QF609 Risk Analysis
Lecture Notes 8

Dr. Tony Wong

Adjunct Professor of MQF
Head of Market Risk Model Validation, Maybank

tonywong@smu.edu.sg

Overview

Main topics:

- Recap
- Dimension Reduction Techniques
- Demo of Dimension Reduction Techniques

Recap

In the last lecture, we have looked at 2 dimension reduction techniques.

A Linear Regression Model

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} \beta_{1,0} & \beta_{1,1} & \cdots & \beta_{1,m} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{n,0} & \beta_{n,1} & \cdots & \beta_{n,m} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ Y_1 \\ \vdots \\ Y_m \end{pmatrix}$$

where $X_1, \cdots, X_n$ are the original set of risk factors and $Y_1, \cdots, Y_m$ are the reduced set of risk factors.

Recap

Using a historical sample of X and Y, denoted by $M_X = \left(\hat{X}_{i,j}\right)_{i=1,\cdots,K,\ j=1,\cdots,n}$ and $M_Y = \left(\hat{Y}_{i,j}\right)_{i=1,\cdots,K,\ j=1,\cdots,m}$, respectively, where K is the number of historical observations, one can estimate the model parameters using linear regression:

$$\begin{pmatrix} \hat{\beta}_{1,0} & \hat{\beta}_{1,1} & \cdots & \hat{\beta}_{1,m} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{\beta}_{n,0} & \hat{\beta}_{n,1} & \cdots & \hat{\beta}_{n,m} \end{pmatrix}^T = \left(\tilde{M}_Y^T \tilde{M}_Y\right)^{-1} \tilde{M}_Y^T M_X$$

where

$$\tilde{M}_Y = \begin{pmatrix} 1_K & M_Y \end{pmatrix}$$

and 1_K is a K-dimensional column vector with each element equal to 1.

Recap

PCA

PCA attempts to identify a new set of data axes so that the projections of the original data onto the new axes orderly preserve as much variance of the original data as possible. The hope is that the projections onto the first few principle components can sufficiently explain the original data variance.

The steps to run PCA on a data set $M_X = (X_{i,j})_{i=1,\cdots,K,\ j=1,\cdots,n}$:

a. Standardize the raw dataset M_X to obtain $M_Z = (X_{i,j})_{i=1,\cdots,K,\ j=1,\cdots,n}$:

$$Z_{i,j} = \frac{X_{i,j} - \mu_j}{\sigma_j}, \quad \mu_j = \frac{1}{K}\sum_{i=1}^{K} X_{i,j}, \quad \sigma_j^2 = \frac{1}{K}\sum_{i=1}^{K} X_{i,j}^2 - \mu_j^2.$$

b. Calculate the correlation matrix V of the dataset M_Z:

$$V = \frac{1}{K} \cdot M_Z^T \cdot M_Z$$

c. Compute the eigenvectors $w_{(1)}, w_{(2)}, \cdots, w_{(n)}$ of V and order them according to the size of their associated eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, starting from the largest first.

d. The principle components are then given by $PC_i = w_{(i)}$, $i = 1, \cdots, n$.

Recap

Once all PCs are obtained, one can obtain the projected PCA coordinates of the original data:

$$M_P \quad = \quad M_Z \cdot W, \qquad W \; = \; \left(\; w_{(1)} \quad w_{(2)} \quad \cdots \quad w_{(n)} \; \right)$$

where row i of M_P, $(P_{i,1}, \cdots, P_{i,n})$, is effectively the coordinates (or representation) of the original data point $(Z_{i,1}, \cdots, Z_{i,n})$ in the PCA space.

Next, how do we decide how many PCs we need to explain the original data?

One way to assess this is by setting a target percent of orignal data variance one wants to explain with the PCs and determine the minimum number of PCs that can achieve that. Let's assume the target is α (e.g. $\alpha = 95\%$). Then the number of PCs to keep is given by:

$$\min \left\{ i \; = \; 1, 2, \cdots, n \;\middle|\; \frac{\lambda_1 + \cdots + \lambda_i}{\lambda_1 + \cdots + \lambda_n} \geq \alpha \right\}$$

Recap

Data Reconstruction:

- **Full reconstruction.** If we keep all the PCs, the original data M_Z can be recovered via:

$$\begin{aligned} M_Z &= M_P \cdot W^T, \quad W = \begin{pmatrix} w_{(1)} & w_{(2)} & \cdots & w_{(n)} \end{pmatrix} \\ (Z_{i,1}, \cdots, Z_{i,n})^T &= P_{i,1} \cdot w_{(1)} + \cdots + P_{i,n} \cdot w_{(n)} \end{aligned}$$

- **Compression.** If we determine that the first m ($m < n$) PCs are sufficient to represent the original data features, one can represent the original data features with the first m PCs via:

$$M_Z^{(m)} = M_P^{(m)} \cdot W_{(m)}^T, \quad W_{(m)} = \begin{pmatrix} w_{(1)} & w_{(2)} & \cdots & w_{(m)} \end{pmatrix}$$

  where $M_P^{(m)}$ is obtained by removing the last $n - m$ columns from the original matrix M_P. In this case, $M_Z^{(m)}$ resprents a compression of the original data which clearly suffers some extent of information loss.

In both cases, the very original data X can be reconstructed by revert the data standardization using the sample mean and variances of the X data.

PCA (cont.)

It is sometimes suggested that PCA can be used as an alternative to resolve collinearity issue in linear regression. But how? To see this, let us observe the followings:

- Recall the matrix $M_P = M_Z \cdot W$. We may write:

$$M_P = \begin{pmatrix} P_1 & P_2 & \cdots & P_n \end{pmatrix}, \quad M_Z = \begin{pmatrix} Z_1 & Z_2 & \cdots & Z_n \end{pmatrix}$$

$$W = \begin{pmatrix} w_{(1)} & w_{(2)} & \cdots & w_{(n)} \end{pmatrix} = \begin{pmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,n} \\ \vdots & \vdots & \vdots & \vdots \\ w_{n,1} & w_{n,2} & \cdots & w_{n,n} \end{pmatrix}$$

Then, it is easy to verify the following:

$$Z_i = w_{1,i}P_1 + w_{2,i}P_2 + \cdots + w_{n,i}P_n \quad i = 1, \cdots, n.$$

Hence, each original data feature Z_i can be expressed as a linear combination (i.e. a weighted sum) of the n new data features in the PCA space.

Dimension Reduction Techniques

- Since each P_i has 0 mean (why?), the covariance matrix of the data M_P is:

$$V_P = \frac{1}{K} \cdot M_P^T \cdot M_P = \frac{1}{K} \cdot W^T \cdot M_Z^T \cdot M_Z \cdot W = W^T \cdot V \cdot W = Diag(\lambda_i)_{i=1}^n$$

  where $Diag(\lambda_i)_{i=1}^n$ is a diagonal matrix with diagonal entries equal to $\lambda_1, \cdots, \lambda_n$. Hence, the data $P_1, \cdots, P_n$ are uncorrelated.

Now, imagine that one attempts to run a regression of some dependent variables $Y_1, \cdots, Y_m$ against the explanatory variables $Z_1, \cdots, Z_n$. The issue of collinearity may arise (e.g. the resulted estimates of the regression coefficients may have large variances etc.) if the $Z_1, \cdots, Z_n$ data are correlated. One potential solution to this issue is, instead, one can consider running the regression against the equivalent data representation of Z in its PCA space, i.e. $P_1, \cdots, P_n$, since they are uncorrelated and hence may avoid the issue of collinearity.

Rank Reduction of a Correlation Matrix

A valid correlation matrix A is of **full rank** if there exists a decomposition $A = \Gamma \cdot \Gamma^T$ such that the factor loading matrix Γ is invertible.

What does this effectively mean?

Imagine that the matrix is A is n-by-n and we have n independent standard normal RVs $z_1, \cdots, z_n$. If A is of **full rank**, that means

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} := \Gamma \cdot \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$$

has a covariance matrix equal to A. Since Γ is invertible, it means that none of x_i can be expressed as a linear combination of the other factors $\{x_j\}_{j \neq i}$. In this sense, the underlying factors $x_1, \cdots, x_n$ are at most partially redundant (no perfect redundancy among the n factors), and the n factors are indeed driven by n independent sources which cannot be further cut down without a loss of information or distorting the original dependence structure.

Dimension Reduction Techniques

If the correlation matrix A is of a high dimension, e.g. 100-by-100, one may still want to cut down the number of independent sources driving the underlying factors without a minimal distortion of the original dependence structure. This is effectively to identify a rank-reduced version of the original correlation matrix A, say $\hat{A}$, which is not of a full-rank, yet provides a good approximation to A.

So how can we do this?

The trick is to take the **eigen decomposition** of A and applies some processing:

- Suppose that A admits the following eigen decomposition:

$$A \quad = \quad H \cdot H^T, \quad H \; = \; E \cdot \sqrt{\Lambda}$$

where E is an $n \times n$ matrix with its columns being the eigen-vectors of A and A is a diagonal matrix with its diagonal entries being its eigenvalues, i.e.

$$E \quad = \quad (e_1 \;\; e_2 \;\; \cdots \;\; e_n), \quad e_i \; = \; \begin{pmatrix} e_{1,i} \\ e_{2,i} \\ . \\ \vdots \\ e_{n,i} \end{pmatrix}, \qquad \Lambda \; = \; Diag(\lambda_i)_{i=1,\cdots,n},$$

where λ_i is the eigenvalue associated with the eigenvector e_i, and $\{e_i\}$ have been ordered accordingly starting from the one with the largest eigen value.

Dimension Reduction Techniques

- assess at the eigenvalues in the same way as we did for PCA to decide the target rank $m < n$ we want
- truncate the matrices E and Λ to obtain an n-by-m matrix $\hat{E} = (e_1 \ e_2 \ \cdots \ e_m)$ and an m-by-m diagonal matrix $\hat{\Lambda} = Diag(\lambda_i)_{i=1,\cdots,m}$
- let C be a n-by-n diagonal matrix with diagonal entries given by:

$$c_{i,i} \quad = \quad \frac{1}{\sum_{k=1}^{m} e_{i,k}^2 \cdot \lambda_k}, \quad i \ = \ 1, \cdots, m,$$

- set $\hat{H} = \sqrt{C} \cdot \hat{E} \cdot \sqrt{\hat{\Lambda}}$ which is an n-by-m matrix effectively giving the rank-reduced factor loadings.

Clearly, with rank reduction, one will not be able to recover the correlation matrix Γ perfectly, i.e. $\hat{A} = \hat{H} \cdot \hat{H}^T \neq A$. However, we shall that see the resulted approximation should reasonably capture the original correlation structure.

Dimension Reduction Techniques

Some applications of the rank reduction technique:

- It is commonly used in the setting up of a Libor-Market Model (LMM). To set up an LMM for derivative pricing, one needs to specify the number of independent driving brownian motions and the instantaneous correlation between them. In general, it is not practical to setup a full rank LMM as it will be computationally intensive. That's where the rank-reduction technique can be used to simply the problem.
- The technique can also be used to fix an invalid (i.e. non positive semi-definite) correlation matrix. To do that, one simply looks at the eigen decomposition of the correlation matrix and performs the rank reduction procedure by removing all eigen vectors with negative eigen values. The resulted rank-reduced correlation matrix should be a valid correlation matrix.

Demo of Dimension Reduction Techniques

Linear Regression Model

- Data: historical daily changes of SOFR curve
- Goal: run dimension reduction based on the linear regression model using daily changes of 1D/1M/6M/1Y/5Y/10Y/20Y/30Y/40Y as a reduced set of factors to drive the daily changes of all 30 zero rates on the curve

Demo of Dimension Reduction Techniques

PCA

- Data: historical daily changes of SOFR curve
- Goal:

Demo of Dimension Reduction Techniques

Rank Reduction of Correlation Matrix