $\zeta(2)_i^{n,m}$.

Lemma 2. *Given assumption (V_1)*

$$\zeta(1)_i^{n,m} = \frac{1}{\sqrt{n}}\Big\{g_{in}(t_{in}^{*m}(W), s_{in}^{*m}(W)) + \tilde{g}_{in}^m\Big\},$$

where

$$\mathbb{E}\big[|\tilde{g}_{in}^m|^p\big] = \mathrm{o}(1), \tag{39}$$

for all $p > 0$ and uniformly in i.

Lemma 3. *If $q \geq 2$, it then holds that*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(\mathbb{E}\left[|\zeta(2)_i^{n,m}|^q\right]\Big)^{\frac{1}{q}} \to 0,$$

for all $t > 0$.

Note that $(t_{in}^{*m}(W), s_{in}^{*m}(W)) = (s_{in}^{*m}(-W), t_{in}^{*m}(-W))$. Moreover, as $(W, B') \stackrel{d}{=} -(W, B')$ and $\nabla g_1\big(\beta_{i,1}^n\big)$ is an even functional of W:

$$\mathbb{E}\left[\nabla g_1\left(\beta_{i,1}^n\right) g_{in}^k(t_{in}^{*m}(W), s_{in}^{*m}(W)) \mid \mathcal{F}_{\frac{i-1}{n}}\right] = 0,$$

for $k = 1, 2, 3$. Hence

$$\mathbb{E}\left[\nabla g_1\left(\beta_{i,1}^n\right) g_{in}(t_{in}^{*m}(W), s_{in}^{*m}(W)) \mid \mathcal{F}_{\frac{i-1}{n}}\right] = 0. \tag{40}$$

Using Hölder's inequality, it follows that

$$
\begin{aligned}
|\,\mathbb{E}\left[\vartheta_i^{n,m} \mid \mathcal{F}_{\frac{i-1}{n}}\right]| &= \frac{1}{\sqrt{n}}\rho_{\sigma_{\frac{i-1}{n}}}(g_2)\,|\,\mathbb{E}\left[\nabla g_1\left(\beta_{i,1}^n\right)\left(\zeta(1)_i^{n,m}+\zeta(2)_i^{n,m}\right) \mid \mathcal{F}_{\frac{i-1}{n}}\right]| \\
&\leq \frac{1}{\sqrt{n}}\rho_{\sigma_{\frac{i-1}{n}}}(g_2)\left(|\,\mathbb{E}\left[\nabla g_1\left(\beta_{i,1}^n\right)\zeta(1)_i^{n,m} \mid \mathcal{F}_{\frac{i-1}{n}}\right]|\right. \\
&\left. + \left(\mathbb{E}\left[|\nabla g_1\left(\beta_{i,1}^n|\right)^p\right]\right)^{\frac{1}{p}}\left(\mathbb{E}\left[|\zeta(2)_i^{n,m}|^q\right]\right)^{\frac{1}{q}}\right),
\end{aligned}
$$

for some $p > 1, q \geq 2$ with $(q_1 - 1)\,p > -1$ and $1/p + 1/q = 1$. Note that $\mathbb{E}\left[\left(\nabla g_1\left(\beta_{i,1}^n\right)\right)^p\right] \leq C_p < \infty$, when $(q_1 - 1)\,p > -1$. Finally, by combining Lemma 2 and 3 with Eq. (39) and (40), we get the AN property of the sequence $\mathbb{E}\left[\vartheta_i^{n,m} \mid \mathcal{F}_{\frac{i-1}{n}}\right]$. Hence, the CLT has been proven. ■

References

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.

———, 2011, "Analyzing the spectrum of asset returns: Jump and volatility components in high frequency data," *Journal of Economic Literature*, (Forthcoming).

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang, 2011, "Edgeworth expansions for realized volatility and related estimators," *Journal of Econometrics*, 160(1), 190–203.

Alizadeh, S., M. W. Brandt, and F. X. Diebold, 2002, "Range-based estimation of stochastic volatility models," *Journal of Finance*, 57(3), 1047–1092.

Andersen, T. G., L. Benzoni, and J. Lund, 2002, "An empirical investigation of continuous-time equity return models," *Journal of Finance*, 57(4), 1239–1284.

Andersen, T. G., and T. Bollerslev, 1997, "Intraday periodicity and volatility persistence in financial markets," *Journal of Empirical Finance*, 4(2), 115–158.

———, 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.

———, 2010, "Parametric and nonparametric volatility measurement," in *Handbook of Financial Econometrics*, ed. by L. P. Hansen, and Y. Ait-Sahalia. North-Holland, pp. 67–138.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys, 2003, "Modeling and forecasting realized volatility," *Econometrica*, 71(2), 579–625.

Andersen, T. G., T. Bollerslev, and X. Huang, 2011, "A reduced form framework for modeling volatility of speculative prices based on realized variation measures," *Journal of Econometrics*, 160(1), 176–189.

Andersen, T. G., D. Dobrev, and E. Schaumburg, 2008, "Jump robust volatility estimation using nearest neighbour truncation," Working paper, Northwestern University.

Back, K., 1991, "Asset prices for general processes," *Journal of Mathematical Economics*, 20(4), 371–395.

Bannouh, K., D. van Dijk, and M. Martens, 2009, "Range-based covariance estimation using high-frequency data: The realized co-range," *Journal of Financial Econometrics*, 7(4), 341–372.

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-Verlag, pp. 33–68.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard, 2008, "Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise," *Econometrica*, 76(6), 1481–1536.

———, 2009, "Realized kernels in practice: trades and quotes," *Econometrics Journal*, 12(3), 1–32.

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.

———, 2007, "Variation, jumps, market frictions and high frequency data in financial econometrics," in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, ed. by R. Blundell, P. Torsten, and W. K. Newey. Cambridge University Press.

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.

Bates, D. S., 1996, "Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche Mark options," *Review of Financial Studies*, 9(1), 69–107.

———, 2011, "U.S. stock market crash risk, 1926 - 2010," *Journal of Financial Economics*, (Forthcoming).

Bollerslev, T., T. H. Law, and G. Tauchen, 2008, "Risk, jumps, and diversification," *Journal of Econometrics*, 144(1), 234–256.

Bollerslev, T., and V. Todorov, 2011, "Tails, fears and risk premia," *Journal of Finance*, (Forthcoming).

Brandt, M. W., and F. X. Diebold, 2006, "A no-arbitrage approach to range-based estimation of return covariances and correlations," *Journal of Business*, 79(1), 61–73.

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, "Realised quantile-based estimation of the integrated variance," *Journal of Econometrics*, 159(1), 74–98.

———, 2011, "Fact or friction: Jumps at ultra high frequency," Working paper, Aarhus University, CREATES.

Christensen, K., and M. Podolskij, 2007, "Realized range-based estimation of integrated variance," *Journal of Econometrics*, 141(2), 323–349.

Christensen, K., M. Podolskij, and M. Vetter, 2009, "Bias-correcting the realized range-based variance in the presence of market microstructure noise," *Finance and Stochastics*, 13(2), 239–268.

Christie, A. A., 1982, "The stochastic behavior of common stock variances: Value, leverage and interest rate effects," *Journal of Financial Economics*, 10(4), 407–432.

Corsi, F., D. Pirino, and R. Renò, 2010, "Threshold bipower variation and the impact of jumps on volatility forecasting," *Journal of Econometrics*, 159(2), 276–288.

Corsi, F., and R. Renò, 2010, "Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling," Working paper, available at http://ssrn.com/abstract=1316953.

Delbaen, F., and W. Schachermayer, 1994, "A general version of the fundamental theorem of asset pricing," *Mathematische Annalen*, 300, 463–520.

Dobrev, D., 2007, "Capturing volatility from large price moves: Generalized range theory and applications," Working paper, Kellogg School of Management, Northwestern University.

Dobrev, D., and P. Szerszen, 2010, "The information content of high-frequency data for estimating equity return models and forecasting risk," Working paper, Board of Governers of the Federal Reserve System.

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.

Gallant, A. R., C.-T. Hsu, and G. E. Tauchen, 1999, "Using daily range data to calibrate volatility diffusions and extract the forward integrated variance," *Review of Economics and Statistics*, 81(4), 617–631.

Gonçalves, S., and N. Meddahi, 2009, "Bootstrapping realized volatility," *Econometrica*, 77(1), 283–306.

———, 2011, "BoxCox transforms for realized volatility," *Journal of Econometrics*, 160(1), 129–144.

Hansen, P. R., and A. Lunde, 2006, "Realized variance and market microstructure noise," *Journal of Business and Economic Statistics*, 24(2), 127–161.

Heston, S. L., 1993, "A closed-form solution for options with stochastic volatility with applications to bond and currency options," *Review of Financial Studies*, 6(2), 327–343.

Huang, X., and G. Tauchen, 2005, "The relative contribution of jumps to total price variance," *Journal of Financial Econometrics*, 3(4), 456–499.

Hull, J., and A. White, 1987, "The pricing of options on assets with stochastic volatilities," *Journal of Finance*, 42(2), 281–300.

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 A. N. Shiryaev, 2003, *Limit Theorems for Stochastic Processes*. Springer-Verlag, 2 edn.

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.

Karatzas, I., and S. E. Shreve, 1991, *Brownian Motion and Stochastic Calculus*. Springer-Verlag, 2 edn.

Large, J., 2010, "Estimating quadratic variation when quoted prices change by a constant increment," *Journal of Econometrics*, 160(1), 2–11.

Lee, S. S., and P. A. Mykland, 2008a, "Jumps in equilibrium prices and market microstructure noise," Working paper, Georgia Institute of Technology.

———, 2008b, "Jumps in financial markets: A new nonparametric test and jump dynamics," *Review of Financial Studies*, 21(6), 2535–2563.

Li, J., 2011, "Testing for jumps: A delta-hedging perspective," Working paper, Princeton University.

Mancini, C., 2004, "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.

Martens, M., and D. van Dijk, 2007, "Measuring volatility with the realized range," *Journal of Econometrics*, 138(1), 181–207.

Mykland, P. A., 2010, "A Gaussian calculus for inference from high frequency data," *Annals of Finance*, (Forthcoming).

Mykland, P. A., N. Shephard, and K. Sheppard, 2010, "Econometric analysis of financial jumps using efficient bipower variation," Working paper, Oxford-Man Institute, University of Oxford.

Oomen, R. C. A., 2006, "Properties of realized variance under alternative sampling schemes," *Journal of Business and Economic Statistics*, 24(2), 219–237.

Parkinson, M., 1980, "The extreme value method for estimating the variance of the rate of return," *Journal of Business*, 53(1), 61–65.

Patton, A. J., and K. Sheppard, 2009, "Optimal combinations of realised volatility estimators," *International Journal of Forecasting*, 25(2), 218–238.

Podolskij, M., and M. Vetter, 2009a, "Bipower-type estimation in a noisy diffusion setting," *Stochastic Processes and their Applications*, 119(9), 2803–2831.

———, 2009b, "Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps," *Bernoulli*, 15(3), 634–658.

Protter, P. E., 2004, *Stochastic Integration and Differential Equations*. Springer-Verlag, 1 edn.

Rényi, A., 1963, "On stable sequences of events," *Sankhya: The Indian Journal of Statistics; Series A*, 25(3), 293–302.

Rossi, E., and F. Spazzini, 2009, "Finite sample results of range-based integrated volatility estimation," Working paper, University of Pavia.

Tauchen, G., and H. Zhou, 2011, "Realized jumps on financial markets and predicting credit spreads," *Journal of Econometrics*, 160(1), 102–118.

Todorov, V., 2009, "Estimation of continuous-time stochastic volatility models with jumps using high-frequency data," *Journal of Econometrics*, 148(2), 131–148.

———, 2010, "Variance risk-premium dynamics: The role of jumps," *Review of Financial Studies*, 23(1), 345–383.

Todorov, V., and G. Tauchen, 2010, "Activity signature functions for high-frequency data analysis," *Journal of Econometrics*, 154(2), 125–138.

Wasserfallen, W., and H. Zimmermann, 1985, "The behavior of intraday exchange rates," *Journal of Banking and Finance*, 9(1), 55–72.

Woerner, J. H. C., 2005, "Estimation of integrated volatility in stochastic volatility models," *Applied Stochastic Models in Business and Industry*, 21(1), 27–44.

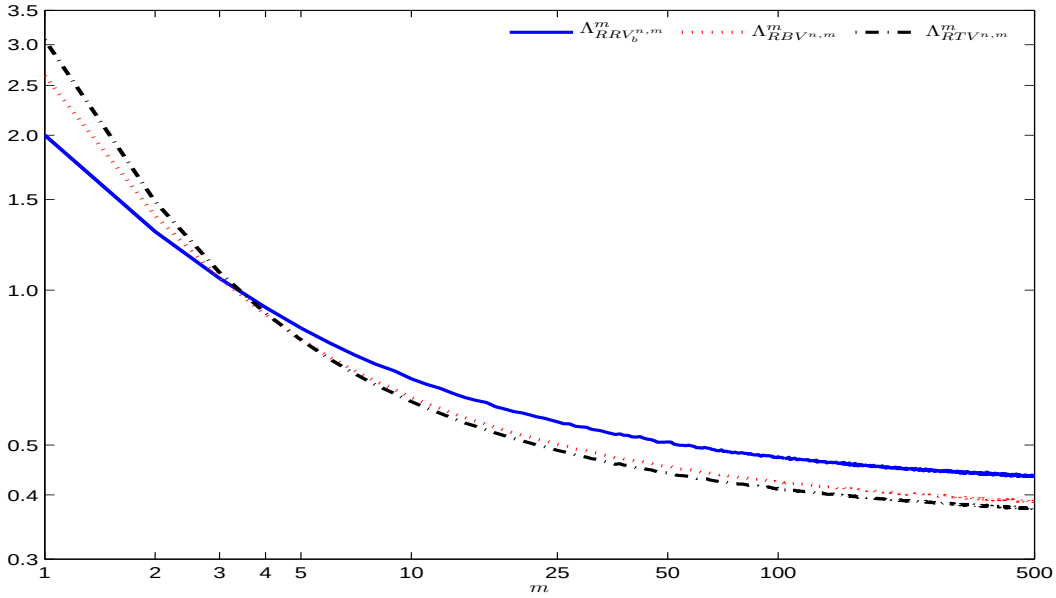
———, 2006, "Power and multipower variation: Inference for high-frequency data," in *Proceedings of the International Conference on Stochastic Finance 2004*, ed. by A. N. Shiryaev, M. do R. Grossihno, P. Oliviera, and M. Esquivel, pp. 343–364. Springer-Verlag.

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.

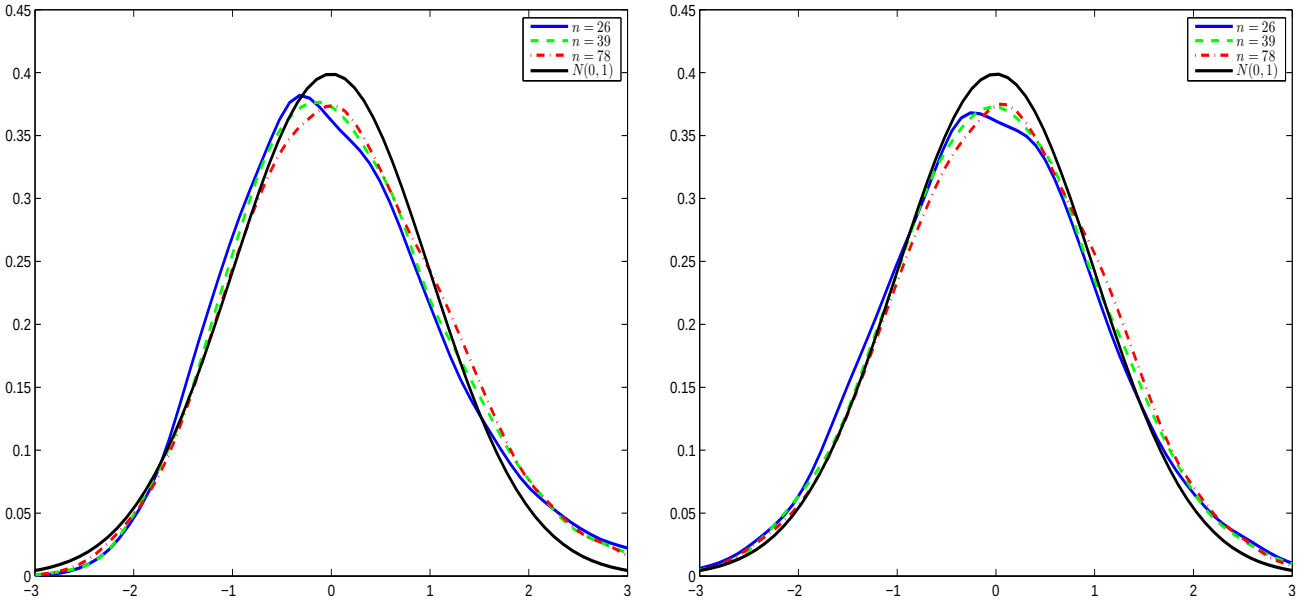
Figure 1: Asymptotic variance factor of realised range-based estimators.



Note. The figure shows the asymptotic variance factor of the range-based realised, bipower and tripower variance, as a function of m.

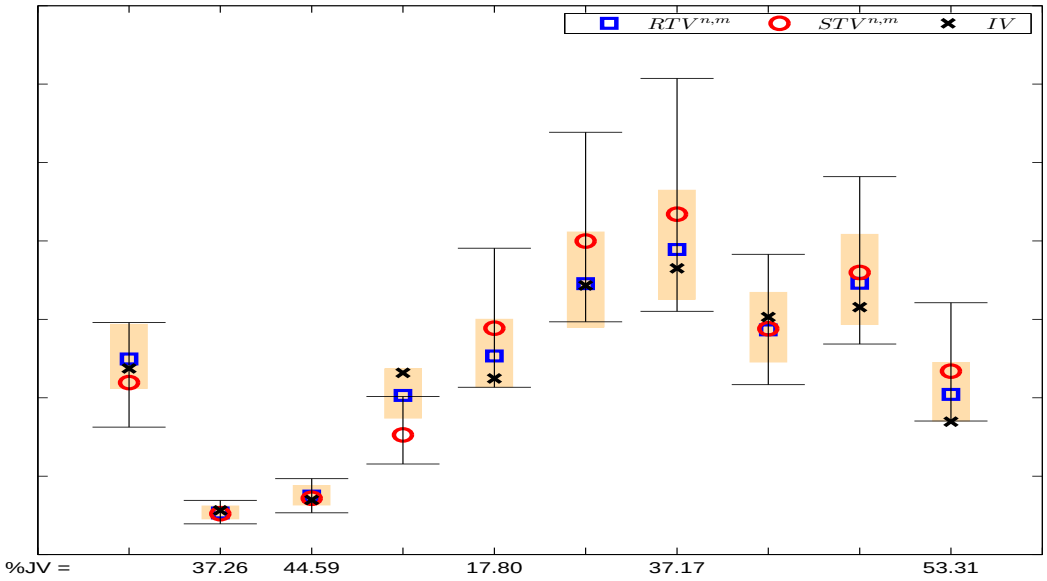
Figure 2: Asymptotic approximation of the realised range-based tripower variance.

Panel A: $RTV^{n,m}$ Panel B: $\ln RTV^{n,m}$



Note. We present smoothed density plots of the raw and log-based sampling distribution of the standardized realised range-based tripower variance estimator. The sampling frequency runs through $n = 26, 39, 78$, which represents 15-, 10- and 5-minute sampling and corresponds to $m = 90, 60, 30$.

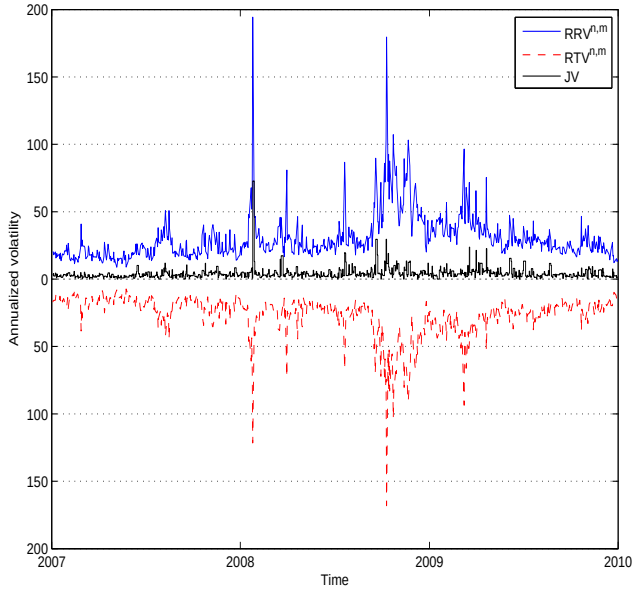
Figure 3: Jump-robust 95% confidence intervals for the integrated variance.



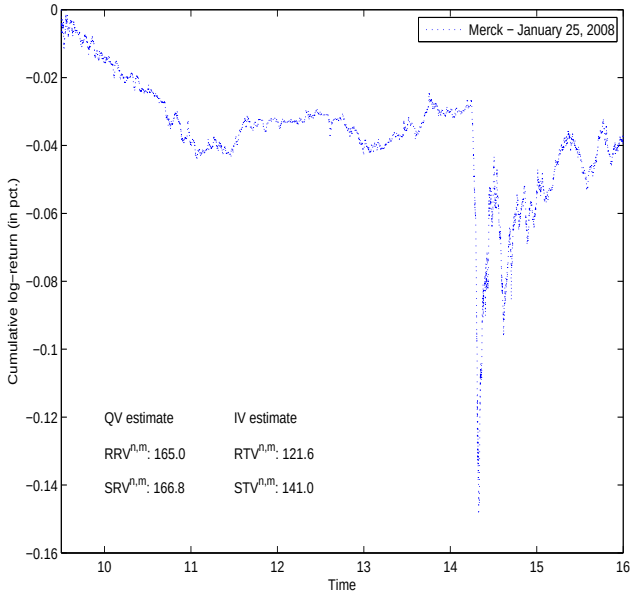
Note. This chart shows point estimates of $RTV^{n,m}$ and $STV^{n,m}$ and 95% confidence intervals for the integrated variance, using $n = 78$. The box ("whisker") is the range-based (return-based) confidence interval. In both cases, the integrated quarticity is proxied with a tripower estimator. **x** marks the true value of integrated variance. The labels on the x-axis report the proportion of quadratic variation induced by jumps (blank if zero).

Figure 4: Illustration using MRK.

Panel A: $RRV^{n,m}$, $RTV^{n,m}$ and jump variation



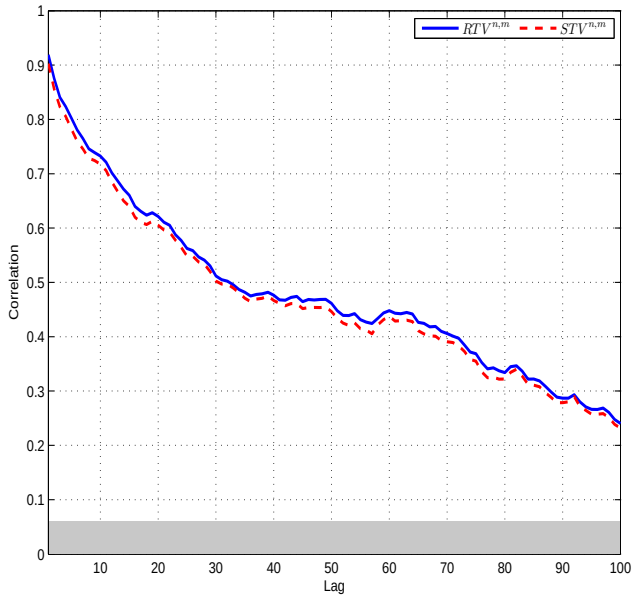
Panel B: MRK transaction data, 20080125



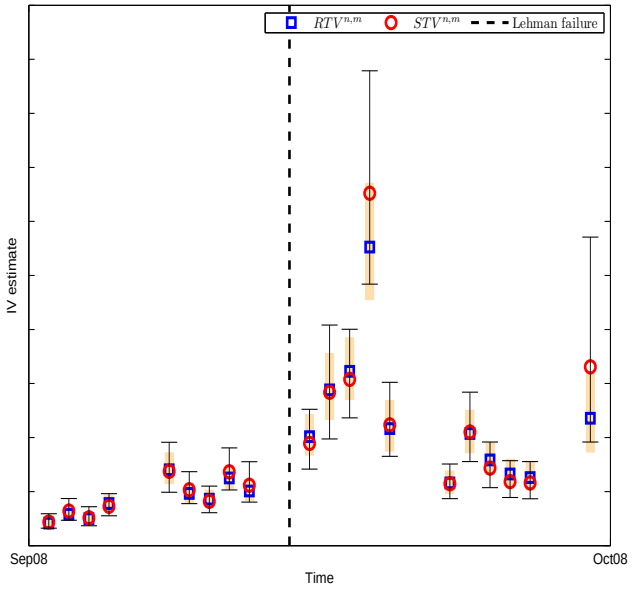
Note. In Panel A, we plot the complete time series of $RRV^{n,m}$ and $RTV^{n,m}$ for MRK. The latter has been reflected in the x-axis to improve the visual layout. Jump variation is the defined as $\max\left(RRV^{n,m} - RTV^{n,m}, 0\right)$. The series have been converted to an annualized standard deviation measure. In Panel B, we plot the filtered transaction data for MRK on January 25, 2008.

Figure 5: Comparison of return- and range-based estimator of the integrated variance, SPY data.

Panel A: ACF of $RTV^{n,m}$ and $STV^{n,m}$



Panel B: Estimates of IV, September 2008



Note. To the left, we plot the autocorrelation function (ACF) of the jump-robust $RTV^{n,m}$ and $STV^{n,m}$ from lags 1 – 100. The shaded area defines the Bartlett's two standard error bands for testing a white noise hypothesis. To the right, we display the $RTV^{n,m}$ and $STV^{n,m}$ point and interval estimates of the integrated variance for the month of September, 2008.

Table 1: Relative bias and root mean squared error.

		Absence of noise			*Presence of noise*		
	$(n,m)=$	$(78,30)$	$(39,60)$	$(26,90)$	$(78,30)$	$(39,60)$	$(26,90)$
Panel A: Range-based							
$RRV^{n,m}$	[QV]	1.011	1.022	1.033	1.018	1.027	1.037
		(0.349)	(0.522)	(0.667)	(0.353)	(0.524)	(0.669)
$RBV^{n,m}$	[IV]	1.019	1.023	1.025	1.031	1.031	1.032
		(0.298)	(0.398)	(0.485)	(0.310)	(0.405)	(0.492)
$RTV^{n,m}$	[IV]	1.010	1.011	1.010	1.022	1.019	1.017
		(0.281)	(0.378)	(0.469)	(0.291)	(0.384)	(0.476)
Panel B: Return-based (with subsampling)							
$SRV^{n,m}$	[QV]	0.999	0.999	0.997	1.001	0.999	0.997
		(0.467)	(0.663)	(0.816)	(0.467)	(0.662)	(0.817)
$SBV^{n,m}$	[IV]	1.034	1.039	1.040	1.036	1.040	1.040
		(0.539)	(0.749)	(0.901)	(0.540)	(0.749)	(0.901)
$STV^{n,m}$	[IV]	1.021	1.023	1.022	1.023	1.023	1.022
		(0.528)	(0.737)	(0.895)	(0.529)	(0.737)	(0.895)

Note. We report the relative bias and rmse of the estimators included in the simulation study. The bias measure is equal to 1 for an unbiased estimator. The number reported in parenthesis is 1000 $\times$ rmse. The square bracket to the right of each estimator shows its theoretical limit, against which the numbers are computed.

Table 2: Relative bias and mean squared error with irregularly spaced data.

$(n, m) = (78, 30)$	return-based		range-based	
	$SRV^{n,m}$	$STV^{n,m}$	$RRV^{n,m}$	$RTV^{n,m}$
Relative bias	1.000	0.990	0.965	0.947
Mse$\times n$	1.399	1.665	0.643	0.681

Note. We report the relative bias and mse of the estimators using irregularly spaced data. The bias measure is 1 for an unbiased estimator. The mse has been normalized by multiplying with n.

Table 3: Descriptive statistics of the equity data.

	AXP				**MRK**			
	2007	2008	2009	all	2007	2008	2009	all
nobs(avg. 1000)	8.3	12.3	11.2	10.6	9.4	11.8	10.5	10.6
$RRV^{n,m}$	24.5	59.8	51.4	45.3	20.7	38.9	30.3	30.0
$RTV^{n,m}$	21.8	55.4	47.2	41.5	18.1	34.6	26.7	26.5
JV(range)	10.9	7.5	8.2	8.4	12.4	11.1	12.0	11.7
$SRV^{n,m}$	23.2	56.0	49.2	42.8	19.4	36.3	28.6	28.1
$STV^{n,m}$	21.7	53.7	46.6	40.7	18.0	34.1	26.8	26.3
JV(return)	6.5	4.2	5.3	5.1	7.3	6.2	6.4	6.5
	SPY				**XOM**			
	2007	2008	2009	all	2007	2008	2009	all
nobs(avg. 1000)	15.3	19.4	19.2	17.9	14.3	16.8	14.6	15.2
$RRV^{n,m}$	11.7	26.5	19.6	19.3	21.1	36.4	22.4	26.6
$RTV^{n,m}$	11.0	25.6	18.5	18.4	19.9	34.4	20.9	25.1
JV(range)	5.5	3.4	5.9	4.6	5.6	5.5	6.6	5.8
$SRV^{n,m}$	11.5	26.0	19.2	18.9	20.2	34.9	21.5	25.6
$STV^{n,m}$	10.9	25.2	18.1	18.1	19.5	33.8	20.5	24.6
JV(return)	4.5	3.2	5.8	4.3	3.8	3.2	4.4	3.7

Note. We report descriptive statistics for AXP, MRK, SPY and XOM. The sample period is January, 2007 – December, 2009, both included. nobs is the average number of transaction data after cleaning (in 1000s). Volatility is reported as annualized standard deviation, in percent. JV is one minus the average ratio of the jump-robust integrated variance estimate to the quadratic variance estimate.

Research Papers 2011

2011-31: Rasmus Tangsgaard Varneskov: Generalized Flat-Top Realized Kernel Estimation of Ex-Post Variation of Asset

2011-32: Christian Bach: Conservatism in Corporate Valuation

2011-33: Adrian Pagan and Don Harding: Econometric Analysis and Prediction of Recurrent Events

2011-34: Lars Stentoft: American Option Pricing with Discrete and Continuous Time Models: An Empirical Comparison

2011-35: Rasmus Tangsgaard Varneskov: Flat-Top Realized Kernel Estimation of Quadratic Covariation with Non-Synchronous and Noisy Asset Prices

2011-36: David F. Hendry and Søren Johansen: The Properties of Model Selection when Retaining Theory Variables

2011-37: Torben G. Andersen, Tim Bollerslev, Peter F. Christoffersen and Francis X. Diebold: Financial Risk Measurement for Financial Risk Management

2011-38: Malene Kallestrup-Lamb: The Role of the Spouse in Early Retirement Decisions for Older Workers

2011-39: Torben Schmith, Søren Johansen and Peter Thejll: Statistical analysis of global surface air temperature and sea level using cointegration methods

2011-40: Søren Johansen and Bent Nielsen: Asymptotic theory for iterated one-step Huber-skip estimators

2011-41: Luc Bauwens, Arnaud Dufays and Jeroen V.K. Rombouts: Marginal Likelihood for Markov-switching and Change-point Garch Models

2011-42: Manuel Lukas: Utility-based Forecast Evaluation with Multiple Decision Rules and a New Maxmin Rule

2011-43: Peter Christoffersen, Ruslan Goyenko, Kris Jacobs, Mehdi Karoui: Illiquidity Premia in the Equity Options Market

2011-44: Diego Amaya, Peter Christoffersen, Kris Jacobs and Aurelio Vasquez: Do Realized Skewness and Kurtosis Predict the Cross-Section of Equity Returns?

2011-45: Peter Christoffersen and Hugues Langlois: The Joint Dynamics of Equity Market Factors

2011-46: Peter Christoffersen, Kris Jacobs and Bo Young Chang: Forecasting with Option Implied Information

2011-47: Kim Christensen and Mark Podolskij: Asymptotic theory of range-based multipower variation