QF609 Risk Analysis
Lecture Notes 7

Dr. Tony Wong

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

tonywong@smu.edu.sg

Overview

Main topics:

- Hints on Group Assignment #1
- Dimension Reduction Techniques

Hints on Group Assignment #1

Simply follow each step involved in VaR implementation:

Identifying the portfolio

The portfolio has been provided clearly in the problem statement: 1 SOFR swap + 4 Stocks.

Identifying the risk factors

- For the SOFR swap, you can assume all the 30 zero rates (1D, 1M, 2M, 3M, 6M, 9M, 1Y, 2Y, 3Y, 4Y, 5Y, 6Y, 7Y, 8Y, 9Y, 10Y, 11Y, 12Y, 13Y, 14Y, 15Y, 16Y, 17Y, 18Y, 19Y, 20Y, 25Y, 30Y, 35Y, 40Y) on the SOFR curve are the underlying risk factors. This is because the value of the swap depends only on discount factors of the SOFR curve which are all determined by the 30 zero rates.
- For each of the 4 stocks, the underlying risk factor is simply the stock price itself.

As a result, in total, we have 34 risk factors for the portfolio.

Hints on Group Assignment #1

Modelling the joint distribution of the daily risk factor changes

- For each risk factor, specify how you want to model their changes. As I suggested, one way is to model the absolute daily changes in the zero rate risk factors and model the daily relative returns of the stock prices.
- Once the above is specified, follow the approach of each VaR model (parametric/Monte Carlo/historical) to model the joint distribution of the daily risk factor changes.

For each risk factor, you would need to use the 1-year historical data provided to compute their daily changes over the 1-year historical period. Using the historical daily changes of the risk factors, one can then determine the joint distribution of 1-day risk factor changes under each of the 3 VaR models:

- **Parametric VaR.** The 1-day risk factor changes are assumed to be multivariate-normally distributed. One can simply estimate the means and covariance matrix of the multivariate normal distribution using the sample means and convariance matrix calculated from the historical daily changes.
- **Monte Carlo VaR.** The 1-day risk factor changes are also assumed to be multivariate-normally distributed. Hence, one simply follow the same appproach as the parametric VaR to estimate the means and covariance matrix of the multivariate normal distribution.
- **Historical VaR.** The joint distribution of the 1-day risk factor changes is simply given by the empirical distribution of the 1-year historical sample of daily risk factor changes.

Hints on Group Assignment #1

Building the distribution of the 1-day portfolio P&L

Since the distribution of the 1-day risk factor changes under each of the 3 VaR models has been determined from above, one can then build the distribution of the the 1-day portfolio P&L :

- **Parametric VaR.** In this case, we have:

$$L \quad = \quad P_{1d} - P_0 \; = \; \sum_{k=1}^{m} a^{(k)} \cdot \Delta x_{0,1d}^{(k)}$$

  where $a^{(k)}$ is the sensitivity of the portfolio value to the k-th risk factor. So the trick is to figure out what is $a^{(k)}$ for each risk factor. For the zero rate risk factor, this should just be the corresponding partial PV01 of the SOFR swap w.r.t the corresponding zero rate. What about the $a^{(k)}$'s for the stock price risk factors? Once the $a^{(k)}$'s are determined, the distribution of L should become very obvious - it's normally distributed, and its mean and variance can be calculated readily.

- **Monte Carlo VaR.** In this case, instead of using an analytical distribution of L as in the parametri VaR model, a sample of L is obtained from Monte Carlo simulation. First, the 1-day risk factor changes are simulated. For each simulated scenario of daily risk factor changes, we can compute the corresponding scenario of L either using a risk-based method (which uses the same formula as in parametric VaR) or using a full revaluation approach.

Hints on Group Assignment #1

- **Historical VaR.** In this case, one starts with the empirical distribution of the 1-year 1-year historical sample of daily risk factor changes. The risk factor changes on each particular historical date can be used to calculate one scenario of L using either the risk-based method of the full revaluation method. Repeating this for all historical dates of the 1-year historical data set of daily risk factor changes, one effectively obtain a sample of L scenarios, and consequently an empirical distribution of L.

Calcualting the VaR

Once the distribution of L is obtained, it becomes trivial to calculate the VaR under each of the 3 models.

Dimension Reduction Techniques

Dimension reduction techniques are used in various financial applications including VaR sometimes. One of the reasons that such techniques are important is that financial data is typically high-dimensional. Such techniques provide a more concise view of the data and also isolate the key components that drive the data dynamics, hence allow for a more efficient and systematic analysis of the underlying problem.

For example, a yield curve is typically characterized by zero rates at many maturities, e.g. more than 20 tenors ranging from 1D to 40Y. It will be cumbersome to analyze movements of all zero rates on the curve as there will be many of them. Instead, one may try to look at only a few important factors that drive the the movement of the whole curve, e.g. level, slope, and curvature.

Another example is that when the VaR of a portfolio may have hundreds of risk factors, one may want to reduce the number of risk factors to make the VaR calculation more efficient.

A Simple Linear Regression Approach

Assume that we are interested in n risk factors $X_1, \cdots, X_n$. Furthermore, assume that these n risk factors are effectively driven by a small number of risk factors $Y_1, \cdots, Y_m$, where $m < n$ via a linear model:

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} \beta_{1,0} & \beta_{1,1} & \cdots & \beta_{1,m} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{n,0} & \beta_{n,1} & \cdots & \beta_{n,m} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ Z_1 \\ \vdots \\ Z_m \end{pmatrix}$$

Suppose that a historical sample of X and Z are avaliable, 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.

Dimension Reduction Techniques

Then, we may estimate the model parameters using linear regression:

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

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.

Once the model parameters are estimated, the model can be used to map/project observations of Y's into those of X's. Clearly, the projected X observations may not be exactly equal to their realized values in general. But in certain cases, they are just accurate enough for the applications in hand. We shall examine some numerical examples on this later.

Principle Component Analysis (PCA)

The purpose of PCA is to identify a rotation of the original data axes such that the projection of the data onto the first new axis preserve the largest part of the variance of the data and the second new axis keeps the next largest and so on.

So, what does it mean by a rotation of data axes?

Take the usual $x - y$ axes on a plane as an example. The two axes can be represented by the two unit vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. A data point $\begin{pmatrix} a \\ b \end{pmatrix}$ in this coordinate system can be represented as:

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} = a \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

So the projections (or coordinates) of the data point in the x-y axes are a and b, respectively.

Next, we consider a rotation of x-y axes to two new axes represented by the unit vectors $\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$ and $\begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$, respectively. They are effectively the two axes represented by the 45-degree and 135-degree lines in the x-y coordinate system, respectively.

So what are the projections (or coordinates) of the original data point $\begin{pmatrix} a \\ b \end{pmatrix}$ in the two new axes? Let us denote them by $\tilde{a}$ and $\tilde{b}$, respectively. Then, they can be obtained via:

$$\begin{aligned} \tilde{a} &= \begin{pmatrix} a \\ b \end{pmatrix}^T \cdot \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{\sqrt{2}} a + \frac{1}{\sqrt{2}} b, \\ \tilde{b} &= \begin{pmatrix} a \\ b \end{pmatrix}^T \cdot \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = -\frac{1}{\sqrt{2}} a + \frac{1}{\sqrt{2}} b. \end{aligned}$$

One can verify that:

$$\begin{pmatrix} a \\ b \end{pmatrix} = \tilde{a} \cdot \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} + \tilde{b} \cdot \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}.$$

Dimension Reduction Techniques

Given an n-dimensional data set, the goal of PCA is to identify n new data axes represented by n column vectors $w_{(1)}, w_{(2)}, \cdots, w_{(n)}$ so that the projections of the data onto the n new axes will orderly preserve the largest variance of the original data.

Here, $w_{(k)}$ is a unit n-dimensional vector representing the direction of the corresponding new axis and is known as the k-th principle component (PC) of the data. One can decide how many PCs to keep for representing the data based on how much they account for the data variance.

Let us consider a data set $M_X = (X_{i,j})_{i=1,\cdots,K,\ j=1,\cdots,n}$ where K is the sample size and n is data dimension. We standardize the data to obtain a data set $M_Z = (Z_{i,j})_{i=1,\cdots,K,\ j=1,\cdots,n}$ which have mean 0 and standard deviation 1 at each dimension, i.e.

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

If we consider the value of the data point at each dimension as a random variable, so the data point can be represented by an n-dimensional random variable $Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$. Then, the the projection of the data Z onto the k-th new axis, $w_{(k)}$, is simply given by $A_k = w_{(k)}^T \cdot Z$.

Dimension Reduction Techniques

In the current context, the joint distribution of $Z_1, \cdots, Z_n$ is given by the empirical distribution of the dataset (i.e. a discrete distribution assigning a probability of $\frac{1}{K}$ to each data point in dataset). Consequently, the variance of A_k is given by:

$$v_k \quad = \quad w_{(k)}^T \cdot V \cdot w_{(k)}, \qquad V \; = \; \frac{1}{K} \cdot M_Z^T \cdot M_Z$$

where V is effectively the correlation matrix for the data M_Z as the dataset has been standardized.

Following the above, we may now look at how the PCs of the dataset can be derived.

1st PC:

This is to determine a unit n-dimensional column vector $w_{(1)}$ such that the variance of the projection of the data M_Z onto $w_{(1)}$ gives the largest variance among all other unit column vector w. Mathematically, this means:

$$w_{(1)} \quad = \quad \underset{||w||=1}{\arg\max} \left[w^T \cdot V \cdot w \right], \qquad V \; = \; \frac{1}{K} \cdot M_Z^T \cdot M_Z$$

Since the maximization here is subject to a constraint $||w|| = 1$. We can turn the problem into an unconstrained maximization problem which looks for the w that maximize:

$$L(w, \lambda) \quad = \quad w^T \cdot V \cdot w - \lambda \cdot (w^T \cdot w - 1)$$

where the scalar λ is the so-called Largrange multiplier.

This boils down to finding w and λ that solves the below

$$\begin{aligned} \frac{\partial L}{\partial w} &= 2 \cdot V \cdot w - 2\lambda w = 0 \\ \frac{\partial L}{\partial \lambda} &= w^T \cdot w - 1 = 0 \end{aligned}$$

or equivaluently,

$$\begin{aligned} V \cdot w &= \lambda \cdot w \\ w^T w &= 1 \end{aligned}$$

It becomes clear that any eigenvector (a n-by-n matrix has at most n different eigenvector) of the correlation matrix V. Therefore we need to choose the eigenvector that gives the maximum variance of the projected data which is equal to

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

As a result, the solution $w_{(1)}$ will be the eigenvector associated with the largest eigenvalue of the correlation matrix V.

Dimension Reduction Techniques

Let us also recall the following fact:

- A symmetric matrix with non-negative diagonal entries is positive semi-definite and all its eigenvalues are non-negative.

The correlation matrix V is one of them. To this end, let us denote its eigenvalues in descending order by:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n.$$

The eigenvectors associated with the eigenvalues above are denoted by

$$w_{(1)},\ w_{(2)},\ \cdots, w_{(n)}.$$

With these notions above, therefore, the 1st PC is $w_{(1)}$ and the projections of the data M_Z onto this direction has a variance λ_1 which is the largest.

2nd PC:

Once the 1st PC $w_{(1)}$ is identified, we can proceed to look for the 2nd PC, following exactly the same logic. The only difference now is that the candidates for search will be in the (n-1)-dimensional subspace that are orthogonal to $w_{(1)}$. Since $w_{(1)}$ is an eigenvector of the correlation matrix V, the search space for the 2nd PC is the sub-eigenspace spanned by the remaning eigenvectors of V, i.e.

$$\Gamma = Subspace\{w_{(2)}, w_{(3)}, \cdots, w_{(n)}\}$$

We have the same problem formulation as before (except that w needs to be in Γ):

$$\underset{||w||=1, w\in\Gamma}{\arg\max} \left[w^T \cdot V \cdot w \right]$$

Applying the same mathematics, we conclude that the 2nd PC is the eigenvector $w_{(2)}$ and the projections of the data M_Z onto this direction has a variance λ_2.

Dimension Reduction Techniques

Following the mathematical derivation above, it should be clear now the PCA solution for the data matrix M_Z is simply given by the the eigenvectors of the correlation matrix of M_Z with the principle components ordered according to the magnitude of the eigenvalues $(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n)$:

Two additional remarks:

- In deriving each PC, the problem of maximizing the variance of data projection is equivalent to minimizing the Mean Squared Error (MSE) of the projection residuals, e.g. for PC1:

$$
\begin{aligned}
w_{(1)} &= \underset{||w||=1}{\arg\min}\, Trace\left\{\left[M_Z - (M_Z \cdot w)\cdot w^T\right]\cdot\left[M_Z - (M_Z \cdot w)\cdot w^T\right]^T\right\} \\
&= \underset{||w||=1}{\arg\min}\, Trace\left\{M_Z \cdot M_Z^T - M_Z \cdot w \cdot w^T \cdot M_Z^T\right\} \\
&= \underset{||w||=1}{\arg\max}\, Trace\left\{M_Z \cdot w \cdot w^T \cdot M_Z^T\right\} = \underset{||w||=1}{\arg\max}\, Trace\left\{w^T \cdot M_Z^T \cdot M_Z \cdot w\right\}
\end{aligned}
$$

- For V (symmetric, positive semidefinite), we have $Trace(V) = \sum_{i=1}^{n} \lambda_i$. The projected data preserves the total variance of the original data.

Dimension Reduction Techniques

Let us summarize how 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 eigenvalues $w_{(1)}, w_{(2)}, \cdots, w_{(n)}$ 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 $PC_i = w_{(i)}$, $i = 1, \cdots, n$.

Dimension Reduction Techniques

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, $(P_{i,1}, \cdots, P_{i,n})$, gives the N coordinates of the original sample data point $(Z_{i,1}, \cdots, Z_{i,n})$ in the PCA space.

Next question: how do we decide how many PC's we should use 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:

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

Dimension Reduction Techniques

Data Reconstruction:

One of the main purposes for PCA is to represent our original data features using a small set of independent variables. This brings us to the topic of 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 further reconstructed by revert the data standardization using the sample mean and variances of the X data.

Dimension Reduction Techniques

In our presentation, standardization is applied to the raw data before running the PCA. But the traditional form of PCA only requires subtracting the mean from the data and PCA is run on the variance-covariance matrix instead. A natural question: should we apply PCA to a correlation matrix or a covariance matrix?

The general rule is that if your raw data features are in the same measurement unit or scale, you can run PCA using either correlation or variance-covariance matrix. However, if the data features are in different units or scale, correlation matrix is preferred. The reason is that in the latter case the PCA results can be biased to put more weights on the variables with larger variances. My view is that it's always a good idea to run PCA on both as a cross check on the robustness of one choice against the other.