Leture 6: Multivariate Volatility Modelling - Part III - PCA and Orthogonal GARCH

FM321: Risk Management and Modelling

Linyan Zhu

1 November 2022

LSE Finance

Review

- Previously, we wrote down several models for Σ_t

  - EWMA: $\Sigma_t = (1-\lambda) r_{t-1} r'_{t-1} + \lambda \Sigma_{t-1}$
  - BEKK: $\Sigma_t = \Omega'\Omega + A r_{t-1} r'_{t-1} A' + B \Sigma_{t-1} B'$

  - Correlation models: $\Sigma_t = D_t C_t D_t$, where $D_t = diag\{\sigma_{1,t}, ..., \sigma_{N,t}\}$ and C_t correlation.

    - Constant conditional correlation models (CCC): $C_t = C$
    - Dynamic conditional correlation models (DCC)

- What benefit does CCC or DCC give us relative to BEKK?

- Why do we use $\hat{q}_t$ instead of r_t to estimate C_t?

- What is a common issue with the models above?

The curse of dimensionality

- Managers deal with large universes.
- MSCI World All-Country Index has over 2400 stocks.
- Around 2,800 stocks trade in the NYSE.
- Financial institutions can have tens of thousands of assets, if not more.
- None of the models we have discussed so far can be scaled to that number of securities.

The idea of factor models

- Correlations among asset returns arise from a small number of common sources of risk (called risk factors).

$$r_t = \beta f_t + \epsilon_t$$

where

- r is an $N \times 1$ vector of (excess) returns.
- β is an $N \times K$ matrix of factor exposures (containing the exposure of each security to each factor).

- f is an $K \times 1$ vector of factor returns.
- ϵ is an $N \times 1$ vector of residual returns for the securities.

The idea of factor models

- If we assume $cov(f_t, \epsilon_t) = 0$, we have

$$\Sigma_t = Var_{t-1}(r_t) = \beta \Sigma_{f,t} \beta' + \Sigma_{\epsilon,t}$$

- Then we can just model $\Sigma_{f,t}$ and $\Sigma_{\epsilon,t}$.

- The parameters to estimate are:
  - The $N \times K$ factor exposures in β (can be constant or time-varying);
  - The $\frac{1}{2}K(K+1)$ elements of $\Sigma_{f,t}$; and
  - the N non-zero elements of $\Sigma_{\epsilon,t}$.

- The total number of parameters is $N(K+1) + \frac{1}{2}K(K+1)$, which is linear in N.

How to find the risk factors f_t?

- Examples motivated by economic or finance theory:
  - Market risk
  - Country risk
  - Sector risk
  - Macroeconomic risk factors
  - Risk factors derived from anomalies (value, size, momentum)
- We are going to discuss one **statistical** way of constructing risk factors, i.e. principal component analysis (PCA)

Principal component analysis

a method of dimensional reduction

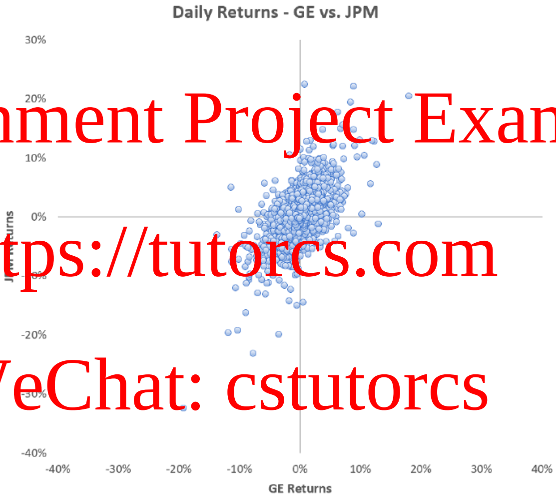

Figure: Scatterplot of daily returns of GE and JPM stocks

- By plotting the returns on this coordinate, a natural basis to express the data is a set of two directions given by vectors $\{(1,0),(0,1)\}$.

$$r_t = \begin{pmatrix} r_{1,t} \\ r_{2,t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} r_{1,t} \\ r_{2,t} \end{pmatrix}$$

- The naive basis reflects the way we gathered the data, but it may fail to uncover the simpler structure that underlie the data.

- Is there another basis, which is a linear combination of the naive basis, that best re-expresses the data?

The idea of PCA

- Let's define a new direction by vector (w_1, w_2) which yields

$$\tilde{r}_t = \begin{pmatrix} w_1 & w_2 \end{pmatrix} \begin{pmatrix} r_{1,t} \\ r_{2,t} \end{pmatrix} \quad t = 1, \ldots, T$$

as a projection of r_t on to that direction.

- Let's define a new direction by vector (w_1, w_2) which yields

$$\tilde{r}_t = \begin{pmatrix} w_1 & w_2 \end{pmatrix} \begin{pmatrix} r_{1,t} \\ r_{2,t} \end{pmatrix} \quad t = 1, \ldots, T$$

as a projection of r_t on to that direction.

- What do we mean by "best re-express"?

The idea of PCA

- Let's define a new direction by vector (w_1, w_2) which yields

$$\tilde{r}_t = \begin{pmatrix} w_1 & w_2 \end{pmatrix} \begin{pmatrix} r_{1,t} \\ r_{2,t} \end{pmatrix} \quad t = 1, \ldots, T$$

  as a projection of r_t on to that direction.

- What do we mean by "best re-express"?

- We assume that the direction with largest variance of data contains the most interesting dynamics.

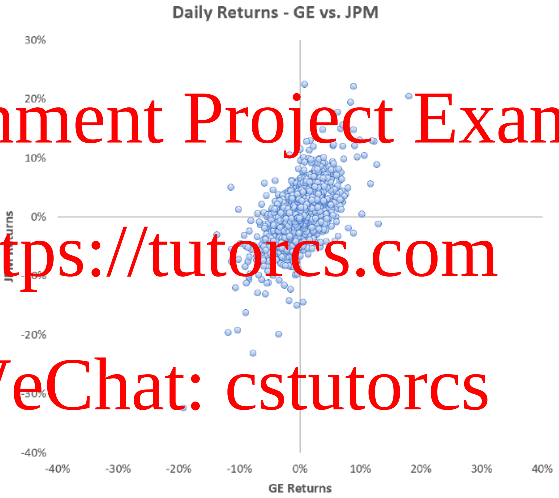

Figure: Scatterplot of daily returns of GE and JPM stocks

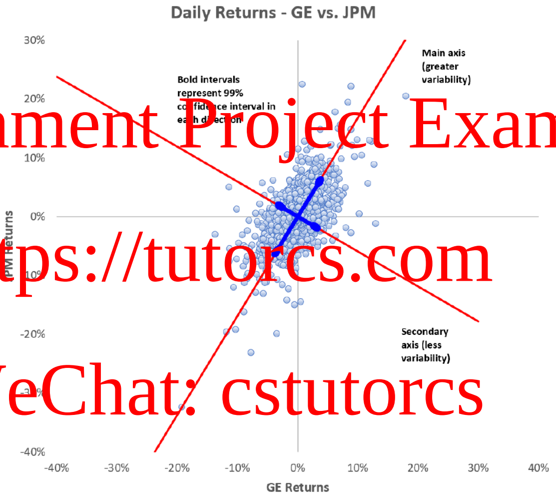

Figure: Scatterplot of daily returns of GE and JPM stocks

- More generally, we have N assets.

$$r = \begin{pmatrix} r_{1,1} & \dots & r_{1,T} \\ \dots & \dots & \dots \\ r_{N,1} & \dots & r_{N,T} \end{pmatrix} = \begin{pmatrix} r_1 & \dots & r_T \end{pmatrix}$$

- First look for a direction in the N-dimensional space denoted with an N-by-1 vector p_1 so that data has the largest variance in that direction.

- Find another direction along which variance is maximized, however, restrict the search to all directions perpendicular to all previous selected directions. Save this vector as p_i.

- Repeat this procedure until K vectors ($K \leq N$) are selected.

- Put all these vectors into a matrix P

$$P = \begin{pmatrix} p_1 \\ \vdots \\ p_K \end{pmatrix} = \begin{pmatrix} p_{1,1} & \cdots & p_{1,N} \\ \vdots & \ddots & \vdots \\ p_{K,1} & \cdots & p_{K,N} \end{pmatrix}$$

P is orthonormal because $p_i p_j' = 0$ if $i \neq j$ and $p_i p_i' = 1$.

- P transforms original return data into

$$\begin{pmatrix} \tilde{r}_{1,t} \\ \vdots \\ \tilde{r}_{K,t} \end{pmatrix} = \begin{pmatrix} p_{1,1} & \cdots & p_{1,N} \\ \vdots & \ddots & \vdots \\ p_{K,1} & \cdots & p_{K,N} \end{pmatrix} \begin{pmatrix} r_{1,t} \\ \vdots \\ r_{N,t} \end{pmatrix}$$

so that $\tilde{r}_t$ has a diagonal variance-covariance matrix and its diagonal elements decline in value from the top-left to the bottom-right.

- PCA can be implemented simply by eigen decomposition.

- It turns out that to diagonalize $Var(\tilde{r}_t)$, we can simply set p_i to be the eigenvector of $Var(r_t)$ associated with its i-th largest eigenvalue.

- $p_1, \ldots, p_K$ are called the principal components.

- The variance associated with p_i, or the i-th eigenvalue, quantifies how important the direction p_i is for capturing the dynamics of data.

- Consider a portfolio with four stocks: C, AAPL, MSFT, JPM.

Principal Component	1	2	3	4
C	0.536	0.471	0.151	0.684
AAPL	0.414	-0.690	0.593	0.019
MSFT	0.482	-0.384	-0.785	0.060
JPM	0.556	0.393	0.093	-0.726
Pct. Total Var.	0.542	0.221	0.148	0.089

- What does each column mean?

- What does Pct. Total Var. mean?

- From the point of view of risk modeling, the procedure has a few benefits from a computational and conceptual perspective:

  - As mentioned above, it makes it possible to identify factor structures on the basis of historical data alone, so the informational requirements are low.

  - PCA achieves dimensional reduction and can greatly simplify many high-dimensional problems.

  - For example for multivariate volatility modelling, the principal components (factors in the model) are portfolios of the underlying securities, so their variance can be readily computed.
  - The factors are orthogonal by construction, so their covariance is zero, simplifying some of the estimation of covariance matrix.

Challenges associated with PCA

- One needs to decide how many principal components to include in the factor model, which is usually done on the basis of:
  - Relevance of the factor (does it capture meaningful amount of common variance across securities?)
  - Economic interpretation (does it seem to reflect a recognizable reason of an economic or financial nature that would cause comovement across securities?)
- While there are statistical techniques to help with this decision, there are no definite rules, and judgement often plays a part in that process.
- PCA requires large data, i.e. a large number of securities, relative to the number of factors in order to properly identify the factor structure.
- The principal components and the factors can sometimes be hard to interpret.

Orthogonal GARCH (O-GARCH)

= PCA + univariate GARCH

- The Orthogonal GARCH procedure consists of the following steps:

  - Determine the first K factors from PCA, and compute their historical returns;

$$\Sigma_t = \beta \Sigma_{f,t} \beta' + \Sigma_{\epsilon_i,t}$$

  - Estimate a univariate volatility model for factor variances, which are diagonal elements of $\Sigma_{f,t}$ (usually GARCH(1, 1));

  - Compute the model's estimate of conditional variance for each of the factors over time;
  - For each period t, use the estimated betas, residual variances (time-invariant or time-varying), and estimated factor variances to compute an estimate of the conditional variance matrix $\hat{\Sigma}_t$

Applying O-GARCH to our example with two stocks, GE and JPM), we obtain the following estimates for GE volatility:

The following are the estimates for JPM volatility:

As expected, correlations vary over time in a way that reflects the time-varying volatility of the common factor:

Implication of factor models for correlations

- In the market model,

$$Var(r_i) = \beta_i^2 \sigma_M^2 + \sigma_{\epsilon_i}^2$$

$$Cov(r_i, r_j) = \beta_i \beta_j \sigma_M^2$$

which implies

$$Corr(r_i, r_j) = \frac{\beta_i \beta_j \sigma_M^2}{\sqrt{(\beta_i^2 \sigma_M^2 + \sigma_{\epsilon_i}^2)(\beta_j^2 \sigma_M^2 + \sigma_{\epsilon_j}^2)}}$$

Implication of factor models for correlations

- The higher the ratio of idiosyncratic risk to market risk in either asset, the lower the correlation between them.

- If we use a time-varying model for market risk (e.g., GARCH), this will endogenously generate higher correlations between securities when σ_M^2 rises relatively to the $\sigma_{\epsilon_i}^2$, which is what typically happens in times of market crisis.

- Therefore, time-varying correlations do not necessarily have to be modeled as a separate phenomenon in the context of factor models, if time-varying volatility is already incorporated in the model.