FORECASTING OF VALUE AT RISK BY USING PERCENTILE OF CLUSTER METHOD

HAE-CHING CHANG
Department of Business Administration, National Cheng Kung University
No.1, University Road, Tainan City 701, Taiwan (R.O.C.)

CHENG-TE CHEN †
Department of Business Administration, National Cheng Kung University
No.1, University Road, Tainan City 701, Taiwan (R.O.C.)

CHIN-SHAN HSIEH‡
Graduate Institute of Management, National Kaohsiung First University of Science and Technology
No 2, Jhuoyue Rd., Nanzih District, Kaohsiung City 811, Taiwan (R.O.C.)

Abstract
The estimation of VAR (Value at Risk) includes the Historical Simulation, Variance-covariance, and Monte Carlo Simulation Methods. These methods first calculate the risk distribution of asset returns, and then use the percentile of statistics method to estimate the value at risk. The Percentile of Cluster Method was proposed and adopted to replace the percentile of statistics method in estimating VAR. The Percentile of Cluster Method is simple and not restricted to data with normal distribution. The empirical results demonstrated that the Percentile of Cluster Method was more accurate and conservative than the percentile of statistics method, and thus has certain advantages of the latter.

Keywords: Value at Risk (VAR), Cluster Method, Monte Carlo Simulation Method

* Associate Professor, chc5728@ms54.hinet.net
† Ph.D. candidate at the National Cheng Kung University and the Lecturer of Far East University, nike@cc.feu.edu.tw
‡ Ph.D. candidate at the National Kaohsiung First University of Science and Technology and the Lecturer of Kao Yuan University, u9327901@ccms.nkust.edu.tw
Introduction

Estimating the distribution of future asset price has become the focus of VAR studies. Currently, it involves taking the logarithm of stock prices and assuming the derived values possess a normal distribution. This study has the following two aims: (1) to modify the method for converting asset prices into probability distribution; (2) to compare the differences between the percentile statistics and percentile cluster methods of converting asset prices into probability distribution.

The Percentile of Cluster Method was proposed and adopted as a replacement for the percentile of statistics method in estimating VAR. The Percentile of Cluster Method is simple and not restricted to data with normal distributions. This method also has a directive function.

What is VAR

According to Jorion (1996), VAR summarizes the worst expected loss over a target horizon within a given confidence interval. Estimating VAR requires determining two parameters, the holding period and confidence level.

Duffie & Pan (1997) developed the notion of using VAR to compare risks of stocks or investment portfolios, market price and historic prices, and risks in different markets. Dowd (1998) showed that VAR is a simple numeric value that represents total investment portfolio risks in different markets.

Although most financial returns exhibit a fat-tail distribution, conventional VAR estimation models assume that returns are normal distribution. Fama (1965) proposed that the distribution of stock price and stock return is usually leptokurtic and fat-tailed. Hull & White (1998) investigated the statistical properties of asset returns and generated systematic errors, and especially returns with a fat-tailed distribution.

Empirical studies of VAR

Alexander & Leigh (1997) used the simple weighted average, exponential weighted moving average, and GARCH models to estimate VAR. Maximum likelihood estimate (MLE), root mean square error (RMSE), back-testing, and forward testing methods were applied to test the models, demonstrating that the exponential weighted moving average method tended to underestimate VAR. Moreover, although showing no significant difference in statistical verifications, GARCH obtained a more accurate 99% VAR estimate.

Hull & White (1998) observed that when performing Historical Simulation, better VAR estimates could be derived when the daily change of market factors
estimated by the GARCH or index weighted moving average models was used to adjust historical changes. Papageorgiou & Paskov (1999) compared the speed and accuracy of the Monte Carlo and quasi-Monte Carlo methods by estimating 34 European stock indexes options and foreign exchange call options.

**Percentile of Cluster Method**

Clustering analysis can be classified into hierarchical and nonhierarchical forms. Hierarchical procedures involve the construction of a tree-like hierarchical structure. The method of Ward is a commonly used hierarchical method, and is also known as the minimum variance method, since it first views each objective as a cluster and decides how to group clusters given a minimum total variance.

The Percentile of Cluster Method divides a cluster of ordered values into 100 proportions and represents values via percentiles. This method is a measurement of position, and denotes the cumulative relative percentage of a value among the entire sample in order or assigns values relative to a given percentage. VAR is estimated to obtain a value relative to a specific percentage (such as 95%). The procedure includes the following steps: (1) clustering data into 100 clusters; (2) extracting a representative value from each cluster; (3) ranking all the representative values to derive a percentile table.

**Data and Methodology**

The research data comprised daily data from the S&P 500 index spanning the period 2004/01/06~2005/01/03, with the forecasting period spanning 2005/01/04~2005/12/31. The rolling-window method was adopted for VAR estimation, and data were sourced from the CRSP US Stock Database.

Previous studies of VAR all used the percentile statistics method for VAR estimation. This study will be the first to use the Percentile of Cluster Method to determine VAR. This study applies multiple asset value estimation methods to compare performance between the percentile of statistics and Percentile of Cluster Methods.

**Percentile of statistics method**

The percentile is a value relative to a given percentage (such as 95%). Given the sample size, the percentile could be defined as:

\[
\text{Percentile} = \begin{cases} 
  x\left[(np/100)+1\right] & \text{when } np/100 \text{ is not an integer} \\
  \frac{1}{2}\left[x(np/100)+x(np/100)+1\right] & \text{when } np/100 \text{ is an integer}
\end{cases}
\]  

(1)
where $x$ denotes the position of the data, $n$ represents sample size, and $p$ is the position of the percentile.

The first of the distribution of VAR was ordered and then the percentile of statistics method was applied to estimate VAR according to the percentage of maximum tolerable losses arising from investor default. This study set the maximum tolerable loss to 5%.

**Percentile of Cluster Method**

The Percentile of Cluster Method is used to cluster the distribution of VAR into 100 clusters, this method relative to the percentile of statistics method. Cluster analysis has the following advantages: (1) cluster analysis can be processed without any predefined assumptions; (2) similar objects are grouped in single clusters, while dissimilar objects are grouped in different clusters. This method considers internal consistency and external differences.

**Assessment methods for VAR models**

Performance comparison between the percentile of statistics and Percentile of Cluster Methods, were compared in terms of conservativeness, accuracy, and efficiency (Engel & Gizyck, 1999), using a confidence level of 95%. The testing method of Kupiec was also used in Goorbergh & Vlaar (1999), Billio & Pelizzon (2000), Guermat & Harris (2002), and Lin, Chang Chien & Chen (2005).

**Empirical Results**

Figure 1 shows the VAR derived by various methods given maximum tolerable loss of 5%. HS (Historical simulation) and HS_Cluster (cluster analysis of Historical simulation method) forecasted similar VAR. In relation to MCS (Monte Carlo simulation method) and MCS_Cluster (cluster analysis of Monte Carlo simulation method), MCS_Cluster produced excessively conservative VAR. Meanwhile, Var-CoVar (Variance-Covariance Methodology) did not provide VAR estimates significantly different from those obtained from HS and MCS.
The Kupiec (1995) testing method was employed to assess VAR model accuracy. The accumulated failure times of each VAR model are listed in Table 1, in which the Percentile of Cluster Method has relatively fewer failures. The Kupiec (1995) likelihood ratio (LR$_{PP}$) calculated based on the accumulated failures is listed in Table 2. Which result was further compared with $\chi^2$ from Table 3 to verify the model accuracy.

### Table 1. Accumulated failures of each model

<table>
<thead>
<tr>
<th>Model</th>
<th>Percentile of Statistics method</th>
<th>Percentile of Cluster Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HS</td>
<td>MCS</td>
</tr>
<tr>
<td>Accumulated failures</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

### Table 2. Kupiec (1995) Likelihood Ratio of each model

<table>
<thead>
<tr>
<th>Percentile of Statistics method</th>
<th>Model</th>
<th>T</th>
<th>N</th>
<th>P=0.01</th>
<th>P=0.05</th>
<th>p=0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS</td>
<td>HS</td>
<td>251</td>
<td>8</td>
<td>7.689</td>
<td>1.982</td>
<td>17.168</td>
</tr>
<tr>
<td>MCS</td>
<td>MCS</td>
<td>251</td>
<td>8</td>
<td>7.689</td>
<td>1.982</td>
<td>17.168</td>
</tr>
<tr>
<td>Var_CoVar</td>
<td>Var_CoVar</td>
<td>251</td>
<td>8</td>
<td>7.689</td>
<td>1.982</td>
<td>17.168</td>
</tr>
<tr>
<td>Percentile of Cluster method</td>
<td>HS_Cluster</td>
<td>251</td>
<td>5</td>
<td>1.937</td>
<td>6.134</td>
<td>25.803</td>
</tr>
<tr>
<td>Cluster method</td>
<td>MCS_Cluster</td>
<td>251</td>
<td>0</td>
<td>5.028</td>
<td>25.728</td>
<td>52.869</td>
</tr>
</tbody>
</table>

Given the failure rate $p=0.01$, only HS_Cluster (1.937) and MCS_Cluster (5.028) had LR$_{PP}$ smaller than $\chi^2(1, \alpha=0.01)$, indicating that the Percentile of Cluster Method was more accurate than the percentile of statistics method in the LR$_{PP}$ test. Regardless of whether it was for $\chi^2(1, \alpha=0.01)$, $\chi^2(1, \alpha=0.05)$, or $\chi^2(1, \alpha=0.1)$, HS_Cluster (1.937) could not reject the null hypothesis. Given the
failure rate $p=0.05$, MCS_Cluster (25.728) had $LR_{pf}$ exceeding $\chi^2(1, \alpha=0.01)$, indicating the method had lower accuracy. When $p=0.1$, all the $LR_{pf}$ exceeded the Chi-square value, resulting in low accuracy.

Table 3. Test of $\chi^2$ distribution

<table>
<thead>
<tr>
<th></th>
<th>$\chi^2(1, \alpha=0.01)$</th>
<th>$\chi^2(1, \alpha=0.05)$</th>
<th>$\chi^2(1, \alpha=0.1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2$</td>
<td>6.6349</td>
<td>3.8415</td>
<td>2.7055</td>
</tr>
</tbody>
</table>

Table 4. RMSRB of each model

<table>
<thead>
<tr>
<th>Model</th>
<th>Percentile of Statistics method</th>
<th>Percentile of Cluster Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HS</td>
<td>MCS</td>
</tr>
<tr>
<td>RMSRB</td>
<td>0.950</td>
<td>0.900</td>
</tr>
</tbody>
</table>

The Root Mean Squared Relative Bias (RMSRB) proposed by Hendricks (1996) was adopted to assess the conservativeness of VAR models. RMSRB is a negative indicator. Smaller RMSRB is associated with higher conservativeness. Table 4 lists the RMSRB of each VAR model, and reveals that models based on the Percentile of Cluster Method were more conservative than those based on percentile of statistics method.

The Mean Relative Scaled Bias (MRSB) is used to assess VAR model efficiency. MRSB can be used to identify the VAR model with the smallest VAR given a theoretical failure rate. MRSB is a negative indicator, with smaller MRSB indicating higher efficiency. Table 5 shows the MRSB of each VAR model, and showed that models based on the percentile of statistics method were more efficient than those based on the Percentile of Cluster Method.

Table 5. MRSB of each model

<table>
<thead>
<tr>
<th>Model</th>
<th>Percentile of Statistics method</th>
<th>Percentile of Cluster Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HS</td>
<td>MCS</td>
</tr>
<tr>
<td>MRSB</td>
<td>0.0380</td>
<td>0.0337</td>
</tr>
</tbody>
</table>

Conclusion

The Percentile of Cluster Method is simple and not restricted to data with a normal distribution. Furthermore numerous packaged software systems are available for processing data clustering, meaning applying this method does not involve increased workload, this method also has directive functions in practice.
Empirical results demonstrated that VAR estimated via the Percentile of Cluster Method was slightly higher than when estimated by the percentile of statistics method. In terms of accuracy, the Percentile of Cluster Method was superior to the percentile of statistics method in the LRₚₚ test, suggesting that the accumulated failure rate of the Percentile of Cluster Method closely approximated the assumed failure rate. In the test of conservativeness, the Percentile of Cluster Method was more conservative than the percentile of statistics method, indicating that the Percentile of Cluster Method would obtain higher VAR estimates. In terms of efficiency, the Percentile of Cluster Method was slightly inferior to the percentile of statistics method.

To sum up, the Percentile of Cluster Method has certain advantages over the statistic percentile method in estimating VAR, especially in terms of estimation accuracy. As a result, we recommend that investors use the Percentile of Cluster Method to estimate VAR of asset returns.

References

