ASPECTS OF STATISTICS IN TERMS OF FINANCIAL MODELLING AND RISK

Daniel Buc – Tomáš Klieštik

Abstract

Financial modelling is the task of building an abstract representation - a model of a financial decision-making situation. This is a mathematical model designed to represent the performance of a financial asset or portfolio of a business, project, or any other investment in a simplified version. Financial modelling is a general term that means different things to different users; the reference usually relates either to accounting and corporate finance applications, or to quantitative finance applications. It is necessary to model the financial assets development in continuous time to derive relations and perform relevant econometric calculations such as data simulation for different financial scenarios. Continuous time is a prerequisite for financial model that simulate changes in prices of financial assets, despite the fact that these prices are observed in discrete time moments. The random process is defined as a set of random variables on the same likelihood level indexed by time values from a set of real time line. Random development in time is often described as a stochastic process. For models in continuous time, the additions are in form of Wiener process, which belongs to Markov processes and represents an essential element of all stochastic processes of financial models.

Key words: random variable, model, asset, development, distribution

JEL Code: C53, G10, G17

Introduction

The accuracy of risk analysis depends largely on the correct application of the statistical distribution that accurately represents the uncertainty and variability of a particular problem. The improper use of the statistical distribution is a common disorder of risk analysis models. It is partially based on incorrect or insufficient understanding of the function and meaning of the statistical distribution and also of negligence of a specific chain effect that represents a result of improper distribution. Probability partition of a random variable is a rule whereby to each phenomenon described by this variable is associated a particular probability. This rule is also called the statistical distribution. The specific statistical distribution of a random variable is obtained by assigning a probability to each value of the discrete random variable, respectively to interval values of the continuous random variable. Four basic general methods which can be utilized for generating discrete rv's were introduced several decades ago, and since then intensive research has been dedicated to developing efficient algorithms implementing these methods. These four basic methods are: the table lookup method, the alias method, the inverse transform method and the acceptance–rejection method. Other methods which have been utilized can be viewed as variations (modifications) or combinations of these four. (Shmerling, 2013)

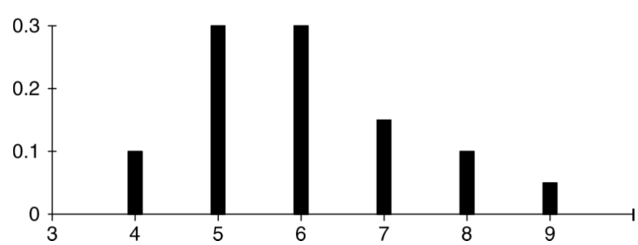
1 Useful statistical distributions for financial modelling

Statistical distribution can also be seen as a view that assigns to each elementary phenomenon a particular real number that determines the likelihood of this phenomenon. A continuous random variable (CRV) on a dcpo X is a pair (v; f), where v is a continuous valuation on Ω, and f is a continuous map from supp v to X. (Larrecq & Varacca, 2011) Statistical distributions of random variables are used as probabilistic models in describing the specific substantive issues, for example they are necessary to calculate future price changes of yields.

1.1 Discrete statistical distribution

This kind of distribution can reach one of several identifiable values, where each can calculate the probability of occurrence of a random phenomenon. It could be used as the parameter modelling in determining the number of key staff to be employed or the number of people visiting the particular place within an hour. The variables in the statistical distribution reach specific values-integers, it is not practicable to speculate for example with 2.5 persons.

Fig. 1: Visualization of a discrete statistical distribution



Source: self processed according to Vose, 2009

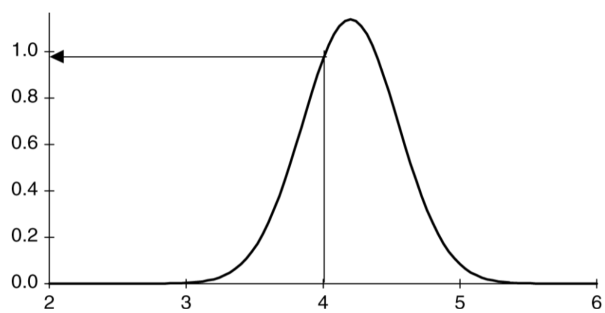
The vertical axis represents the actual probability of occurrence-random variables. The sum of these values must be 1.The horizontal axis is the number of examination subjects (persons, cash units, etc.). There is an example of a discrete statistical distribution of the number of bridges to be built on a particular section of highway in the Figure 1. There is a probability 0.30, that it will be built six bridges, 0.10 probability of building eight bridges etc. The counting result of these probabilities (0.10 + 0.30 + 0.30 + 0.15 + 0.10 + 0.50) is 1.

1.2 Continuous statistical distribution

Random variable X has a uniform distribution if the constant probability density is for all $x \in M$. If $M=\{x, \alpha<x<\beta\}$. Continuous statistical distribution is used to represent a random variable that can reach any value within a specified range (interval). For example, when determining the person´ s height a continuous statistical distribution could be used, because the particular height of person is infinitely divisible. It can be measured accurate to centimeters, millimeters, tenth of a millimeter etc. The scale can be divided again and will create more possible values. Random variables such as time, distance and weight, i.e. those that can be infinitely divided, are modelled by continuous statistical distributions. In practice, the continuous probability distribution is used also for modelling of variables that are essentially discrete, but the difference between the permissible values is negligible.

Fig. 2: Visualization of a continuous statistical distribution

2



Source: self processed according to Vose

1.3 Normal statistical distribution

Normal probability distribution (Gauss distribution) is one of most used and most important statistical distributions. It has an important role in probability theory and mathematical statistics. It is used as a probabilistic model serves of a large number of random events behaviour. It is used where random variable fluctuation is caused by sum of a large number of mutually independent and minor impacts. It sense is that under particular conditions it approximate well a number of other probability distributions (continuous and discrete). Normal probability distribution is denoted as N (μ, σ2), where μ is a mean value of random variable characterizing the location of this distribution and σ2 is the variance of the distribution of a random variable r_t around the mean. It is being understood that the random variable r_t has the normal probability distribution, if its probability density function has the form:

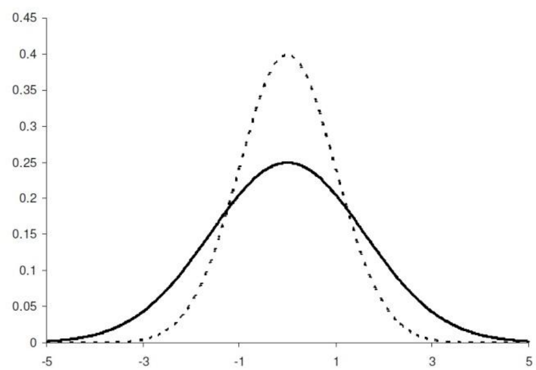
$$f(r_t) = \frac{1}{\sqrt{2\pi\sigma^2}} * e^{-\frac{1}{2\sigma^2}*(r_t-\mu)^2} \qquad (1)$$

where the mean μ also represents the median and mode of a normal distribution and belongs to the interval <- ∞, ∞> and σ^2 and belongs to the interval <0, ∞>. Normal distribution parameters have this mathematical form:

$$\mu = E[r_t], \qquad (2)$$

$$\sigma^2 = E\lfloor(r_t - \mu)^2\rfloor. \qquad (3)$$

Fig. 3: Visualization of a normal statistical distribution



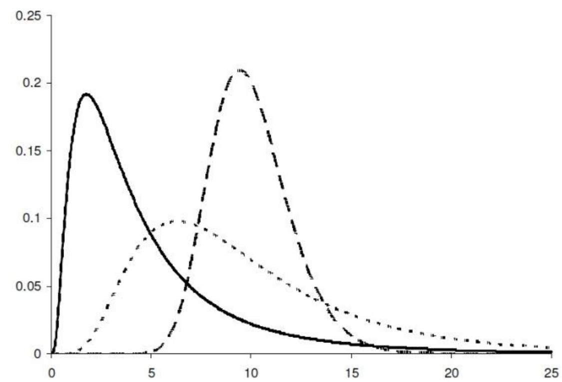
Source: self processed according to Klieštik, 2013

Most modelling of jointly Gaussian (normal) random variables involves the specification of a structure on the mean and the covariance matrix K. However, models which specify structure on K^{-1} have also been developed, although they are seemingly less popular. (Speed and Kiiveri, 1986)

1.4 Log-normal statistical distribution

Lognormal distributions often provide a good representation for a physical quantity that extend from zero to + infinity and is positively skewed, perhaps because some Central limit Theorem type of process is determining the variable's size. Lognormal distributions are also very useful for representing quantities that are thought of in orders of magnitude. For example, if a variable can be estimated to within a factor of 2 or to within an order of magnitude, the Lognormal distribution is often a reasonable model. (Vose, 2009)

Fig. 4: Visualization of a log-normal statistical distribution



Source: self processed according to Vose, 2009

Log-normal distribution is the distribution used for modelling in finance, where the random variable reaches only non-negative values and the shape of the distribution is asymmetric to its peak.

Random variable r_t has log-normal probability distribution, where the random variable *ln* (r_t) has distribution *N (μ; σ2)*. Random variable *ln* (r_t) represents the logarithmic price P_t, where the price P_t is determined by the price P_{t-1}. Exponential form of a probability density of a random variable r_t has the shape:

$$f(P_t) = \frac{1}{P_{t-1}\sqrt{2\pi\sigma^2}} * e^{-\left[\frac{(P_{t-1}-\mu)^2}{2\sigma^2}\right]} \tag{4}$$

where $P_{t-1} > 0$. Log-normal distribution parameters have this mathematical form:

$$\mu = E[P_t] = e^{\left(u + \frac{\sigma_t^2}{2}\right)} \tag{5}$$

$$\sigma^2(P_t) = \left\lfloor e^{2\mu} * e^{2\sigma_t^2} - e^{\sigma_t^2} \right\rfloor \tag{6}$$

By the practical use of this distribution practice the process is as follows: Random variable r_t is firstly converted into the form of *ln* (r_t) and then it continues as by the normal distribution.

2 Modelling the development of financial prices and yields

There are a lot of principles, methods and models used to predict financial prices and yield. One of the most favourite is the methodology CorporateMetrics which includes models such as Random Walk, Mean-reversion. Recent studies show that mean-reversion in equity returns induces positive hedging demands and economically significant market timing opportunities (Koijen, Rodriguez and Sbuelz, 2009). Valuation models for derivative and fixed-income securities have changed risk management and investment practice in significant ways. Examples include the Black and Scholes (1973), the Cox, Ross, and Rubinstein (1979) binomial option-pricing model, the Vasicek (1977), the Brennan and Schwartz (1979), the Cox, Ingersoll, and Ross (1985), and the Heath, Jarrow, and Morton (1992) bond valuation models. (Chen and Bakshi) The next part includes description and specifications of Random Walk and Mean-reversion.

2.1 Random Walk definition

The random walk theory is based on the fact that market and securities prices are random and not influenced by past events. This theory also states that all methods of predicting stock prices are futile in the long run. Professor Burton G. Malkiel calls the notion of intrinsic value[1] undependable because it relies on subjective estimates of future earnings using factors like expected growth rates, interest rates, dividend payouts, estimated risk, and others. The random walk model is generally not consistent with the stochastic behaviour of weekly returns, especially for the smaller capitalization stocks (Lo and MacKinlay, 1988).

2.2 Random walk features

In modern econometric and financial theory, stochastic differential equations are successfully used to describe the financial market and macroeconomic indicators development. Stochastic (linear trend) component of the differential equation is represented by Brownian motion or otherwise mathematically Wiener process belonging to Lévy processes which various modifications due to their features excellently approximate volatility and trend of asset prices. Random path model is named as Random Walk or a geometric Brownian motion.

2.3 Mathematical explanation of Random Walk

It is possible to model for example a random prices and exchange rates development through Random Walk model. This way of modelling has no tendency to return to its mean value. Since the sum of independent normal variables is normal, from normal invariants follows a normally distributed random walk, i.e. a Brownian motion. (Meucci, 2009)

If the considered asset P with initial price P_t where the random component is taken into account, we express the dynamics of asset prices at the time, which is expressed by the stochastic differential equation

[1] Intrinsic value is calculated by summing the future income generated by the asset and discounting it to the present value.

$$dP_t = \mu * P_t * dt + \sigma * P_t * dz, \qquad (7)$$

that represents a geometric Brownian motion defined as Ito process for $t \geq 0$, μ is the growth rate (average yield), σ is the volatility. This process can be written in a way which indicates that the yields of particular asset are modelled by using the deterministic (drift) and a random (diffusion) component in the form

$$\frac{dP_t}{P_t} = \mu * dt + \sigma * dz \qquad (8)$$

In some methodologies that are focused on risk quantification and diversification the logarithmic prices $p_t = \ln P_t$ are modeled that have normal distribution in the form $N(0; \sigma^2)$. Geometric Brownian motion with logarithmic prices can be then described as

$$dp_t = d \ln P_t = \tau * dt + \sigma * dz, \qquad (9)$$

where the logarithmic price has the deterministic coefficient τ in form $\tau = \mu - \frac{\sigma^2}{2}$ and random coefficient σ.

Deterministic coefficient τ represents the average rate of profit resulting from price growth of relevant asset. It is estimated as a regressive model through statistical estimation – least squares method,

$$\sum_{t=1}^{T} \varepsilon_t^2 \rightarrow min, \qquad (10)$$

where ε_t is residual deviation which indicates the difference between the actual and estimated yield

$$\varepsilon_t^2 = \sum_{t=1}^{T} (r_t - \tau * P_t)^2 \qquad (11)$$

The random coefficient σ determining the standard deviation is calculated as a square root from residual deviation,

$$\sigma = \sqrt{\frac{1}{N} * \sum_{t=1}^{T} \varepsilon_t^2} \qquad (12)$$

where N is the number of observations. The modelled price development according to Random Walk model is following

$$P_t = P_{t-1} * e^{\tau * dt} \qquad (13)$$

To estimate the price it is necessary to know the logarithmic price simulation, its mean value and dispersion. In case of Geometric Brownian motion with logarithmic prices, these calculations are relevant

$$P_t = P_{t-1} * e^{\tau * dt + \sigma * dz} \qquad (14)$$

$$E(P_t) = P_0 * e^{\tau * dt * n} = P_0 * e^{\tau * T} \qquad (15)$$

$$\sigma^2(P_t) = P_0^2 * e^{2 * \tau * T} * (e^{\sigma^2 * T} - 1) \qquad (16)$$

The value of the quantile on the likelihood level α from log-normal distribution, which determines the limits within which should the random variables move, has the formula:

$$P_T^{\gamma} = P_0 * e^{\tau * dt * n + \Phi^{-1} * \alpha * \sigma \sqrt{T}} \qquad (17)$$

Asset prices in financial markets behave randomly and independently of previous development, Brownian motion is thus an ideal tool to describe the behaviour of asset prices.

2.4 Mean-reversion model

For commodity prices and interest rates, sometimes also for exchange rates, the model Mean-reversion has better explanation and economic sense than geometric Brownian model. Interest rates and commodity prices have certain specifications, e.g. moving in a certain range, usually do not grow to infinity and even decrease to zero, they tend to return to particular equilibrium value.

Mean-reversion model is a stochastic process, which tends to return in time to the long-term equilibrium or stay close to long-run equilibrium. We can include to this mean value the historical average prices or yields, or other reasonable average such as economic growth. The basis of Mean-reversion models is the assumption that the process r_t is Ito´ s (diffusion) process

$$dr_t = a * (r_t, t) * dt + b * (r_t, t) * dz \qquad (18)$$

These models contain specific Wiener process with drift coefficient $a * (r_t, t)$ which also represents the approaching parameter speed to the long-term equilibrium, and with diffusion coefficient $b * (r_t, t)$ which reflects the long-term equilibrium.

If the selected model of financial data can be estimated, it is possible to estimate the continuous yield curves from limited number of data, or perform various simulations of yield curves, the instantaneous rate of profit and others.

Mean reversion contributes strongly to reducing long-horizon variance, but it is more than offset by various uncertainties faced by the investor, especially uncertainty about the expected return. (Pástor and Stambaugh, 2012)

In practice, the most commonly used stochastic models of interest rates are the following:

- special cases of single-factor interest rate model that includes only one interest factor, for example Vasicek model, Cox-Ingersoll-Ross, Hull-White model, Black-Karasinski model, and so on,
- binomial tree models, for example Randleman-Bartter model, Jarrow-Rud model,
- multifactor interest rate models involving more models at once, e.g. Brennan-Schwartz model, Fong-Vasicek model, Longstaff-Schwartz model.

Conclusion

There are objective criteria and procedures used to select the model, when each choice can be quantitatively evaluated and compared with another choice. The most common criteria include: information criteria, predictive criteria and selective iterative methods. The most frequently used criteria are predictive criteria that are used to select the model by comparing the forecasting ability of each selection variant.

Acknowledgment

The article is an output of scientific project VEGA 1/0357/11 Klieštik, T. and col.: Research on the possibility of applying fuzzy-stochastic approach and CorporateMetrics as tools of quantification and diversification of business risk.

References

[1] CHEN, Z., BAKSHI, G. 2005. *Stock Valuation in Dynamic Economics*. Journal of Financial Markets, Volume 8, Issue 2: 122-130.

[2] CISKO, Š., KLIEŠTIK, T. 2013. *Finančný manažment podniku II.* (1st ed.). Žilina: EDIS Publishers: 50-63.

[3] KOIJEN S.J. R., RODRIGUEZ, J.C., SBUELZ, A. 2009. *Momentum and Mean Reversion in Strategic Asset Allocation*, Management Science, vol. 55, no. 7: 1199-1213.

[4] LARRECQ J.G., VARACCA D. 2011. *Continuous Random Variables*. LICS: 55-70.

[5] LO, W. A., MACKINLAY, C. A. 1988: *Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test*, The Review of Financial Studies, Volume 1, number 1: 41-66.

[6] MEUCCI, A. 2009. *Review of Discrete and Continuous Processes in Finance: Theory and Applications*, SYMMYS: 32-49.

[7] PÁSTOR Ľ., STAMBAUGH F.R. 2012. *Are Stocks Really Less Volatile in the Long Run?* The Journal of Finance, Volume 67, Issue 2: 81-96.

[8] SHMERLING E. 2013. *A range reduction method for generating discrete random variables.* Statistics & Probability Letters Volume 83, Issue 4: 203-217.

[9] SPEED, T.P., KIIVERI, H.T. 1986 *Gaussian Markov Distributions over Finite Graphs.* Annals of Statistics, Volume 14, Number 1: 64-78.

[10] VOSE, D. 2009. *A Compendium of Distributions.* [ebook]. Vose Software, Available from www.vosesoftware.com, 38-63.

Contact

Ing. Daniel Buc
Department of Economics
Faculty of Operation and Economics of Transport and Communications
University of Žilina in Žilina
010 26 Žilina, Slovakia
E-mail: daniel.buc@fpedas.uniza.sk

doc. Ing. Tomáš Klieštik, PhD.
Department of Economics
Faculty of Operation and Economics of Transport and Communications
University of Žilina in Žilina
010 26 Žilina, Slovakia
E-mail: tomas.kliestik@fpedas.uniza.sk