# Outlyingness weighted covariation

Kris Boudt[a,b,*] Christophe Croux[a] Sébastien Laurent[c]

[a] *Faculty of Business and Economics, K.U.Leuven*
[b] *Department of Business Studies, Lessius University College*
[c] *Maastricht University and CORE*

**Abstract**

Quadratic covariation is a popular descriptive measure for the volatility of a multivariate price process. It is consistently estimated by the sum of outer products of high-frequency returns. The proposed Realized Outlyingness Weighted Covariation (ROWCov) is a *weighted* sum of outer products of high-frequency returns and downweights returns that, because of jumps or other reasons, are outliers under the Brownian semimartingale model. The ROWCov is positive semidefinite and remains consistent for the integrated covariance in the presence of a finite activity jump process. We illustrate the usefulness of the estimator on 5-minute returns on the transaction prices of the Dow Jones Industrial Average constituents.

Keywords: Continuous-time methods, high-frequency data, jump robustness, quadratic covariation, realized covolatility.

JEL: C22, C32.

---

*Correspondence to: Kris Boudt, Faculty of Business and Economics, 69 Naamsestraat, B-3000 Leuven, Belgium. E-mail: Kris.Boudt@econ.kuleuven.be. Tel: +32 16 326728 Fax: +32 16 326624
Financial support from the FWO, National Bank of Belgium, Research Fund K.U.Leuven (GOA/2007/04) and Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State Prime Minister's Office, science policy programming, is gratefully acknowledged. We thank the Editor (George Tauchen), Associate Editor and two anonymous referees for their extensive and insightful suggestions. We are also grateful to Luc Bauwens, Jonathan Cornelissen, Geert Dhaene, Dobrislav Dobrev, Thomas Ghys, Irène Gijbels, Cecilia Mancini, Neil Shephard and the participants of the Financial Econometrics and Vast Data Conference at the Oxford-Man Institute of Quantitative Finance for helpful comments. The usual disclaimer applies.

# 1 Introduction

The literature on volatility measurement of multivariate price processes is shifting its attention from the estimation of conditional covariance matrices under the multivariate GARCH or stochastic volatility approach (see Bauwens et al., 2006, and Asai et al., 2006, for a review) to the ex post estimation of *Quadratic Covariation* (QCov) under the high-frequency approach. A major reason is the simplicity of the estimation of the QCov, which does not require any optimization nor high-level assumptions on the price diffusion process. Also forecasting the QCov is computationally convenient (Andersen et al., 2007, and Corsi, 2009). Moreover, under the assumption that prices are generated by a continuous-time price diffusion, quadratic covariation is the natural price volatility measure for asset pricing and hedging (Andersen et al., 2001).

The *Realized Covariation* (RCov), defined as the sum of outer products of high-frequency returns, is a consistent estimator for the QCov (Andersen et al., 2003). For the purpose of volatility forecasting and hedging, it is important to be able to disentangle the quadratic covariation due to the continuous diffusion component and due to the jump component (Andersen et al., 2007). A second estimator is thus needed to estimate the contribution of each component to the QCov of the price process.

In their seminal paper "Power and bipower variation with stochastic volatility and jumps", Barndorff-Nielsen and Shephard (2004c) note that for achieving this goal, some structure must be imposed on the continuous time log-price process. They assume that it consists of a conditionally normal component with time-varying covariance matrix and a finite activity jump process. This is the *Brownian Semi-Martingale with Finite Activity Jumps* (BSMFAJ) model for log-prices. Under this model, the RCov is an estimator of the sum of the integral of the spot covariance matrix process, called the *Integrated Covariance matrix* (ICov), and the variability of the jump process. They proposed the *Realized BiPower Covariation* (RBPCov) as a consistent estimator for the ICov. By construction, this estimator of the ICov is asymptotically robust to finite activity jumps. Barndorff-Nielsen et al. (2006) show that the univariate version of the RBPCov is in some cases also robust to infinite activity jumps. However, in finite samples, jumps induce an upward bias in the RBPCov, especially if jumps affect two or more contiguous returns (Andersen et al., 2009; Corsi et al., 2009). Furthermore, in a multivariate setting, the RBPCov is not a completely satisfactory covolatility estimator because it is not affine equivariant and not always positive semidefinite.[1] A final practical disadvantage is that the implied realized correlation estimate given by the ratio of the realized bipower covariation of two assets and the square root of the products of the univariate realized bipower variation does not always lie between -1 and 1.

This paper proposes the *Realized Outlyingness Weighted Covariation* (ROWCov) as an alternative to the RBPCov that has none of these shortcomings. It is defined as the classical RCov applied to weighted, instead of raw, high-frequency returns. We downweight returns with a large local outlyingness, where we say that a return is a local outlier if it has an extreme value relatively to its neighboring returns. Returns affected by jumps will have a large local outlyingness and will therefore receive a lower weight. Our Monte Carlo study shows that there exist weight functions for which the ROWCov is both more efficient and more robust to jumps than the RBPCov.

The remainder of the paper is organized as follows. In Sections 2 and 3 we describe the BSMFAJ model, the existing integrated covariance estimators and the

[1]The affine equivariance property relates to the effect of linear transformations of the return series on the covariance estimate. Let $A$ be a $N \times N$ matrix of constants. If the $N$-dimensional return series, with affine equivariant covariance estimate equal to $Z$, is premultiplied by $A$, then the covariance estimate of the transformed series equals $AZA'$.

proposed ROWCov estimator. Its properties are studied in Sections 4 and 5. In Section 6 we illustrate the ROWCov estimator on 5-minute returns of transaction prices of the Dow Jones Industrial Average constituents. Finally, Section 7 summarizes our findings and makes some suggestions for further research. An accompanying webappendix sketches the proofs and contains additional simulation results.[2]

## 2 Definitions

This section reviews elements of the limit theory presented in Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2004a,c) of the realized covariation and bipower covariation of a Brownian semimartingale log-price diffusion process with jumps. We consider a $N$-dimensional log-price process $p(s)$ over the period $[0,1]$ (typically one day). Assume further that the value of this process is observed every $\Delta$ units of time and synchronously across all $N$ assets. This means that there are $\lfloor 1/\Delta \rfloor$ equidistant observations of $p(s)$ in the time interval $[0,1]$.[3] Denote by $r_{i,\Delta} = p(i\Delta) - p((i-1)\Delta)$ the $i$-th intraday return in the period $[0,1]$. The Quadratic Covariation (QCov) of the log-price process over $[0,1]$ is

$$\text{QCov} = \operatorname*{plim}_{\Delta \to 0} \sum_{i=1}^{\lfloor 1/\Delta \rfloor} r_{i,\Delta} r'_{i,\Delta}. \tag{2.1}$$

By construction, QCov is consistently estimated by the Realized Covariation,

$$\text{RCov}_\Delta = \sum_{i=1}^{\lfloor 1/\Delta \rfloor} r_{i,\Delta} r'_{i,\Delta}. \tag{2.2}$$

As common in the literature on nonparametric estimation of the continuous and

[2]The webappendix is available at `www.econ.kuleuven.be/kris.boudt/public`.
[3]The function $\lfloor \cdot \rfloor$ returns the largest integer less than or equal to its argument.

jump components in the QCov, we suppose that the log-price process belongs either to the Brownian SemiMartingale (BSM) or the BSM with Finite Activity Jumps (BSMFAJ) families of models. Under the BSMFAJ model, the diffusion component captures the smooth variation of the price process, while the jump component accounts for the rare, large discontinuities in the observed prices. Andersen et al. (2007a) cite the work of several authors who found that this is a realistic model for the price series of many financial assets.

Under the BSMFAJ model, the observed returns are assumed to be generated by a $N$-dimensional log-price diffusion $dp(s)$ consisting of a conditionally normal random process with mean $\mu(s)ds$ and covariance $\Sigma(s)ds$ and of an additive jump process. The occurrence of jumps is governed by a vectorial counting process $q(s)$. The jump process is supposed to be independent of the Brownian semimartingale diffusion and to have finite activity, which means that the change in the counting process over any interval of time is finite with probability 1. Denote by $w(s)$ a vector of $N$ independent Brownian motions. Write $\Omega(s)$ the $N \times N$ càdlàg process such that $\Sigma(s) = \Omega(s)\Omega'(s)$ is the spot covariance matrix process of the continuous component of the price diffusion. Let $K(s)$ be the $N \times N$ process controlling the magnitude and transmission of jumps such that $K(s)dq(s)$ is the contribution of the jump process to the price diffusion. The log-price process follows the BSMFAJ model if it can be decomposed as follows:

$$\text{BSMFAJ:} \quad dp(s) \quad = \quad \mu(s)ds + \Omega(s)dw(s) + K(s)dq(s). \tag{2.3}$$

The integrated covariance matrix (ICov) over $[0, 1]$ is the matrix

$$\text{ICov} = \int_0^1 \Sigma(s)ds. \tag{2.4}$$

Denote by $\kappa_j$ the contribution of the $j$-th jump in $[0,1]$ to the log-price diffusion. Andersen et al. (2003) have shown that the RCov is a consistent estimator for the sum of the ICov and the realized jump variability

$$\operatorname*{plim}_{\Delta \to 0} \mathrm{RCov}_\Delta = \mathrm{ICov} + \sum_{j=1}^{J} \kappa_j \kappa_j', \tag{2.5}$$

where $J = \int_0^1 dq^*(s)$, with $q^*(s)$ the univariate counting process derived from $q(s)$ such that $q^*(s)$ increases by 1 whenever $q(s)$ changes. For disentangling the continuous and jump components in the RCov, we need an additional estimator for the ICov that is robust to jumps.

For this purpose, Barndorff-Nielsen and Shephard (2004b) introduce the Realized BiPower Covariation (RBPCov) as the estimator whose value over period $[0,1]$ is the $N$-dimensional square matrix with $k,l$-th element equal to

$$\frac{\pi}{8}\Bigg( \sum_{i=2}^{\lfloor 1/\Delta \rfloor} \left| r_{(k)i,\Delta} + r_{(l)i,\Delta} \right| \left| r_{(k)i-1,\Delta} + r_{(l)i-1,\Delta} \right| - \left| r_{(k)i,\Delta} - r_{(l)i,\Delta} \right| \left| r_{(k)i-1,\Delta} - r_{(l)i-1,\Delta} \right| \Bigg), \tag{2.6}$$

where $r_{(k)i,\Delta}$ is the $k$-th component of the return vector $r_{i,\Delta}$. The factor $\pi/8$ ensures that the RBPCov converges to the ICov under the BSMFAJ model:

$$\operatorname*{plim}_{\Delta \to 0} \mathrm{RBPCov}_\Delta = \int_0^1 \Sigma(s) ds. \tag{2.7}$$

As noted in the introduction, the RBPCov has several disadvantages such as a relative large finite sample bias in the presence of jumps and lack of positive definiteness and affine equivariance. In the next section, we show that the inclusion of a weight function in the RCov giving a zero weight to extreme returns, leads to an estimator

(i) that is affine equivariant and yields positive semidefinite matrices; (ii) is consistent for the ICov and has a high efficiency under the BSM and BSMFAJ models and (iii) for which the finite sample bias remains small if jumps affect contiguous returns.

# 3 Realized outlyingness weighted covariation

## 3.1 Definition

The ROWCov is a weighted version of the RCov, where outlying returns receive a zero weight. The mechanics are simple and based on two steps.

*Step 1: estimation of local outlyingness.* Assume the goal is to estimate the ICov over the period $[0,1]$. First we measure for each high-frequency return in our dataset the local outlyingness. We suppose that the spot volatility process is sufficiently smooth such that we can compute an accurate first step jump robust estimate of the spot covariance matrix at time $(i-1)\Delta$. Denote this estimate $\hat{\Sigma}_{i,\Delta}$. More details on this estimator are given in the next subsection. We will be operating with sufficiently high-frequency return series such that the mean process can be safely ignored. We measure the local outlyingness of the multivariate return vector $r_{i,\Delta}$ by the squared Mahalanobis distance between $r_{i,\Delta}$ and zero in terms of $\hat{\Sigma}_{i,\Delta}\Delta$:

$$d_{i,\Delta} = \frac{r'_{i,\Delta}\hat{\Sigma}^{-1}_{i,\Delta}r_{i,\Delta}}{\Delta}. \tag{3.1}$$

*Step 2: computation of the ROWCov.* The ROWCov is computed as a weighted version of the RCov where returns with a high outlyingness value are downweighted:

$$\text{ROWCov}_{\Delta} = c_w \sum_{i=1}^{\lfloor 1/\Delta \rfloor} w(d_{i,\Delta}) r_{i,\Delta} r'_{i,\Delta}. \tag{3.2}$$

The weight function $w : \mathbb{R}^+ \to \mathbb{R}^+$ needs to be continuous almost everywhere and needs to satisfy that the function $w(z)z$ is bounded. Popular weight functions that satisfy these conditions are the Hard Rejection weight function

$$w_{\text{HR}}(z) = \begin{cases} 1 & \text{if } z \leq k \\ 0 & \text{otherwise,} \end{cases} \tag{3.3}$$

and the Soft Rejection weight function

$$w_{\text{SR}}(z) = \min\{1, k/z\}, \tag{3.4}$$

with threshold $0 < k < \infty$ a tuning parameter to be selected. Under the BSM model and if the first step covariance estimate is consistent for the spot covariance, the outlyingness measure is asymptotically chi-square distributed with $N$ degrees of freedom ($\chi^2_N$). It is thus natural to set the threshold $k$ to $\chi^2_N(\beta)$, the $\beta$ quantile of the $\chi^2_N$ distribution function. On the basis of the simulation study of Section 5, we recommend $\beta = 99.9\%$. Alternatively, more conservative thresholds can be chosen. In a intraday jump test setting, Andersen et al. (2007b) set the threshold such that $1-\beta$ is the probability of wrongly classifying a return within a day. Then $\beta = (1-\alpha)^{\Delta}$, with e.g. $\alpha = 2.5\%$.

The correction factor $c_w$ in (3.2) ensures that the ROWCov is consistent for the ICov under the BSM and BSMFAJ models (as will be further discussed in Section 4). It depends on the dimension $N$ of the process and on the weight function used. The correction factor $c_w$ is then given by

$$c_w = N/E\left[w(z)z\right], \tag{3.5}$$

Table 1: Correction factor $c_w$ for the ROWCov with hard and soft rejection weight functions with threshold $k$, when the dimension of the series is $N$.

| | $k = \chi^2_N(0.95)$ | | | | $k = \chi^2_N(0.99)$ | | | | $k = \chi^2_N(0.999)$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | 1 | 2 | 5 | 10 | 1 | 2 | 5 | 10 | 1 | 2 | 5 | 10 |
| HR | 1.387 | 1.250 | 1.157 | 1.119 | 1.092 | 1.059 | 1.036 | 1.027 | 1.013 | 1.008 | 1.005 | 1.003 |
| SR | 1.095 | 1.053 | 1.026 | 1.015 | 1.018 | 1.010 | 1.005 | 1.003 | 1.002 | 1.001 | 1.000 | 1.000 |

with $z$ a chi-square random variable with $N$ degrees of freedom. Let $F_{\chi^2_{N+2}}(\cdot)$ be the $\chi^2_{N+2}$ distribution function. For the hard rejection weight function, $c_w$ is $1/F_{\chi^2_{N+2}}(k)$ (Croux and Haesbroeck, 1999; Pison et al., 2002). For the soft rejection weight function, we show in the webappendix that for $N = 1$ and $k = \chi^2_1(\beta)$, $c_w$ is $1/[-2\sqrt{k}\phi(\sqrt{k}) + \beta + k(1-\beta)]$, where $\phi$ is the standard univariate Gaussian density function. For $N > 1$, we calculate the factor $c_w$ of the ROWCov with soft rejection weight function by numerical integration. These correction factors are reported in Table 1 for $N = 1, 2, 5$ and 10 and for $\beta = 0.95$, 0.99 and 0.999.

## 3.2 First step covariance estimation

The ROWCov requires a first step covariance estimate $\hat{\Sigma}_{i,\Delta}$ for every intraday return. This estimator needs to be an accurate estimator of the spot covariance estimator, both in the absence and presence of jumps.

In practice we obtain the first step covariance estimate as follows. We divide the time interval $[0, 1]$ into $\lfloor 1/\lambda \rfloor$ contiguous subintervals having time length $\lambda$. We call these intervals "local windows" and assume volatility is almost constant in these intervals.[4] We then estimate $\hat{\Sigma}_i$ as the Minimum Covariance Determinant (MCD) covariance estimate of the standardized returns $r_{j,\Delta}\Delta^{-1/2}$ belonging to the local

[4] Alternatively, sliding local windows centered around $r_{i,\Delta}$ could be used for the estimation of $\hat{\Sigma}_i$. This requires however to compute $\lfloor 1/\Delta \rfloor$ first step covariance estimates per day, rather than $\lfloor 1/\lambda \rfloor$ estimates. For computational reasons, we use non sliding local windows.

window containing $r_{i,\Delta}$.[5] The MCD estimator was initially proposed by Rousseeuw and van Driessen (1999) and is defined as follows. Consider all the subsamples of the sample of returns in the local window that contain 75% of the return observations. Compute for each of these subsamples the sample covariance. The MCD covariance estimate is the sample covariance matrix of the subsample for which the determinant of the covariance matrix is the smallest, multiplied by the consistency factor $0.75/F_{\chi^2_{N+2}}(\chi^2_N(0.75))$. For multivariate normal data, the MCD covariance estimate has been shown to be consistent and asymptotically normal by Butler et al. (1993) and Cator and Lopuhaä (2011). We conjecture that under smoothness conditions on the spot covolatility process and provided jumps have finite activity, this estimator is consistent for the spot covariance matrix under the double asymptotics that there is a sufficiently large number of observations in the local window ($\lambda/\Delta \to \infty$) and that the window is short enough such that the change in the volatility process is negligible over the local window ($\lambda \to 0$). The reason is that if jumps have finite activity, none of the returns affected by jumps end up in the subset from which the MCD covariance is computed almost surely. Moreover, the proportion of returns affected by jumps is infinitesimal and therefore does not affect the correction factor needed for consistency.

We show in the webappendix that the efficiency of the ROWCov is significantly improved by multiplying the weights $w(d_{i,\Delta})$ in (3.2) with the ratio between the expected value of the weights and the sample average weight in the local window where $r_{i,\Delta}$ belongs to.[6] We therefore implement the ROWCov with this modified weight function.

[5] The univariate counterpart of the MCD covariance estimator is the least trimmed squares scale estimator, but many other jump robust variance estimators that are consistent at the normal distribution can be used, such as the median absolute deviation or the Qn estimator (Rousseeuw and Croux, 1993).

[6] For the hard rejection weight with threshold $k = \chi^2(\beta)$, $E[w(z)]$ is of course $\beta$, while for the soft rejection weight function and $N = 1$, $E[w(z)] = \beta + 2\sqrt{k}\phi(\sqrt{k}) - k(1-\beta)$.

The choice of the length of the local windows must be such that the number of observations in the local window is high enough, but not too high (otherwise the approximation that the returns in the local window that are not affected by jumps come from the same normal distribution may no longer be acceptable). Simulation evidence in the webappendix indicates that, when sampling at the 5 or 15-minute frequency, jumps are more accurately detected using local windows of one day rather than shorter windows. For this reason, we recommend using a window length of 1 day.

# 4 Properties of the ROWCov

Under appropriate regularity conditions and the double asymptotics described above, we have that the ROWCov is consistent for the ICov under the BSMFAJ model:

$$\operatorname*{plim}_{\lambda\to 0,\lambda/\Delta\to\infty} \text{ROWCov}_{\Delta} = \int_0^1 \Sigma(s)ds. \tag{4.1}$$

Under the BSM model without leverage effect and appropriate regularity conditions, one has further that for $\lambda \to 0$ and $\lambda/\Delta \to \infty$

$$\Delta^{-1/2}\,\text{vec}\left(\text{ROWCov}_{\Delta} - \text{ICov}\right) \xrightarrow{d} N\left(0, \int_0^1 G(s)\Theta G'(s)ds\right), \tag{4.2}$$

where $G(s) = \Omega(s)\otimes\Omega(s)$ and $\Theta$ is the asymptotic covariance matrix of the ROWCov computed for i.i.d. $N$-dimensional standard normal random variables. If $w(z) = 1$ for all $z$, the ROWCov coincides with the RCov and $\Theta$ is 2 times the $N^2 \times N^2$ identity matrix, corresponding to the result of Barndorff-Nielsen and Shephard (2004a). A sketch of a proof of the results in (4.1)-(4.2) is given in the webappendix. Even though the central limit theorem in (4.2) assumes jumps away, it is a useful result since it allows testing the null of no (co)jump contribution in the RCov.

Note also that, by construction, the ROWCov is an affine equivariant estimator that yields symmetric and positive semidefinite matrices. By the Cauchy-Schwarz inequality, we further have that for any two assets $k$ and $l$, the Realized Outlyingness Weighted Correlation (ROWCor) lies between -1 and 1:

$$-1 \leq \text{ROWCor}_{(k,l)\Delta} \equiv \frac{\text{ROWCov}_{(k,l)\Delta}}{\sqrt{\text{ROWCov}_{(k,k)\Delta}\text{ROWCov}_{(l,l)\Delta}}} \leq 1. \tag{4.3}$$

Unlike the ROWCov, the realized correlation estimate based on the RBPCov does not have this property.

Under the BSMFAJ model, there are two drivers of volatility: the continuous, Brownian diffusion and the jump process. As can be seen in (2.5), the RCov estimates the sum of the ICov and the realized jump variability, but many authors have stressed the importance of being able to decompose the RCov into its continuous and jump components. According to Aït-Sahalia and Jacod (2004) the ability to disentangle the continuous and jump components of volatility is the essence of downside risk management, which should focus on large risks leaving aside the day-to-day Brownian fluctuations. Andersen et al. (2007a) empirically verified that the continuous and jump components of price variability have very different dynamics, the first one being very persistent, the latter one being much less predictable. By the robustness of the ROWCov to jumps, we have that the difference between the RCov and ROWCov can be used to estimate the contribution of jumps to the RCov.

## 4.1 Relationship to the RTCov estimator

In independent and concurrent work, Mancini and Gobbi (2009) proposed the Realized Threshold Covariance (RTCov) estimator, which is similar to the ROWCov implemented with a hard rejection weight function. Unlike the ROWCov, this estimator uses univariate truncation rules however. It estimates the $k, l$th element of the ICov as

$$\sum_{i=1}^{\lfloor 1/\Delta \rfloor} \tilde{w}(r_{(k)i,\Delta})\tilde{w}(r_{(l)i,\Delta})r_{(k)i,\Delta}r_{(l)i,\Delta}, \tag{4.4}$$

where $\tilde{w}(z) = 1$ if $z^2 \leq h(\Delta)$ and the threshold is a deterministic function $h$ of $\Delta$ satisfying the following properties

$$\lim_{\Delta \to 0} h(\Delta) = 0, \; \lim_{\Delta \to 0} \frac{\Delta \log \frac{1}{\Delta}}{h(\Delta)} = 0.$$

The authors recommend to use a threshold function having the functional form $c\Delta^{\xi}$, where $c$ and $\xi$ are constants. In the empirical applications of threshold variance estimation, $\beta$ is typically set to a value that is very close to one. Mancini and Gobbi (2009) recommend $\beta = 0.99$. The constant $c$ is usually calibrated at a multiple (often 9) of a first step estimate of the daily volatility (Aït-Sahalia and Jacod, 2009), such as the realized bipower variation (Jacod and Todorov, 2009) or the GARCH volatility prediction (Mancini and Renó, 2011). In the simulation study and empirical application, we implement the RTCov with $c$ equal to 9 times the daily realized bipower variation.

In the univariate case, this implementation of the threshold covariance estimator is thus very similar to the proposed hard rejection ROWCov. Note however that, even in the case where the spot volatility process is constant over the day, the RTCov is not unbiased at all sampling frequencies. Due to the correction factor $c_w$ in (3.2), the ROWCov does have this property. The use of a univariate threshold is clearly appealing from a computational viewpoint since it does not require a first step multivariate scale estimate. This comes at the price that the resulting covariance estimate is not affine equivariant. Because it uses univariate truncation rules, it has also less robustness to modest-sized cojumps which are not detectible as univariate jumps (Bollerslev et al., 2008). Mancini and Gobbi (2009) prove consistency of the RTCov for the ICov under the BSM model with both finite and infinite activity jumps. We only show consistency of the ROWCov for the ICov under the BSM model with finite activity jumps. Like the ROWCov and RCov, the RTCov is positive semidefinite and the realized correlations based on the RTCov are always between -1 and 1.

# 5 Monte Carlo study

In this section we first compare the accuracy of the RCov, RBPCov, RTCov and ROWCov applied to series of 1, 5 and 15-minute returns generated according to a 2-variate factor stochastic volatility model allowing for rather strong variations of volatility within the day, as in Barndorff-Nielsen et al. (2010).[7] A finite activity jump process is added to investigate the robustness of the various estimators. We then study the impact of microstructure noise and non-synchronous trading on the bias of the estimators. Finally, we study the finite sample accuracy of the normality approximation for variance, covariance and correlation statistics derived from the ROWCov.

## 5.1 Setup

The simulation setup is as follows. We generate a bivariate price series with $p^{(i)}(s)$ the associated log-price expressed in percentage points, for $i = 1, 2$. Let $b^{(i)}(s)$ and $w(s)$ be independent Brownian motions. The Brownian motion $w(s)$ has no

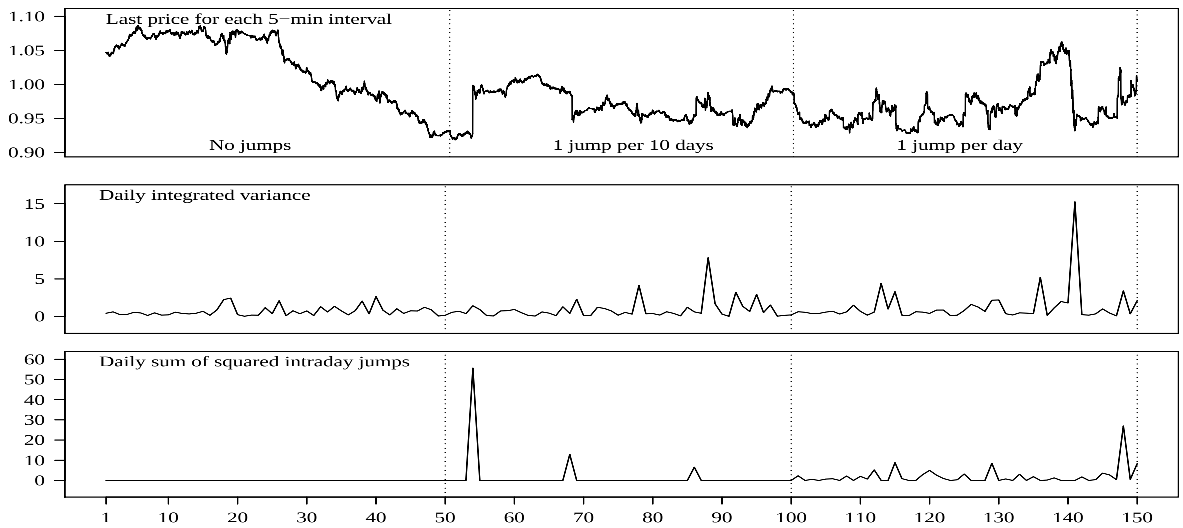
[7] We also did the analysis for a GARCH diffusion process and obtained similar results.

subscript to indicate that it plays the role of a common factor. We simulate a hypothetical log-price series with one observation every second from the log-price diffusion given by

$$dp^{(i)}(s) = \mu ds + \gamma\sigma^{(i)}(s)db^{(i)}(s) + \sqrt{1-\gamma^2}\sigma^{(i)}(s)dw(s) + \kappa^{(i)}(s)dq^{(i)}(s), \qquad (5.1)$$

$i = 1, 2$. The spot volatility $\sigma^{(i)}(s) = \exp(\beta_0 + \beta_1\rho^{(i)}(s))$, with $d\rho^{(i)}(s) = \alpha\rho^{(i)}(s)ds + db^{(i)}(s)$. Note that the innovations in the spot volatility and return series are correlated (the so-called leverage effect). As in Barndorff-Nielsen et al. (2010), the parameters $(\mu, \beta_0, \beta_1, \alpha, \gamma)$ are calibrated at (0.03,$-5/16$, 1/8, $-1/40$, $-0.3$). The process $\rho^{(i)}(s)$ is initialized each day by a random draw from its stationary distribution $N(0, 1/(-2\alpha))$. Because of the constant exposure to the common factor, the continuous log-returns have a constant correlation equal to $(1-\gamma^2) = 0.91$. Jump occurrences are governed by the Poisson processes $q^{(i)}(s)$. They are either independent across assets or identical (cojumps). The intensity is constant and depends on the parameter $\kappa^*$, defined as the expected number of jumps per day. We model the size of the jumps $\kappa^{(i)}(s)$ as the product between the realization of a uniformly distributed random variables on $\sqrt{m/\kappa^*}([-2,-1]\cup[1,2])$ and the mean value of the stochastic volatility process $\sigma^{(i)}$ of that day, for $i = 1, 2$. The parameter $m$ models the magnitude of the jumps. Note that the lower the intensity of the jump process, the larger the jumps are. If volatility is constant and equals $\sigma$ for each series, then the daily variance of each series equals $\sigma^2 + m(7/3)\sigma^2$. In case of cojumps, the sign of the jumps is specified to be the same for all assets. We consider the case of on average 0, 1 or 5 jumps per day ($\kappa^*$) and medium-sized ($m = 0.5$) or large ($m = 1$) jumps.

Figure 1: First series of simulated 5-minute prices, daily integrated variance and sum of squared intraday jumps over 150 days. The parameter $\kappa^*$ is 0 for days 1-50, 1 per 10 days for days 51-100 and 1 per day for days 101-150.

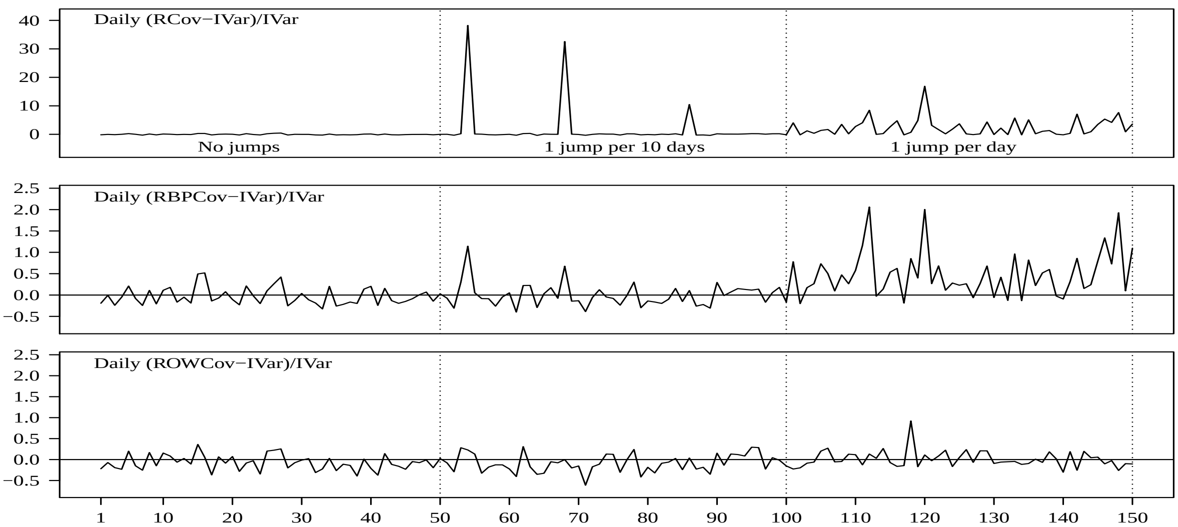



## 5.2 Illustration

We first illustrate the bias effect of jumps on the bivariate RCov, RBPCov and the ROWCov. We do this through an experiment in which we generate three times 50 days of bivariate 5-minute returns and where we let the parameter $\kappa^*$ (the number of expected jumps per day) vary from no jumps on days 1-50, 1 jump per 10 days on days 51-100 and 1 jump per day on the last 50 days. The parameter $m = 1$ and the ROWCov is implemented with the hard rejection weight function and threshold $k = \chi^2_2(0.999)$. In the next subsection, we show that this choice of weight function offers an attractive efficiency/robustness trade-off. For the ROWCov, the outlyingness values are computed using the MCD covariance estimator on local windows of one day, as described in Subsection 3.2. Jump occurrences are independent across assets.

In Figure 1 we plot the simulated values of the 5-minute prices, the daily integrated variance and the daily sum of squared intraday jumps of the first price

Figure 2: Percentage estimation error when using RCov, RBPCov and ROWCov based on 5-minute returns to estimate the IVar of the first price series. The parameter $\kappa^*$ is 0 for days 1-50, 1 per 10 days for days 51-100 and 1 per day for days 101-150.

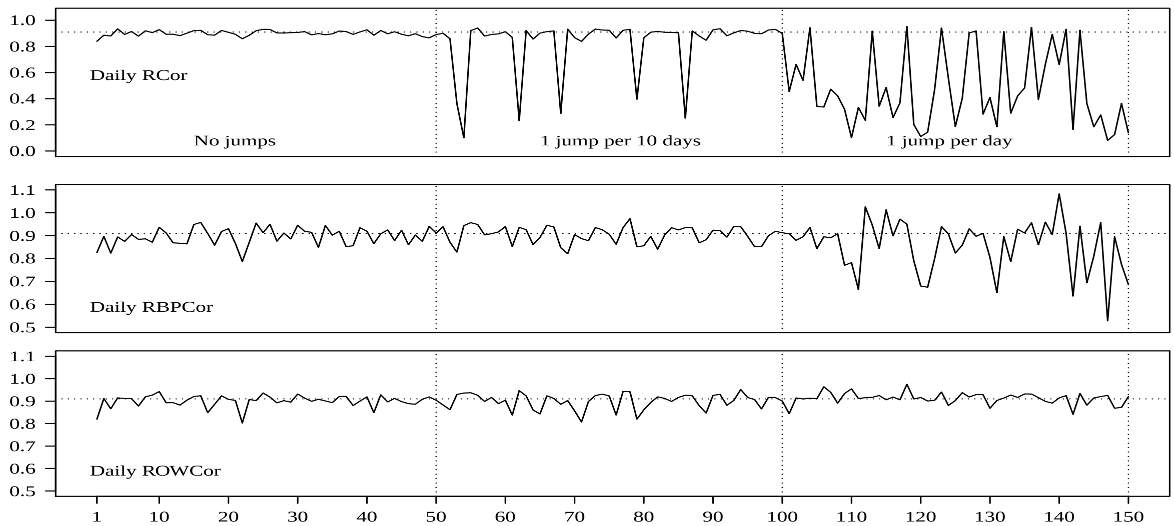



component $p^{(1)}(s)$.[8] In Figure 2 we report the percentage estimation error when using the univariate RCov, RBPCov and ROWCov as estimators for the integrated variance of the first price series. In the absence of jumps, the RCov is the most precise estimator for the IVar. But on days where jumps occur, its bias becomes large and is approximately equal to the sum of squared intraday jumps. As we can see in Figure 1, both the RBPCov and ROWCov are much less affected by the jumps. Note that the bias and variance of the ROWCov in the presence of jumps are much smaller than for the RBPCov. The ROWCov thus seems to be a more robust to jumps estimator of the integrated variance than the RBPCov.

Call RCor, RBPCor and ROWCor the daily realized correlation estimates based on the RCov, RBPCov and ROWCov. These series are plotted in Figure 3. We

[8] Recall that the daily volatility measures and its estimators are based on the log-prices expressed in percentage points. The daily volatility of the 5-minute price series in the first panel of Figure 1 is thus of the order of the daily volatility measures in the second and third panel divided by 10,000.

Figure 3: Daily realized correlation, bipower correlation and outlyingness weighted correlation. The integrated correlation is equal to 0.91. The parameter $\kappa^*$ is 0 for days 1-50, 1 per 10 days for days 51-100 and 1 per day for days 101-150.




see that in the absence of jumps, the RCor is a very accurate correlation estimator. However, in the presence of jumps, the RCor is strongly biased. Recall that we simulated independent jump occurrences, causing the corresponding RCor to be biased towards zero. Simultaneous jumps (caused cojumps) with the same sign would lead to an overestimation of the correlation. In the presence of jumps, the RBPCor and ROWCor remain centered around the true value of 0.91, but the ROWCor is more stable than the RBPCor and is therefore a more precise correlation estimate. Finally, note that the RBPCor has values above unity for observations 112, 115 and 140, while the RCor and ROWCor lie by construction always between -1 and 1.

## 5.3 Accuracy

Next we compare the accuracy of the RCov, RBPCov, RTCov and the ROWCov estimators with Hard and Soft rejection Weight function (called HR and SR ROWCov, respectively). The thresholds are set to the 95%, 99% and 99.9% quantiles of the $\chi^2_2$ distribution. The RTCov is implemented as described in Subsection 4.1 with $c$ set to 9 times the daily realized bipower variation. Table 2 reports the root mean squared (relative) error of these estimators, defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N_* T} \sum_{t=1}^{T} \left\| \text{vech}\left(\widehat{\text{ICov}}_{t,\Delta} - \text{ICov}_t\right) / \text{vech}\left(\text{ICov}_t\right) \right\|^2}, \tag{5.2}$$

where vech$(A)$ is the vector in which the $N_*$ lower-diagonal elements of $A$ are stacked one under the other, / stands for element by element division of two vectors and $\|\cdot\|$ is the Euclidian norm operator. The number of simulated days is $T = 5000$.

Let us first of all consider the case without jumps. From Table 2, we see that the SR ROWCov is always more efficient than the HR ROWCov with the same threshold. The HR ROWCov with 95% threshold is less efficient than the RBPCov, but all other versions of the ROWCov considered have a lower RMSE than the RBPCov. The RMSE of the RTCov is higher than the SR ROWCov estimators and similar to the RMSE of the HR ROWCov estimators with a high threshold. When using 5 or 15-minute returns, the RMSE is approximately $\sqrt{5}$ and $\sqrt{15}$ times the RMSE of the estimators based on 1-minute returns.

In the presence of jumps, the RMSE of the RBPCov increases sharply. This increase is higher for cojumps than for independent jumps. The adverse effect of jumps on the RMSE of the RBPCov becomes more important for larger values of the number of jumps per day and the jump magnitude. In contrast, the RMSE of the RTCov and ROWCov estimators computed on 1-minute return data is little affected

Table 2: RMSE of Realized Covariance (RCov), BiPower Covariation (RBPCov), Threshold Covariance (RTCov), Outlyingness Weighted Covariation (ROWCov) with hard and soft rejection weight functions and thresholds set to the $\beta$-quantile of the chi-square distribution.

| | | Independent jump occurrences | | | | All jumps are cojumps | | | |
|---|---|---|---|---|---|---|---|---|---|
| Jumps per day $\kappa^*$: | 0 | 1 | 1 | 5 | 5 | 1 | 1 | 5 | 5 |
| Magnitude of jumps $m$: | | 0.5 | 1 | 0.5 | 1 | 0.5 | 1 | 0.5 | 1 |
| 1-minute returns ($\Delta = 1/390$) | | | | | | | | | |
| RCov | 0.076 | 0.127 | 0.187 | 0.133 | 0.200 | 1.878 | 3.711 | 1.462 | 2.898 |
| RBPCov | 0.086 | 0.160 | 0.218 | 0.227 | 0.325 | 0.266 | 0.376 | 0.377 | 0.579 |
| RTCov | 0.092 | 0.088 | 0.087 | 0.089 | 0.086 | 0.087 | 0.085 | 0.084 | 0.081 |
| HR, $\beta = 0.95$ | 0.103 | 0.104 | 0.104 | 0.104 | 0.102 | 0.105 | 0.104 | 0.103 | 0.104 |
| HR, $\beta = 0.99$ | 0.086 | 0.085 | 0.086 | 0.085 | 0.085 | 0.086 | 0.085 | 0.086 | 0.086 |
| HR, $\beta = 0.999$ | 0.078 | 0.078 | 0.079 | 0.078 | 0.078 | 0.078 | 0.077 | 0.078 | 0.079 |
| SR, $\beta = 0.95$ | 0.079 | 0.078 | 0.079 | 0.079 | 0.079 | 0.081 | 0.080 | 0.105 | 0.108 |
| SR, $\beta = 0.99$ | 0.076 | 0.076 | 0.077 | 0.077 | 0.077 | 0.081 | 0.081 | 0.128 | 0.132 |
| SR, $\beta = 0.999$ | 0.076 | 0.076 | 0.076 | 0.076 | 0.077 | 0.086 | 0.087 | 0.171 | 0.176 |
| 5-minute returns ($\Delta = 1/78$) | | | | | | | | | |
| RCov | 0.170 | 0.292 | 0.431 | 0.303 | 0.444 | 1.956 | 3.822 | 1.587 | 3.122 |
| RBPCov | 0.195 | 0.328 | 0.440 | 0.374 | 0.552 | 0.584 | 0.845 | 0.818 | 1.320 |
| RTCov | 0.187 | 0.179 | 0.181 | 0.271 | 0.256 | 0.180 | 0.179 | 0.309 | 0.295 |
| HR, $\beta = 0.95$ | 0.227 | 0.221 | 0.228 | 0.238 | 0.235 | 0.227 | 0.225 | 0.237 | 0.235 |
| HR, $\beta = 0.99$ | 0.191 | 0.188 | 0.192 | 0.197 | 0.197 | 0.192 | 0.192 | 0.220 | 0.201 |
| HR, $\beta = 0.999$ | 0.175 | 0.174 | 0.177 | 0.188 | 0.187 | 0.177 | 0.177 | 0.287 | 0.214 |
| SR, $\beta = 0.95$ | 0.176 | 0.174 | 0.178 | 0.191 | 0.193 | 0.210 | 0.210 | 0.431 | 0.470 |
| SR, $\beta = 0.99$ | 0.171 | 0.170 | 0.174 | 0.190 | 0.193 | 0.233 | 0.234 | 0.570 | 0.637 |
| SR, $\beta = 0.999$ | 0.170 | 0.169 | 0.173 | 0.197 | 0.202 | 0.285 | 0.287 | 0.769 | 0.897 |
| 15-minute returns ($\Delta = 1/26$) | | | | | | | | | |
| RCov | 0.290 | 0.514 | 0.741 | 0.518 | 0.770 | 2.123 | 4.006 | 1.889 | 3.658 |
| RBPCov | 0.326 | 0.510 | 0.706 | 0.540 | 0.768 | 0.985 | 1.482 | 1.321 | 2.265 |
| RTCov | 0.311 | 0.333 | 0.326 | 0.487 | 0.617 | 0.489 | 0.509 | 1.202 | 1.981 |
| HR, $\beta = 0.95$ | 0.378 | 0.380 | 0.389 | 0.543 | 0.634 | 0.390 | 0.380 | 0.868 | 1.037 |
| HR, $\beta = 0.99$ | 0.326 | 0.328 | 0.337 | 0.474 | 0.588 | 0.345 | 0.327 | 0.946 | 1.260 |
| HR, $\beta = 0.999$ | 0.302 | 0.310 | 0.319 | 0.460 | 0.597 | 0.381 | 0.316 | 1.148 | 1.659 |
| SR, $\beta = 0.95$ | 0.300 | 0.311 | 0.322 | 0.444 | 0.567 | 0.499 | 0.508 | 1.158 | 1.765 |
| SR, $\beta = 0.99$ | 0.291 | 0.307 | 0.319 | 0.447 | 0.587 | 0.618 | 0.642 | 1.323 | 2.117 |
| SR, $\beta = 0.999$ | 0.290 | 0.314 | 0.327 | 0.464 | 0.626 | 0.808 | 0.869 | 1.497 | 2.494 |

by the presence of jumps. If these estimators are computed using 5 or 15-minute return series, their RMSE increases significantly but much less than the RBPCov. Compared to the HR ROWCov, the RTCov and SR ROWCov have a higher RMSE in the case of 5 jumps per day. Overall, Table 2 seems to indicate that the HR ROWCov with threshold set to the 99.9% quantile of the chi-square distribution with $N = 2$ degrees of freedom strikes a good balance between efficiency in the absence of jumps and presence of jumps. In the remainder of the paper, we use this version of the ROWCov.

## 5.4 Bias under microstructure noise and non-synchronous trading

Up to now, we have focussed on the accuracy of the estimators under the assumption that prices are synchronous and reflect directly the efficient price. In reality, prices are observed with some measurement error, often called microstructure noise, and transactions are not synchronized across assets. It is well known that microstructure noise leads to an upward bias in the variance elements of the RCov, while due to non-synchronous trading the off-diagonal elements in the RCov are biased towards zero. In Table 3 we investigate the sensitivity of the ROWCov to these effects.

Consider first columns 1-5 of Table 3, where there is no microstructure noise and trading is synchronous across assets. In column 1, there are no jumps. Note that, compared to the RCov, RBPCov and ROWCov, the RTCov variance estimates have a small downward bias, because the RTCov has no correction factor for the fact that some of the returns that are not affected by jumps are also downweighted. In this table, all jumps are cojumps having the same sign across the assets. Hence, they cause an upward bias both in the variance and covariance RCov estimates. Like in Corsi et al. (2009), we find that jumps induce a relatively large finite sample bias in the RBPCov variance estimates. In the presence of cojumps with the same sign, a similar bias is found for the RBPCov covariance estimates. The bias of the RTCov and ROWCov in the presence of jumps is relatively similar. It is negligible when using 1-minute returns. At lower sampling frequencies, the returns from the diffusion are more difficult to distinguish from the returns that have been affected by jumps, especially for small jumps (Aït-Sahalia, 2002). Since the jump size is inversely related to the number of jumps per day, this explains the relatively large bias of all estimators in the case of 5 jumps per day, compared to the situation of 1 jump per day, and the bias when using 15-minute returns, relatively to the use of 5-minute returns.

In columns 6-10, the observed price series is contaminated with i.i.d. normal microstructure noise. More precisely, the observed prices are no longer $p_t^{(1)}$ and $p_t^{(2)}$, but $p_t^{(1)} + \varepsilon_t^{(1)}$ and $p_t^{(2)} + \varepsilon_t^{(2)}$, with $\varepsilon_t^{(1)} \overset{i.i.d.}{\sim} N(0, \sigma_\varepsilon^2)$, $\varepsilon_t^{(2)} \overset{i.i.d.}{\sim} N(0, \sigma_\varepsilon^2)$, and $\varepsilon_t^{(1)} \perp \varepsilon_t^{(2)}$. The noise variance is calibrated to 0.001 times the daily integrated variance. This value is of the same order as the noise ratios reported in Table 1 of Christensen et al. (2010) for S&P 500 stocks in 2006. We further assume that $\varepsilon_t^{(1)}$ and $\varepsilon_t^{(2)}$ are independent, such that the contamination with microstructure noise only generates a bias in the variances. We therefore study its impact only on the bias of a variance element of the estimators.

Consider first the case without jumps. We see that the bias caused by microstructure noise is similar for all estimators. When computed using 1-minute returns, the bias is extremely large. It is modest-sized when 5-minute returns are used and negligible when the estimators are constructed from the 15-minute return series. For the ROWCov, this is not surprising since, if market microstructure noise is additive and normally distributed, the outlyingness statistic is still chi-square distributed with $N$ degrees of freedom. In the presence of noise, the ROWCov thus estimates the ICov

Table 3: Bias of a variance (upper panel) and covariance (lower panel) element in the Realized Covariance (RCov), BiPower Covariation (RBPCov), Threshold Covariance (RTCov) and Outlyingness Weighted Covariation (ROWCov). All jumps are cojumps.

| Jumps per day $\kappa^*$: | 0 | 1 | 1 | 5 | 5 | 0 | 1 | 1 | 5 | 5 |
|---|---|---|---|---|---|---|---|---|---|---|
| Magnitude of jumps $m$: | | 0.5 | 1 | 0.5 | 1 | | 0.5 | 1 | 0.5 | 1 |
| **Bias variance** | No microstructure noise | | | | | Microstructure noise | | | | |
| 1-minute returns ($\Delta = 1/390$) | | | | | | | | | | |
| RCov | 0 | 1.264 | 2.479 | 1.300 | 2.550 | 0.780 | 2.057 | 3.325 | 2.047 | 3.336 |
| RBPCov | -0.002 | 0.137 | 0.193 | 0.300 | 0.449 | 0.818 | 1.000 | 1.077 | 1.195 | 1.387 |
| RTCov | -0.032 | -0.024 | -0.023 | -0.024 | -0.020 | 0.733 | 0.740 | 0.744 | 0.737 | 0.740 |
| ROWCov | -0.006 | -0.005 | -0.005 | -0.004 | -0.002 | 0.776 | 0.775 | 0.777 | 0.775 | 0.777 |
| 5-minute returns ($\Delta = 1/78$) | | | | | | | | | | |
| RCov | -0.002 | 1.285 | 2.514 | 1.362 | 2.677 | 0.157 | 1.431 | 2.705 | 1.417 | 2.719 |
| RBPCov | -0.015 | 0.288 | 0.421 | 0.631 | 1.014 | 0.144 | 0.462 | 0.627 | 0.784 | 1.180 |
| RTCov | -0.035 | -0.029 | -0.028 | 0.176 | 0.119 | 0.120 | 0.117 | 0.122 | 0.371 | 0.295 |
| ROWCov | -0.014 | -0.009 | -0.009 | 0.097 | 0.022 | 0.144 | 0.143 | 0.144 | 0.242 | 0.170 |
| 15-minute returns ($\Delta = 1/26$) | | | | | | | | | | |
| RCov | -0.006 | 1.320 | 2.559 | 1.522 | 2.993 | 0.047 | 1.325 | 2.629 | 1.328 | 2.604 |
| RBPCov | -0.048 | 0.466 | 0.721 | 0.962 | 1.674 | 0.001 | 0.502 | 0.790 | 0.909 | 1.575 |
| RTCov | -0.043 | 0.082 | 0.042 | 0.906 | 1.485 | 0.008 | 0.110 | 0.100 | 0.891 | 1.504 |
| ROWCov | -0.033 | 0.026 | -0.013 | 0.748 | 0.964 | 0.020 | 0.049 | 0.038 | 0.482 | 0.670 |
| **Bias covariance** | Synchronous transaction times | | | | | Non-synchronous transaction times | | | | |
| 1-minute returns ($\Delta = 1/390$) | | | | | | | | | | |
| RCov | 0 | 1.270 | 2.492 | 1.293 | 2.558 | -0.083 | 1.076 | 2.204 | 1.100 | 2.262 |
| RBPCov | -0.002 | 0.142 | 0.200 | 0.311 | 0.467 | -0.091 | 0.039 | 0.102 | 0.198 | 0.339 |
| RTCov | -0.046 | -0.035 | -0.032 | -0.032 | -0.023 | -0.134 | -0.122 | -0.117 | -0.116 | -0.111 |
| ROWCov | -0.006 | -0.005 | -0.005 | -0.005 | -0.002 | -0.089 | -0.088 | -0.087 | -0.087 | -0.088 |
| 5-minute returns ($\Delta = 1/78$) | | | | | | | | | | |
| RCov | -0.001 | 1.293 | 2.526 | 1.360 | 2.691 | -0.014 | 1.233 | 2.459 | 1.307 | 2.624 |
| RBPCov | -0.013 | 0.301 | 0.440 | 0.654 | 1.056 | -0.026 | 0.270 | 0.441 | 0.624 | 1.004 |
| RTCov | -0.049 | -0.039 | -0.036 | 0.083 | 0.016 | -0.063 | -0.056 | -0.051 | 0.068 | -0.013 |
| ROWCov | -0.013 | -0.010 | -0.010 | 0.108 | 0.025 | -0.026 | -0.027 | -0.024 | 0.098 | 0.003 |
| 15-minute returns ($\Delta = 1/26$) | | | | | | | | | | |
| RCov | -0.008 | 1.331 | 2.572 | 1.524 | 3.016 | 0 | 1.290 | 2.554 | 1.508 | 2.989 |
| RBPCov | -0.050 | 0.490 | 0.752 | 0.991 | 1.743 | -0.038 | 0.461 | 0.738 | 0.996 | 1.751 |
| RTCov | -0.061 | 0.013 | -0.017 | 0.814 | 1.325 | -0.055 | -0.007 | -0.021 | 0.804 | 1.309 |
| ROWCov | -0.034 | 0.031 | -0.012 | 0.791 | 1.031 | -0.025 | 0.014 | -0.016 | 0.800 | 1.024 |

plus the cumulated noise covariance. In the presence of jumps, the bias of the estimators is approximately equal to the sum of the bias of the estimator in the case of no microstructure noise, but jumps and the bias in the case of microstructure noise, and no jumps. One exception is the ROWCov computed using 15-minute returns in the case of 5 jumps a day. It seems that in this case the additional variability of the microstructure noise increases the power to detect the returns affected by relatively big jumps and hence reduces the bias of the estimator.

In the second panel of columns 6-10, we focus on the effect of synchronous versus non-synchronous transaction times on the bias in the covariance estimates. Non-synchronous transaction times are generated by using independent Poisson sampling schemes such that the inter-transaction times are exponentially distributed with on average one transaction every 5 seconds. We align the price series to a regular grid, using the previous tick approach.

Non-synchronous observations are known to lead to a downward bias in covariance and correlation estimates based on high frequency data. This bias is known as the Epps effect (Epps 1979). We see in Table 3 that at the 1-minute frequency, non-synchronicity leads to a large bias in the estimates, but has only a small effect when the covariances are computed using 5 or 15-minute returns. The bias is similar for the RCov, RBPCov and ROWCov, and slightly higher for the RTCov. Because the cojumps are no longer always realized at the same time across assets, non-synchronicity reduces the bias due to jumps in the RCov and RBPCov. It has little impact on the jump robustness of the RTCov and ROWCov.

The bottom line of this additional simulation study is that in the presence of a modest sized microstructure noise, the ROWCov computed using 5-minute returns is preferable over the ROWCov using 1 or 15-minute returns. It strikes the best balance between accuracy under the diffusion model (Table 2) and a small finite sample bias

Table 4: Empirical normal-based coverage probabilities for the variance, log variance, covariance, beta, correlation and Fisher transformed correlation statistics computed from the ROWCov using 15, 5 and 1-minute returns, over 5000 simulations.

| Asymptotic coverage probability | 95% | | | 99% | | |
|---|---|---|---|---|---|---|
| | 15-min | 5-min | 1-min | 15-min | 5-min | 1-min |
| variance | 0.878 | 0.918 | 0.944 | 0.927 | 0.963 | 0.983 |
| log-variance | 0.920 | 0.934 | 0.946 | 0.972 | 0.983 | 0.987 |
| covariance | 0.880 | 0.922 | 0.944 | 0.931 | 0.967 | 0.983 |
| beta | 0.915 | 0.937 | 0.937 | 0.970 | 0.985 | 0.986 |
| correlation | 0.900 | 0.937 | 0.951 | 0.945 | 0.975 | 0.988 |
| Fisher transformed correlation | 0.928 | 0.950 | 0.954 | 0.978 | 0.987 | 0.992 |

in presence of jumps, microstructure noise and non-synchronous trading (Table 3).

## 5.5 Normality

The aim of this subsection is to assess the accuracy of the normal approximation for the distribution of the variance, log-variance, covariance, beta, correlation and Fisher transformed correlation estimates based on the ROWCov computed on 1, 5 and 15-minute returns. We simulate from the process described in Subsection 5.1. As before, we take a local window of one day. It is then natural to estimate the asymptotic covariance matrix in (4.2) by $(\text{ROWCov}_\Delta^{1/2} \otimes \text{ROWCov}_\Delta^{1/2})\Theta(\text{ROWCov}_\Delta^{1/2} \otimes \text{ROWCov}_\Delta^{1/2})'$. An explicit expression for the matrix $\Theta$ is given in the webappendix. The log transformation of the variance estimate and the Fisher transformation of the correlation estimate are considered because Barndorff-Nielsen and Shephard (2004a) showed that, for the RCov, they clearly improve the accuracy of the asymptotic approximation.

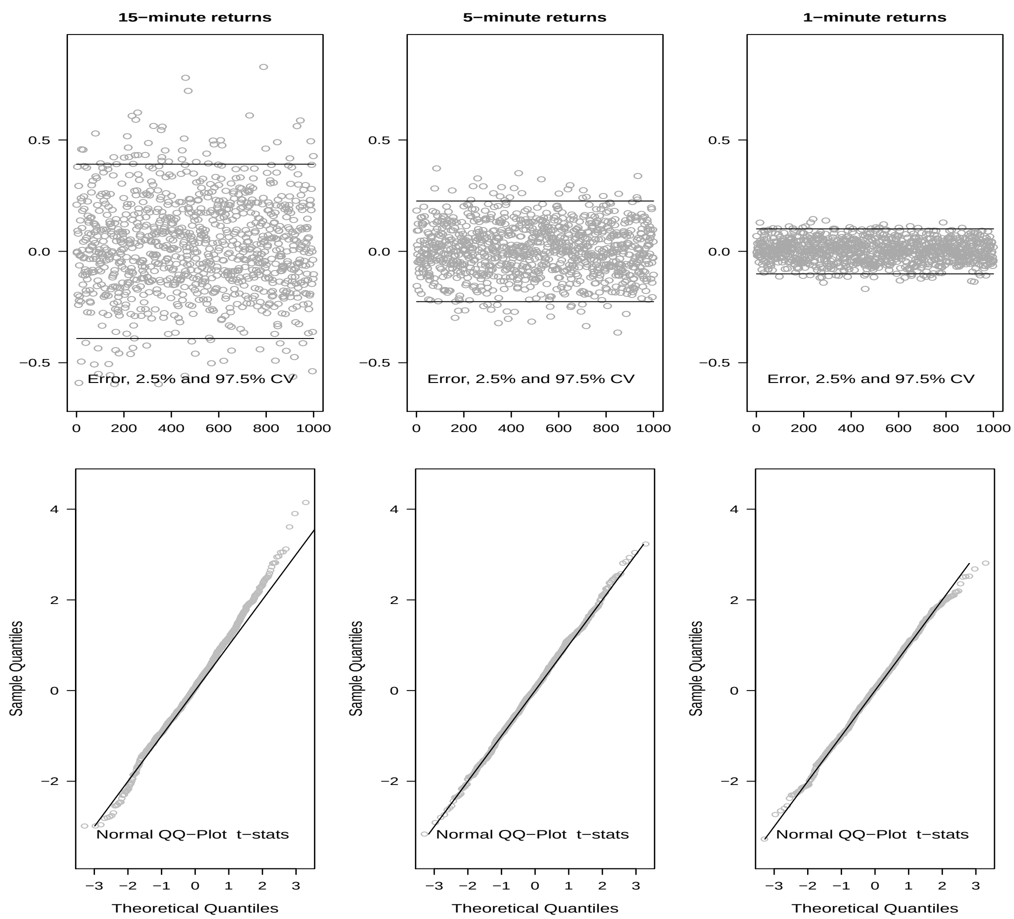
Table 4 reports the normal-based empirical coverage probability of the variance, log-variance, covariance, beta, correlation and Fisher correlation estimates using the ROWCov with 99.9% threshold and computed on equispaced 15, 5 and 1-minute returns over 5000 simulated days.[9] The asymptotic coverage probability is 95% and 99%. Note that for the ROWCov computed using 1-minute returns, the empirical coverage probability is close to the asymptotic one. For all statistics, the coverage is rather poor for the ROWCov using 15-minute returns. The deterioration of the normality approximation when decreasing the sampling frequency used to compute the ROWCov is not surprising. Similar findings for the RCov have been documented in Barndorff-Nielsen and Shephard (2004a, 2005), Dovonon et al. (2011), Gonçalves and Meddahi (2009), Huang and Tauchen (2006), Veraart (2010) and Zhang et al. (2011), among others. The empirical coverages for the log-variance, beta and Fisher transformed correlation computed on the 5-minute returns are very close to their asymptotic values. The log transformation of the variance estimate and the Fisher transformation of the correlation estimate improve thus also for the ROWCov the accuracy of the asymptotic approximation.

The accuracy of the normal approximation can be directly evaluated through the QQ-plots of the t-scores associated to the statistics. To save space, the normal QQ-plots of all statistics are reported in the webappendix and we limit us in Figure 4 to the analysis of the Fisher transformed correlation for the first one thousand simulations. The upper part plots the series of estimation errors, together with the estimated 2.5% and 97.5% critical values based on the normal approximation. As we move from the left hand side across the page, we increase the sampling frequency and we can see the decrease in the spread of these errors. The lower panel of Figure 4 reports the normal QQ-plots for the t-score of the Fisher transformed correlation. The normality approximation is rather poor for the ROWCov using 15-minute returns, but it improves considerably when sampling at higher frequencies.

We may thus conclude from this simulation study that normality is a good ap-

[9] Explicit expressions for the estimators of these statistics and of their asymptotic standard errors are given in the webappendix.

Figure 4: Difference between estimated and true Fisher transformed correlation (upper panel, with estimated asymptotic 2.5% and 97.5% critical values as solid lines) and normal QQ-plots of standardized estimation error (lower panel) for the ROWCov estimator using 15, 5 and 1-minute returns over 1000 simulated days.




proximation for the distribution of the ROWCov computed on 1 and 5-minute returns.

# 6 Empirical application

## 6.1 Data

The data consists of the intraday transaction prices from the Trade and Quotes database (TAQ) of the New York Stock Exchange for the 30 Dow Jones Industrial Average constituents in January 2008. The sample ranges from January 2, 2008 to May 29, 2009. We excluded the AIG and GM stocks because for over 25% of the days, there are more than 30% of zero returns.[10] The ROWCov is computed with the hard rejection weight function and threshold set to the 99.9% quantile of the chi-square distribution with $N$ degrees of freedom. Accurate estimation of the outlyingness of the returns requires to account for the intraday pattern in volatility. We follow Boudt et al. (2011) by computing the outlyingness statistic on the returns divided by their jump robust weighted standard deviation periodicity estimate.

## 6.2 An illustrative look at AXP and JPM

In the midst of the credit crisis, the House of Representatives rejected on September 29 2008 USD 700 billion rescue of the financial industry. That day, many stock prices cojumped causing a bias in the realized correlation estimate. Figures 5 and 6 illustrate this for the stock prices of American Express (AXP) and JPMorgan Chase & Co (JPM). Figure 5 plots the correlation estimates based on the RCov and the ROWCov for the period August 4 - October 30, 2008, together with the 95% confidence bands. The confidence bands are based on the normal approximation of the Fisher-transformed correlation coefficients. The ROWCov and RCov correlation

[10]The tickers of the stocks in the sample are: AA, AXP, BA, C, CAT, DD, DIS, GE, HD, HON, HPQ, IBM, INTC, JNJ, JPM, KO, MCD, MMM, MO, MRK, MSFT, PFE, PG, T, UTX, VZ, WMT, XOM. We deleted half trading days. Over these 2.5 years there are 352 trading days. Prior to the analysis, we apply the filters presented in Barndorff-Nielsen et al. (2009) to remove the obvious outliers from the data.

between AXP and JPM are both equal to 36% on September 26. But on September 29, the RCov correlation jumps to 55%, while it remains at 34% for the ROWCov. From the intraday price series of the stocks in Figure 6, it is clear that this difference can be directly attributed to rumors about the outcome of the vote, causing a cojump of the two price series in the 13:40-13:45 EST time interval.

### 6.3 Positive semidefiniteness RBPCov

A major advantage of the ROWCov and RTCov with respect to the RBPCov is that by construction they are positive semidefinite. The RBPCov computed on 5-minute return data is usually positive semidefinite in small dimensions, but for moderate and high dimensions, it is not. We illustrate this in Figure 7. For each day, we compute the smallest eigenvalue of the RBPCov of 500 randomly chosen combinations of $N$ DJIA constituents, with $N = 2, \ldots, 28$. We report the proportion of RBPCov estimates with negative smallest eigenvalue. This proportion is 0 for $N = 2$ but increases fast with the dimension $N$ of the series. For $N = 10, 15$ and 20, this proportion is approximately 6%, 50% and 95%, respectively. Another advantage of the ROWCov is that, unlike for the RBPCov, the ROWCov based correlations are always between -1 and 1. Interestingly, there is one day in our sample where the correlation estimate based on the RBPCov is larger than one in absolute value: on March 25, 2008 the RBPCor between the Disney and Home Depot stock equals 1.161.

### 6.4 Application to BEKK models

We conclude this application by analyzing the effect of adding the lagged RCov, RBPCov or ROWCov as an explanatory variable to the scalar BEKK model of Engle and Kroner (1995) for the conditional covariance matrix $H_t$ of the close-

Figure 5: Daily AXP and JPM correlation estimates implied by the RCov and ROWCov computed on 5-minute returns for the period August 4 - October 30, 2008, together with the 95% confidence bands. The vertical dashed line corresponds to September 29 2008.

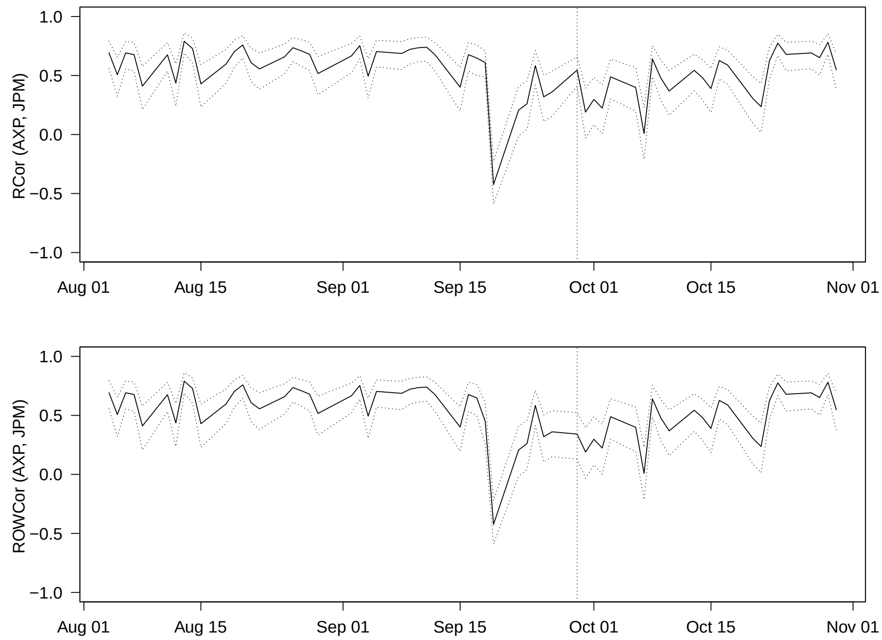



Figure 6: Intraday plot of 5-minute AXP and JPM stock price series on September 29 2008.

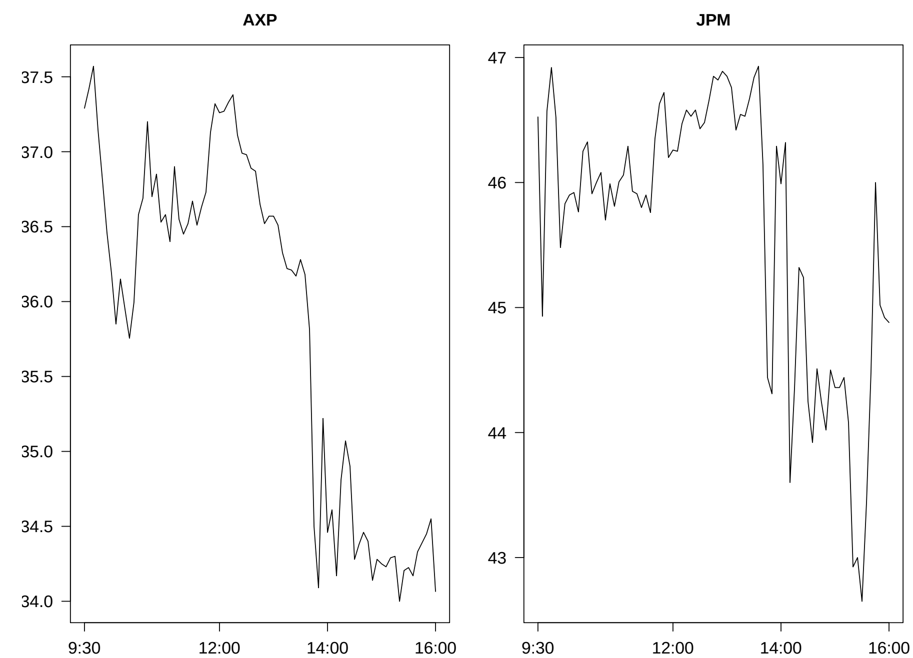
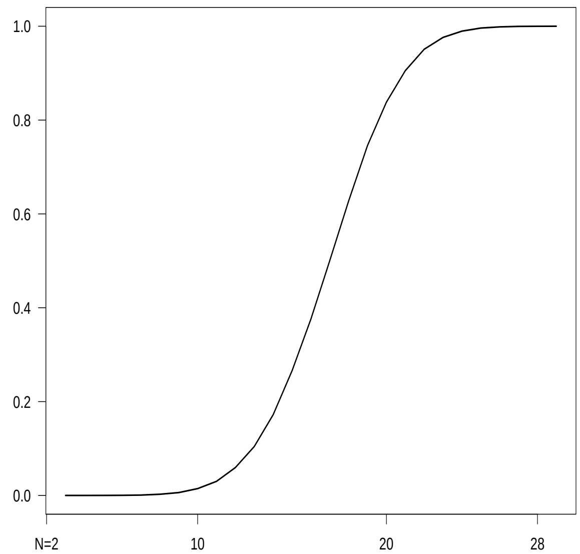


Figure 7: Proportion of daily RBPCov estimates with negative smallest eigenvalue as a function of the dimension $N$ of the series. The RBPCov estimates are computed on 500 arbitrarily chosen combinations of $N$ DJIA constituents for each day in the sample.




to-close return $r_t$ on the selected 28 DJIA constituents on day $t$. Let $H$ be the sample covariance matrix of $r_t$ and $\eta_t$ the overnight return between the closing of the markets on day $t-1$ and the opening on day $t$. Denote by $K^1_{t-1}$, $K^2_{t-1}$ and $K^3_{t-1}$, the RCov, RBPCov and ROWCov ex post covariance estimator for day $t-1$, respectively. Similarly as in Barndorff-Nielsen et al. (2010) and Fleming et al. (2003), we consider the following scalar BEKK model with targeting:

$$H_t = (1 - \alpha - \beta - \sum_j \gamma_j)H + \beta H_{t-1} + \alpha r_{t-1}r'_{t-1} + \sum_j \gamma_j (K^j_{t-1} + \eta_{t-1}\eta'_{t-1}), \quad (6.1)$$

where $\alpha, \beta, \gamma_j \geq 0$ and , $\alpha + \beta + \sum_j \gamma_j \leq 1$. Inference is done with the standard Gaussian quasi-likelihood $-\frac{1}{2}\sum_{t=101}^{T}(\log |H_t| + r'_t H_t^{-1} r_t)$, with a burn-in period of 100 days and $T = 352$. The resulting estimates are reported in Table 5. Asymptotic

Table 5: Scalar BEKK models for the conditional covariance matrix of close-to-close returns on 28 DJIA constituents over the period January 2, 2008 to May 29, 2009.

| | M1 | M2 | M3 | M4 | M5 | M6 | M7 | M8 | M9 | M10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $H_{t-1}$ | 0.701 | 0.539 | 0.716 | 0.667 | 0.575 | 0.727 | 0.669 | 0.608 | 0.594 | 0.603 |
| | (2.9e-4) | (3.0e-4) | (1.5e-4) | (2.2e-4) | (3.4e-4) | (1.7e-4) | (2.3e-4) | (3.5e-4) | (3.0e-4) | (3.1e-4) |
| $r_{t-1}r'_{t-1}$ | 0.014 | | | | 0.021 | 0.017 | 0.011 | 0.021 | 0.013 | 0.013 |
| | (1.2e-5) | | | | (3.2e-5) | (2.2e-5) | (1.4e-5) | (2.9e-5) | (1.7e-5) | (1.8e-5) |
| $\text{RCov}_{t-1}$ | | 0.195 | | | 0.173 | | | 0.121 | 0.084 | 0.069 |
| $+\eta_{t-1}\eta'_{t-1}$ | | (1.8e-4) | | | (1.6e-4) | | | (1.9e-4) | (9.5e-5) | (1.0e-4) |
| $\text{RBPCov}_{t-1}$ | | | 0.120 | | | 0.133 | | 0.045 | | 0.015 |
| $+\eta_{t-1}\eta'_{t-1}$ | | | (9.7e-5) | | | (8.4e-5) | | (1.1e-4) | | (4.7e-5) |
| $\text{ROWCov}_{t-1}$ | | | | 0.168 | | | 0.161 | | 0.133 | 0.128 |
| $+\eta_{t-1}\eta'_{t-1}$ | | | | (1.0e-4) | | | (9.6e-5) | | (1.1e-4) | (1.1e-4) |
| $\log L$ | -8626 | -8362 | -8379 | -8358 | -8352 | -8366 | -8349 | -8350 | -8319 | -8319 |

standard errors are in parenthesis.

Model M1 is the standard BEKK model using only daily returns. Model M10 is the full model, including all variables. Comparing model M1 with all other models, we see that adding the realized covariation estimators increases the log-likelihood from -8626 to at least -8379. Models that contain both the lagged returns and a realized covariance estimator have a significantly higher likelihood than models with either the realized covariance estimate or the outer product of lagged returns. A relatively smaller (but significant) increase in the log-likelihood is achieved by using the realized covariance estimate and the outer product of lagged daily returns. The inclusion of the RBPCov in model M9, that has the outer product of lagged daily returns, the RCov and the ROWCov, does not yield a statistically significant increase in the log-likelihood (the likelihood ratio statistic has a p-value of 38%). Note that the usage of the ROWCov (rather than the RBPCov) in scalar BEKK models has also the advantage that the resulting conditional covariance estimate is affine equivariant and guaranteed to be positive semidefinite.

# 7 Conclusion

We study the problem of disentangling the continuous and jump components in the realized covariation of multivariate log-price processes. We show that the inclusion in the realized covariation of a weight function that gives a zero weight to local outliers leads to an estimator that is consistent for the integrated covariance matrix of a Brownian semimartingale with finite activity jump process. This Realized Outlyingness Weighted Covariation (ROWCov) estimator has the advantage that, unlike the realized bipower covariation, it is little affected by jumps affecting two contiguous returns and it yields positive semidefinite matrices. An extensive simulation study confirms the high efficiency of the ROWCov under the Brownian semimartingale model and its robustness to price jumps. The usefulness of the ROWCov is illustrated in an application to daily covariance estimation using 5-minute stock return series. The ROWCov estimator is implemented in the G@RCH and RTAQ software packages of Laurent (2009) and Cornelissen and Boudt (2010).

In the empirical application we avoided market microstructure problems such as bid-ask bounce and non-synchronous trading by using returns of liquid assets sampled at 5 minutes. A noise-robust estimator for the realized covariance estimator has been recently proposed in Aït-Sahalia and Jacod (2010), Christensen et al. (2010) and Zhang et al. (2011). It is clearly an interesting topic of further research to extend these analysis to the ROWCov and to modify the ROWCov estimator such that it is also robust to microstructure noise, non-synchronous trading and infinite activity jumps. Another interesting direction for future research is to create a data-driven method for optimally selecting the length of the local window, as in e.g. Fan and Gijbels (1995) and Mercurio and Spokoiny (2004). Finally, we believe that the decomposition of the realized covariance in its continuous and jump components using the ROWCov can be successfully applied to a variety of applications such as the measurement and prediction of portfolio volatility and testing for jumps and cojumps.

# References

Aït-Sahalia, Y. (2002). Telling from discrete data whether the underlying continous-time model is a diffusion. *Journal of Finance 53*, 2075–2112.

Aït-Sahalia, Y. and J. Jacod (2004). Disentangling diffusion from jumps. *Journal of Financial Economics 74*, 487–528.

Aït-Sahalia, Y. and J. Jacod (2009). Testing for jumps in a discretely observed process. *Annals of Statistics 37*, 184–222.

Aït-Sahalia, Y. and J. Jacod (2010). High frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association*, forthcoming.

Andersen, T. G., T. Bollerslev, and F. Diebold (2007a). Roughing it up: including jump components in the measurement, modelling and forecasting of return volatility. *Review of Economics and Statistics 89*, 701–720.

Andersen, T. G., T. Bollerslev, F. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association 96*, 42–55.

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

Andersen, T. G., D. Dobrev, and E. Schaumburg (2009). Jump-robust volatility estimation using nearest neighbor truncation. *NBER Working Paper No. 15533*.

Asai, M., M. McAleer, and J. Yu (2006). Multivariate stochastic volatility: a review. *Econometric Reviews 25*, 145–175.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2009). Realized kernels in practice: Trades and quotes. *Econometrics Journal 12*, C1–C32.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2010). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, forthcoming.

Barndorff-Nielsen, O. E. and N. Shephard (2004a). Econometric analysis of realised covariation: high frequency covariance, regression and correlation in financial economics. *Econometrica 72*, 885–925.

Barndorff-Nielsen, O. E. and N. Shephard (2004b). Measuring the impact of jumps in multivariate price processes using bipower covariation. *Discussion paper, Nuffield College, Oxford University*.

Barndorff-Nielsen, O. E. and N. Shephard (2004c). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics 2*, 1–37.

Barndorff-Nielsen, O. E. and N. Shephard (2005). How accurate is the asymptotic approximation to the distribution of realized variance? In D. Andrews and J. Stock (Eds.), *Identification and inference for econometric models. A Festschrift in honour of T.J. Rothenberg*, pp. 306–311. Cambridge University Press, Cambridge, UK.

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*, 796–806.

Bauwens, L., S. Laurent, and J. Rombouts (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics 21*, 79–109.

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

Boudt, K., C. Croux, and S. Laurent (2011). Robust estimation of intraweek periodicity in volatility and jump detection. *Journal of Empirical Finance 18*, 353–367.

Butler, R., P. Davies, and M. Jhun (1993). Asymptotics for the minimum covariance determinant estimator. *Annals of Statistics 21*, 1385–1400.

Cator, E. A. and H. Lopuhaä (2011). Central limit theorem and influence function for the MCD estimators at general multivariate distributions. *Bernouilli*, forthcoming.

Christensen, K., R. Oomen, and M. Podolskij (2010). Pre-averaging estimators of the ex-post covariance matrix. *Journal of Econometrics 159*, 116–133.

Cornelissen, J. and K. Boudt (2010). *RTAQ: Tools for the analysis of trades and quotes in R.* R package version 0.3.

Corsi, F. (2009). A simple approximate long memory model of realized volatility. *Journal of Financial Econometrics 7*, 174–196.

Corsi, F., D. Pirino, and R. Renó (2009). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics 159*, 276–288.

Croux, C. and G. Haesbroeck (1999). Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. *Journal of Multivariate Analysis 71*, 161–190.

Dovonon, P., S. Gonçalves, and N. Meddahi (2011). Bootstrapping realized mutivariate volatility measures. *Journal of Econometrics*, forthcoming.

Engle, R. and F. Kroner (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory 11*, 122–150.

Epps, T. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association 74*, 281–298.

Fan, J. and I. Gijbels (1995). Data-driven bandwith selection in local polynomical fitting: variable bandwith and spatial adaptation. *Journal of the Royal Statistical Society B 57*, 371–364.

Fleming, J., C. Kirby, and B. Ostdieka (2003). The economic value of volatility timing using "realized" volatility. *Journal of Financial Economics 67*, 473–509.

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

Huang, X. and G. Tauchen (2006). The relative contribution of jumps to total price variance. *Journal of Financial Econometrics 3*, 456–499.

Jacod, J. and V. Todorov (2009). Testing for common arrivals of jumps for discretely observed multidimensional processes. *Annals of Statistics 37*, 1792–1838.

Laurent, S. (2009). *G@RCH 6, Estimating and Forecasting ARCH Models.* Timberlake Consultants.

Mancini, C. and F. Gobbi (2009). Identifying the covariation between the diffusion parts and the co-jumps given discrete observations. *Mimeo*.

Mancini, C. and R. Renó (2011). Threshold estimation of Markov models with jumps and interest rate modeling. *Journal of Econometrics 160*, 77–92.

Mercurio, D. and V. Spokoiny (2004). Statistical inference for time-inhomogenous volatility models. *Annals of Statistics 32*, 577–602.

Pison, G., S. Van Aelst, and G. Willems (2002). Small sample corrections for LTS and MCD. *Metrika 55*, 111–123.

Rousseeuw, P. and C. Croux (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association 88*, 1273–1283.

Rousseeuw, P. and K. van Driessen (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics 41*, 212–223.

Veraart, A. E. (2010). How precise is the finite sample approximation of the asymptotic distribution of realised variation measures in the presence of jumps? *CREATES research paper 2010-65*.

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