Roughing it up some more: Jumps and Co-Jumps in Vast-Dimensional Price Processes
(Positive Semidefinite Integrated Covariance Estimation, Factorizations and Asynchronicity)

Kris Boudt, Sébastien Laurent and Rogier Quaedvlieg

OxMetrics User Conference, CASS Business School

September, 2014

Objective

We propose an estimator of the integrated covariance of log-returns under

- asynchronicity;
- microstructure noise;
- and finite activity jumps.

Objective

We propose an estimator of the integrated covariance of log-returns under

- asynchronicity;
- microstructure noise;
- and finite activity jumps.

The estimator is

- based on the CholCov, introduced in BLLQ (2014);
- positive semidefinite by construction;
- very efficient;
- useful for forecasting the covariance of log-returns.

DGP: multivariate BSMFAJ

We assume that the true d-dimensional log-price process is:

$$dY(s) = \mu(s)ds + \sigma(s)dW(s) + K(s)dq(s),$$

- μ is a $d \times 1$ predictable locally bounded drift process;
- W is a vector of independent Brownian motions;
- σ is a $d \times d$ càdlàg process such that $\Sigma(s) = \sigma(s)\sigma'(s)$;
- $K(s)$ is a $d \times d$ process controlling the magnitude and transmission of jumps, such that $K(s)dq(s)$ is the contribution of the jump process to the change in log-price series at time t.

DGP: multivariate BSMFAJ with noise

$$Y(t) = \int_0^t \mu(s)ds + \int_0^t \sigma(s)dW(s) + \sum_{j=1}^{N} K_j,$$

At very high frequencies, microstructure noise leads to a departure from this model. $\rightarrow$ Instead of observing Y, X is observed, where

$$X_t^{(d)} = Y_t^{(d)} + \epsilon_t^{(d)},$$

where $\epsilon_t^{(d)}$ is microstructure noise and $Y_t^{(d)}$ is the d-th component of Y. $\epsilon_t = (\epsilon_t^{(1)}, ..., \epsilon_t^{(d)})'$ is an i.i.d. process ($\perp Y$, satisfying $\mathbb{E}(\epsilon_t) = 0, \quad \mathbb{E}\left(\epsilon_t \epsilon_t'\right) = \Psi$, with Ψ a PD $d \times d$ matrix.

Main Goal

Our parameter of interest is the integrated covariance over the unit interval:

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

Cholesky decomposition

The spot covariance matrix can be split into

$$\Sigma(s) = H(s)G(s)H(s)',$$

where $H(s)$ is a lower diagonal matrix with ones on the diagonal, and $G(s)$ a diagonal matrix. More specifically,

$$H(s) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ h_{21}(s) & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_{d1}(s) & h_{d2}(s) & \cdots & 1 \end{bmatrix} \quad G(s) = \begin{bmatrix} g_{11}(s) & 0 & \cdots & 0 \\ 0 & g_{22}(s) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & g_{dd}(s) \end{bmatrix}$$

CholCov

- The elements of H and G can estimated sequentially;

CholCov

- The elements of H and G can estimated sequentially;
- Each element requires only a subset of the data series;

CholCov

- The elements of H and G can estimated sequentially;
- Each element requires only a subset of the data series;
- Let r be a vector of high-frequency intraday returns such that $\mu(t) = 0$;

CholCov

- The elements of H and G can estimated sequentially;
- Each element requires only a subset of the data series;
- Let r be a vector of high-frequency intraday returns such that $\mu(t) = 0$;
- In absence of jumps, $r \sim N(0, \Sigma)$, where $\Sigma = HGH'$ and therefore $f = H^{-1}r \sim N(0, G)$;

CholCov

- The elements of H and G can estimated sequentially;
- Each element requires only a subset of the data series;
- Let r be a vector of high-frequency intraday returns such that $\mu(t) = 0$;
- In absence of jumps, $r \sim N(0, \Sigma)$, where $\Sigma = HGH'$ and therefore $f = H^{-1}r \sim N(0, G)$;
- Since H is triangular, we obtain that each component in the vector of returns $r = Hf$, is an explicit function of the components with a lower index.

CholCov

More precisely, $f^{(1)} = r^{(1)} \sim N(0, g_{11})$, and for $k = 2, ..., d$,

$$r^{(k)} = h_{k1} f^{(1)} + h_{k2} f^{(2)} + ... + h_{k(k-1)} f^{(k-1)} + f^{(k)}$$
$$f^{(k)} \sim N(0, g_{kk}).$$

The g_{kk} and h_{kl} elements are therefore nothing but residual variances and regression coefficients of the factors.

CholCov, case $d = 3$, no noise, no jumps, synchronous tradings

$$\Sigma = HGH' = \begin{bmatrix} g_{11} & h_{21}g_{11} & h_{31}g_{11} \\ h_{21}g_{11} & h_{21}^2 g_{11} + g_{22} & h_{21}h_{31}g_{11} + h_{32}g_{22} \\ h_{31}g_{11} & h_{21}h_{31}g_{11} + h_{32}g_{22} & h_{31}^2 g_{11} + h_{32}^2 g_{22} + g_{33} \end{bmatrix}.$$

CholCov, case $d = 3$, no noise, no jumps, synchronous tradings

$$
\begin{aligned}
\text{Step 1: Set:} \quad & f_{t_j}^{(1)} &&= r_{t_j}^{(1)} \\
\text{Set:} \quad & \hat{g}_{11} &&= IV(f^{(1)}) \\
\text{Step 2: Regress:} \quad & r_{t_j}^{(2)} &&= h_{21} f_{t_j}^{(1)} + \eta_{t_j}^{(2)} \\
\text{Set:} \quad & f_{t_j}^{(2)} &&= r_{t_j}^{(2)} - \hat{h}_{21} f_{t_j}^{(1)} \\
\text{Set:} \quad & \hat{g}_{22} &&= IV(f^{(2)}) \\
\text{Step 3: Regress:} \quad & r_{t_j}^{(3)} &&= h_{31} f_{t_j}^{(1)} + h_{32} f_{t_j}^{(2)} + \eta_{t_j}^{(3)} \\
\text{Set:} \quad & f_{t_j}^{(3)} &&= r_{t_j}^{(3)} - \hat{h}_{31} f_{t_j}^{(1)} - \hat{h}_{32} f_{t_j}^{(2)} \\
\text{Set:} \quad & \hat{g}_{33} &&= IV(f^{(3)}) \\
\text{Step d: Regress:} \quad & r_{t_j}^{(d)} &&= \sum_{l=1}^{d-1} h_{dl} f_{t_j}^{(l)} + \eta_{t_j}^{(d)} \\
\text{Set:} \quad & f_{t_j}^{(d)} &&= r_{t_j}^{(d)} - \sum_{l=1}^{d-1} \hat{h}_{dl} f_{t_j}^{(l)} \\
\text{Set:} \quad & \hat{g}_{dd} &&= IV(f^{(d)}).
\end{aligned}
$$

Maximum likelihood estimator under no microstructure noise, constant volatility and a multivariate sampling grid.

- $r_{t_j} \sim N(0, \Delta_j \Sigma)$, where $\Delta_j = t_j - t_{j-1}$.

Maximum likelihood estimator under no microstructure noise, constant volatility and a multivariate sampling grid.

- $r_{t_j} \sim N(0, \Delta_j \Sigma)$, where $\Delta_j = t_j - t_{j-1}$.
- $f_{t_j} = H^{-1} r_{t_j} \sim N(0, \Delta_j G)$.

Maximum likelihood estimator under no microstructure noise, constant volatility and a multivariate sampling grid.

- $r_{t_j} \sim N(0, \Delta_j \Sigma)$, where $\Delta_j = t_j - t_{j-1}$.
- $f_{t_j} = H^{-1} r_{t_j} \sim N(0, \Delta_j G)$.
- $\mathcal{L} = -\frac{1}{2}\left[Nd \ln(2\pi) + \sum_{i=1}^{d} \sum_{j=1}^{N} \ln \Delta_j g_{ii} + \sum_{i=1}^{d} \sum_{j=1}^{N} \frac{f_{t_j}^{(i)2}}{\Delta_j g_{ii}} \right].$

Maximum likelihood estimator under no microstructure noise, constant volatility and a multivariate sampling grid.

- $r_{t_j} \sim N(0, \Delta_j \Sigma)$, where $\Delta_j = t_j - t_{j-1}$.
- $f_{t_j} = H^{-1} r_{t_j} \sim N(0, \Delta_j G)$.
- $\mathcal{L} = -\frac{1}{2}\left[Nd \ln(2\pi) + \sum_{i=1}^{d} \sum_{j=1}^{N} \ln \Delta_j g_{ii} + \sum_{i=1}^{d} \sum_{j=1}^{N} \frac{f_{t_j}^{(i)2}}{\Delta_j g_{ii}} \right]$.
- From the FOC, it can easily be shown that the MLE estimators of the elements of q are

$$\hat{g}_{kk}^{MLE} = \sum_{j=1}^{N} f_{t_j}^{(k)2} = RV(f^{(k)}),$$

$$\hat{h}_{kl}^{MLE} = \frac{\sum_{j=1}^{N} r_{t_j}^{(k)} f_{t_j}^{(l)}}{\sum_{j=1}^{N} f^{(l)2}} = RBeta(r^{(k)}, f^{(l)}).$$

Properties of CholCov

Let $q = (\text{diag}(G), \text{lvech}(H))'$.

Properties of CholCov

Let $q = (\text{diag}(G), \text{lvech}(H))'$.

Assumption 1: Assume that $\hat{q} \xrightarrow{p} q$, as $N \to \infty$.

This is a very mild assumption, as for a wide variety of assumptions on the underlying DGP, such an estimator exists.

Properties of CholCov

Let $q = (\text{diag}(G), \text{lvech}(H))'$.

Assumption 1: Assume that $\hat{q} \xrightarrow{p} q$, as $N \to \infty$.

This is a very mild assumption, as for a wide variety of assumptions on the underlying DGP, such an estimator exists.

Theorem 1: Let Assumption 1 hold. Then, in absence of jumps,

$$CholCov \xrightarrow{p} \int_0^1 \Sigma(s)ds$$

as $N \to \infty$.

Properties of CholCov

Assumption 3: Assume $\Sigma(s) = \Sigma$, $\mu(s) = 0$ and $\Psi = 0$.

Assumption 4: Assume a multivariate sampling grid and that all d assets trade at the resulting sampling times.

Properties of CholCov

Assumption 3: Assume $\Sigma(s) = \Sigma$, $\mu(s) = 0$ and $\Psi = 0$.

Assumption 4: Assume a multivariate sampling grid and that all d assets trade at the resulting sampling times.

Theorem 2: Let Assumption 1-4 hold. Then, in absence of jumps, the realized variance and realized beta estimates $\hat{g}_{kk}^{MLE}$ and $\hat{h}_{kl}^{MLE}$ are the Maximum Likelihood estimators of the elements in q, and therefore

$$\sqrt{N}(\hat{q} - q) \xrightarrow{d} \mathcal{N}(0, \Phi),$$

where Φ is the inverse Fisher information matrix.

Properties of CholCov

In the three dimensional case,

$$\Phi = \text{Hessian}^{-1} = \begin{bmatrix} 2g_{11}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2g_{22}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2g_{33}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{g_{22}}{g_{11}} & \frac{h_{32}g_{22}}{g_{11}} & 0 \\ 0 & 0 & 0 & \frac{h_{32}g_{22}}{g_{11}} & \frac{g_{33}+h_{32}^2g_{22}}{g_{11}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{g_{33}}{g_{22}} \end{bmatrix}.$$

Properties of CholCov

Assumption 5: There exists $\alpha > 0$ such that $N^{\alpha}(\hat{q} - q) \xrightarrow{d} \mathcal{N}(0, \Phi)$, when $N \to \infty$. Applying the delta method, the following limiting distribution for the CholCov can be obtained.

Properties of CholCov

Assumption 5: There exists $\alpha > 0$ such that $N^{\alpha}(\hat{q} - q) \xrightarrow{d} \mathcal{N}(0, \Phi)$, when $N \to \infty$. Applying the delta method, the following limiting distribution for the CholCov can be obtained.

Theorem 2: Let Assumption 5 hold. Then

$$N^{\alpha}(\text{vech}(CholCov) - \text{vech}(ICov)) \xrightarrow{d} \mathcal{N}(0, \nabla\Phi\nabla'),$$

when $N \to \infty$ and where ∇ is the $d^* \times d^*$ matrix whose ijth element is the derivative of $\text{vech}(CholCov)_i$ with respect to $\hat{q}$.

Properties of CholCov

In the three dimensional case,

$$\nabla = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ h_{21} & 0 & 0 & g_{11} & 0 & 0 \\ h_{31} & 0 & 0 & 0 & g_{11} & 0 \\ h_{21}^2 & 1 & 0 & 2h_{21}g_{11} & 0 & 0 \\ h_{21}h_{31} & h_{32} & 0 & h_{31}g_{11} & h_{21}g_{11} & g_{22} \\ h_{31}^2 & h_{32}^2 & 1 & 0 & 2h_{31}g_{11} & 2h_{32}g_{22} \end{bmatrix}$$

$\rightarrow$ $\nabla\Phi\nabla'$ is identical to the one of vech($RCOV$).

Strip and replace

The strip and replace method ensures that the diagonal elements of the CholCov are estimated using all available observations. More precisely, denote by $\hat{D}$ the diagonal matrix with the square root of univariate Integrated Variance estimates. Then the strip and replace version of the CholCov is

$$CholCov^* = \hat{D}\,\text{diagonal}(CholCov)^{-1/2}\,CholCov\,\text{diagonal}(CholCov)^{-1/2}\,\hat{D}.$$

Simulation

We generate hypothetical prices, with $Y^{(k)}(s)$ the associated log-price of asset k, from the log-price diffusion given by

$$
\begin{aligned}
dY_t^{(k)} &= \mu^{(k)} ds + dV_t^{(k)} + dF_t^{(k)} + dJ_t^{(k)} \\
dV_t^{(k)} &= \rho^{(k)}\ \sigma_t^{(k)} dB_t^{(k)} \\
dF_t^{(k)} &= \sqrt{1 - (\rho^{(k)})^2}\sigma_t^{(k)} dW_t,
\end{aligned}
$$

with $k = 1, \dots, d$. All $B^{(k)}$ as well as W are independent Brownian motions. $F^{(k)}$ denotes the common factor, scaled by $\sqrt{1-\rho^2}$ to determine its strength.

Simulation: diffusive SV

- Each $Y^{(k)}$ is a diffusive SV model with drift $\mu^{(k)}$.
- Their random spot volatility are given by $\sigma^{(k)} = \exp(\beta_0^{(k)} + \beta_1^{(i)} \varrho^{(k)})$, with $d\varrho^{(k)} = \alpha^{(k)} \varrho^{(k)} dt + dB^{(k)}$.
- We calibrate the parameters $(\mu, \beta_0, \beta_1, \alpha, \rho)$ at $(0.03, -5/16, 1/8, -1/40, -0.3)$.
- Equicorrelation of 0.91.
- The parameter choice ensures that $\mathrm{E}\left(\int_0^1 \sigma^{(k)2}(u)du\right) = 1$.

Simulation: Microstructure noise

Microstructure noise is added to the return log-prices as $X^{(k)} = Y^{(k)} + \epsilon^{(k)}$ with

$$\epsilon^{(k)} \mid \sigma, X \overset{iid}{\sim} N(0, \omega^2) \quad \text{with} \quad \omega^2 = \xi^2 \sqrt{N^{-1} \sum_{j=1}^{N} \sigma^{(k)4}(j/N)}.$$

Hence, the variance of the noise increases with the variance of the underlying process, in line with evidence from Bandi and Russell (2006). We set $\omega^2 = 0.001$.

Simulation: jumps

- Occurrence of jumps: to each series $dX^{(k)}$ (independently), we add a single jump at a random point in the day.
- Size of the jumps: product between the realization of a uniformly distributed random variable on $\sigma_J([-2,-1] \cup [1,2])$ and the mean of the spot volatility process of that day.
- $\sigma_J = 0 \rightarrow$ no jumps while $\sigma_J = 1 \rightarrow$ moderately large jumps.

Pre-averaging

To make our estimators robust to microstructure noise, we do pre-averaging.

Define the series of intra-day returns $r^{(1)}_{\tau_j}, \dots, r^{(d)}_{\tau_j}$, where $r^{(k)}_{\tau_j} = X^{(k)}_{\tau_j} - X^{(k)}_{\tau_{j-1}}$. Pre-averaged returns are defined as

$$\bar{r}^{(k)}_{\tau_j} = \sum_{h=1}^{k_N - 1} g\left(\frac{h}{k_N}\right) r^{(k)}_{\tau_{j+h}},$$

with $g(x) = \min(x, 1 - x)$.

Robust CholCov

1. Set $f^{(1)} = r^{(1)}$
2. For $k = 1, ..., d$
   - Projection: $r^{(k)} = \sum_{l<k} h_{lk} f^{(l)} + f^{(k)}$
   - $g_{kk} = IV(f^{(k)})$

For $IV(f^{(k)})$, we use pre-averaged

$$BPV = \frac{N}{N-2k_n+2} \frac{\pi}{2} \sum_{j=0}^{N-2k_n+1} |r_{\tau_j}||r_{\tau_{j+k_n}}|$$

$$MEDRV = \frac{N}{N-3k_n+2} \frac{\pi}{6-4\sqrt{3}+\pi} \sum_{j=0}^{N-3k_n+1} med(|r_{\tau_j}|, r_{|\tau_{j+k_n}|}, r_{\tau_{j+2k_n}})^2$$

Robust CholCov

1. Set $f^{(1)} = r^{(1)}$
2. For $k = 1..., d$
   - Projection: $r^{(k)} = \sum_{l<k} h_{lk} f^{(l)} + f^{(k)}$
   - $g_{kk} = IV(f^{(k)})$

For the h_{kl} elements we use robust realized betas, that is $h_{kl} = \widehat{ICOV}(u^{(k)}, f^{(l)})/\widehat{IV}(f^{(l)})$ and the Polarization result $\widehat{ICOV}(u^{(k)}, f^{(l)}) = \frac{1}{4}\left(\widehat{IV}((u^{(k)} + f^{(l)})) - \widehat{IV}(u^{(k)} - f^{(l)})\right)$ with either BPV or $MEDRV$ for $\widehat{IV}((u^{(k)} + f^{(l)}))$ and $\widehat{IV}(u^{(k)} - f^{(l)})$

Finally do strip and replace.

Results: Bivariate case

Element	(0,0)		(0,1)		(1,1)	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
			$\sigma_J = 0$			
RCov	0.463	0.463	-0.058	0.058	0.464	0.464
BPCov	0.492	0.492	-0.090	0.090	0.491	0.491
CholCov RV	0.002	0.026	0.001	0.026	0.001	0.030
CholCov BPRV	0.001	0.028	0.000	0.028	0.001	0.032
CholCov MEDRV	0.000	0.030	-0.002	0.029	-0.001	0.033
			$\sigma_J = 1$			
RCov	1.002	4.265	-0.064	0.022	0.993	4.152
BPCov	0.538	1.190	-0.083	0.033	0.534	1.165
CholCov RV	0.534	1.370	-0.038	0.028	0.534	1.354
CholCov BPRV	0.090	0.049	0.041	0.020	0.110	0.073
CholCov MEDRV	0.015	0.008	0.003	0.009	0.018	0.010

Note: Bivariate simulation, with poisson process parameters $\lambda = \{10, 20\}$.

Results:10-Dimensional case

	Variances		Covariances	
	Bias	RMSE	Bias	RMSE
		$\sigma_J = 0$		
	Bias	RMSE	Bias	RMSE
RCov	0.058	0.061	-0.044	0.049
BPCov	0.042	0.049	-0.064	0.066
CholCov RV	0.003	0.039	-0.023	0.043
CholCov BPRV	-0.001	0.042	-0.026	0.045
CholCov MEDRV	-0.002	0.044	-0.032	0.049
		$\sigma_J = 1$		
RCov	0.601	1.679	-0.043	0.038
BPCov	0.122	0.080	-0.035	0.016
CholCov RV	0.551	1.495	-0.039	0.077
CholCov BPRV	0.126	0.101	-0.020	0.024
CholCov MEDRV	0.086	0.206	-0.022	0.028

Note: $\lambda = \{10, 20, \ldots, 100\}$.

Application: Data

- We consider 20 stocks with tickers AET, AFL, AIG, ALL, AXP, BAC, BBT, BK, COF, GS, HIG, MET, PNC, PRU, RF, STI, STT, UNH, USB, WFC.
- Observations last from January 3rd 2007 till December 21st 2012, for a total of 1499 observations.
- We cleaned the data.
- We used open-to-close returns obtained from the TAQ data

HEAVY

The HEAVY model of Noureldin, Shephard, and Sheppard, (2012):

$$\text{HEAVY: } H_{t|t-1} = (1-\alpha-\beta)\Omega + \alpha \widehat{CholCov}^{RV}{}_{t-1} + \beta H_{t-1|t-2},$$

where $\widehat{CholCov}^{RV}{}_{t-1}$ is the CholCov estimate using RV on pre-averaged returns. To reduce the number of parameters to be estimated, we apply covariance targeting, where Ω is the average Cov. The HEAVY model additionally has a correction term, to match the unconditional variance of the model to that of daily returns.

HEAVY-Jumps

$$\hat{J}_t = Chol\hat{Cov}^{RV}{}_t - Chol\hat{Cov}^{MedRV}{}_t.$$

HEAVY-Jumps

$\hat{J}_t = Chol\hat{C}ov^{RV}{}_t - Chol\hat{C}ov^{MedRV}{}_t$.

We incorporate the jump component in the HEAVY model as

(HEAVY-J1):

HEAVY-J1: $H_{t|t-1} = (1-\alpha-\beta)\Omega + \alpha Chol\hat{C}ov_{RV\,t-1} + \beta H_{t-1|t-2} + \gamma \hat{J}_{t-1}$

HEAVY-Jumps

$$\hat{J}_t = Chol\hat{C}ov^{RV}{}_t - Chol\hat{C}ov^{MedRV}{}_t.$$

We incorporate the jump component in the HEAVY model as

(HEAVY-J1):

HEAVY-J1: $H_{t|t-1} = (1-\alpha-\beta)\Omega + \alpha Chol\hat{C}ov_{RV\,t-1} + \beta H_{t-1|t-2} + \gamma \hat{J}_{t-1}$

HEAVY-J2: $J_t^{(ij)} = \begin{cases} F^{(ij)} & \text{if } F^{(ii)} > 0 \text{ and } F^{(jj)} > 0 \\ 0 & \text{otherwise.} \end{cases}$

In-sample Results

	α	β	γ	Likelihood
HEAVY	0.398	0.597		-25273
HEAVY-J1	0.638	0.766	-0.043	-25269
HEAVY-J2	0.614	0.786	-0.165	-25234