Detecting Correlation Structure of Stock Returns by Network Clustering

Takashi Isogai (Bank of Japan, JAIST) E-mail: takashi.isogai@boj.or.jp

**Motivation:** The correlation of returns is a key concept for portfolio risk management. We propose an approach to detect the high dimensional correlation structure of market-wide stock returns by network clustering for reducing portfolio risk (VaR, Expected Shortfall) with effective diversification of investment.

**Method:** GARCH filtering and a hierarchical network clustering with modularity maximization is combined to deal with the fat-tailness of return distribution and the high dimensional correlation structure.

**Findings:** Hierarchical groups are detected. The current sector classification is partially effective; stocks in some sectors are grouped almost together. The group properties are identified by classification tree analysis.

1. Correlation of fat-tail stock returns

Portfolio risk = Variance (\(r\)) + Covariance (\(r\))

Stock return feature: Fat tail and volatility clustering

Volatility of Nikkei by GARCH(1,1)

\[ r_t = \sigma_z z_t \quad z_t \sim IID(0,1) \]

- Focus on covariance of \(z_t\), rather than \(r_t\)
- GARCH filtering to separate volatilities \(\sigma_t\) and i.i.d innovations \(z_t\) with correlation matrix \(P\)

Multivariate GARCH for stock returns \(i_t\)

- for over 1400 Stocks, listed at Tokyo Stock Exchange, 1st section

\[
\begin{bmatrix}
\sigma_{i,t}^2 \\
0 & \sigma_{j,t}^2
\end{bmatrix}
\begin{bmatrix}
z_{i,t} \\
z_{j,t}
\end{bmatrix} = \begin{bmatrix}
\sigma_{i,j,t} \\
0
\end{bmatrix} \Rightarrow Z_t \text{ has correlation } P_t
\]

- Hard to estimate parameters due to high dimensionality
- Exclude cross effects (no volatility spillovers, diagonal specifications)

**Vector GARCH(1,1) volatility equations:**

\[
\begin{align*}
\sigma_{i,t}^2 &= \alpha_0 + \alpha_1 r_{i,t-1}^2 + \gamma_1 \sigma_{i,t-1}^2 + \beta_1 \sigma_{i,t-1}^2 \\
\sigma_{j,t}^2 &= \alpha_2 + \beta_2 \sigma_{j,t-1}^2 + \beta_3 \sigma_{j,t-1}^2
\end{align*}
\]

- Simplify time varying \(P_t\) as constant over time \(P\)
- CCC-GARCH (Bollerslev(1990)) cc: constant conditional correlation

2. Clustering stock returns

**Current 33 sector classification: the best grouping?**

Heat map of correlation matrix of stock returns: corr(\(r_t\), \(r_j\))

- Find more data-oriented grouping using correlation matrix \(P\) = \(\{P_i\}\)
- Correlation matrix \(P_i\) = adjacent matrix \(A_i\) = Network clustering
  - Divise, hierarchical
  - Modularity maximization + spectral clustering (Newman(2006))
- Resolution limit problem… work around by recursive clustering; simple but it works.

3. Building hierarchical group structure

**Macro view of clustering**

**Understanding group properties**
- how stocks are divided into groups at each level; which factor plays a key role in determining the subdivision at every layer.
- shared properties of a group reflect investors’ views of those stocks.
- merits: forecast group ID of newly listed stocks, stocks with limited price data, etc.

**Classification tree at every hierarchy**
- TOPX beta, size, PBR (Fama and French), %/rate beta, overseas sales, sectors, ...
- TOPX beta, size, PBR (Fama and French), %/rate beta, overseas sales, sectors, ...
- TOPX beta, size, PBR (Fama and French), %/rate beta, overseas sales, sectors, ...

** Conditional sequential tree model for stock classification**
- e.g., B1 – G27: L1 Tree * L2 Tree * L3 Tree * L4 tree
- merits: forecast group ID of newly listed stocks, stocks with limited price data, etc.

4. Further topics

- Quantify risk (VaR, ES) reduction effects by random portfolio simulation
- Multivariate GARCH with non-diagonal specification
- Market-wide analysis by reduced size of GARCH models
- Volatility spillovers and dynamic conditional correlation (DCC GARCH) ...

**References**


Views expressed here are those of the author and do not necessarily reflect the official views of the Bank of Japan.