Risk analysis in the economics through R Language

METODI TRAYKOV  
Department of Electrical Engineering, Electronics and Automatics  
University Center for Advanced Bioinformatics Research  
South-West University "Neofit Rilski"  
66 Ivan Mihaylov Str., 2700 Blagoevgrad  
BULGARIA  
metodi_043@abv.bg

MIGLENA TRENCHEVA  
Department of Finance and Accounting  
South-West University "Neofit Rilski"  
66 Ivan Mihaylov Str., 2700 Blagoevgrad  
BULGARIA

ELENA STAVROVA  
Department of Finance and Accounting  
South-West University "Neofit Rilski"  
66 Ivan Mihaylov Str., 2700 Blagoevgrad  
BULGARIA

RADOSLAV MAVREVSKI  
Department of Electrical Engineering, Electronics and Automatics  
University Center for Advanced Bioinformatics Research  
South-West University "Neofit Rilski"  
66 Ivan Mihaylov Str., 2700 Blagoevgrad  
BULGARIA

IVAN TRENCHEV  
Department of Electrical Engineering, Electronics and Automatics  
University Center for Advanced Bioinformatics Research  
South-West University "Neofit Rilski"  
66 Ivan Mihaylov Str., 2700 Blagoevgrad

Abstract: Usually the risk management leads to improved company. Many often are cases, big risk decisions are being made too low in organizations, with staff who don't stimulated to make the right decisions for the organization. The aim of our study is to demonstrate the effectiveness and benefits of good risk management. This article is based on the latest techniques for measuring and managing on risks in various sectors of business. Using the programming language R Language, we show effective way to evaluate and analysis risk. Apply the comparative analysis in the continuing quest to find and adapt better practices for management risk which leads to increased profits and competitiveness of firms. We showed a good and easy risk management using R Language, which can be useful for a happy and successful career.

Key-Words: - Risk analysis, R Language, economics

1 Introduction

The classical theories say that investors can eliminate specific risk through diversification of their companies in order to include many different assets. Specific risk can be avoided through diversification, but exposure to it may be rewarded in the market. Instead, investors should hold combination of risk-free asset and marketing portfolio, as the exact combination depends on investors’ appetite for risk. In this basic setup, companies must not squander resources for risk management, as investors are not concerned about the specific risk that takes a company [1, 2].
From the famous theorem of Modigliani and Miller we know that the value of firm is independent of its structure: the company just needs to maximize expected profits, despite the risk that is associated with it. Holders of securities may carry risk transfers through appropriate portfolio. It is clear, however, that in practice routinely violate the strict conditions necessary for the theorem of Modigliani and Miller. In particular, capital market imperfections, such as taxes and costs of financial problems lead to the failure of the theorem and create conditions for risk management.

Are risk management leads to improved company? The general answer to this question is yes. The aim of our study is to demonstrate the effectiveness and benefits of good risk management. Based on several basic methods and techniques will show how good for business is good risk management. Analysis of the practice of risk management in the industry for gold show that stock prices are less sensitive to price movements of gold after risk management. Similarly, in the gas industry, better risk management has concluded that the variables share prices are less. One study also found that risk management in a wide group of companies leads to a decrease in interest rates and movements in exchange rates.

2 Problem Formulation

2.1 Historical Simulation

We consider approaches Historical Simulation (HS) [2, 3, 4]. Let today be a day t. We consider portfolio of n assets. Let we denote our today units or share of asset i whit $M_{i,t}$, then the value of the portfolio for today is

$$V_{PF,t} = \sum_{i=1}^{n} M_{i,t} S_{i,t}$$

Using today's company portfolio, but with past asset prices, we can calculate the old value of the portfolio (pseudo value), which will occur if during that time has been used for today's distribution of the portfolio. For example, yesterday's value of pseudo value is

$$V_{PF,t-1} = \sum_{i=1}^{n} M_{i,t} S_{i,t-1}.$$  

This is pseudo value, that is represent as shares of each asset, which typically undergo changes over time. We can calculate the pseudo log-return as

$$Pr_{PF,t} = \ln \left( \frac{V_{PF,t}}{V_{PF,t-1}} \right).$$

Armed with this definition, we can give a definition of approaches Historical Simulation (HS) for risk management. HS technique is deceptively simple [1, 4, 5, 7]. Consider the existence of past sequence of m hypothetical portfolio of daily returns, calculated by past prices of the underlying assets of the portfolio, but using the weights for today portfolio: to say \( \{ Pr_{PF,t+1-\varphi} \}_{\varphi=1}^{m} \).

HS technique simply assumes that the distribution of returns from the tomorrow's portfolio, \( Pr_{PF,t+1-\varphi} \), will be closer to the empirical distribution of past observations \( m \), \( \{ Pr_{PF,t+1-\varphi} \}_{\varphi=1}^{m} \). In other words, the distribution of \( Pr_{PF}\varphi+1 \) is taken from the histogram for \( \{ Pr_{PF,t+1-\varphi} \}_{\varphi=1}^{m} \). The variable \( \gamma \) with size of range \( p \), is calculated simply by 100p% of the sequence of returns from past portfolio. We write

$$VaR_{t+1}^P = -\text{Percentile} \left( \{ Pr_{PF,t+1-\varphi} \}_{\varphi=1}^{m} \right).$$

Thus, after simply sort of returns \( \{ Pr_{PF,t+1-\varphi} \}_{\varphi=1}^{m} \) in ascending order and we can choose \( VaR_{t+1}^P \) to be such number that only 100p% of the observations are smaller than \( VaR_{t+1}^P \). Since VaR usually falls between two observations, we can use linear interpolation to calculate the exact number [5, 6]. Standard quantitative software packages will need to have Percentile function or we need to make such functions so that linear interpolation can be performed automatically.

2.2 IBNR techniques

In the statistical literature often mentions heuristics to complete the IBNR triangle generated by IBNR (Incurred But Not Reported – IBNR) techniques. We will briefly describe the two most commonly used techniques for IBNR predictions: Chain Ladder method and Bronhueter-Ferguson technique [2, 6]. There are many fundamental Generalized Linear Models, for which some evaluators can be calculated with the Chain ladder method. On the other hand, it is possible to obtain pattern which comprises at least rigid statistical structure and calculations in which the calculation on Chain ladder method for optimal price are in the sense of mean square error [2, 7, 8].
2.2.1 Chain Ladder method

The idea of the Chain Ladder method is that in one year will be definite the same overall percentage of the benefits of each starting year. In other words, in the run-off triangle the columns are proportional [2]. Form table 1 we can see how in the Chain Ladder method are calculated predictions for unnoticed part of the run-off rectangle to. Note that in most texts given run-off amounts accumulate in rows. This is a sign of the time when the calculations had to be made by hand.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
 & 1 & 2 & 3 & 4 \\
\hline
01 & A & A & A & B & \cdot \\
02 & A & A & A & B & \\
03 & b & B & b & * \\
04 & D & D & \tilde{D} & ** \\
05 & \cdot & \cdot & \cdot & \cdot & \\
\hline
\end{tabular}
\caption{Completed run-off rectangle with CL predictions}
\end{table}

We consider the element (3, 4) in Table 1, denoted by * and representing payments in respect of the policy for the third year of the four years of development. This is the amount of compensation for six year. This is amount for the first calendar year and the future amount is just around the edge of the observed amounts [6, 9]. Because of committed proportionality, the ratio of elements \( b \) will be equal to the ratio \( B : A \). Therefore, the prediction \( \tilde{X}_{34} \) for the element * is

\[ \tilde{X}_{34} = b \times \frac{B}{A} \]  \hspace{1cm} (1)

Here \( B \) means the total amount of \( B \)-elements in Table 1, the observed values [2, 6, 11]. The prediction \( \tilde{D} \) is calculated in the same way, multiply the total amount of increasing payments to the left of it with the total amount over it and divide the total amount of losses on policies from previous years and years of development. The prediction ** for \( \tilde{X}_{44} \) can be calculated using the same development factor \( B \)/\( A \):

\[ \tilde{X}_{44} = D \times \frac{B}{A} \]  \hspace{1cm} (2)

where the amount \( D \) comprises \( \tilde{D} \), which is not actual observation, but the prediction is constructed as above. Using the fact that \( * = b \times \frac{B}{A} \) we can see that you are getting exactly the same prediction

\[ \tilde{X}_{44} = \frac{D \times (B + *)}{A} \]  \hspace{1cm} (3)

Here, we follow the same procedure as at observation for the next calendar year. Thus starting from row 2 and going from left to right, we can fill it with prediction all lower part of the triangle.

**Note 1. (Mirror property of Chain Ladder method).** We note that if you swap the roles of the first year of developed and starting year this procedure will output the same estimates for the completion of the square, i.e. take the mirror image of the triangle around the NW-SE diagonal [2, 10, 11].

**Note 2.** One way to describe the Chain Ladder method is as follows: find the numbers, \( \tilde{a}_i, \tilde{\beta}_j \), where \( i, j = 1, ..., t \) such that the product, \( \tilde{a}_i \tilde{\beta}_j \) (entered value) for "observed" \( (i, j) \) combinations with \( i + j - 1 \leq t \), has the same amounts in columns and rows, as actual observations:

\[ \sum_{i(t)} \tilde{a}_i \tilde{\beta}_j = \sum_{j(t)} X_{ij}, \forall i(j). \]  \hspace{1cm} (4)

Then predict future values of \((i, j)\) with \( i + j - 1 > t \) of \( \tilde{a}_i \tilde{\beta}_j \).

We will show why this procedure leads to the same predictions as the Chain Ladder method as look at Table 1. First we see that \( B \) and \( B \) are amounts of rows and columns and from these values we can calculate the necessary amounts values of benefits \( A \). For example, \( A_3 = R_1 + R_2 \) (entered value) for "observed" \( (i, j) \) combinations with \( i + j - 1 \leq t \), has the same amounts in columns and rows, as actual observations:

\[ \sum_{i(t)} \tilde{a}_i \tilde{\beta}_j = \sum_{j(t)} X_{ij}, \forall i(j). \]  \hspace{1cm} (4)

Then note that if we replace recent losses \( X_{ij} \) with their entered values, \( \tilde{a}_i \tilde{\beta}_j \), amounts in rows and columns remain unchanged, therefore values as \( A \) also remain unchanged [2, 11]. When applying the above-described Chain Ladder method to the new triangle, the numbers \( \tilde{a}_i \tilde{\beta}_j \) are result of future predictions [12].

The basic principle of this method allows many variations (13, 11). Without doubt, this is determined by the effects work on the axis, describing the initial year of claims. The Chain Ladder method only covers the run-off pattern as all other factors that affect the ratio of established claims remain unchanged over time. This method
has only one algorithm, deterministic method. But there are also stochastic models for generating processes that emphasize the run-off triangle in which these same calculations result in a sense to an optimal prediction.

2.2.2 Bornhuetter-Ferguson method

One of the difficulties in using the Chain Ladder method is that predictions for reserve estimates can be quite unstable. In Table 1, the change of p% in $b_2$ of the variability on sample will generate the same change in all predictions for this order. So, this method on experiment with variable benefits wills variable prognosis. This instability will be shown by changes in the calculation of the reserve for each year when the triangle to add new diagonal observations. The Bornhuetter-Ferguson method (1972) provides a procedure to stabilize the calculations [13, 14, 15].

Suppose, that we have some preliminary expectations regarding final loss to get out of any unfortunate period $i$, i.e. $E[X_1 + \cdots + X_{t_i}] = M_i$ for some know value $M_i$. This value is often explained as inventory or final budget losses. In particular, we can have preview of the loss ratio $M_i/P_i$, where $P_i$ is the premium income from the occasional year $i$. Combining these estimates with the factors for the development of the Chain Ladder method we can be able to make the evaluation of the whole scheme of development of losses [2, 12, 15, 16, 17, 18, 19].

The Bornhuetter-Ferguson method could be interpreted as Bayesian method. Estimates are in the form of a trust evaluator.

3 Problem Solution

We will show research analyzes of data for stock prices. Also, we will illustrate the use of R Language [20] to generate a portfolio of risks in food industry, which we will use for testing purposes. For the analysis of portfolio we use Generalized Linear Models (GLM).

To illustrate how we can use R Language for risk management, suppose that we are interested of buying on portfolio of shares of company for production on meat and other company for production milk. Our historical data consist of weekly share prices over the last three calendar years, and on based of that we want to predict the future behavior of portfolio consisting of combination of two types of shares. R Language has great functionality when working with the dataset, but here we will use the standard functions. The prices are as follows:

The prices are as follows:

$\text{Meat} <- \text{scan(n=157)}$


$\text{Milk} <- \text{scan(n=157)}$

$$771 774 747 731 718 716 733 736 758 742 748 734 769 786 770 805 781 769 792 787 790 850 863 844 845 836 814 823 828 741 817 837 849 843 906 915 928 939 928 939 990 993 1008 1012 1016 1009 1027 1016 959 1001 984 1026 1020 1008 1005 996 1000 960 963 973 974 970 970 986 965 956 977 1018 1037 1044 1030 1025 1047 1091 1080 1084 1098 1092 1094 1095 1002 1088 1103 1066 1050 1044 1047 1068 1067 1081 1083 1069 1076 1054 1124 1157 1207 1145 1161 1164 1120 1208 1240 1319 1340 1324 1308 1305 1300 1288 1260 1379 1419 1340 1367 1356 1414 1445 1465 1468 1458 1420 1399 1397 1404 1399 1392 1376 1372 1382 1381 1435 1438 1452 1464 1472 1467 1455 1515 1537 1528 1426 1455 1444 1450 1466 1511 1578 1577 1529 1564 1527 1508 1520 1490 1544 1522 1488$$

To show their behavior in time we need to look at some standard graphics (Figure 1).

![Fig. 1: The share prices of Meat and Milk](image_url)
To determine whether the marginal distributions are really normal, we can do visual test for normality by checking normality with Q-Q graphics (Figure 3).

The Q-Q graphics plotted sample of Quintiles against theoretic Quintiles. If the normal Q-Q graphics are close to straight line, perhaps the marginal values are normal. In our case, the fitting is not very good: the diagonal line in the first and third point of comparison is a quarter below or above the median (obtained by using the qqline function). It also follows from the fact that the sample quintiles in tails is larger than in the case of normality, while at normal distributions, the tails of distributions are too "thick".

Another graphical tool with which to assess normality is plotting histogram and compared to appropriate normal density (Figure 4).

We can easily visualize the evaluation of the density of the core. For function of the core take for example the normal density with suitably choose standard deviation divided by $n$. Put it into the center of each of the $n$ data points and then just add them all together to form the evaluation of the density of the core. In fact, if $\{y_1, \ldots, y_n\}$ are realizations of random samples $Y_1$, $\ldots$, $Y_n$ from continuous random variable $Y$ and $X$ is discrete random variable having probability $Pr[X = y_i] = \frac{1}{n}$, $i = 1, \ldots, n$, then core density function ($cd$) for $X$ is an empirical function for $Y_1$, $\ldots$, $Y_n$. Additionally, if, $U$ is standard normal random variable, independent of $X$, then $cd$ function $X + \sigma U$ is the assessment of core density for $cd$ function of $Y$. If the evaluation of the core density is very similar with the assessment of the appropriate normal density concluded that logarithmic returns may be normal. For logarithmic gains of weekly stock prices, the normal assumption is more plausible than the daily prices due to the central limit theorem.

The statistical test for normality on sample $X_1$, $\ldots$, $X_n$ is test known as test of Jarque and Bera (Jarque-Bera test – JB). This statistical test is defined as

$$JB = \frac{n}{6} \left( S^2 + \frac{K^2}{4} \right). \quad (5)$$

Here $n$ is the number of observations (or degrees of freedom), $S$ is asymmetrical sample and $K$ is the excess of sample. They are calculated as

$$S = \frac{\hat{\mu}_3}{\hat{\sigma}^3} = \frac{\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^3}{\left( \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \right)^{3/2}},$$

$$K = \frac{\hat{\mu}_4}{\hat{\sigma}^4} - 3 = \frac{\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^4}{\left( \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \right)^{2}} - 3, \quad (6)$$

where $\bar{X} = \frac{1}{n} \sum X_i$ is the sample mean and $\hat{\sigma}^2 = \frac{1}{n} \sum (X_i - \bar{X})^2$. Statistical $JB$ has approximately $\chi^2(2)$ distribution: asymptotically, for the two conditions under the null hypothesis we can show that are independent, raised square, standard, normal, random variables. If the asymmetry or the excess deviate from their "normal" value (0), $JB$ will be big. The critical level of 95% is 6.0.

Logarithmic returns are not just quite normal value, and in fact are bivariate normal values. From the graph on Figure 3 we cannot understand anything about the capitals having joint normal...
distribution. But to see whether this assumption is valid, we can look at scatter plots of logarithmic returns in Figure 2. Using R Language we can determine the parameters $\mu_{\text{Meat}}, \sigma_{\text{Meat}}, \mu_{\text{Milk}}, \sigma_{\text{Milk}},$ and $\rho_{\text{Meat,Milk}}$ of the bivariate normal distribution.

Suppose that we want to buy portfolio of equal parts of Meat shares at the current price 474 euro and Milk shares at price of 1498 euro. Then, we are interested in their future action over an interval of two calendar years, or 104 weeks. Based on our data we can assume that the weekly log-profit $(X_i, Y_i), i = 158, \ldots, 261$ has bivariate normal distribution with parameters similar to those, calculated above and independent of each other $i$. We write

$$X = X_{158} + \cdots + X_{261}; Y = Y_{158} + \cdots + Y_{261},$$

(7)

random variable that must predict is

$$S = 474e^X + 1498e^Y.$$

(8)

From the fact that $\text{Cov}(\sum X_i, \sum Y_i) = \sum\text{Cov}(X_i, Y_i)$, if the pair $(X_i, Y_i)$ is independent, we can calculate the parameter estimates for $(X, Y)$.

Since $X$ and $Y$ are the bivariate normal value, parameter $S$ at Figure 5 is the sum of the dependent logarithmic normal random variables. The calculation of the $cd$ function and quartiles of $S$ is difficult problem. One way to proceed is to simulate many results for $S$ and seek examples instead theoretical quintiles of $S$. To do this we need method that to generate drawings of $(X, Y)$.

To generate sample of multivariate, random, normal accompanying variables with some average and variation matrix, we can use the `mvrnorm` function in R Language. It is found in the MASS library, consisting objects related to the book "Modern Applied Statistics with S". Below, as illustration, we will explain the method of Cholesky decomposition of bivariate average with $n = 2$.

Let $U$ and $V$ are independent standard normal. Then for all real $\alpha$, we have

$$r(U, U + aV) = \frac{\text{Cov}[U, U + aV]}{\sigma_U \sigma_U + aV} = \frac{1}{\sqrt{1 + \alpha^2}},$$

(9)

which means that for $r > 0$, the correlation of $U$ with $U + aV$ is equal to $r$, if we take

$$\alpha = \sqrt{1/r^2 - 1}.$$  

(10)

Then $W = r(U + aV)$ is also standard normal value. Here, also we have correlation between $r$ and $U$ in the case where $r < 0$: for $r = 0$, we get $W = V$. Finally, let

$$X' = E[X] + U\sqrt{\text{Var}[X]}; Y' = E[Y] + W\sqrt{\text{Var}[Y]},$$

(11)

then $S' := 474e^{X'} + 1498e^{Y'}$ ~ $S$, with $S$ as in Figure 5.

4 Conclusion

We reduced the problem of drawing pseudo-random results from the distribution of S/100 index to generate streams of univariate independent, standard, normal, random variables $U$ and $V$. We can do this in R Language by simply calling the `rnorm` function. In its standard mode, it applies the inverse of the standard rate of $cd$ function $\Phi^{-1}(Q)$, to unify the arguments $U_1, U_2, \ldots$: the latter may be obtained by calling the `runif` function. As a result, we can generate sample of thousand results for $S$, where $(X, Y)$ have bivariate normal distribution with parameters as calculated above.

So, judging from the obtained quintiles of $S$, on every two years, the profit will by approximately 50% ± 25%. Portfolio of one Meat share and one Milk share with price $474 + 1498 = 1972$, with probability of 50% will be worth at least 152% from current value for two years, with probability of 90% will be with amount between 109% and 207%, and this will reduce the amount of about one in forty cases. However, for one of the 200 cases, the portfolio is sold for 2.5 times more from the initial value. All this is on condition, that future share prices will follow our expected model for last prices. Aside the fact that the normality of the logarithmic return is firmly rejected, Figure 1 shows an example that of Meat shares is quite possible that growing trend in the price to be stopped at 100% for about week.

Fig. 5: Histogram for $S=1000$, with evaluating the core density and normal distribution
The Risk Analysis is a proven way of identifying and assessing factors that could negatively affect the success of a business. It allows to examine the risks that the organization face and helps decide whether or not to move forward with a decision. This is a difficult task for the staff, but using R Language can become easy and fast task, as shown in this study.

References:


