Comparison of high-frequency covolatility forecasts for high-dimensional portfolio allocation

with N. Hautsch

My visit at the University of Pennsylvania was sponsored by the Austrian Ministry for Digital and Economic Affairs and by the Slovak Ministry of education, science, research and sport.
Goal:
construction of global minimum variance portfolios for S&P 500.

Crucial input:
forecasts of high-dimensional covariance matrices.

Hautsch, Kyj and Malec (JAE, 2015):
→ H-F forecasts beat L-F in terms of portfolio volatility,
→ H-F methods lead to higher portfolio turnover,
→ H-F methods are worth up to 199 bp for risk-averse investor,
→ H-F methods perform better for moderate-size portfolios
  (e.g., 30-100 most traded stocks).

Next move:
would the efficient H-F approach further increase investor’s utility?

How to solve practical issues with H-F data . . . as the portfolio grows:
→ different sample sizes,
→ asynchronous observations . . . loss of information,
→ ill-conditioned/in-definite covariance matrix estimators,
→ lack of computational memory,
→ computation takes too long for short-term portfolio management.
Content

- Framework
  - Global minimum variance portfolio
  - Portfolio characteristics
  - Economic criterion

- Local method of moments estimator (LMME)
  - Spectral approach and efficiency
  - Clustering and regularization
  - Quantifying information loss

- Other high-frequency estimators
  - Regularization and blocking estimator (RnB)
  - Cholesky factorization estimator (CHOLCOV)

- Empirical comparison results
  - LOBSTER data
  - Portfolio characteristics
  - Economic criterion
Global minimum variance portfolio

Finding optimal weights at time $t$ for portfolio of $m$ stocks at time $t + 1$

$$\omega_{t+1}^* = \arg \min_{\omega \in \mathbb{R}^m} \omega^\top \Sigma_{t+1} \omega, \quad \text{s.t.} \quad \omega^\top l = 1,$$

we obtain the solution

$$\omega_{t+1}^* = \frac{\Xi_{t+1} l}{l^\top \Xi_{t+1} l}, \quad \text{with} \quad \Xi = \Sigma^{-1}.$$

Criterion for cov-mat forecasts, Patton and Sheppard (2008): pluggin $\hat{\Sigma}_{t+1|t} \neq \Sigma_{t+1}$, we get $\hat{\omega}_{t+1|t}$ and conditional portfolio variance

$$\hat{\omega}_{t+1|t}^\top \Sigma_{t+1} \hat{\omega}_{t+1|t} > \omega_{t+1}^* \Sigma_{t+1} \omega_{t+1}^*.$$

→ Focus on day-ahead prediction of cov-mat,
   ... we do not predict conditional mean,
→ evaluation depends on choice of ex-post measure for $\Sigma_{t+1}$,
   ... sample portfolio variance is biased criterion, see Voev (2009),
→ comparison based on
   ... conditional portfolio variance, turnover and shortselling,
   ... performance fees a risk-averse investor would be willing to pay.
Portfolio characteristics

→ Conditional portfolio variance

\[ \hat{\omega}_{t+1|t}^{\top} \text{MRK}_{t+1} \hat{\omega}_{t+1|t}, \]

→ portfolio turnover according to de Pooter, Martens and Dijk (2008),

\[ \sum_{l=1}^{m} \left| \hat{\omega}_{t+1|t}^l - \hat{\omega}_{t|t-1}^l \frac{1 + r_t^l}{1 + p r_t} \right|, \]

→ amount of short-selling

\[ \sum_{l=1}^{m} \omega_{t+1|t}^l \mathbb{I} \{ \omega_{t+1|t}^l < 0 \}. \]
Example: Assume investor invests 10$ into 2 stocks:

Why is the turnover from day $t$ to day $t + 1$ computed as

$$\sum_{l=1}^{m} |\hat{\omega}'_{t+1|t} - \hat{\omega}'_{t|t-1} \frac{1 + r'_{t}}{1 + p r_{t}}|$$

→ on day $t$ at 9:30 am, investor has/knows
- 10$ capital,
- optimal weights $\hat{\omega}_{t|t-1} = (0.6, 0.4)$ for day $t$ obtained on day $t - 1$ (after 4 pm),
- opening prices of the 2 stocks $(1.5$, 0.5$).

→ so investor has 4 shares of stock $a$ and 8 shares of stock $b$.

→ on day $t$ after 4 pm, investor knows/has
- closing prices of the 2 stocks are $(2$, 0.1$).
- $r_{t} = (\frac{1}{3}, -\frac{4}{5})$ and $p r_{t} = -\frac{3}{25}$.
- 8.8$ of capital in stocks.

→ on $t + 1$ just before rebalancing according to $\hat{\omega}_{t+1|t}$ (9:30 am sharp)
- prices of the 2 stocks are still the same $(2$, 0.1$).
- but the weights of stocks $a, b$ are not $(0.6, 0.4)$ but

$$(\frac{4}{5}, \frac{8}{9}) = (0.6, 0.4) \odot (\frac{1+\frac{1}{3}}{2}, \frac{1-\frac{4}{5}}{2})$$.
We assess economic significance of a lower (conditional) portfolio variance by the approach of Fleming, Kirby and Ostdiek (JoF, 2001).

→ If investor has quadratic utility

\[ U(\text{pr}_{t+1}) = 1 + \text{pr}_{t+1} - \frac{\gamma}{2(1 + \gamma)} (1 + \text{pr}_{t+1})^2, \]

→ \( \Delta_\gamma \) is defined by the equality

\[ \sum_{t=1}^{T-1} \mathbb{E} [U(\text{pr}_{t+1}^I)|\mathcal{F}_t] = \sum_{t=1}^{T-1} \mathbb{E} [U(\text{pr}_{t+1}^{II} - \Delta_\gamma)|\mathcal{F}_t]. \]

→ Interpret \( \Delta_\gamma \) as fee the investor is willing to pay in order to switch from allocation implied by \( \hat{\Sigma}_{t+1|t}^I \) to \( \hat{\Sigma}_{t+1|t}^{II} \).

→ assuming \( \mathbb{E} [r_{t+1}|\mathcal{F}_t] = \text{const.}, \Delta_\gamma > 0 \text{ iff } \overline{\text{pv}}_{t+1}^{2,I} > \overline{\text{pv}}_{t+1}^{2,II}. \)

\(^{1}\gamma = 1, 10\) as in Fleming et al. (2003).
Assume a continuous latent $d$-dimensional true log-price process:

$$y_t^* = y_0^* + \int_0^t \Sigma^{1/2}(s) \, db_s, \quad t \in [0, 1].$$

We are estimating integrated covolatility

$$ICov = \int_0^1 \Sigma(t) \, dt,$$

Observations are non-synchronous and polluted by iid noise:

$$y_{t_i}^{(l)} = y_{t_i}^{*(l)} + \varepsilon_i^{(l)}, \quad \varepsilon_i^{(l)} \sim N(0, \eta_i^2), \quad i = 1, \ldots, n_l, \quad l = 1, \ldots, m.$$

Efficient LMM estimator of Bibinger, Hautsch, Malec, Reiss (2014):

$\rightarrow$ partition the trading day into $K$ intervals of length $h = 1/K$,

$\rightarrow$ assume $\Sigma^k(t) = \Sigma^k$ and define $S_{jk} \sim N_m(0, \Sigma^k + \text{bias})$,

$\rightarrow$ compute $S_{jk}^{(l)} = \sqrt{2/h} \sum_{i=1}^{n_l} \Delta y_i^{(l)} \sin \left( j \pi h^{-1}(\bar{t}_i^{(l)} - kh) \right)$.

$$\text{LMME} = \sum_{k=1}^B h \sum_{j=1}^J \hat{W}_{jk} \text{vec}(S_{jk}^{\otimes 2} - \hat{\text{bias}}).$$

Local method of moments estimator (LMME)
Clustering and regularization

→ Computational memory and speed issues\(^2\)
   
   \[
   \hat{W}_{jk} = I_k^{-1} I_{jk} \in \mathbb{R}^{m^2 \times m^2}
   \]
   
   require \(J \times K \times 123 \text{ GB}\) for S&P 500,
   
   \[
   \hat{W}_{jk}
   \]
   
   can not be decomposed and estimated piecewise,
   
   \[
   \text{if } m > 50 \text{ LMME takes days to obtain (if feasible)}.
   \]

→ Application to portfolio allocation related issues
   
   \[
   \hat{\Sigma}
   \]
   
   has to be positive-definite and well-conditioned,
   
   \[
   \text{adjust LMME to be feasible and well-conditioned}.
   \]

→ Divide and conquer by Hsieh et al. (2012) and Tan et al. (2015):
   
   \[
   \text{Hierarchical correlation clustering & regularization}
   \]

→ GLASSO estimator for the \(\Sigma^{-1}\) denoted \(\Xi\):

\[
\hat{\Xi}_\lambda = \arg\min_{\Xi > 0} \left( \text{Tr}(\hat{\Xi} \hat{R}) - \log \det(\Xi) + \|L \circ (\Xi - \text{diag}(\Xi))\|_1 \right),
\]

→ SEC estimator by Cui, Leng and Yu (CSDA, 2016):

\[
\hat{R}_\lambda = \arg\min_{R - \epsilon E > 0} \frac{1}{2} \left\| R - \hat{R} \right\|_F^2 + \|L \circ R\|_1,
\]

\[
L_{ij} = \lambda > 0 \text{ should assets } i \text{ and } j \text{ be in the same cluster or } \lambda \to \infty.
\]

\(^2\)LMME impl. in Matlab is feasible for \(m \leq 20\) on Mac, \(m \leq 150\) on VSC 256 GB.
GLASSO

Local method of moments estimator (LMME)
Local method of moments estimator (LMME)
GLASSO - average absolute correlation

Local method of moments estimator (LMME)
SEC - average absolute correlation

Local method of moments estimator (LMME)
GLASSO - conditional numbers

Local method of moments estimator (LMME)
Local method of moments estimator (LMME)
Divide and conquer step-by-step

On every intra-day interval $1, \ldots, K$:

→ Estimate PILOT spectral estimator, Bibinger and Reiss (2014),
→ use the scaled estimator as adjacency matrix $1 - |R|$ for HCC,
... Step 0, each asset is a cluster, ... Steps $1, \ldots, m - 1$ merge two nearest (linkage) clusters,
... search dendrogram top-down and find “height” $\lambda$, such that each cluster $c = 1, \ldots, C$ has $\leq 50$ assets,
→ GLASSO/SEC on each cluster of assets with parameter $\lambda$.
→ with $\hat{\Xi}_\lambda^{k,c}$ or $\hat{R}_\lambda^{k,c}$, compute local-cluster-LMME (now feasible),
→ put all clusters back together & replace 0’s by respective PILOT entries,

\[
\text{LMME-GLASSO} = \sum_{k=1}^{B} h \cdot (\hat{\Xi}_\lambda^{k})^{-1}
\]

\[
\text{LMME-SEC} = \sum_{k=1}^{B} h \cdot (\hat{\Sigma}_\lambda^{k})^{-1}
\]
<table>
<thead>
<tr>
<th>Linkage</th>
<th>Description</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>$\frac{1}{|c_1|_0|c_2|<em>0} \sum</em>{c_1 \in C_1, c_2 \in C_2} \text{corr}(c_1, c_2)$</td>
<td>Tan et al. (2015)</td>
</tr>
<tr>
<td>Complete</td>
<td>$\max {\text{corr}(c_1, c_2); c_1 \in C_1, c_2 \in C_2}$</td>
<td>Tumminello et al. (2010)</td>
</tr>
<tr>
<td>Single</td>
<td>$\min {\text{corr}(c_1, c_2); c_1 \in C_1, c_2 \in C_2}$</td>
<td>leads to trailing clusters</td>
</tr>
</tbody>
</table>

Local method of moments estimator (LMME)
Clustering and regularization lead to loss of information:
→ information distance of GLASSO and SEC from efficient LMME,
→ Tumminello et al. (Phys. Rev. 2007) use Kulback-Leibler distance:

\[ K(R_1, R_2) = \frac{1}{2} \left( \log \left( \frac{\det R_2}{\det R_1} \right) + \text{tr} \left( R_2^{-1} R_1 \right) - m \right). \]

and showed that

\[ E \left[ K(\hat{R}, R) \right] = \frac{1}{2} \left( m \log \left( \frac{n}{2} \right) - \sum_{t=n-m+1}^{n} \psi(t/2) \right). \]

<table>
<thead>
<tr>
<th>Data</th>
<th>cluster</th>
<th>glasso</th>
<th>sec</th>
<th>identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP 100</td>
<td>21.23</td>
<td>12.21</td>
<td>30.75</td>
<td>65.12</td>
</tr>
<tr>
<td>SP 500</td>
<td>3.13</td>
<td>2.76</td>
<td>6.07</td>
<td>6.04</td>
</tr>
</tbody>
</table>
start: multivariate realized kernel estimator proposed by Barndorf-Nielsen et al. (2011),

\[ \text{MRK} = \sum_{h=-H}^{H} k \left( \frac{h}{H+1} \right) \Gamma, \quad \text{with bandwidth } H, \]

... refresh-time sampling for synchronization of prices,
... sample if all assets have been (re-)traded at least once,
... large loss of data - frequency is driven by least liquid asset,
... it gets worse as portfolio grows (heterogeneity).

Hautsch et al. (2012): blocking (clustering) based on liquidity,
... set fixed number of liquidity groups (e.g., 4), apply refresh-time on each group,
... estimate \( G(G + 1)/2 \) blocks by MRK,
... put blocks back together in hierarchical order.

Need regularization (beside smoothing)
... Laloux et al. (1999) propose eigenvalue cleaning.
... idea: eigenvalues which are close to 0 are noisy and can be inflated without loosing information.
RnB - average absolute correlation

Other high-frequency estimators
RnB - conditional numbers

Other high-frequency estimators
Boudt et al. (2017): reorder assets according to their liquidity,
... estimate $\Sigma$ hierarchically element-by-element,
... hierarchical Cholesky factorization $\Sigma_k = BVB^\top$,

$$B = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ b_{21} & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \ldots & 1 \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} v_{11} & 0 & \ldots & 0 \\ 0 & v_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & v_{mm} \end{bmatrix}.$$  

... factorization of the returns

$$f_{\tau_i} = B^{-1} x_{\tau_i} \sim N(0, (\tau_i - \tau_{i-1}) V),$$  

... where $\tau_1, \ldots, \tau_n$ are synchronized observation times,
... convenient hierarchical regression representation

$$x_{l,\tau} = f_{l,\tau} b_{l,1} + \ldots, + f_{l-1,\tau} b_{l,l-1} + f_{l,\tau}$$  

... regression coefficients $b_{ij} = E[r_i, f_j]/E[f_j^2], j < i$ and $v_{ii} = E[f_i^2],$

$\rightarrow$ We: uni/bi- local spectral estimator of Bibinger and Reiss (2014).
CHOLCOV - average absolute correlation

Other high-frequency estimators
→ mid-quote data on 100 and 426 most liquid assets of S&P 500,
... sample period 1/1/2012 – 31/12/2015, i.e. 994 trading days,
... raw quotes cleaned as in Barndorf-Nielsen et al. (2009),
... quotes revised for 0 returns, reduction of comp. burden.
... e.g., for AAPL on 1/Jan/2012: 104 711 → 39 843, i.e. 62%,
... the least liquid asset on this day was XRX with only 436
observations after revisions,
→ compute PILOT, GLASSO, SEC, RnB and CHOLCOV for all
\( t = 1 \ldots, 993 \), ... estimators are smoothed across 1, 2, 5, 20 days,
... day \( t \) estimates are forecasts for \( t + 1 \)
→ performance depends on choice of nuisance parameters,
... for spectral approach, \( B, J \) according to Bibinger et al. (2018),
... for RnB, \( G, H \) and others according to Hautsch et al. (2015),
... for clustering and regularization: number of clusters, size,
... arbitrary choice: smoothing after regularization, MRK as
ex-post approximation.
Empirical comparison results

**Portfolio characteristics**

- **PV**: A) SP 100 - PV, B) SP 500 PT C) SP 100, D) SP 500
**Economic criterion**

CHOLCOV: A) SP 100 B) SP 500 RV5min: C) SP 100 D) SP 500

Empirical comparison results
Economic criterion

CHOLCOV: A) SP 100  B) SP 500  PILOT: C) SP 100  D) SP 500

Empirical comparison results
Thank you!