MATEC Web of Conferences **387**, 06002 (2023) https://doi.org/10.1051/matecconf/202338706002
Power Transmissions 2023

Monte Carlo simulation in risk assessment in mathematical generation of long data

Anton Antonov [1]* and *Siyka* Demirova[2]

[1]Technical University of Varna, Faculty of Manufacturing Engineering and Technologies, 9010, Varna, Bulgaria
[2]Technical University of Varna, Faculty of Manufacturing Engineering and Technologies, 9010, Varna, Bulgaria
*E-mail: a.antonov@tu-varna.bg
E-mail: s_demirova@tu-varna.bg

Abstract. The report will address the main problem in risk measurement, namely the lack of a sufficiently long series of data for the various variables, so that the trend of their development can be formed with negligible error. A model will be made through a mathematical calculation in order to derive a long enough series with statistics reflecting the correlation between the individual variable indices and the standard deviation (volatility) to be able to obtain results with greater accuracy.

Introduction

Monte Carlo simulation was created in the late 1940s by Stanislav Ulam in partnership with his colleague John von Neumann. The Monte Carlo method allows all possible outcomes of decisions to be seen and the impact of risk to be assessed. This allows for more motivated decision-making under conditions of uncertainty.

The main idea of the Monte Carlo simulation in calculating the volume of risk (VaR - Value at Risk) is to construct a detailed picture of the portfolio risk by computer simulation of a huge number of random numbers possessing the main characteristics of the empirical portfolio return distribution for a certain period [1,2]. One of the main statistical characteristics of asset returns is their random distribution. The change in their future prices is unknown, and usually the best guess for the next period's price is the last known price [3,4,5]. When such a variety of random changes in asset prices is observed, one logical solution is to determine the size of this change (market risk) by simulating a large number of random variables with certain characteristics. In this way, a complete picture of the price dispersion is constructed with an emphasis on the last known price. The use of random variables accompanied by the corresponding probabilities allows to calculate all possible variations in the price of the asset. This is precisely the main idea behind the Monte Carlo method for calculating VaR. The main stages through which the calculation of VaR passes when using Monte Carlo simulation are the following:

- Acquisition of up-to-date historical data for the assets with a sampling horizon of 1-3 years;
- Investigating the shape of the empirical di stribution and constructing a mathematical model;
- Calculation of changes in production indices, arithmetic mean, standard deviation and correlation-variation matrix;
- Computer simulation of a minimum of 1,000 to 10,000 random numbers with defined statistical characteristics;
- Converting the random variables into sampling, by using certain statistical distributions;
- Simulating (calculating) the resource change in the production portfolio;
- Simulating the dynamics in the value of the entire portfolio of production costs and finding its difference compared to the value planned at the initial stage;
- Determining the value of VaR as a certain percentage with a certain percentage of accuracy;
- Analysis of the results and selection of a scenario to influence the factors with the highest volatility.

Essentially, simulation is a computerized mathematical technique that allows people to account for risk in a quantitative analysis. It provides all possible outcomes for a given choice of actions and shows how likely they are to occur. Applied to a production portfolio of activities and the indices on which they depend, this means that it is possible to use Monte Carlo simulation to help analyze all the risk factors. It shows a range of results from the best possible index change scenario to the worst possible one. And of course, it can also show what would happen if decisions were made "in the middle of the road". This is especially useful for managers who want to analyze possible options [6,7,8].

Monte Carlo simulation builds models of potential outcomes by substituting a range of values for each uncertain factor. This is known as *a probability distribution.* The simulation then cycles through all possible outcomes, using a different set of random values each time. This can take tens of thousands of calculations. During a Monte Carlo simulation, values are drawn randomly from the input probability distributions. Each set of samples is defined as *a* *replicate*. The result obtained from each sample is then recorded.

1. Research Methodology and Results

For the purpose of the study, four indices will be taken for possible changes in production costs, as we have data on their movement for a period of two years. The data is displayed for weekly periods that correspond to the weekly periods of the planned cost matrix. Example data can be seen in Table 1:

Table 1. Historical changes in manufacturing indices by week.

HISTORICAL VALUES			
Index1	***Index2***	***Index3***	***Index4***
100.00	100.00	100.00	100.00
101.31	101.10	99.88	99.30
99.04	100.54	100.72	99.12
100.73	101.03	99.95	97.74
97.57	100.32	100.57	101.81

The initial plan value for each index is taken as 100%, and the rest are formed according to the formula:

*Index 1 = 2 period expenditure/1 period expenditure * 100* (1)
etc.

In order to calculate the standard deviation of the assumed historical statistics, it is necessary to calculate an index of relative change, which is calculated as follows:

LN (Index12;index11) = (index 12-index 11)/index11 (2)
etc.

This results in an array of data for the weeks presented in Table 2.

Table 2. Historical changes in manufacturing indices by week.

SIMULATED HISTORICAL CHANGES			
Index1	***Index2***	***Index3***	***Index4***
0.0025	0.0073	0.0029	-0.0134
0.0132	0.0026	0.0015	0.0178
0.0129	-0.0119	-0.0083	0.0028
-0.0584	-0.0038	0.0006	-0.0321

Examining the obtained data on historical changes between indices 1 to 4, two essential statistical indicators for the future model can be derived:
-Coefficients of Correlation between Indices (Correlation Matrix);
- Volatility (standard deviation of each index) .

1.1. Correlation and correlation matrix

A correlation is a statistical relationship between two random variables, whether it is intentionally caused or not. In the broadest sense, correlation is any statistical association, although it usually refers to the degree to which a pair of variables are linearly related.

In industry, correlation is an indicator that measures the dependence in the price movement of two resources or assets. A correlation coefficient is a statistical measure of the strength of the relationship between the relative movements of two variables. Its values range between -1.0 and 1.0:
- A correlation coefficient of -1 means a perfect negative correlation, i.e., asset prices move perfectly in opposite directions.
- A correlation coefficient of 1 means a perfect positive correlation, i.e., asset prices move perfectly in the same direction.
- A correlation coefficient of 0 means there is no relationship between asset price movements.

When it comes to types of correlation between two indices, they are basically three:
- Straight Correlation (Positive Correlation);
- Inverse correlation (negative correlation);
- Neutral correlation.

Direct correlation
Two assets have a direct correlation (positive correlation) when their values move in the same direction. Therefore, to be able to speak of a direct correlation, it must have a positive correlation coefficient (greater than 0).

Inverse correlation
An inverse correlation is observed when the parameter values move in opposite directions. So, to have an inverse correlation, you must have a negative correlation coefficient (less than 0).

Neutral correlation
When the prices of two assets on the financial markets have no relation to each other, the two assets can be said to be neutrally correlated. This happens when they have a correlation coefficient around 0.
In order to determine the correlation coefficients between the indices of relative changes, a correlation matrix was calculated in Excel.

Table 3. Correlation coefficients between indices of relative changes

CORRELATION MATRIX				
	Index1	***Index2***	***Index3***	***Index4***
Index1	*1.0000*	*0.5761*	*0.0060*	*0.0711*
Index2	*0.5761*	*1.0000*	*0.5608*	*0.0933*
Index3	*0.0060*	*0.5608*	*1.0000*	*0.4368*
Index4	*0.0711*	*0.0934*	*0.4384*	*1.0000*

Note: The diagonal is always one because it shows the index ratio to himself

1.2. Volatility – standard deviation

Standard deviation is a measure of the spread of a series or the distance from the standard. In 1893 Carl Pearson introduced the idea of standard deviation, which is arguably the most widely used measure, into scientific research.

It is the square root of the mean of the squares of the deviations from their mean. In other words, for a given set of data, the standard deviation is the root-mean-square deviation from the arithmetic mean. For the whole population it is denoted by the Greek letter "sigma (σ)" and for a sample it is represented by the Latin letter "s".

Standard deviation is a measure that quantifies the degree of dispersion of a set of observations. The further the data points are from the mean, the greater the variance within the data set, indicating that the data points are spread over a wider range of values, and vice versa.

Here is a quick overview of the steps involved:

Step 1: Find the mean of the array.

Step 2: For each value, find the square of the difference between the particular value in the data set and the mean value.

Step 3: Sum the values from step 2.

Step 4: Divide t by the number of values in the data set.

Step 5: Calculate the square root of the resulting amount.

Formula for calculating standard deviation:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n}} \qquad (3)$$

The syntax for calculating the standard deviation in Excel is as follows:

= STDEV (Index11:Index1n) *(4)*

The sampling extent of the relative deviation data is illustrated in Table 4.

Table 4. Example of Volatility of production indexes by periods

Average	*0.0001*	*-0.0001*	*0.0000*	*-0.0001*	
Volatility	*0.0222*	*0.0105*	*0.0054*	*0.0151*	*weekly*
Volatility	*0.0461*	*0.0217*	*0.0112*	*0.0312*	*monthly*
Volatility	*0.1596*	*0.0753*	*0.0388*	*0.1081*	*annually*

Volatility, standard deviation results are calculated on a weekly basis. If for the purposes of the model they need to be transformed into data on a monthly or annual basis, the following transposition formulas can be used :

*per month=(Volatility Index n weekly)*SQRT(1/7*30)* *(5)*

*for a year =(Volatility Index n monthly)*SQRT(12)* *(6)*

Having the volatility for the selected time slice (for example e week) and the correlation between the factors is a base to use standart excel formulas to generate the adequate long line of variables that can run reliable Monte Carlo simulation. 1st step of the final part of the model is generates long line of Non-correlated normal distributed randoms using:

=NORM.S.INV(RAND()) *(7)*

It is a pre-built integrated function in Excel that is categorized under statistical functions in Excel. The NORM.S.INV:

- The normal distribution is the most widely used in statistics. It is also called a “Bell curve” or “Gaussian curve.”

- We can fully describe the normal distribution based on its mean and standard deviation (SD) values.

- A normal distribution is called “standard normal distribution” when its mean value is “0” or zero, and the standard deviation value is equal to 1.

The normal distribution can be standardized by using the below-mentioned formula:

z = (x – mean)/ sd *(8)[9]*

- *z = Z-score*
- *x = the value being evaluated*
- *μ = the mean*
- *sd = the standard deviation*

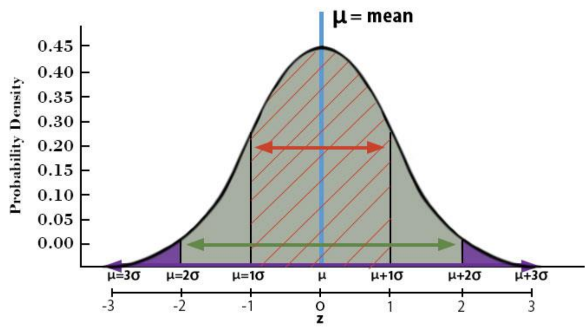

Fig 1. Bell Curve (Standart Deviation) [9].

Here, the red arrowed line in the curve indicates where the standard deviation of the mean value is within 1, whereas the green arrowed line in the curve indicates where the standard deviation of the 0mean value is within 2.

A set of normal distributed randoms (for example 5000) are generated for every risk index and time point by random generator, the result is a scenario matrix: (Indexes x time points) x scenarios, for example 4 indexes x 12-time points x 5000 scenarios. The matrix is then multiplied by Cholesky decomposition of the covariance matrix to receive correlated small changes of basic factors on time points. The simulation produces a large set of possible developments for the choosen indexes.

As a result, the model generates long enough Monte Carlo line of scenarios that contain volatility based correlated randoms for each time slice of the selected Indexes, pulled from the limited historical data.

Conclusions

In conclusion, it can be concluded that Monte Carlo simulation is particularly applicable to the manufacturing business, as it is often associated with random variables. It is used to assess the likelihood of cost overruns on large projects. Having enough historical data for each variable is a luxury for the management. That is why, analytics is using such mathematical models to create long enough line of data to run reliable Simulations that can help to determine the probability of the cost change and can help reduce the risk.

The use of Monte Carlo simulation can be useful as a window into the potential future of production costs. But it should not be taken as an absolute truth. It is a tool that would help in making asset allocation decisions, but it is important to remember that markets can and likely will be volatile and unpredictable.

References

1. E. Indarwati and R. Kusumawati. Estimation of the Portfolio Risk from Conditional Value at Risk Using Monte Carlo Simulation. Jurnal Matematika Statistika dan Komputasi, vol.17(3), 2021, pp.370-380, DOI: 10.20956/j.v17i3.11340
2. J. Mrkonja and A. Zaimovic. Monte Carlo simulacija u funkciji mjerenja tržišnog rizika Value-At-Risk metodom Using a Monte Carlo simulation in measuring market risk with Value-at-Risk method. Book. Publisher: Ekonomski fakulet u Sarajevu, Univerzitet u Sarajevu, 2021
3. Z. Zhang and X. Tu. The improvement of Monte Carlo method in VAR calculating. Journal of Computational Information Systems, vol. 4(5), 2008, pp.2375-2380
4. K. Vania and R. Gor. Back-testing approaches for validating var models. International Journal of Engineering Science Technologies, vol.6(6), 2022, pp. 9–18
5. A. Benhamed and M. Gassouma. Investigation and Modelling of Economic Systematic Risk and Capital Requirement: A Monte Carlo Simulation, Journal of Risk and Financial Management 16(4):241, 2023, DOI: 10.3390/jrfm16040241
6. P. Brandimarte. Handbook in Monte Carlo Simulation Applications in Financial Engineering, Risk Management, and Economics. Publ. Wiley, 2022
7. A. Pereira and A. Gomes. Assessing Budget Risk with Monte Carlo and Time Series Bootstrap. U Porto Journal of Engineering 9(1), 2023, pp:1-15
8. D. Sinha and S. Chakrabarty. A review of efficient Multilevel Monte Carlo algorithms for derivative pricing and risk management, MetodsX, Vol.10, 2023, DOI: 10.1016/j.mex.2023.102078
9. https://www.wallstreetmojo.com/norm-s-inv-function-in-excel/