Appendix to
Outlyingness Weighted Covariation

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

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

This version: February 2011

This report contains technical details and supplementary material for the article "Outlyingness Weighted Covariation". The first section covers the sketches of the proofs of the results that are presented in the main paper. More specifically, we show consistency of the ROWCov estimator under the Brownian semimartingale with finite activity jumps model and asymptotic normality under the Brownian semimartingale model. The second section contains simulation results regarding the sensitivity of the ROWCov estimator to the choice of window length. Furthermore, we validate the normal approximation for the finite sample distribution of the ROWCov estimator. References to equations, tables or figures are either to the main article, or to this webappendix if starting with I or II.

*Correspondence to: Kris Boudt, Faculty of Business and Economics, 69 Naamsestraat, B-3000 Leuven, Belgium. E-mail: Kris.Boudt@econ.kuleuven.be.

I Proofs

I.1 Correction factor

The correction factors in (3.5) and Footnote 6, using the soft rejection function, for $N = 1$ and for k equal to the β quantile of the χ_1^2 distribution, have been obtained as follows:

$$\begin{aligned}
E[w(u^2)] &= 2\int_0^{\sqrt{k}} \phi(u)du + 2k\int_{\sqrt{k}}^{\infty} u^{-2}\phi(u)du = \beta - 2k\int_{\sqrt{k}}^{\infty}\phi(u)d(u^{-1}) \\
&= \beta - 2k\phi(u)u^{-1}|_{\sqrt{k}}^{\infty} + 2k\int_{\sqrt{k}}^{\infty} u^{-1}\phi'(u)du \\
&= \beta + 2\sqrt{k}\phi(\sqrt{k}) - 2k\int_{\sqrt{k}}^{\infty}\phi(u)du \\
&= \beta + 2\sqrt{k}\phi(\sqrt{k}) - k(1-\beta) \\
E[w(u^2)u^2] &= 2\int_0^{\sqrt{k}} u^2\phi(u)du + 2k\int_{\sqrt{k}}^{\infty}\phi(u)du \\
&= = -2\int_0^{\sqrt{k}} ud\phi(u) + k(1-\beta) \\
&= -2u\phi(u)|_0^{\sqrt{k}} + 2\int_0^{\sqrt{k}}\phi(u)du \\
&= -2\sqrt{k}\phi(\sqrt{k}) + \beta + k(1-\beta).
\end{aligned}$$

I.2 Consistency under the BSMFAJ model

Here we give a sketch of the proof of consistency of the ROWCov estimator under the BSMFAJ model. The consistency result for the BSM model follows as a special case.

Divide the unit time interval $[0, 1]$, corresponding to one day, into $\lfloor 1/\lambda \rfloor$ contiguous local windows of length λ: $[0, \lambda[$, $[\lambda, 2\lambda[$, ..., $[1-\lambda, \lambda]$. Log-prices are observed at the equispaced times 0, Δ, ..., $\lfloor 1/\Delta \rfloor$. Denote the corresponding continuously compounded high-frequency returns as $r_{l,i,\Delta} = p[(l-1)\lambda + i\Delta] - p[(l-1)\lambda + (i-1)\Delta]$, with $i = 1, \ldots, \lfloor \lambda/\Delta \rfloor$, and with l the index of the window.

Assume that within each local window, we can consider the continuous volatility dynamics to be a locally constant process, i.e. $\Omega(s) = \Omega((l-1)\lambda)$ for $s \in [(l-1)\lambda, l\lambda[$. For λ small enough, this is a reasonable approximation. Let $\Omega_l = \Omega((l-1)\lambda)$ and

$\Sigma_l = \Sigma((l-1)\lambda)$. The standardized returns $u_{l,i,\Delta} = \Omega_l^{-1} r_{l,i,\Delta}\Delta^{-1/2}$ are standard normal distributed, except for a proportion ε_Δ that are affected by jumps. Because of the assumption of finite activity, $\varepsilon_\Delta \to 0$ as $\Delta \to 0$.

Let $\hat{\Sigma}_{l,i,\Delta}$ be the first step affine equivariant covariance estimate of $r_{l,i,\Delta}\Delta^{-1/2}$, computed as detailed in Subsection 3.2. We consider the ROWCov estimator as defined in (3.2) with the weights adjusted by multiplying with the ratio between the expected value of the weights and the sample average weight in the local window where the return belongs to, as discussed in the last paragraph of page 10:

$$\text{ROWCov}_\Delta = d_w \sum_{l=1}^{\lfloor 1/\lambda \rfloor} \left(\frac{\sum_{i=1}^{\lfloor \lambda/\Delta \rfloor} w(r'_{l,i,\Delta}\hat{\Sigma}_{l,i,\Delta}^{-1} r_{l,i,\Delta}\Delta^{-1}) r_{l,i,\Delta} r'_{l,i,\Delta}}{\frac{1}{\lfloor \lambda/\Delta \rfloor}\sum_{i=1}^{\lfloor \lambda/\Delta \rfloor} w(r'_{l,i,\Delta}\hat{\Sigma}_{l,i,\Delta}^{-1} r_{l,i,\Delta}\Delta^{-1})} \right),$$

with $d_w = NE[w(z)]/E[w(z)z]$ and z a chi-square random variable with N degrees of freedom. The results for the ROWCov estimator with the unadjusted weights follow as a special case.

Because of the equivariance of the initial scale estimator, $\hat{S}_{l,i,\Delta} = (\Omega_l^{-1})\hat{\Sigma}_{l,i,\Delta}(\Omega_l^{-1})'$ is the first step covariance estimate of $u_{l,i,\Delta}$. Because of the equivariance of the ROWCov, we can thus equivalently rewrite the ROWCov as:

$$\text{ROWCov}_\Delta = d_w \sum_{l=1}^{\lfloor 1/\lambda \rfloor} \Omega_l \left(\frac{\sum_{i=1}^{\lfloor \lambda/\Delta \rfloor} w(u'_{l,i,\Delta}\hat{S}_{l,i,\Delta}^{-1} u_{l,i,\Delta}) u_{l,i,\Delta} u'_{l,i,\Delta}}{\sum_{i=1}^{\lfloor \lambda/\Delta \rfloor} w(u'_{l,i,\Delta}\hat{S}_{l,i,\Delta}^{-1} u_{l,i,\Delta})} \right) \Omega'_l \lambda, \quad \text{(I.1)}$$

with $d_w = NE[w(z)]/E[w(z)z]$ and z a chi-square random variable with N degrees of freedom.

Let $\iota_{l,i,\Delta}$ be an indicator function that is one when $u_{l,i,\Delta}$ is normally distributed with mean 0 and covariance matrix I and zero otherwise. We can rewrite the ROWCov as:

$$\begin{aligned}\text{ROWCov}_\Delta = d_w \sum_{l=1}^{\lfloor 1/\lambda \rfloor} \Omega_l \left(\frac{\sum_{i=1}^{\lfloor \lambda/\Delta \rfloor} w(u'_{l,i,\Delta}\hat{S}_{l,i,\Delta}^{-1} u_{l,i,\Delta}) u_{l,i,\Delta} u'_{l,i,\Delta}\iota_{l,i,\Delta}}{\sum_{i=1}^{\lfloor \lambda/\Delta \rfloor} w(u'_{l,i,\Delta}\hat{S}_{l,i,\Delta}^{-1} u_{l,i,\Delta})} \right) \Omega'_l \lambda \\ + d_w \sum_{l=1}^{\lfloor 1/\lambda \rfloor} \Omega_l \left(\frac{\sum_{i=1}^{\lfloor \lambda/\Delta \rfloor} w(u'_{l,i,\Delta}\hat{S}_{l,i,\Delta}^{-1} u_{l,i,\Delta}) u_{l,i,\Delta} u'_{l,i,\Delta}(1-\iota_{l,i,\Delta})}{\sum_{i=1}^{\lfloor \lambda/\Delta \rfloor} w(u'_{l,i,\Delta}\hat{S}_{l,i,\Delta}^{-1} u_{l,i,\Delta})} \right) \Omega'_l \lambda.\end{aligned}$$

The first step covariance estimate must be chosen such that, for $\Delta \to 0$, we have that for all $u_{l,i,\Delta}$ for which $\iota_{l,i,\Delta} = 0$ (i.e. all returns affected by jumps),

$w(u'_{l,i,\Delta}\hat{S}^{-1}_{l,i,\Delta}u_{l,i,\Delta}) = 0$ almost surely. This is the case if the MCD covariance estimate described in Subsection 3.2 is used. The reason is that for $\varepsilon_\Delta \to 0$, the probability that an outlying return is included in the subset from which the MCD covariance is computed, is zero, almost surely. It follows from Theorem 3 in Butler et al. (1993) that the MCD covariance of $u_{l,1,\Delta}, \ldots, u_{l,1/\Delta,\Delta}$ is $\sqrt{\lambda/\Delta}$-consistent for I. Then, for Δ tending to zero, returns affected by jumps have distance $u'_{l,i,\Delta}\hat{S}^{-1}_{l,i,\Delta}u_{l,i,\Delta}$ converging to infinity, and will get a weight arbitrarily close to zero. It follows that the ROWCov has asymptotically the same probability limit as the ROWCov computed on the returns that are not affected by jumps:

$$Z_\Delta = d_w \sum_{l=1}^{\lfloor 1/\lambda \rfloor} \Omega_l \left(\frac{\sum_{i=1}^{\lfloor \lambda/\Delta \rfloor} w(u'_{l,i,\Delta}\hat{S}^{-1}_{l,i,\Delta}u_{l,i,\Delta}) u_{l,i,\Delta}u'_{l,i,\Delta}\iota_{l,i,\Delta}}{\sum_{i=1}^{\lfloor \lambda/\Delta \rfloor} w(u'_{l,i,\Delta}\hat{S}^{-1}_{l,i,\Delta}u_{l,i,\Delta})\iota_{l,i,\Delta}} \right) \Omega'_l \lambda.$$

Note that, by the consistency of the first step covariance estimator, the outlyingness statistic of a return that is not affected by jumps is asymptotically χ^2_N distributed. It is a property of the multivariate normal distribution that, by the choice of d_w,

$$\operatorname*{plim}_{\Delta \to 0} Z_\Delta = \lambda \sum_{l=1}^{\lfloor 1/\lambda \rfloor} \Sigma_l.$$

Under smoothness conditions on Ω, this convergence result under the locally constant volatility model implies that the ROWCov is consistent for the ICov under the BSMFAJ model if $\lambda \to 0$, while $\lambda/\Delta \to \infty$, by integrating the spot covariance over the unit interval.

I.3 Asymptotic normality under the BSM model

Take the same notation as in Subsection I.2. Assuming locally constant variance, the ROWCov coincides with the sum of reweighted MCD estimates on local windows, see (I.1).

Let $M_{l,\Delta}$ be the reweighted MCD of the standardized returns $u_{l,i,\Delta}$ on the l-th local window, i.e.

$$M_{l,\Delta} = d_w \frac{\sum_{i=1}^{\lfloor \lambda/\Delta \rfloor} w(u'_{l,i,\Delta}\hat{S}^{-1}_{l,i,\Delta}u_{l,i,\Delta}) u_{l,i,\Delta}u'_{l,i,\Delta}}{\sum_{i=1}^{\lfloor \lambda/\Delta \rfloor} w(u'_{l,i,\Delta}\hat{S}^{-1}_{l,i,\Delta}u_{l,i,\Delta})}.$$

In the absence of jumps, it follows from Theorem 4.1 in Lopuhaä (1999) and Corol-

Table I.1: Asymptotic variance of a diagonal and off-diagonal element of the reweighted MCD for dimensions $N = 1$, 2, 5 and 10.

	$N=1$	$N=2$		$N=5$		$N=10$	
k	θ^{on}	θ^{on}	θ^{off}	θ^{on}	θ^{off}	θ^{on}	θ^{off}
ROWCov with standardized hard rejection weights							
$\chi^2_N(0.95)$	4.137	3.827	2.081	2.870	1.414	2.534	1.249
$\chi^2_N(0.99)$	2.785	2.524	1.256	2.238	1.103	2.139	1.062
$\chi^2_N(0.999)$	2.152	2.087	1.036	2.038	1.015	2.021	1.009
ROWCov with simple hard rejection weights							
$\chi^2_N(0.95)$	4.955	4.310	2.081	3.143	1.414	2.726	1.249
$\chi^2_N(0.99)$	2.996	2.646	1.256	2.304	1.103	2.185	1.062
$\chi^2_N(0.999)$	2.179	2.102	1.036	2.046	1.015	2.027	1.009

lary 4.1 in Cator and Lopuhaä (2011) that, when $\Delta \to 0$,

$$\sqrt{\lambda/\Delta}(M_{l,\Delta} - I) \xrightarrow{d} N\left(0, \Theta\right). \tag{I.2}$$

Since the reweighted MCD is affine equivariant, it follows from Corrolary 13.1 in Bilodeau and Brenner (1999) that there exists constants σ_1 and $\sigma_2 \geq -2\sigma_1/N$ such that the asymptotic covariance of the reweighted MCD is given by

$$\Theta = \sigma_1(I + K_N)(I \otimes I) + \sigma_2 \operatorname{vec}(I)[\operatorname{vec}(I)]',$$

where K_N is the commutation matrix, which is a $N^2 \times N^2$ matrix consisting of $N \times N$ blocks and each (i, j)th block is a $N \times N$ matrix, which is 1 at entry (j, i) and 0 everywhere else. This implies that the covariance between elements (k, i) and (l, j) at the standard normal distribution is given by

$$\sigma_1(I_{ij}I_{kl} + I_{kj}I_{il}) + \sigma_2 I_{ki}I_{lj}.$$

The asymptotic variance of a diagonal and off-diagonal element of the reweighted MCD has been derived by Croux and Haesbroeck (1999) and Cator and Lopuhaä (2011). We tabulate them in Table I.1. Denote these θ^{on} and θ^{off}, respectively. We have $\theta^{\text{on}} = 2\sigma_1 + \sigma_2$ and $\theta^{\text{off}} = \sigma_1$.

Since we assumed no leverage and by (I.2),

$$\begin{aligned}
&\text{ACov}_{\Delta\to 0}\left[\Delta^{-1/2}\,\text{vec}\left(\text{ROWCov}_{\Delta}-\text{ICov}\right)\right]\\
&= \text{ACov}_{\Delta\to 0}\left[\Delta^{-1/2}\sum_{l=1}^{\lfloor 1/\lambda\rfloor}\text{vec}\left(\Omega_l(M_{l,\Delta}-I)\Omega_l'\lambda\right)\right]\\
&= \text{ACov}_{\Delta\to 0}\sqrt{\lambda}\sum_{l=1}^{\lfloor 1/\lambda\rfloor}\sqrt{\lambda/\Delta}\,\text{vec}\left(\Omega_l(M_{l,\Delta}-I)\Omega_l'\right)\\
&= \lambda\sum_{l=1}^{\lfloor 1/\lambda\rfloor}G_l\Theta G_l',
\end{aligned}$$

where $G_l = \Omega_l \otimes \Omega_l$, and ACov indicates the asymptotic covariance matrix.

Under smoothness conditions on Ω, this central limit theorem for the ROWCov under the locally constant volatility model implies the result in (4.2) where we let $\lambda \to 0$, while $\lambda/\Delta \to \infty$. This concludes the sketch of the proof of asymptotic normality of the ROWCov under the BSM model.

To gain further insight in the asymptotic variance, let us focus on the one-dimensional case. The univariate versions of the RCov, RBPCov and ROWCov are called the Realized Variance (RVar), Realized BiPower Variation (RBPVar) and Realized Outlyingness Weighted Variance (ROWVar), respectively. The asymptotic variance of the RVar, RBPVar and ROWVar is a factor θ times the integrated quarticity of the process. The factor θ characterizes the efficiency of the estimator. The lower the value of θ, the more efficient the estimator for the IVar is. For the RVar, one has that $\theta = 2$, while for the RBPVar $\theta = 2.609$ (Barndorff-Nielsen and Shephard, 2006). In the column $N = 1$, we see that for our preferred estimator (the ROWVar with hard rejection weight function and threshold $k = \chi^2_1(0.999)$) $\theta = 2.152$ and is thus more efficient than the RBPVar. It is also more efficient than other recently proposed alternatives to the RBPVar such as the MinRV and MedRV estimators of Andersen et al. (2009) for which θ is 3.81 and 2.96, respectively. Some versions of the Quantile RV estimator of Christensen et al. (2010) are more efficient than our preferred ROWVar estimator, but these versions are based on extreme empirical quantiles (such as 99%) of the returns in the local window. These versions are thus only robust to jumps affecting a small proportion of returns in the local window.

I.4 Efficiency gains of the use of standardized weights

As explained in Subsection 3.2 of the main paper, the weights we use are not the simple hard and soft rejection weights, but the weights, multiplied by the ratio between the expected value of the weights and the sample average of the weights of the returns belonging to the same local window. This modification stabilizes the estimates. In Table I.1 we report the asymptotic variance of the diagonal and off-diagonal elements of the ROWCov with the standardized hard rejection weights (upper panel) and the simple hard rejection weights (lower panel). We see that under the standard multivariate normal, the use of the modified weights leads to a significant increase in the efficiency of the diagonal elements of the ROWCov estimator, while not affecting the off-diagonal elements.

II Additional simulation studies

In the last paragraph of Section 3 we recommend the use of windows of one day. This recommendation is based on the simulation study in Subsection II.1. In Subsection 5.5 of the main paper we study the accuracy of the normal approximation for the finite sample 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. In Subsection II.2 we give the explicit expression for these statistics and the estimators of their standard errors. We further report the time series plot of the estimation error and the QQ-plot of the corresponding t-statistics for 1000 simulated days.

II.1 Choice of local window length

The jump detection with the hard rejection weight function consists in comparing the outlyingness statistic with a threshold, which we take as a high quantile of the chi-square distribution with N degrees of freedom. The outlyingness statistic $d_{i,\Delta}$ is computed using the MCD covariance on a local window of length λ. Practically, the choice of λ must be such that $\lfloor \lambda/\Delta \rfloor$ is large enough (such that the number of observations in the local window is high enough), but not too large (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).

We evaluate in Table II.1 the impact of the length of local window on the size of the multivariate jump detection statistic. We compare $\lambda = 1$ with $\lambda = 0.5$. Standard errors around the reported results are in parenthesis. We consider 95%, 99% and 99.9% threshold and thus expect a 5%, 1% and 0.1% rejection probability, respectively. We see that in all cases considered the test is oversized, but the size distortion becomes small when returns are sampled at high enough frequencies. When sampling at the 5 or 15-minute frequency, jumps are more accurately detected using local windows of one day rather than shorter windows. At the 1-minute frequency, there is little or no impact of the window length on the size of the test. On the basis of these simulation results, we therefore recommend using a window length of 1 day.

Table II.1: Proportion of outlyingness statistics $d_{i,\Delta}$ exceeding the threshold k. Local windows are either the complete day $[0,1]$ or half of the day: $[0,0.5[$ and $[0.5,1]$.

Local window	Complete day			Half of day		
k	$\chi^2_2(0.95)$	$\chi^2_2(0.99)$	$\chi^2_2(0.999)$	$\chi^2_2(0.95)$	$\chi^2_2(0.99)$	$\chi^2_2(0.999)$
15-min	0.076	0.028	0.008	0.089	0.043	0.019
	(1.2e-3)	(0.7e-3)	(0.4e-3)	(1.2e-3)	(0.7e-3)	(0.4e-3)
5-min	0.063	0.019	0.004	0.067	0.022	0.005
	(0.6e-3)	(0.4e-3)	(0.2e-3)	(0.6e-3)	(0.4e-3)	(0.2e-3)
1-min	0.056	0.014	0.002	0.057	0.014	0.002
	(0.3e-3)	(0.1e-3)	(0.1e-3)	(0.3e-3)	(0.1e-3)	(0.1e-3)

II.2 Distribution of the ROWCov in finite samples

The asymptotic normality of the ROWCov is given in (4.2). This subsection complements Subsection 5.5 in assessing how good the normal distribution approximates the finite sample distribution of statistics based on the ROWCov. To shorten notation, we use $\hat{S}_\Delta$ and S to denote the ROWCov and the ICov, respectively. For local windows of one day, an estimate of the asymptotic covariance matrix of $(\hat{S}_{\Delta(11)}, \hat{S}_{\Delta(12)}, \hat{S}_{\Delta(22)})'$ is

$$\hat{C}_\Delta = \begin{pmatrix} \hat{V}_{\Delta(1,1)} & \hat{V}_{\Delta(1,2)} & \hat{V}_{\Delta(1,N+2)} \\ \hat{V}_{\Delta(1,2)} & \hat{V}_{\Delta(2,2)} & \hat{V}_{\Delta(2,N+2)} \\ \hat{V}_{\Delta(1,N+2)} & \hat{V}_{\Delta(2,N+2)} & \hat{V}_{\Delta(N+2,N+2)} \end{pmatrix},$$

where $\hat{V}_\Delta = (\hat{S}_\Delta^{1/2} \otimes \hat{S}_\Delta^{1/2})\Theta_w(\hat{S}_\Delta^{1/2} \otimes \hat{S}_\Delta^{1/2})'$, see equation (4.2) in the main paper.

In particular, the estimated standard error for the variance and log-variance statistics is then given by:

$$(\hat{C}_{\Delta(1,1)}\Delta)^{1/2} \text{ and } (\hat{C}_{\Delta(1,1)}\Delta/\hat{S}^2_{\Delta(1,1)})^{1/2},$$

respectively. The corresponding t-statistics are:

$$\sqrt{\frac{1}{\Delta}}\frac{\left(\hat{S}_{\Delta(1,1)} - S_{(1,1)}\right)}{(\hat{C}_{\Delta(1,1)})^{1/2}} \text{ and } \sqrt{\frac{1}{\Delta}}\frac{\left(\log \hat{S}_{\Delta(1,1)} - \log S_{(1,1)}\right)}{(\hat{C}_{\Delta(1,1)}/\hat{S}^2_{\Delta(1,1)})^{1/2}}.$$

Our main interest is however in the covariance, beta and (Fisher) correlation estimates, with the daily beta of component 2 on component 1 defined as the daily integrated covariance $S_{(1,2)}$ divided by $S_{(1,1)}$, the integrated variance of component 1. For the covariance and the beta estimators, the estimated standard errors are:

$$(\hat{C}_{\Delta(2,2)}\Delta)^{1/2} \text{ and } (g'\hat{C}_{1,2}g\Delta)^{1/2},$$

with $g = (-\hat{S}_{\Delta(1,2)}\hat{S}_{2,2}^{-2}, \hat{S}_{\Delta(1,1)}^{-1}, 0)'$. The corresponding t-statistics are:

$$\sqrt{\frac{1}{\Delta}}\frac{\left(\hat{S}_{\Delta(1,2)} - S_{(1,2)}\right)}{(\hat{C}_{\Delta(2,2)})^{1/2}} \text{ and } \sqrt{\frac{1}{\Delta}}\frac{\left(\hat{S}_{\Delta(1,1)}^{-1}\hat{S}_{\Delta(1,2)} - S_{(1,1)}^{-1}S_{(1,2)}\right)}{(g'\hat{C}_{1,2}g)^{1/2}}.$$

Finally, for the correlation and Fisher transformed correlation estimators, the estimated standard errors are given by:

$$(h'\hat{C}_{1,2}h\Delta)^{1/2} \quad \text{and} \quad ((1-\hat{r}_{(1,2)}^2)^{-2}h'\hat{C}_{1,2}h\Delta)^{1/2},$$

with $\hat{r}_{\Delta(1,2)} = \hat{S}_{\Delta(1,1)}^{-1/2}\hat{S}_{\Delta(2,2)}^{-1/2}\hat{S}_{\Delta(1,2)}$ the correlation estimate. The corresponding t-statistics are:

$$\sqrt{\frac{1}{\Delta}}\frac{\left(\hat{r}_{\Delta(1,2)} - r_{(1,2)}\right)}{(h'\hat{C}_{1,2}h)^{1/2}} \quad \text{and} \quad \sqrt{\frac{1}{\Delta}}\frac{\left(\frac{1}{2}\log\frac{1+\hat{r}_{(1,2)}}{1-\hat{r}_{(1,2)}} - \frac{1}{2}\log\frac{1+r_{(1,2)}}{1-r_{(1,2)}}\right)}{((1-\hat{r}_{(1,2)}^2)^{-2}h'\hat{C}_{1,2}h)^{1/2}},$$

with $r_{(1,2)} = S_{(1,1)}^{-1/2}S_{(2,2)}^{-1/2}S_{(1,2)}$ and

$$h = (-0.5\hat{S}_{\Delta(1,1)}^{-3/2}\hat{S}_{\Delta(2,2)}^{-1/2}\hat{S}_{\Delta(1,2)}, \hat{S}_{\Delta(11)}^{-1/2}\hat{S}_{\Delta(22)}^{-1/2}, -0.5\hat{S}_{\Delta(1,1)}^{-1/2}\hat{S}_{\Delta(2,2)}^{-3/2}\hat{S}_{\Delta(1,2)})'.$$

Figures 1-6 plot the estimation errors and confidence bands of these t-statistics for the ROWCov estimates computed on 15, 5 and 1-minute returns over 1000 simulated days, for $N = 2$. The upper plots are the estimation errors. 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 and the width of the 95% confidence bands of these errors. The figures are similar to the ones obtained by Barndorff-Nielsen and Shephard (2005) for the RCov. Due to the changing volatility in time, the confidence intervals of the variance, covariance, beta and correlation estimates vary a lot over the days.

The lower panel of Figures 1-6 report the normal QQ-plots. For all statistics, the normality approximation is somewhat poor for the ROWCov using 15-minute returns, but it improves considerably when sampling at higher frequencies. The log transformation of the variance estimate and the Fisher transformation of the correlation estimate clearly improve the accuracy of the asymptotic approximation. We see that, also when 5-minute returns are used, the distribution of the log-variance and Fisher transformed correlation estimates is still well approximated by the normal distribution.

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

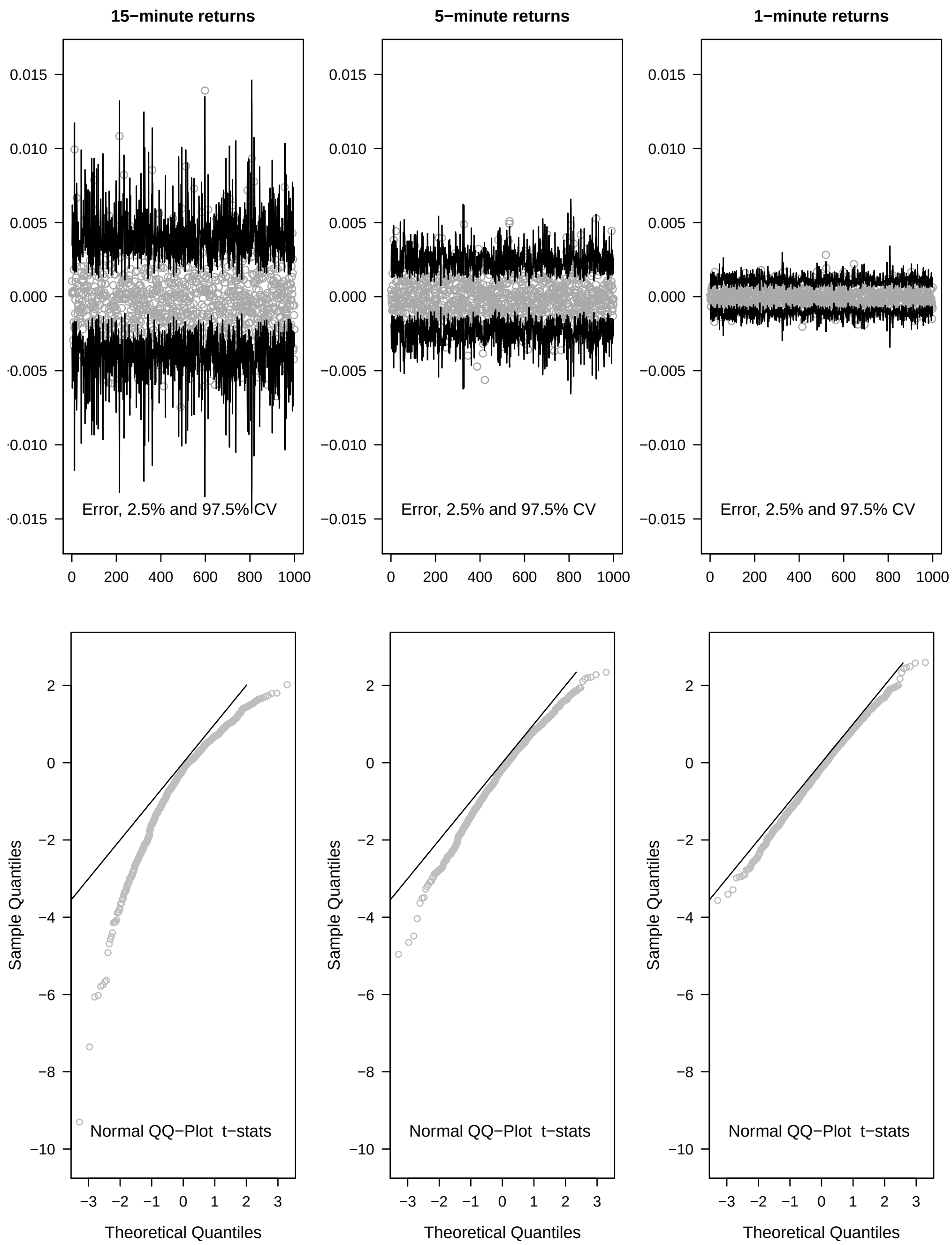


Figure 2: Difference between estimated and true log of daily integrated variance (upper panel, with estimated asymptotic 2.5% and 97.5% critical values as solid lines) and normal QQ-plots of the standardized estimation error (lower panel) for the ROWCov estimator using 15, 5 and 1-minute returns over 1000 simulated days.

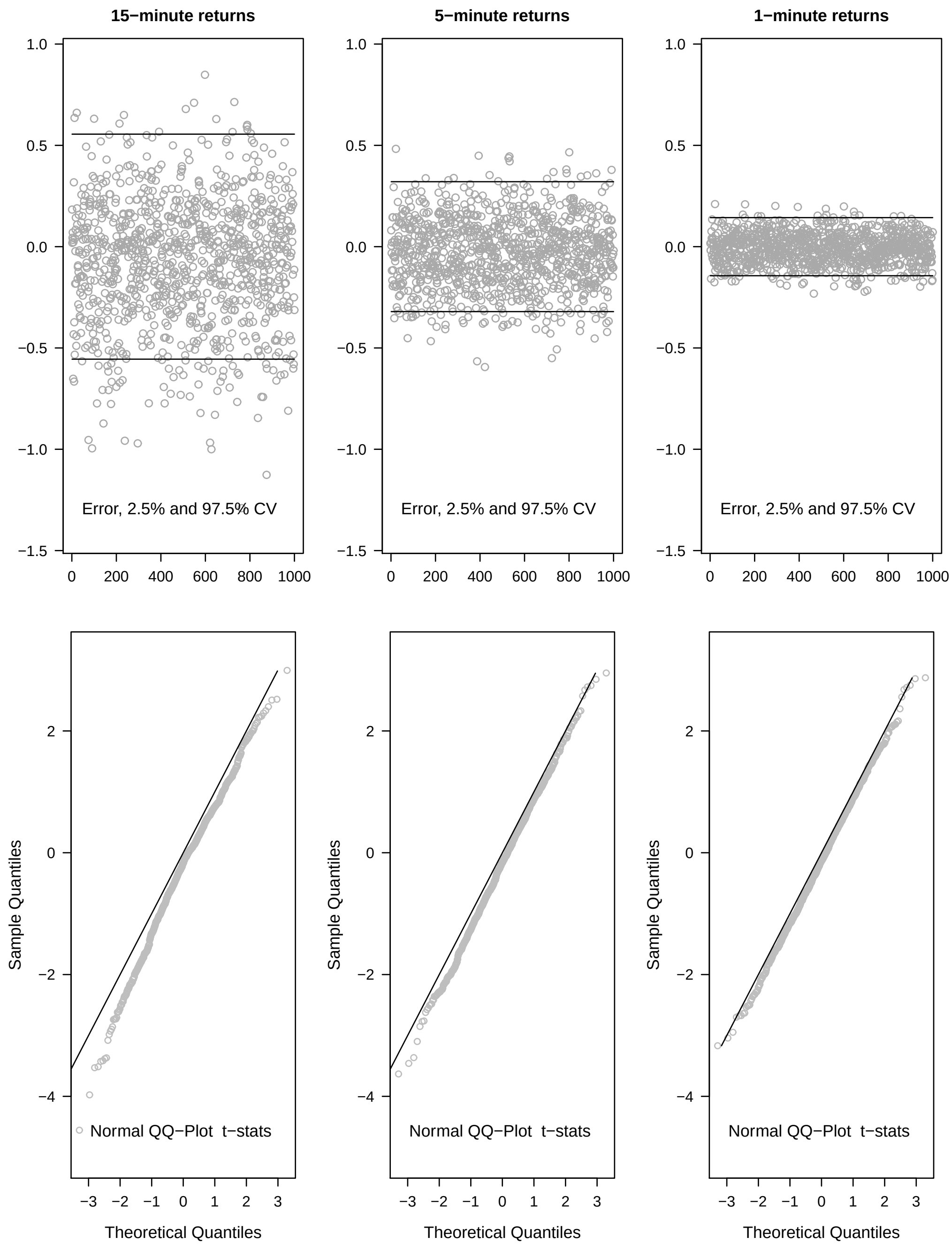


Figure 3: Difference between estimated and true daily integrated covariance (upper panel, with estimated asymptotic 2.5% and 97.5% critical values as solid lines) and normal QQ-plots of standardized estimation error (lower panel) for the ROWCov estimator using 15, 5 and 1-minute returns over 1000 simulated days.

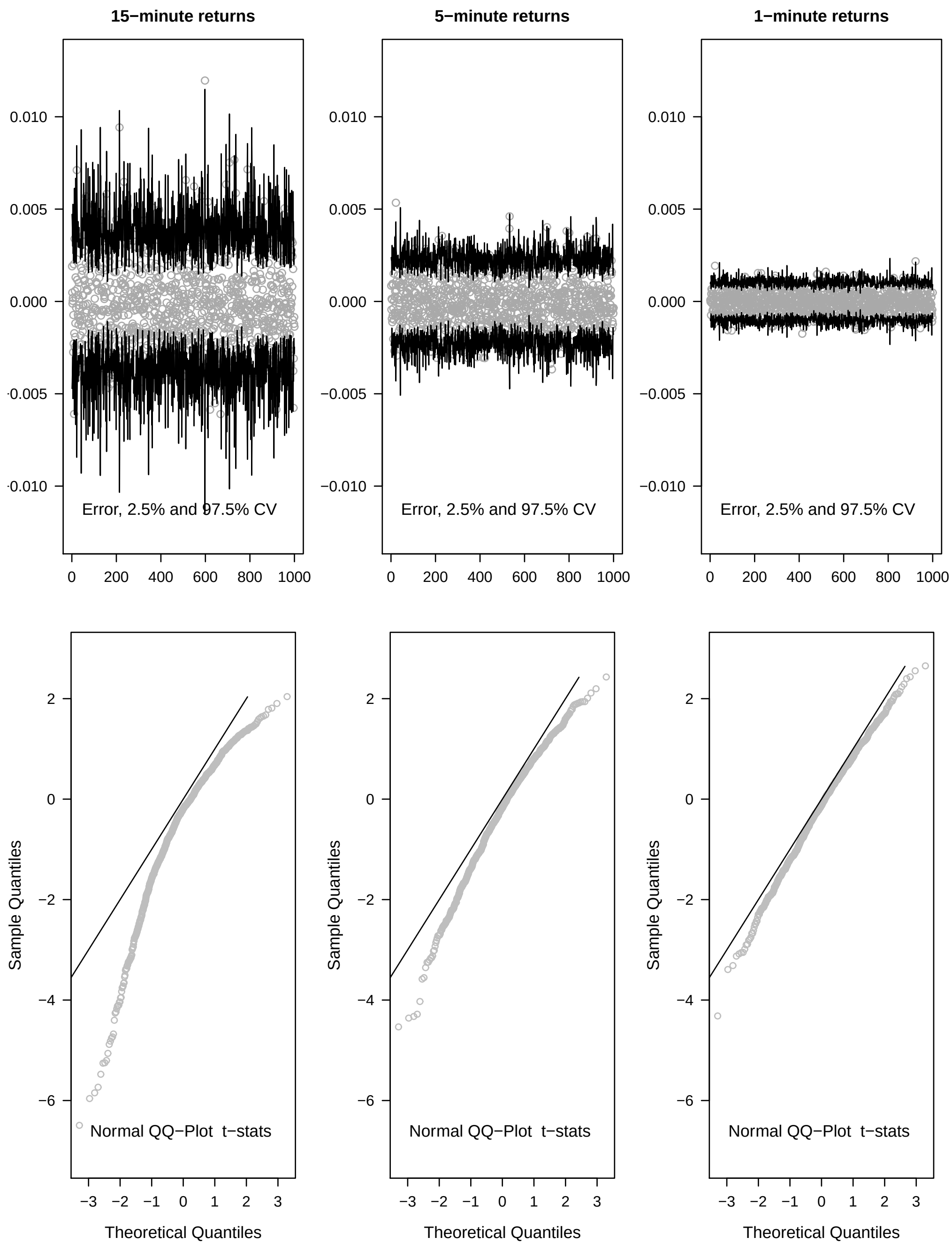


Figure 4: Difference between estimated and true daily beta (upper panel, with estimated asymptotic 2.5% and 97.5% critical values as solid lines) and normal QQ-plots of the standardized estimation error (lower panel) for the ROWCov estimator using 15, 5 and 1-minute returns over 1000 simulated days.

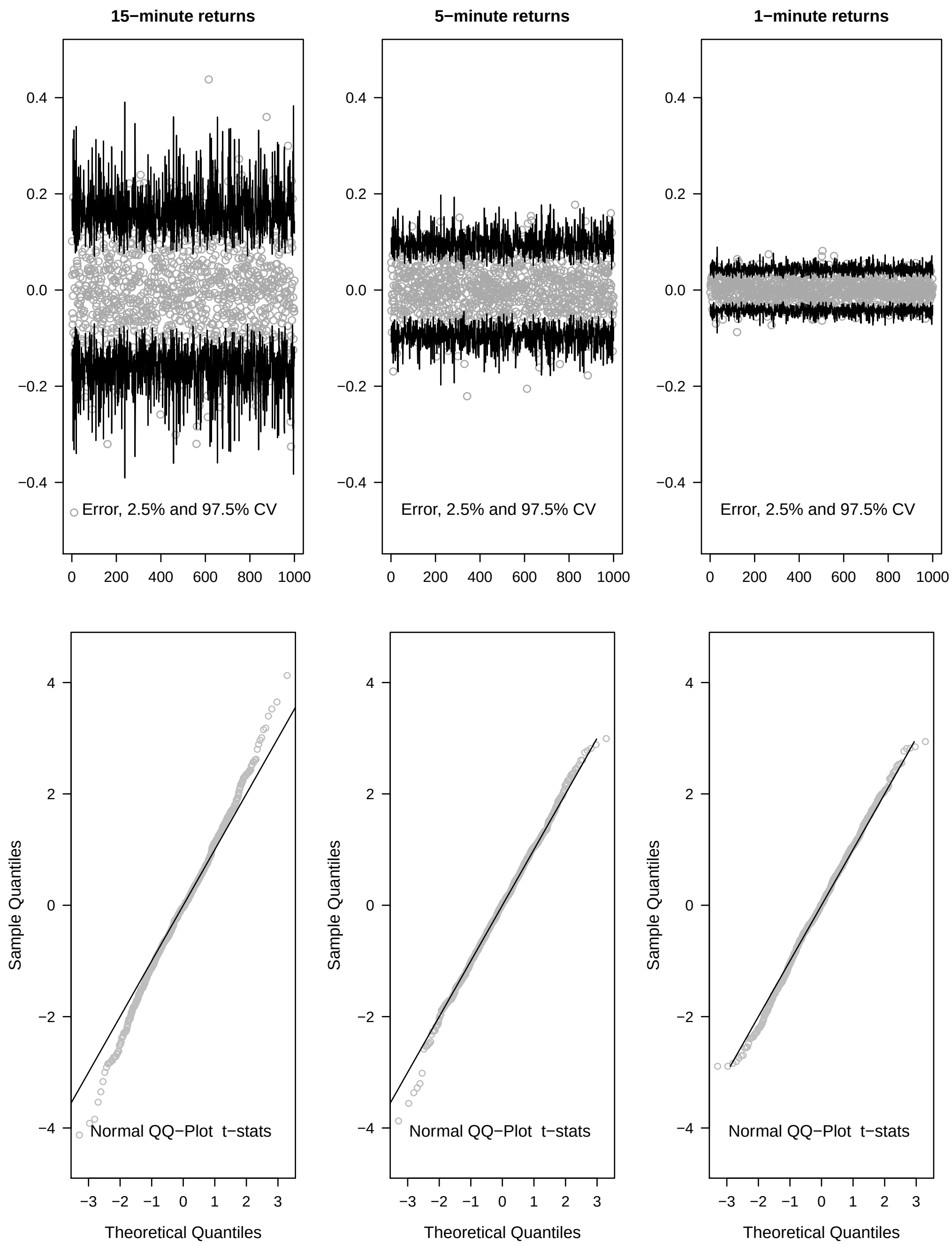


Figure 5: Difference between estimated and true daily integrated correlation (upper panel, with estimated asymptotic 2.5% and 97.5% critical values as solid lines) and normal QQ-plots of the standardized estimation error (lower panel) for the ROWCov estimator using 15, 5 and 1-minute returns over 1000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

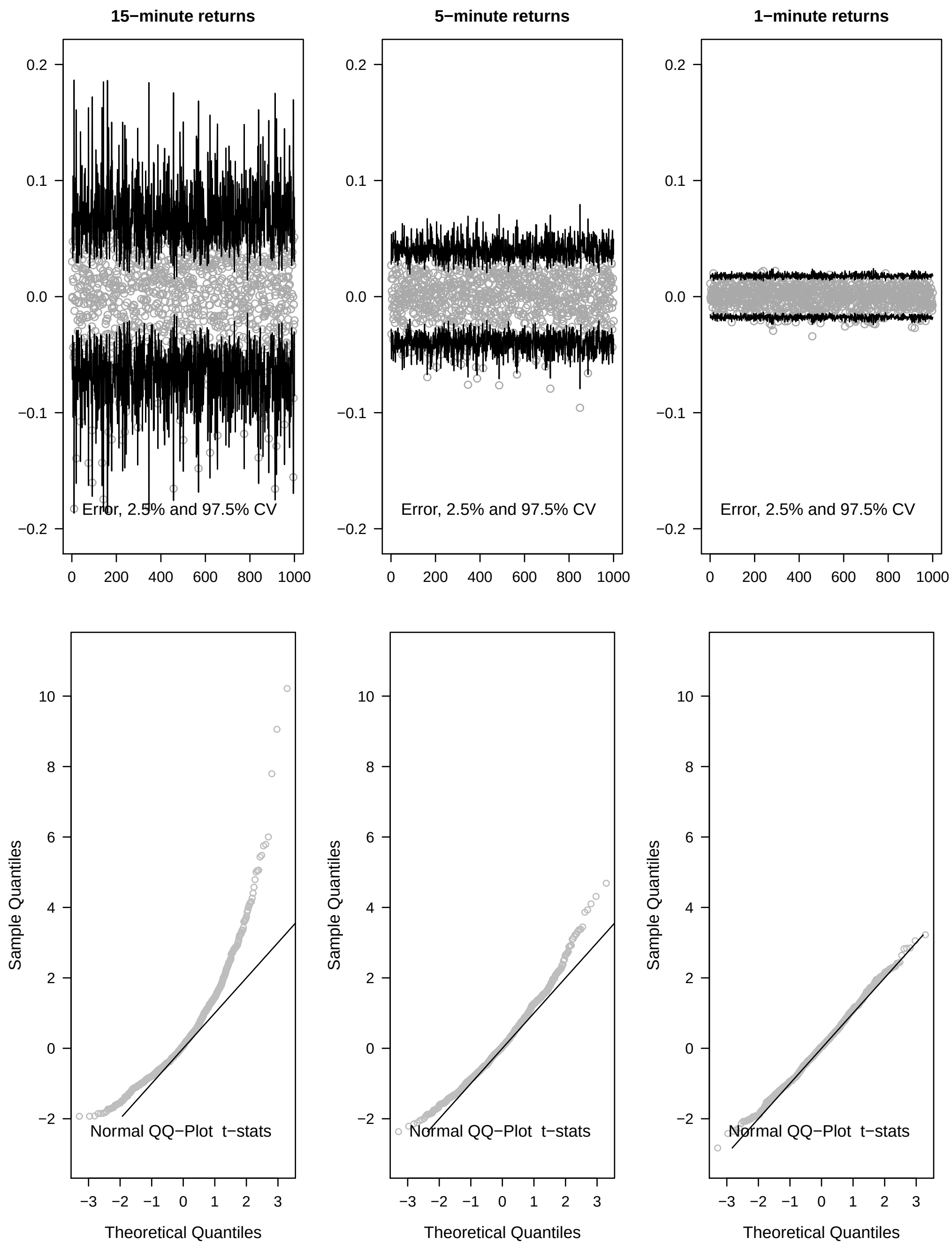
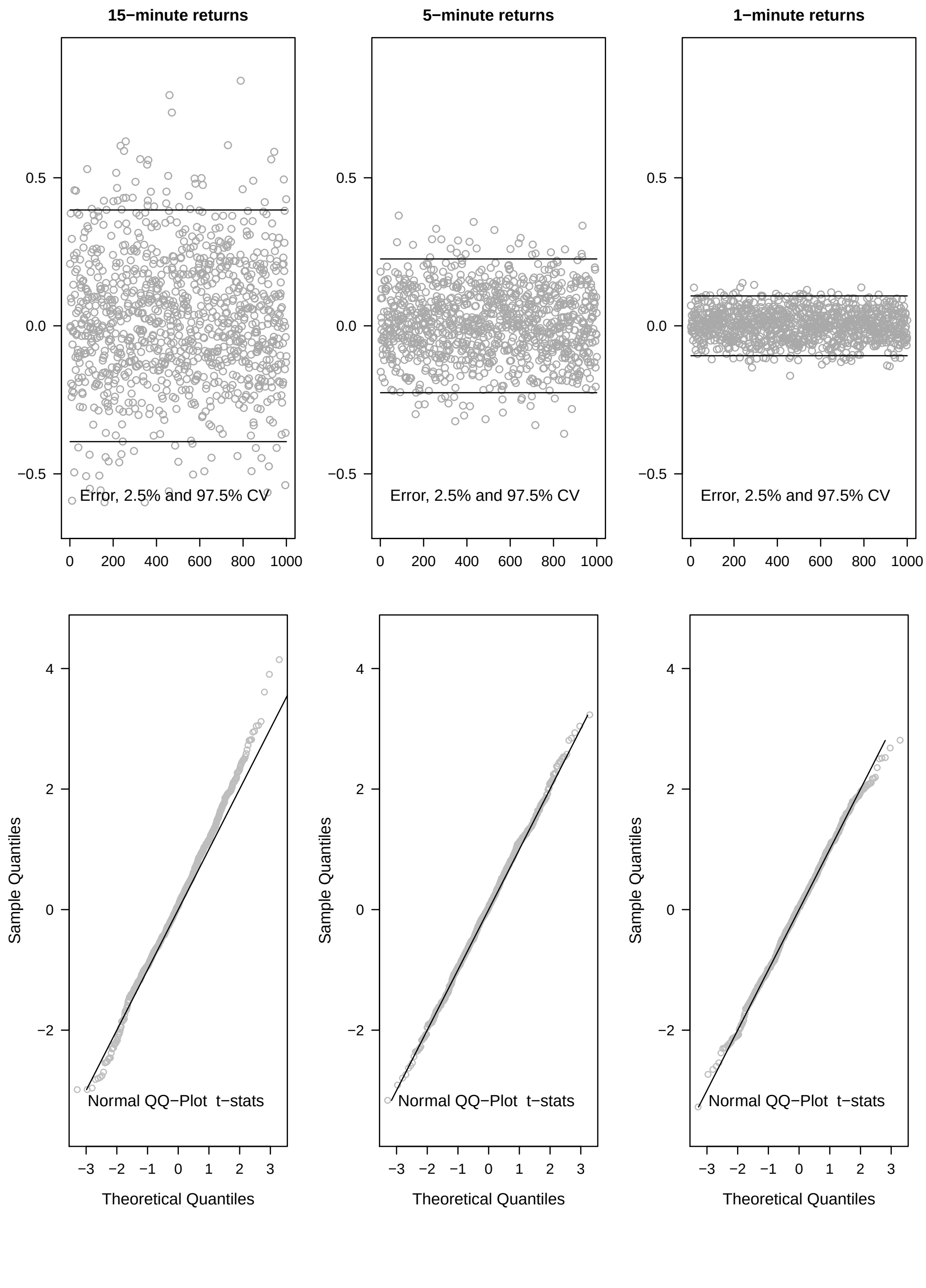


Figure 6: Difference between Fisher transformed estimated and true daily integrated correlation (upper panel, with estimated asymptotic 2.5% and 97.5% critical values as solid lines) and normal QQ-plots of the standardized estimation error (lower panel) for the ROWCov estimator using 15, 5 and 1-minute returns over 1000 simulated days.

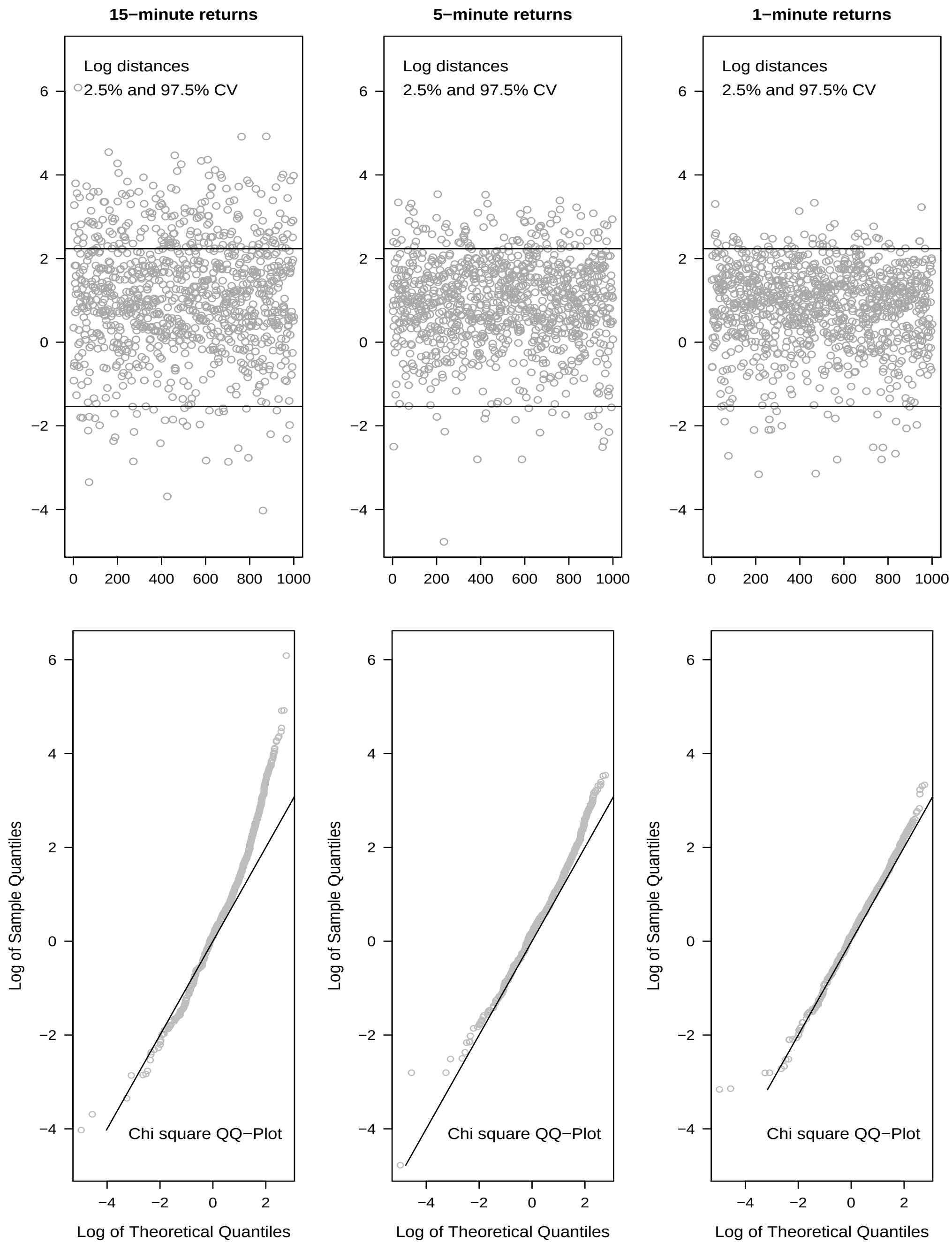


Finally, we also consider for $N = 2$ the distances between the lower diagonal elements of the ROWCov estimates and the ICov:

$$\sqrt{\frac{1}{\Delta}}\text{vech}(\hat{S}_\Delta - \text{ICov})'\hat{C}_\Delta^{-1}\text{vech}(\hat{S}_\Delta - \text{ICov}),$$

where vech$(\cdot)$ denotes the operator that stacks the lower triangular portion of a $N \times N$ matrix as a $N(N+1)/2 \times 1$ vector. Under (4.2) of the main paper, this statistic should be approximately chi-square distributed with 3 degrees of freedom when Δ is small. The upper panel of Figure 7 plots the log of these distances over 1000 simulated days, with the horizontal lines corresponding to the asymptotic 2.5% and 97.5% quantiles. Using 15, 5 and 1-minute returns, there are 23.9%, 10.8% and 7% of the ROWCov distances that are more extreme than these quantiles. This is consistent with the QQ-plots in Figures 1 and 3, showing the fat tails of the standardized ROWCov variance and covariance estimates, and that the normality approximation improves when sampling at higher frequencies. The lower panel of Figure 7 reports the QQ-plot for the log of the distances with respect to the log of the quantiles from a chi-square distribution with 3 degrees of freedom. We see that the chi-square approximation is somewhat poor for the distribution of the ROWCov computed using 15-minute returns, but it improves considerably when sampling at higher frequencies.

Figure 7: Chi-square distribution of distance between ROWCov and ICov (in logs) and 2.5 and 97.5% critical values.



References

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

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. and N. Shephard (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics 4*, 1–30.

Bilodeau, M. and D. Brenner (1999). *Theory of Multivariate Statistics*. Springer.

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). Realized quantile-based estimation of integrated variance. *Journal of Econometrics 159*, 74–98.

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.

Lopuhaä, H. (1999). Asymptotics of reweighted estimators of multivariate location and scatter. *Annals of Statistics 27*, 1638–1665.