High-Dimensional Copula Models with Mixed Frequency Data and Asymmetric Volatility

Kaiji Motegi\textsuperscript{1}  Xiaojing Cai\textsuperscript{1}  Shigeyuki Hamori\textsuperscript{1}  Haifeng Xu\textsuperscript{2}

\textsuperscript{1}Kobe University  \textsuperscript{2}Xiamen University

88th Annual Meeting of Southern Economic Association
Marriott Marquis Washington, DC

November 18, 2018
There are several well-known challenges on modelling serial dependence of asset returns:

1. A large number of assets (i.e. high dimensionality).
2. Nonlinear dependence among asset returns.
3. Seasonality.

Oh and Patton (2016) propose a novel approach that addresses those challenges by combining heterogeneous autoregressive (HAR) models and jointly symmetric copula (JSC) models.

They analyze daily returns of 104 stocks that were ever constituents of S&P 100 Index, using realized variances and correlations based on intraday returns.
Suppose that there are $N$ sectors (or shares) and $T$ days.

Let $r_t = [r_{1t}, \ldots, r_{NT}]^\top$ be a vector of target daily returns.

Suppose that

$$r_t = H_t(\beta)^{1/2} \times e_t,$$

where

$$e_t \overset{i.i.d.}{\sim} F(\delta), \quad \mathbb{E}[e_t] = 0, \quad \mathbb{E}[e_te_t^\top] = I.$$

How should we model $H_t(\beta)$ and $F(\delta)$, and how should we estimate $(\beta, \delta)$ when $N$ is relatively large?
Oh and Patton’s (2016) procedure is as follows:

1. Construct $H_t$ based on realized variances and correlations.
2. Fit a HAR model to each of the diagonal elements of $H_t$.
3. Fit a HAR model to the off-diagonal elements of $H_t$ via pooled OLS.
4. Fit a JSC model to $\hat{e}_t = H_t(\hat{\beta})^{-\frac{1}{2}} \times r_t$ in order to allow for nonlinear dependence among $\{\hat{e}_{1t}, \ldots, \hat{e}_{Nt}\}$ while ensuring $\mathbb{E}[e_t e_t^\top] = I$.

- An issue is that asymmetry in variances, correlations, and standardized errors cannot be captured by HAR or JSC models.
- We address that issue by adding threshold terms to the HAR models.
Let $y_{it} = \ln H_t(i, i)$ be the log realized variance of sector $i$ at time $t$.

Oh and Patton (2016) fit a standard HAR model for each sector (subscript $i$ is omitted):

\[
y_t = \beta^{(c)} + \beta^{(d)} \times y_{t-1}
+ \beta^{(w)} \times \frac{1}{5} \sum_{k=1}^{5} y_{t-k}
+ \beta^{(m)} \times \frac{1}{20} \sum_{k=1}^{20} y_{t-k} + \xi_t.
\]

$\beta^{(d)}$, $\beta^{(w)}$, and $\beta^{(m)}$ measure the persistence of log realized variance at daily, weekly, and monthly levels, respectively.
We propose to add **threshold terms** to each of the daily, weekly, and monthly levels:

\[
y_t = \beta^{(c)} + \beta^{(d)} \times y_{t-1} + \psi^{(d)} I \left( y_{t-1} \geq \hat{\mu}_{t-1}^{(d)} \right) \times y_{t-1} \\
+ \beta^{(w)} \times \left( \frac{1}{5} \sum_{k=1}^{5} y_{t-k} + \psi^{(w)} I \left( \frac{1}{5} \sum_{k=1}^{5} y_{t-k} \geq \hat{\mu}_{t-1}^{(w)} \right) \right) \times \frac{1}{5} \sum_{k=1}^{5} y_{t-k} \\
+ \beta^{(m)} \times \left( \frac{1}{20} \sum_{k=1}^{20} y_{t-k} + \psi^{(m)} I \left( \frac{1}{20} \sum_{k=1}^{20} y_{t-k} \geq \hat{\mu}_{t-1}^{(m)} \right) \right) \times \frac{1}{20} \sum_{k=1}^{20} y_{t-k} + \xi_t,
\]

where

\[
\hat{\mu}_{t}^{(d)} = \frac{1}{\ell + 1} \sum_{\tau=t-\ell}^{t} y_{\tau}, \quad \hat{\mu}_{t}^{(w)} = \frac{1}{\ell + 1} \sum_{\tau=t-\ell}^{t} \left( \frac{1}{5} \sum_{k=1}^{5} y_{\tau-k} \right), \\
\hat{\mu}_{t}^{(m)} = \frac{1}{\ell + 1} \sum_{\tau=t-\ell}^{t} \left( \frac{1}{20} \sum_{k=1}^{20} y_{\tau-k} \right).
\]
Methodology: HAR Model with Thresholds

- We call our proposed model **moving average threshold HAR (MAT-HAR) model.**
- Key features of the MAT-HAR model are as follows.
  - The thresholds are time-varying and easy to interpret.
  - The thresholds do not contain any parameter to estimate (though we need to pick lag length \( \ell \)). Hence **OLS is feasible.**
  - We can include a subset of the three threshold terms based on AIC, out-of-sample forecast MSE, etc.

<table>
<thead>
<tr>
<th></th>
<th>( M_1 )</th>
<th>( M_2 )</th>
<th>( M_3 )</th>
<th>( M_4 )</th>
<th>( M_5 )</th>
<th>( M_6 )</th>
<th>( M_7 )</th>
<th>( M_8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Week</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Month</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>
Methodology: HAR Model with Thresholds

- Let $\rho_{ijt}$ be the realized correlation between sectors $i$ and $j$ at time $t$, and center it as $\tilde{\rho}_{ijt} = \rho_{ijt} - (1/T) \sum_{t=1}^{T} \rho_{ijt}$.

- Oh and Patton (2016) fit a standard HAR model with pooled OLS:

$$\tilde{\rho}_{ijt} = b(d) \times \tilde{\rho}_{ij,t-1} + b(w) \times \frac{1}{5} \sum_{k=1}^{5} \tilde{\rho}_{ij,t-k}$$

$$+ b(m) \times \frac{1}{20} \sum_{k=1}^{20} \tilde{\rho}_{ij,t-k} + \xi_{ijt}, \quad i > j.$$
We propose to add threshold terms to each of the daily, weekly, and monthly levels:

\[
\tilde{\rho}_{ijt} = b^{(d)} \times \tilde{\rho}_{ij,t-1} + c^{(d)} I(\tilde{\rho}_{ij,t-1} \geq \tilde{\mu}_{ij,t-1}^{(d)}) \times \tilde{\rho}_{ij,t-1} \\
+ b^{(w)} \times \frac{1}{5} \sum_{k=1}^{5} \tilde{\rho}_{ij,t-k} + c^{(w)} I \left( \frac{1}{5} \sum_{k=1}^{5} \tilde{\rho}_{ij,t-k} \geq \tilde{\mu}_{ij,t-1}^{(w)} \right) \times \frac{1}{5} \sum_{k=1}^{5} \tilde{\rho}_{ij,t-k} \\
+ b^{(m)} \times \frac{1}{20} \sum_{k=1}^{20} \tilde{\rho}_{ij,t-k} + c^{(m)} I \left( \frac{1}{20} \sum_{k=1}^{20} \tilde{\rho}_{ij,t-k} \geq \tilde{\mu}_{ij,t-1}^{(m)} \right) \times \frac{1}{20} \sum_{k=1}^{20} \tilde{\rho}_{ij,t-k} + \xi_{ijt},
\]

where

\[
\tilde{\mu}_{ij,t}^{(d)} = \frac{1}{\ell + 1} \sum_{\tau=t-\ell}^{t} \tilde{\rho}_{ij,\tau}, \quad \tilde{\mu}_{ij,t}^{(w)} = \frac{1}{\ell + 1} \sum_{\tau=t-\ell}^{t} \left( \frac{1}{5} \sum_{k=1}^{5} \tilde{\rho}_{i,\tau-k} \right), \\
\tilde{\mu}_{ij,t}^{(m)} = \frac{1}{\ell + 1} \sum_{\tau=t-\ell}^{t} \left( \frac{1}{20} \sum_{k=1}^{20} \tilde{\rho}_{i,\tau-k} \right).
\]
Data

- We analyze CSI 300 Sector Indices: five-minute stock price indices of $N = 10$ sectors in China.


  - **Subsample 2**: period of tranquility (Jan. 4, 2010 - Dec. 31, 2014; 1212 days).


- For each day, we have $d = 49$ five-minute observations spanning 9:30am-11:30am and 1:05pm-3:00pm.
Data

**Figure**: Five-minute data of Sector #1 (Consumer Discretionary)

![Price](image1.png)  
![Log return](image2.png)

Motegi, Cai, Hamori & Xu (Kobe & Xiamen)  
High-Dimensional Copula Models  
November 18, 2018
Data

Figure: Log-RV and moving average threshold (#1. Cons. Disc.)

For the daily level, the blue line depicts $y_{1,t-1} = \ln RV_{ar_{1,t-1}}$ while the red line depicts $\hat{\mu}^{(d)}_{1,t-1} = (\ell + 1)^{-1} \sum_{\tau=t-\ell}^{t-1-\ell} y_{1,\tau}$. We include $\ell = 120$ days of lags (roughly half a year).

The weekly and monthly levels are treated analogously.

Figures on Sectors #2–#10 are similar and hence omitted.
For the daily level, the blue line depicts $\tilde{\rho}_{12,t-1}$ while the red line depicts $\hat{\mu}_{12,t-1}^{(d)} = (\ell + 1)^{-1} \sum_{\tau=t-1-\ell}^{t-1} \tilde{\rho}_{12,\tau}$.

We include $\ell = 120$ days of lags (roughly half a year).

The weekly and monthly levels are treated analogously.

Figures on other pairs of sectors are similar and hence omitted.
Empirical Results

**Table**: Optimal HAR models for log-RV based on AIC

<table>
<thead>
<tr>
<th></th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
<th>#7</th>
<th>#8</th>
<th>#9</th>
<th>#10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>$M_4$</td>
<td>$M_6$</td>
<td>$M_6$</td>
<td>$M_4$</td>
<td>$M_2$</td>
<td>$M_3$</td>
<td>$M_2$</td>
<td>$M_4$</td>
<td>$M_1$</td>
<td>$M_2$</td>
</tr>
<tr>
<td>Subprime</td>
<td>$M_6$</td>
<td>$M_1$</td>
<td>$M_1$</td>
<td>$M_2$</td>
<td>$M_1$</td>
<td>$M_1$</td>
<td>$M_2$</td>
<td>$M_3$</td>
<td>$M_1$</td>
<td>$M_1$</td>
</tr>
<tr>
<td>Tranquility</td>
<td>$M_1$</td>
<td>$M_1$</td>
<td>$M_4$</td>
<td>$M_2$</td>
<td>$M_2$</td>
<td>$M_2$</td>
<td>$M_6$</td>
<td>$M_1$</td>
<td>$M_1$</td>
<td>$M_2$</td>
</tr>
<tr>
<td>B &amp; B</td>
<td>$M_4$</td>
<td>$M_1$</td>
<td>$M_8$</td>
<td>$M_2$</td>
<td>$M_4$</td>
<td>$M_4$</td>
<td>$M_3$</td>
<td>$M_2$</td>
<td>$M_5$</td>
<td>$M_2$</td>
</tr>
</tbody>
</table>

**Table**: The number of cases out of 40 where each model is chosen

<table>
<thead>
<tr>
<th></th>
<th>$M_1$</th>
<th>$M_2$</th>
<th>$M_3$</th>
<th>$M_4$</th>
<th>$M_5$</th>
<th>$M_6$</th>
<th>$M_7$</th>
<th>$M_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>12</td>
<td>12</td>
<td>3</td>
<td>7</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Empirical Results

Table: Results of optimal HAR models for log-RV (full sample)

<table>
<thead>
<tr>
<th></th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_{AIC}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_c$</td>
<td>$-0.423^*$</td>
<td>$-0.466^*$</td>
<td>$-0.396^*$</td>
<td>$-0.341^*$</td>
<td>$-0.413^*$</td>
</tr>
<tr>
<td>$\beta_d$</td>
<td>$0.299^*$</td>
<td>$0.304^*$</td>
<td>$0.289^*$</td>
<td>$0.278^*$</td>
<td>$0.368^*$</td>
</tr>
<tr>
<td>$\psi_d$</td>
<td>-</td>
<td>0.006</td>
<td>0.010*</td>
<td>-</td>
<td>0.012*</td>
</tr>
<tr>
<td>$\beta_w$</td>
<td>$0.410^*$</td>
<td>$0.439^*$</td>
<td>$0.494^*$</td>
<td>$0.444^*$</td>
<td>$0.382^*$</td>
</tr>
<tr>
<td>$\psi_w$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\beta_m$</td>
<td>$0.239^*$</td>
<td>$0.198^*$</td>
<td>$0.163^*$</td>
<td>$0.235^*$</td>
<td>$0.198^*$</td>
</tr>
<tr>
<td>$\psi_m$</td>
<td>$0.008^*$</td>
<td>$0.008^*$</td>
<td>0.006</td>
<td>0.008*</td>
<td>-</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.629</td>
<td>0.587</td>
<td>0.621</td>
<td>0.663</td>
<td>0.665</td>
</tr>
<tr>
<td>$P(\psi = 0)$</td>
<td>0.017</td>
<td>0.007</td>
<td>0.003</td>
<td>0.022</td>
<td>0.002</td>
</tr>
</tbody>
</table>
### Empirical Results

**Table**: Results of optimal HAR models for realized correl. based on AIC

<table>
<thead>
<tr>
<th></th>
<th>$M^*$</th>
<th>$b^{(d)}$</th>
<th>$c^{(d)}$</th>
<th>$b^{(w)}$</th>
<th>$c^{(w)}$</th>
<th>$b^{(m)}$</th>
<th>$c^{(m)}$</th>
<th>$R^2$</th>
<th>$P(H_0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>$M_5$</td>
<td>0.123*</td>
<td>0.054*</td>
<td>0.295*</td>
<td>0.021*</td>
<td>0.386*</td>
<td>-</td>
<td>0.279</td>
<td>0.000</td>
</tr>
<tr>
<td>Sub.</td>
<td>$M_8$</td>
<td>0.149*</td>
<td>0.071*</td>
<td>0.383*</td>
<td>$-0.083^*$</td>
<td>0.193*</td>
<td>0.048*</td>
<td>0.208</td>
<td>0.000</td>
</tr>
<tr>
<td>Tranq.</td>
<td>$M_6$</td>
<td>0.093*</td>
<td>0.018*</td>
<td>0.274*</td>
<td>-</td>
<td>0.389*</td>
<td>0.028*</td>
<td>0.182</td>
<td>0.000</td>
</tr>
<tr>
<td>B &amp; B</td>
<td>$M_2$</td>
<td>0.201*</td>
<td>0.066*</td>
<td>0.265*</td>
<td>-</td>
<td>0.306*</td>
<td>-</td>
<td>0.264</td>
<td>0.000</td>
</tr>
</tbody>
</table>

- **Wald test** with respect to $H_0 : c^{(d)} = c^{(w)} = c^{(m)} = 0$ leads to a rejection in each sample period, indicating the existence of threshold effects.
HAR models are a standard tool of analyzing realized variances and correlations (e.g. Oh and Patton, 2016).

We extend HAR by adding time-varying, parameter-free threshold terms at the daily, weekly, and monthly levels.

The threshold is the moving average of previous realizations of the target variable, hence it can be computed from data.

We apply the proposed model to intraday stock price indices of Chinese sectors.

Our empirical results indicate that threshold effects exist in both the realized variances and the realized correlations.

The realized variances and correlations have greater persistence when they are above their moving averages.
References