Heavy-tailed modeling of CROBEX

DANIJEL GRAHOVAC, PhD*
NENAD ŠUVAK, PhD*

Article**
JEL: C15, C22, C51
doi: 10.3326/fintp.39.4.4

* The authors would like to thank two anonymous referees for their helpful comments and suggestions.
** Received: June 1, 2015
Accepted: October 11, 2015

The article was submitted for the 2015 annual award of the Prof. Dr. Marijan Hanžeković Prize.

Danijel GRAHOVAC
J. J. Strossmayer University of Osijek, Department of Mathematics, Trg Ljudevita Gaja 6, 31000 Osijek, Croatia
e-mail: dgrahova@mathos.hr

Nenad ŠUVAK
J. J. Strossmayer University of Osijek, Department of Mathematics, Trg Ljudevita Gaja 6, 31000 Osijek, Croatia
e-mail: nsuvak@mathos.hr

Abstract

Classical continuous-time models for log-returns usually assume their independence and normality of distribution. However, nowadays it is widely accepted that the empirical properties of log-returns often show a specific correlation structure and deviation from normality, in most cases suggesting that their distribution is heavy-tailed. Therefore we suggest an alternative continuous-time model for log-returns, a diffusion process with Student's marginal distributions and exponentially decaying autocorrelation structure. This model depends on several unknown parameters that need to be estimated. The tail index is estimated by the method based on the empirical scaling function, while the parameters describing mean, variance and correlation structure are estimated by the method of moments. The model is applied to the CROBEX stock market index, meaning that the estimation of parameters is based on the CROBEX log-returns. The quality of the model is assessed by means of simulations, by comparing CROBEX log-returns with the simulated trajectories of Student's diffusion depending on estimated parameter values.

Keywords: log-return, heavy-tailed distribution, Student's distribution, diffusion process, geometric Brownian motion

1 INTRODUCTION

CROBEX is the official stock market index of the Zagreb Stock Exchange, first published on September 1, 1997. The index is based on the free float market capitalization and includes the stocks of 25 companies. CROBEX serves as the main indicator for the Croatian stock market and closely describes the economic trends of the country. Like any other stock market index, CROBEX can be studied as the value (price) of a risky asset at some time point. A realistic modeling of a risky asset price through time is of great practical importance, especially in the risk assessment and pricing of financial derivatives. For this purpose, the values of the financial asset in some time interval can be considered as a realization of some stochastic process $(P_t, 0 \leq t \leq T)$. Many different classes of processes have been proposed as models for (P_t), both in discrete and continuous time (see e.g. Tsay, 2010 for an overview). Instead of the price process P_t, $t = 0, 1, ..., T$, the financial time series data are usually investigated through the log-returns of the original series, that is:

$$R_t = \ln \frac{P_t}{P_{t-1}}, \; t = 1, \ldots, T.$$

The classical model for the price of the risky asset is a geometric Brownian motion (GBM), also known as the Black-Scholes model. If the asset value is assumed to follow the GBM model, then the log-returns R_t, $t = 1, ..., T$, form a sequence of independent normally distributed random variables. This feature of the log-returns is nowadays considered unrealistic for many financial data. Contrary to the GBM model, the log-returns of many risky assets exhibit very weak correlation, but are far from independent. Moreover, the distribution of log-returns has tails much heavier than Gaussian, thus showing that extreme events are more probable than in the GBM model. Models taking this into account use the so-called heavy-tailed distributions which have tail probabilities decaying at infinity as slow as the power function. The importance of heavy-tailed distribution lies in the fact that they can realistically quantify the probabilities of extreme events. Such events are especially important in financial modeling while ignoring the possibility of large fluctuations often leads to a severe underestimation of risk. More details on these and other "stylized facts" of log-returns can be found in Cont (2001).

In this paper we analyze the historic data of the CROBEX index. First, we show that the log-returns of CROBEX exhibit behavior characteristic for a risky asset. In particular, we present evidence that the underlying distribution of log-returns is heavy-tailed and far from Gaussian. Awareness of such property is of great importance in risk assessment. In the next step, we claim that the distribution of log-returns can be successfully modeled with the Student's *t*-distribution, which is heavy-tailed. Many empirical studies have confirmed this for other financial data (see e.g. Hurst and Platen, 1997). Since this makes a standard GBM model for the asset price inappropriate, in section 3.2 we propose a new model for the log-returns based on the Student's diffusion process. Diffusion processes have been successfully used before in financial modeling (see Bibby and Sørensen, 1996; and Rydberg, 1999). The model proposed here uses a stationary solution of the diffusion stochastic differential equation which has Student's marginal distribution. Not only is the distribution modeled more realistically, but the dependence structure is also allowed to be more complex, since the constructed Student's diffusion process exhibits a form of weak dependence. In section 4 we estimate the parameters of the proposed model by using some recently introduced techniques. We tackle the statistically challenging problem of estimating the tail index of the log-returns which is the main parameter of any heavy-tailed distribution. It is worth mentioning that the estimation method used is non-parametric, in the sense that the tail index parameter is estimated without an assumption of the particular form of the underlying distribution. Estimation of other parameters of the model is also conducted with a brief discussion on the asymptotic properties of the estimators used. The quality of the proposed model is assessed by the means of simulations. Section 5 contains some concluding remarks and possible improvements with some guidelines and indications of further applications of the results.

The heavy-tailed nature of the CROBEX index has been addressed so far in several references. In Žiković and Pečarić (2010), left and right-hand CROBEX tails are fitted to separate generalized Pareto distributions, and such a model proved to be successful in forecasting some risk measures. The same distribution is advocated in Arnerić, Lolić and Galetić (2012). Since the method we use here in assessing the tail index is non-parametric, it provides a more robust estimate. Most frequently traded stocks included in the CROBEX were modeled by GARCH process with Student's innovations in Arnerić, Jurun and Pivac (2007), confirming the heavy-tailed structure of CROBEX. A similar model has been considered in Miletić and Miletić (2015) for CROBEX and other stock indices of the Central and Eastern European capital markets. On the global level, there is an inexhaustive list of papers dealing with modeling of stock market indices. Comprehensive empirical studies analyzing distribution of log-returns were done in Gray and French (1990), Hurst and Platen (1997), Jondeau and Rockinger (2003); see also Tsay (2010) and references therein.

2 THE CLASSICAL MODEL AND HEAVY-TAILED DISTRIBUTIONS

In empirical finance, the classical model refers to the price process of the risky asset modeled with the GBM. GBM is a continuous time stochastic process

$$S_t = S_0 \exp\left\{\left(\alpha - \frac{\sigma^2}{2}\right)t + \sigma W_t\right\}, 0 \le t \le T.$$

Here S_0 is the initial price and $(W_t, 0 \le t \le T)$ denotes the standard Brownian motion on $[0, T]$, that is a process with stationary independent increments, continuous sample paths and such that W_t is Gaussian (normal), $W_t \sim \mathcal{N}(0, t)$. The parameter $\alpha > 0$ can be interpreted as the expected rate of return and parameter $\sigma > 0$ as volatility, and therefore one of the indicators of the riskiness of the asset.

Instead of the price process P_t, $t = 0, 1, ..., T$, the financial time series data is usually investigated through the log-returns of the original series. More precisely, log-return at time t is defined as

$$R_t = \ln \frac{P_t}{P_{t-1}}, \ t = 1, \ldots, T.$$

Log-returns are scale independent quantities and can usually be plausibly modeled as a stationary sequence. Moreover, there is no loss of information as knowing the log-returns values and the initial price P_0 gives the price at time T by the equation:

$$P_T = P_0 \exp\left\{\sum_{t=1}^{T} R_t\right\}. \tag{1}$$

The advantage over the usual returns $(P_t - P_{t-1})/P_t$ is that R_t is additive in the sense of (1) and usually stationary.

If the asset value is assumed to follow the GBM model, then the log-returns are

$$R_t = \left(\alpha - \frac{\sigma^2}{2}\right) + \sigma \Delta W_t, \ t = 1, \ldots, T, \tag{2}$$

where $\Delta W_t = W_t - W_{t-1}$ are one-step increments of the Brownian motion. This means that R_t, $t = 1, ..., T$, is a sequence of independent normally distributed random variables, more precisely

$$R_t \sim \mathcal{N}\left(\alpha - \frac{\sigma^2}{2}, \sigma^2\right).$$

As discussed in the Introduction, it is highly unlikely to encounter this property in many financial data and there is a need for using heavy-tailed distributions.

Heavy-tailed distributions are of considerable importance in modeling a wide range of phenomena in finance and many other fields of science. Prominent examples of such distributions are Pareto distribution, stable distribution and Student's t-distribution. Distribution of some random variable X is said to be heavy-tailed with index $\alpha > 0$ if its tail probabilities decay as a power law, i.e.

$$P\left(|X|>x\right) = \frac{L(x)}{x^{\alpha}},$$

where $L(t)$, $t > 0$ is a slowly varying function, that is, $L(tx)/L(x) \to 1$ as $x \to \infty$, for every $t > 0$. In particular, this implies that $E|X|^q < \infty$ for $q < \alpha$ and $E|X|^q = \infty$ for $q > \alpha$. The parameter α is called the tail index and measures the "thickness" of the tails. The lower the value of α is, the more probable are extreme values of X. This way extreme events can be modeled and these events are usually the most important as they can generate great profit but also, more importantly, catastrophic loss. On the other hand, the usual Gaussian distribution has tail probabilities that decay exponentially fast as $\sim e^{-x^2/2}$ when $x \to \infty$. For this reason, Gaussian distribution is inadequate for modeling phenomena that can exhibit extreme behavior.

Pioneering work in applying heavy-tailed models to finance was done by B. Mandelbrot in Mandelbrot (1963) where stable distributions have been advocated for describing fluctuations of cotton prices. Stable distributions allow for tail index value $0 < \alpha < 2$, which is nowadays considered an unrealistically small value for most time series data. A richer modeling ability is provided by the Student's t-distribution, which allows for arbitrary tail index parameter.

2.1 THE CROBEX LOG-RETURNS

The data considered in this paper consist of 2524 closing values of the CROBEX stock market index collected in the period from January 3, 2005 until December 31, 2014. The time series of values is shown in figure 1(a). This and all other figures in the paper are made from the publicly available CROBEX data with Wolfram Mathematica software.

The log-returns of the CROBEX index are shown in figure 1(b). From the appearance of the plot, it seems plausible to model the log-returns as a realization of the stationary sequence of random variables $R_1, \ldots, R_T$ with $T = 2523$.

As the first step of the analysis, we investigate the underlying distribution of the sequence $R_1, \ldots, R_T$. For this purpose, a histogram is plotted in figure 2(a) and one can see that it has sharper peak and tails heavier than Gaussian distribution. These characteristics are known as the stylized facts of asset returns and are common for almost all data of this type. Heavy-tails of the underlying distribution are confirmed with the QQ-plot of normal quantiles on the x-axis with respect to the empirical quantiles on the y-axis (figure 2(b)). The left end of the pattern is below the reference line and the right end of the pattern is above the line which indicates tails heavier than Gaussian. Further evidence of the heavy-tailed nature of CROBEX will be given in subsection 4.1 where the tail index will be estimated.

Figure 1

CROBEX data in period January 3, 2005 – December 31, 2014

(a) CROBEX values *(b) Log-returns*

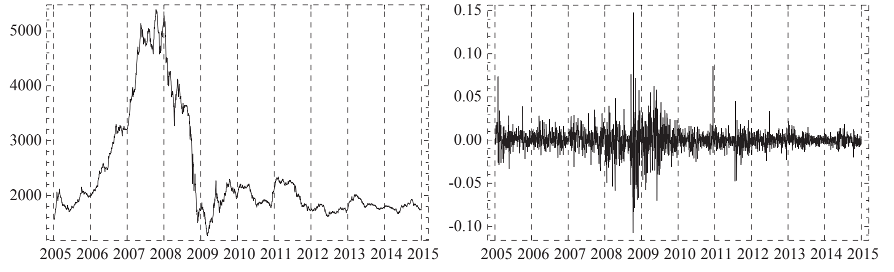
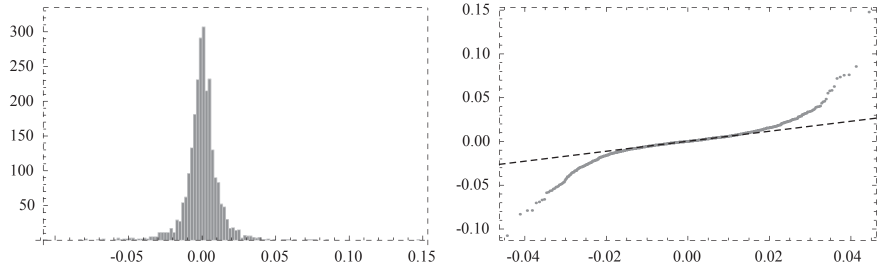


Figure 2

Distribution analysis of CROBEX log-returns

(a) Histogram of log-returns *(b) QQ-plot of log-returns*



3 STUDENT'S DIFFUSION AS MODEL FOR TIME EVOLUTION OF LOG-RETURNS

GBM, the classical model for stock prices and values of the stock market indices, implies both independence and normality of distribution of log-returns (equation (2)). The log-returns on financial markets usually are not in correspondence with these demands, i.e. over a long time period they often show a specific correlation structure and a deviation from normality. In most cases their distribution exhibits heavy tails and for CROBEX this was indicated in section 2. A natural heavy-tailed generalization of the Gaussian distribution is provided by Student's *t*-distribution, which we now introduce.

3.1 STUDENT'S DISTRIBUTION

Student's distribution represents a natural choice for modeling the distribution of log-returns, because of its heavy-tailed characteristics and still close relationship with the normal distribution. In order to capture more information from the realized log-returns, we use Student's distribution with three parameters:

- shape parameter $\upsilon>0$ (also called the number of degrees of freedom),
- scale parameter $\delta>0$,
- location parameter $\mu \in \mathbb{R}$.

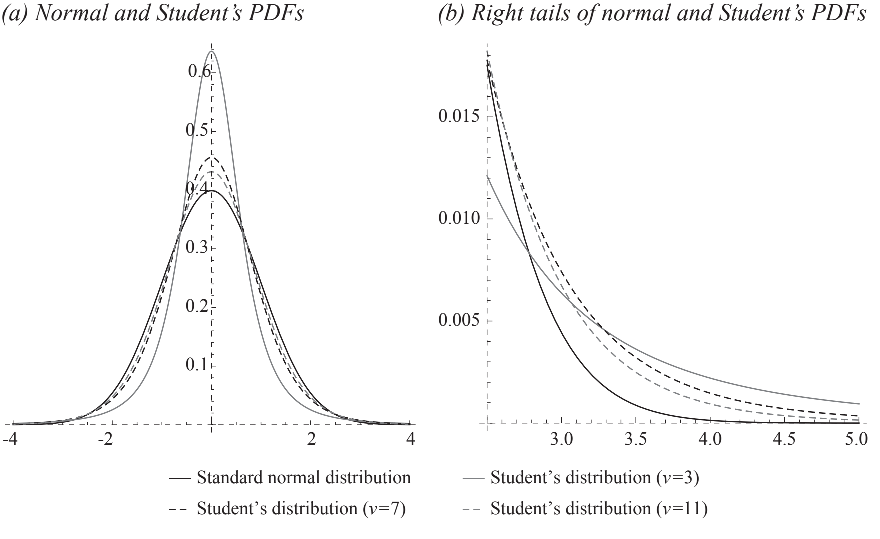
This distribution, usually denoted as $T(\upsilon, \delta, \mu)$, is defined by the probability density function

$$f(x)=\frac{\Gamma\left(\frac{v+1}{2}\right)}{\delta\sqrt{\pi}\,\Gamma\left(\frac{v}{2}\right)}\left(1+\left(\frac{x-\mu}{\delta}\right)^{2}\right)^{-\frac{v+1}{2}},\ x\in\mathbb{R}, \tag{3}$$

where $\Gamma(\cdot)$ denotes the classical gamma function (see Abramowitz and Stegun, 1964). If υ is an integer, then $T(\upsilon, \sqrt{\upsilon}, 0)$ coincides with the usual *t*-distribution widely used in statistics. For large values of the parameter υ Student's $T(\upsilon, \delta, \mu)$ distribution behaves approximately like the normal distribution. Probability density functions (PDFs) of standard normal distribution and Student's distributions with zero mean, unit variance and $\upsilon = 3$, $\upsilon = 7$ and $\upsilon = 9$ degrees of freedom are plotted in figure 3(a), while the right tails of all four PDFs are plotted in figure 3(b).

FIGURE 3

PDFs of standard normal distribution and Student's distributions with zero mean, unit variance and various degrees of freedom v



The left and the right-hand tails of Student's $T(\upsilon, \delta, \mu)$ distribution (3) decrease like $|x|^{-\upsilon-1}$, i.e. this distribution is heavy-tailed and the tail index corresponds to degrees of freedom, that is $\alpha = \upsilon$. In particular, moments of order greater than υ do not exist. The central moment of order n exists under the restriction $n < \upsilon$, $n \in \mathbb{N}$, and it is given by the following expression:

$$E\left[(R-E[R])^n\right]=\begin{cases}\frac{\delta^n}{\sqrt{\pi}}\,\Gamma\left(\frac{n+1}{2}\right)\Gamma\left(\frac{\nu-n}{2}\right)\left(\Gamma\left(\frac{\nu}{2}\right)\right)^{-1}, & n=2k-1\\ 0, & n=2k\end{cases},\quad k\in\mathbb{N}, \tag{4}$$

where R is the random variable with Student's $T(\upsilon, \delta, \mu)$ distribution, i.e. $R \sim T(\upsilon, \delta, \mu)$. We will be mainly interested in its expectation and variance:

$$E[R]=\mu,\ \nu>1;\ Var(R)=\frac{\delta^2}{\nu-2},\ \nu>2. \tag{5}$$

The model developed in this paper uses Student's $T(\upsilon, \delta, \mu)$ distribution as the marginal distribution of the stationary sequence R_t, $t = 1, ..., T$ of log-returns. Student's distribution is heavy-tailed and thus fits the usual empirical properties for the log-returns distribution. Additional parameters μ and σ allow more flexibility in modeling as they describe the mean and the variance when $\upsilon > 2$, i.e. when mean and variance exist. When $\upsilon \to \infty$, Student's distribution reduces to normal distribution and the log-returns would have the same distribution as in the standard GBM model. That distributions of log-returns can often be fitted extremely well by Student's distribution has been confirmed in many empirical studies; see Hurst and Platen (1997), Heyde and Liu (2001), Heyde and Leonenko (2005) and references therein.

3.2 STUDENT'S DIFFUSION MODEL

Here we propose a model that generalizes the model (1) for log-returns and incorporates Student's distribution (3). The obvious step in this direction would be to replace ΔW_t, $t = 1, ..., T$ in (1) by a sequence of independent random variables with Student's distribution. However, this would imply independence of log-returns which is an unrealistic property for the asset returns.

A more advanced model can be built by taking $(R_t, 0 \le t \le T)$ to be a stationary diffusion process such that $R_t \sim T(\upsilon, \delta, \mu)$ for some parameters υ, δ and μ and which exhibits the so-called β-mixing dependence structure with the exponentially decaying rate, meaning that the coefficient which in some way describes the dependence in the process $(R_t, 0 \le t \le T)$ tends to zero exponentially fast. This type of mixing implies another type of mixing called α-mixing or strong mixing, which is more frequently used in the studies of the dependence structures of stochastic processes. For more general details on mixing theory we refer to Bradley (2005), and Abourashchi and Veretennikov (2010). Such a choice will allow $(R_t, 0 \le t \le T)$ to have the exponentially decaying autocorrelation function $\rho(t) = \mathrm{Corr}(R_s, R_{s+t})$, $0 \le s < s + t \le T$, and heavy-tailed marginal distribution.

A stationary diffusion process with prescribed marginal distribution can be constructed as a solution of a particular stochastic differential equation (see Bibby, Skovgaard and Sørensen, 2005). For $v > 2$ Student's diffusion is a stochastic process satisfying the stochastic differential equation

$$dR_t = \theta(\mu - R_t)dt + \sqrt{\frac{2\theta\delta^2}{v-1}\left(1+\left(\frac{R_t-\mu}{\delta}\right)^2\right)}\, dW_t,\ 0 \le t \le T, \tag{6}$$

where $(W_t, 0 \le t \le T)$ is a standard Brownian motion and $\theta > 2$ is the so-called autocorrelation or dependence parameter appearing in the autocorrelation function

$$\rho(t) = \mathrm{Corr}(R_s, R_{s+t}) = e^{-\theta t},\ 0 \le s < s+t \le T. \tag{7}$$

Moreover, if Student's diffusion starts from Student's $T(v, \delta, \mu)$ distribution, i.e. if $R_0 \sim T(v, \delta, \mu)$, then $R_t \sim T(v, \delta, \mu)$ for all $t \in [0, T]$ and the process $(R_t, 0 \le t \le T)$ is said to be strictly stationary (all of its finite-dimensional distributions are invariant to time-shifts).

For R_t interpreted as the log-return at time t, equation (6) could be interpreted in view of the change of the log-return in a small time interval $[t, t + \Delta t]$, $0 \le t < t + \Delta t \le T$:

$$R_{t+\Delta t} - R_t = \theta(\mu - R_t)\Delta t + \sqrt{\frac{2\theta\delta^2}{v-1}\left(1+\left(\frac{R_t-\mu}{\delta}\right)^2\right)}\left(W_{t+\Delta t} - W_t\right). \tag{8}$$

Equation (8) relates the log-return $R_{t+\Delta t}$ at time $(t + \Delta t)$ to the log-return R_t at time t taking into account the increase of time Δt and the change of the value of the Brownian motion $(W_{t+\Delta t} - W_t)$, i.e. its increment between time points t and $(t + \Delta t)$.

Therefore, it could be interpreted as follows: the log-return $R_{t+\Delta t}$ could be obtained from the historical log-return R_t by adding to it the increase of time Δt with the factor $\theta(\mu - R_t)$ and the increment of the Brownian motion $(W_{t+\Delta t} - W_t)$ between time points t and $(t + \Delta t)$ with the factor

$$\sqrt{\frac{2\theta\delta^2}{v-1}\left(1+\left(\frac{R_t-\mu}{\delta}\right)^2\right)}, \tag{9}$$

where both factors depend on the historical log-return R_t, parameters v, δ and μ of the Student's $T(v, \delta, \mu)$ distribution and the autocorrelation parameter θ. Factor (9) can be understood as conditional variance of the Gaussian innovation term $(W_{t+\Delta t} - W_t)$, conditionally on the past values of R_t. This structure resembles the discrete time conditional heteroscedasticity models, e.g. ARCH models. It is well-known that there is an intimate relation between GARCH models and diffusions processes (see e.g. Fornari and Mele, 2000 and references therein).

4 FITTING THE CROBEX LOG-RETURNS TO STUDENT'S DIFFUSION

In view of the empirical properties of log-returns of the risky asset presented in section 2 and remarks on the suitability of Student's distribution for modeling the marginal distribution of log-returns presented in section 3, Student's diffusion (6) seems to be a plausible stochastic model for log-returns. In this section CROBEX log-returns will be fitted to Student's diffusion. More precisely, we derive estimators of parameters ν, δ, μ and θ and calculate their values based on the CROBEX data.

4.1 ESTIMATION OF UNKNOWN PARAMETERS

The parameter estimation problem will be treated in three separate but dependent steps. First, the parameter ν will be estimated as the tail index parameter by using the method based on the empirical scaling function recently introduced in Grahovac et al. (2015). This estimated value of ν will be treated as the known value of this parameter in the estimation of parameters μ and δ by the classical method of moments (see Serfling, 1980). Finally, the autocorrelation parameter θ will be estimated by the generalized method of moments based on Pearson's sample correlation function (see Leonenko and Šuvak, 2010).

4.1.1 Estimation of parameter ν

The shape parameter ν corresponds to the tail index of Student's distribution. Extreme value theory provides many methods for estimating the unknown tail index (see Embrechts, Klüppelberg and Mikosch, 1997 for an overview). Here we will use a novel approach introduced in Grahovac et al. (2015), based on the so-called empirical scaling functions. Suppose that we are given a zero mean sample $X_1, X_2, \ldots, X_n$, coming from some stationary heavy-tailed sequence with strong mixing property with an exponentially decaying rate. Partition function of this sample is defined as

$$S_q(n,t) = \frac{1}{\lfloor n/t \rfloor} \sum_{i=1}^{\lfloor n/t \rfloor} \left| \sum_{j=1}^{\lfloor t \rfloor} X_{(i-1)\lfloor t \rfloor + j} \right|^q,$$

where $q > 0$ and $1 \leq t \leq n$. Using this definition the empirical scaling function at the point q based on the points $s_i \in (0,1)$, $i = 1, \ldots, N$, can be defined by

$$\hat{\tau}_{N,n}(q) = \frac{\sum_{i=1}^{N} s_i \frac{\ln S_q(n, n^{s_i})}{\ln n} - \frac{1}{N} \sum_{i=1}^{N} s_i \sum_{j=1}^{N} \frac{\ln S_q(n, n^{s_j})}{\ln n}}{\sum_{i=1}^{N} (s_i)^2 - \frac{1}{N} \left(\sum_{i=1}^{N} s_i \right)^2}.$$

The estimation method is based on the asymptotic behavior of $\hat{\tau}_{N,n}$. One can show that for each $q > 0$, when $n, N \to \infty$, $\hat{\tau}_{N,n}(q)$ tends in probability to

$$\tau_\alpha^\infty(q) = \begin{cases} \frac{q}{\alpha}, & \text{if } q \le \alpha \text{ and } \alpha \le 2, \\ 1, & \text{if } q > \alpha \text{ and } \alpha \le 2, \\ \frac{q}{2}, & \text{if } 0 < q \le \alpha \text{ and } \alpha > 2, \\ \frac{q}{2} + \frac{2(\alpha-q)^2(2\alpha+4q-3\alpha q)}{\alpha^3(2-q)^2}, & \text{if } q > \alpha \text{ and } \alpha > 2, \end{cases}$$

where α is the tail index. This implies that the shape of the empirical scaling function depends on the value of the tail index. Since $\hat{\tau}_{N,n}(q)$ can be easily computed from the sample, this provides information on the unknown tail index. The asymptotic form τ_N^∞ is plotted in figure 4(a). For the heavy-tailed samples the empirical scaling function will approximately have the shape of the broken line. The break of the line occurs at point α. The limiting case $\alpha \to \infty$ corresponds to non heavy-tailed distributions and the scaling function would be a straight line $q/2$ (dotted in figure 4(a)). This way it is possible to detect heavy tails in data. Estimation can be done by fitting the empirical scaling function to its asymptotic form. Taking some points $q_i \in (0, q_{max})$, $i = 1, ..., M$, tail index can be estimated as

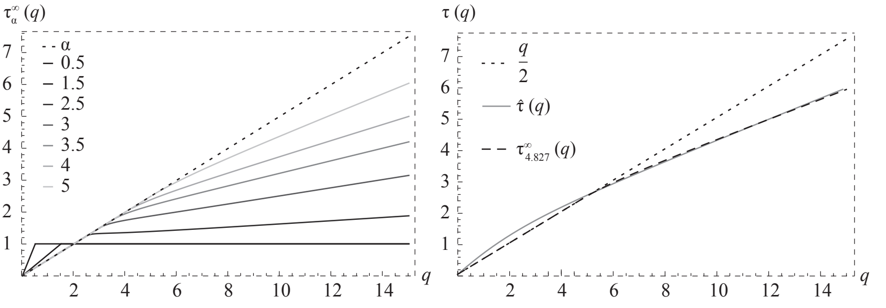
$$\hat{\alpha} = \arg\min_{\alpha \in (0,\infty)} \sum_{i=1}^{M} (\hat{\tau}_{N,n}(q_i) - \tau_\alpha^\infty(q_i))^2. \tag{10}$$

More details on the method can be found in Grahovac et al. (2015). It is important to note that the estimation does not depend on the particular form of the underlying distribution and the only assumption is that the sample comes from the class of heavy-tailed distributions, which in particular also includes Student's distribution.

The empirical scaling function computed on the sample of CROBEX log-returns $R_1, ..., R_n$ with $n = T = 2523$ is shown in figure 4(b). A clear departure from the line $q/2$ confirms that the log-returns are heavy-tailed. The scaling function has a shape of the broken line and breaks at around value 5. Computing the estimator by equation (10) gives the value $\hat{\alpha} = 4.827$. The estimated value appears as a break in the plot of the scaling function in figure 4(a). The plot of τ_N^∞ for $\alpha = 4.827$ (dashed in figure 4(b)) almost coincides with the empirical scaling function, confirming the quality of the estimate. Therefore, the estimated value of the shape parameter υ is $\hat{\upsilon} = 4.827$ and as such is consistent with many other studies that suggest that the tail index value of the asset returns is between 3 and 5 (see e.g. Hurst and Platen, 1997).

FIGURE 4
Scaling functions

(a) Asymptotic form *(b) Empirical for CROBEX data*



4.1.2 Estimation of parameters μ and δ

The problem of estimation of the location parameter $\mu \in \mathbb{R}$ and the scale parameter $\delta > 0$ is approached by assuming that υ is equal to its estimated value 4.827. Suppose that $R_1, ..., R_n$ is a random sample of n log-returns. Parameters μ and δ will be estimated by the classical method of moments in which estimators are obtained as solutions of the system of equations relating the theoretical moments to the corresponding empirical moments. Since μ and δ are parameters of the marginal distribution of Student's diffusion (6), estimators will be obtained by relating the expectation $E[R_t] = \mu$ and the second moment $E[R_t^2] = \delta^2/(\upsilon - 1) + \mu^2$ to the first and the second empirical moments

$$\bar{R}_n = \frac{1}{n}\sum_{k=1}^{n} R_k \text{ and } \overline{R^2}_n = \frac{1}{n}\sum_{k=1}^{n} R_k^2, \tag{11}$$

respectively. Solutions of this system of equations with respect to the unknowns δ and μ are the estimators of these parameters:

$$\hat{\delta} = \sqrt{(v-1)\left(\overline{R^2}_n - \bar{R}_n^2\right)} = \sqrt{(v-1)\left(\frac{1}{n}\sum_{k=1}^{n} R_k^2 - \left(\frac{1}{n}\sum_{k=1}^{n} R_k\right)^2\right)} \tag{12}$$

$$\hat{\mu} = \bar{R}_n = \frac{1}{n}\sum_{k=1}^{n} R_k. \tag{13}$$

Computing the values of estimators $\hat{\delta}$ and $\hat{\mu}$ based on the CROBEX log-returns results in estimated values of parameters δ and μ of Student's diffusion. Estimated values of all three parameters of the marginal distribution of Student's diffusion are given in table 1.

TABLE 1
Estimated values of parameters υ, δ and μ of Student's diffusion

Parameter	υ	δ	μ
Estimated value	4.827	0.025	0.00004

4.1.3 Estimation of parameter θ

Autocorrelation parameter $\theta > 0$ is estimated by the generalized method of moments based on the empirical autocorrelation function. Autocorrelation function (ACF) of Student's diffusion is well defined if $\upsilon > 0$, so its existence is assured by the estimated value 4.827 of the parameter υ (see table 1).

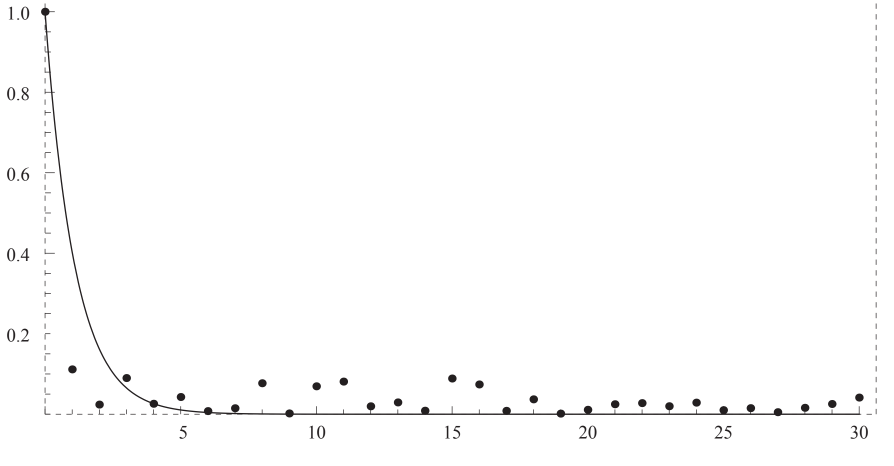
The empirical counterpart of the autocorrelation function (7) is given by the absolute value of Pearson's sample correlation function

$$\hat{\rho}_n(t) = \left| \frac{\frac{1}{n-t}\sum_{i=1}^{n-t} R_i R_{t+i} - \frac{1}{n-t}\sum_{i=1}^{n-t} R_i \cdot \frac{1}{n-t}\sum_{i=1}^{n-t} R_{t+i}}{\sqrt{\frac{1}{n-t}\sum_{i=1}^{n-t} R_i^2 - \left(\frac{1}{n-t}\sum_{i=1}^{n-t} R_i\right)^2} \cdot \sqrt{\frac{1}{n-t}\sum_{i=1}^{n-t} R_{t+i}^2 - \left(\frac{1}{n-t}\sum_{i=1}^{n-t} R_{t+i}\right)^2}} \right|, \quad (14)$$

where the term in the numerator represents the empirical covariance of random variables R_s and R_{t+s}, while the term in the denominator represents the product of the empirical standard deviations of random variables R_s and R_{t+s}, $0 \le s < s + t \le T$. Pearson's sample correlation function (or empirical ACF), plotted in figure 5 for lags t = 0, 1, ..., 30, shows autocorrelation in the time series of log-returns for small values of lag t and suggests the exponential decay of autocorrelations with respect to the lag t. Notice that after lag 10 the estimated correlations stabilize near zero, so the majority of the exponentially decaying autocorrelation structure of Student's diffusion is contained in these first few correlations. For alternative methods of estimation of the autocorrelation parameter based on a small number of lags we refer to Forman (2005).

Figure 5

Empirical ACF (dotted) and theoretical ACF (solid) for θ=0.91



For fixed t, the method of moments estimator for θ is derived by solving, with respect to the unknown parameter θ, the equation that relates the empirical autocorrelation function (14) to the theoretical autocorrelation function $\rho(t) = e^{-\theta t}$ given in (7). The estimator is given by the following expression:

$$\hat{\theta}(\mathrm{t}) = -\frac{1}{t}\ln\,\hat{\rho}_n(t). \tag{15}$$

Notice that for each lag t we obtain a single estimate of θ. Since the majority of the autocorrelation structure is held by the first 10 lags, to obtain just one estimate of the parameter θ we calculate the value of the estimator $\hat{\theta}(t)$ for $t = 1, ..., 10$ and the final estimate 0.91 for θ is obtained as the mean of these 10 values (see table 2). The theoretical autocorrelation function $\rho(t) = e^{-\theta t}$ for $\theta = 0.91$ is also plotted in figure 5.

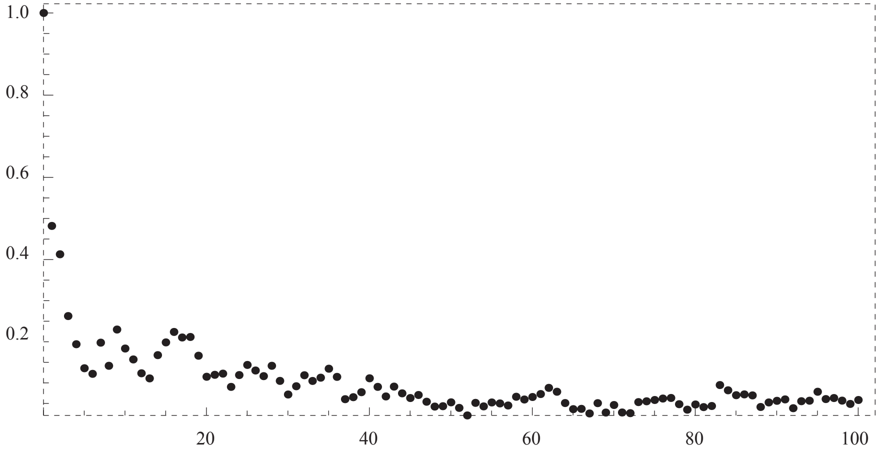
Table 2

Estimation of parameter θ of Student's diffusion

Lag t	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Value of $\hat{\theta}(t)$	2.19	1.87	0.80	0.91	0.63	0.81	0.60	0.32	0.69	0.27
Estimated value of θ	0.91									

Figure 6

Empirical ACF for the squared log-returns



Beside plotting the theoretical and the empirical ACF, the usual graphical method for exploring the dependence structure of log-returns is the ACF of their squares, which are often used as the volatility approximations. From nice theoretical properties of Student's diffusion and the methodology based on its eigenfunctions (orthogonal Routh-Romanovski polynomials, see Leonenko and Šuvak, 2010) it follows that the autocorrelation function of the squared log-returns, which is well defined for $\upsilon > 4$, is given by:

$$\mathrm{Corr}\left(R_s^2, R_{s+t}^2\right) = \frac{\dfrac{2\delta^4(\nu-1)}{(\nu-2)^2(\nu-4)}\,e^{-2\theta t\frac{\nu-2}{\nu-1}} + \dfrac{4\mu^2\delta^2}{\nu-2}\,e^{-\theta t}}{\dfrac{2\delta^4(\nu-1)}{(\nu-2)^2(\nu-4)} + \dfrac{4\mu^2\delta^2}{\nu-2}},\quad 0 \le s < s+t \le T. \tag{16}$$

Autocorrelation function (16) is exponentially decaying function of the lag t. From the empirical ACF of squared CROBEX log-returns, estimated by Pearson's sample correlation function (14) for squared data and plotted in figure 6, we see that it corresponds to the theoretically suggested exponential decay.

4.2 NOTES ON THE ASYMPTOTIC BEHAVIOR OF PARAMETER ESTIMATORS

In this section we briefly discuss the two most commonly analyzed asymptotic properties of estimators – consistency and asymptotic normality of the estimators $\hat{\theta}, \hat{\delta}, \hat{\mu}$ of parameters θ, δ, μ. Generally, estimator $\hat{\kappa}_n$, where n emphasizes its dependence on the number of observations, is a consistent estimator of the unknown parameter κ if the probability that the absolute deviation of $\hat{\kappa}_n$ from κ can be made arbitrarily small by choosing n large enough. Furthermore, estimator $\hat{\kappa}_n$ is asymptotically normal if for large n the standardized estimator $\hat{\kappa}_n$ has an approximately standard normal distribution. For more details on these properties of estimators we refer to Serfling (1980).

4.2.1 Consistency

It is well known that the first and the second empirical moments $\overline{R}_n$ and $\overline{R^2_n}$ given in (11) are consistent estimators of the first and the second theoretical moments and that the Pearson sample correlation function $\hat{\rho}_n(t)$ given by (14) is a consistent estimator of the autocorrelation function $\rho(t) = \mathrm{Corr}(R_s, R_{s+t})$, (see Serfling, 1980). Since estimators $\hat{\mu}, \hat{\delta}$ and $\hat{\theta}$ are continuous transformations of the estimators $\overline{R}_n$ and $\overline{R^2_n}$ and $\rho(t)$, from the continuous mapping theorem (see Serfling, 1980) it follows that $\hat{\mu}, \hat{\delta}$ and $\hat{\theta}$ are consistent estimators of parameters μ, δ and θ, respectively.

4.2.2 Asymptotic normality

Estimators $\hat{\mu}$ and $\hat{\delta}$ are continuous transformations of estimators $\overline{R}_n$ and $\overline{R^2_n}$ which are known to be asymptotically normal due to the β-mixing property of Student's diffusion (see Leonenko and Šuvak, 2010). Therefore, according to the delta-method (see Serfling, 1980) it follows that for $\upsilon > 4$ the bivariate estimator $(\hat{\mu}, \hat{\delta})$ is also asymptotically normal, i.e.

$$\sqrt{n}\left[\Sigma\left(\nu, \hat{\delta}, \hat{\mu}, \hat{\theta}\right)\right]^{-\frac{1}{2}}\left(\hat{\mu}-\mu, \hat{\delta}-\delta\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbf{I}\right),$$

where υ is supposed to be the known value of the tail index of marginal distribution of Student's diffusion and $\hat{\delta}, \hat{\mu}$ and $\hat{\theta}$ are consistent estimators of parameters δ, μ and θ given in (12), (13) and (15), respectively. The covariance matrix $\Sigma(\upsilon, \hat{\delta}, \hat{\mu}, \hat{\theta})$ could be represented as $D\Sigma D^{\tau}$, where

$$D = \begin{bmatrix} 1 & 0 \\ -\dfrac{\hat{\mu}(1-\nu)}{\hat{\delta}} & \dfrac{(\nu-1)}{2\hat{\delta}} \end{bmatrix},$$

and the elements of the matrix $\sum$ are as follows:

$$\sigma_{11}^2 = \frac{\hat{\delta}^2}{\nu - 2}\left(\frac{2}{e^{\hat{\theta}} - 1} + 1\right),$$

$$\sigma_{12}^2 = \frac{2\hat{\mu}\hat{\delta}^2}{\nu - 2}\left(\frac{2\hat{\delta}}{e^{\hat{\theta}} - 1} + 1\right) = \sigma_{21}^2,$$

$$\sigma_{22}^2 = \frac{4\hat{\delta}^4(\nu - 1)}{(\nu - 2)^2(\nu - 4)} \frac{1}{e^{2\hat{\theta}\frac{\nu-2}{\nu-1}} - 1} + \frac{8\hat{\mu}^2\hat{\delta}^2}{\nu - 2}\frac{1}{e^{\hat{\theta}} - 1}$$

$$+ \frac{3\hat{\delta}^4(\nu - 2) + 4\hat{\mu}^2\hat{\delta}^2(\nu - 2)(\nu - 4) - \hat{\delta}^4(\nu - 4)}{(\nu - 2)^2(\nu - 4)}.$$

For more details on the methodology of analysis of asymptotic properties of some of these estimators and calculation of explicit form of the covariance matrix $\sum(\upsilon, \hat{\delta}, \hat{\mu}, \hat{\theta})$ we refer to Leonenko and Šuvak (2010).

4.3 INFERENCE ON THE QUALITY OF THE MODEL

The quality of the model is examined by simulations. To obtain some objective indicators that relate CROBEX log-returns and Student's diffusion as the stochastic model for them, we simulated 1000 independent sample paths of this process by using estimated values of parameters υ, δ, μ and θ (see tables 1 and 2). Student's diffusion can be simulated using the so-called Milstein scheme for simulating paths of solutions of stochastic differential equations (see Iacus, 2009 for more details). The length of each simulated sample path coincides with the number of observed CROBEX log-returns. Several of these sample paths are plotted in figure 7.

In this setting for each time point t we deal with the sample of 1000 simulated data that will be used to describe the CROBEX log-return at t by the sample percentiles. More specifically, for each time point t we calculate the 5^{th} percentile, the lower quartile (25^{th} percentiles), median, upper quartile (75^{th} percentile) and 95^{th} percentile of the sample of 1000 data simulated for this exact time point. The values of these sample percentiles for each t, together with the time series of CROBEX log-returns, are plotted in figure 8.

Furthermore, we estimate probabilities that CROBEX log-returns fall outside the interquartile interval and the interval between the 5^{th} and the 95^{th} percentiles:

- 13.99% of CROBEX log-returns fall outside the interquartile interval – we can say that the probability that the CROBEX log-return is smaller than the lower quartile and greater that the upper quartile is estimated to be 0.1399,
- 2.46% of CROBEX log-returns fall outside the interval between the 5^{th} and the 95^{th} percentiles – we can say that the probability that the CROBEX log-

return is smaller than the 5th percentiles and greater that the 95th percentiles is estimated to be 0.0246.

These probabilities indicate that the simulated paths of Student's diffusion with parameters $\upsilon = 4.827$, $\delta = 0.025$, $\mu = 0.00004$ and $\theta = 0.91$ capture the time evolution of the historical values of CROBEX log-returns quite well. Moreover, there are no values of CROBEX log-returns that fall outside the interval between the minimal and the maximal simulated value of Student's diffusion at time t.

FIGURE 7

Simulated trajectories of Student's diffusion

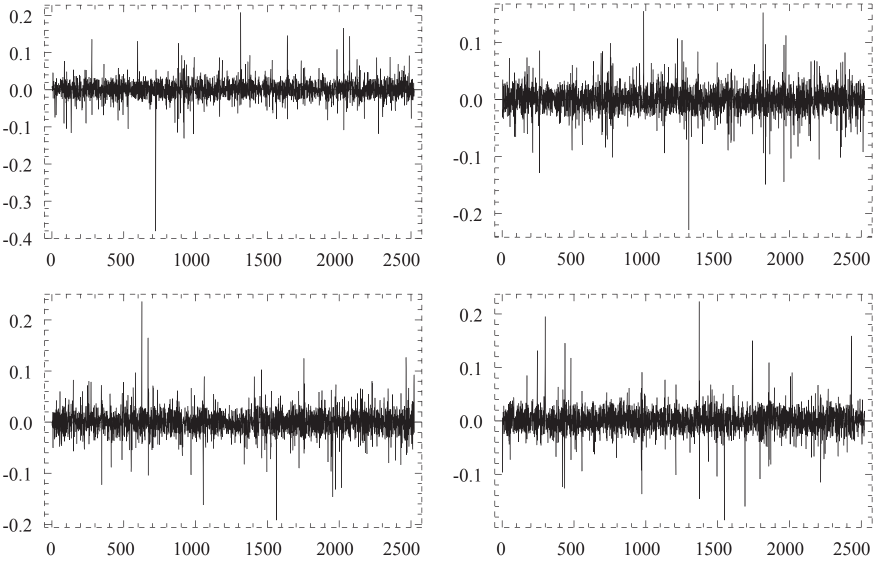
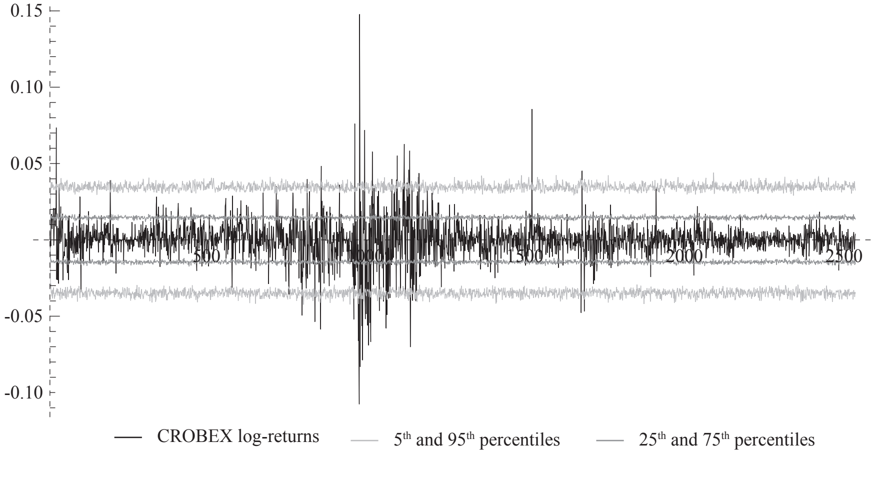


FIGURE 8

CROBEX log-returns and percentiles of 1000 simulated trajectories of Student's diffusion



5 CONCLUSION

In this paper we introduced stationary Student's diffusion as a model for log-returns of stock prices or values of a stock market index. The model captures some main features characteristic for log-returns of risky assets, mainly heavy-tailed marginal distribution and nontrivial dependence structure. The parameters of the diffusion process provide flexibility in fitting the model to data. Here we concentrated on fitting the CROBEX log-returns to the proposed model. Our analysis shows that CROBEX, as well as many other stock market indices, exhibits heavy tails. This important fact must always be taken into account in any serious risk analysis. Simulations of Student's diffusion process show that it can realistically model risky asset returns and reproduce many of their features, like volatility clustering, meaning that periods of low volatility are followed by periods of high volatility, indicating partial predictability of volatility fluctuations.

The main purpose of the proposed model is not to forecast future values; rather it is tailored for a quality risk assessment. Computation of some risk measures, like e.g. value at risk (VaR), in the context of the proposed model has not been considered in this work. However, from the estimated parameters of stationary Student's distributions, it is easy to compute the VaR as the quantile of this distribution. From the comparison with the normal distribution made in figure 3, it is clear that these estimates will tend to give more pessimistic, although realistic predictions. We also did not consider the problem of option pricing which would require a more detailed approach. Option pricing problem can be approached through Monte Carlo simulations but could also be built on the known expressions for the transition density of the Student's diffusion.

The model proposed is flexible enough to cover a wide range of heavy-tailed data. Some extensions of the model may include considering Student processes with prescribed dependence structure. A large class of such models has been proposed in Heyde and Leonenko (2005). Another possible extension of the model would be a diffusion process using the so-called skewed Student's distribution which allows for non-symmetry of the tails through an additional skewness parameter.

REFERENCES

1. Abourashchi, N. and Veretennikov, A. Yu., 2010. On exponential bounds for mixing and the rate of convergence for Student processes. *Theory of Probability and Mathematical Statistics*, 81, pp. 1-13. doi: 10.1090/S0094-9000-2010-00805-2
2. Abramowitz, M. and Stegun, I. A., 1997.) *Handbook of Mathematical Functions*. Washington: The National Bureau of Standards.
3. Arnerić, J., Jurun, E. and Pivac, S., 2007. Theoretical distributions in risk measuring on stock market. *Proceedings of the 8th WSEAS International Conference on Mathematics & Computers in Business & Economics*. Canada, Vancouver, pp. 194-199.
4. Arnerić, J., Lolić, I. and Galetić, J., 2012. Threshold parameter of the expected losses. *Croatian Operational Research Review*, 3(1), pp. 270-279.
5. Bibby, B. M., Skovgaard, I. M. and Sørensen, M., 2005. Diffusion-type models with given marginal distribution and autocorrelation function. *Bernoulli*, 11, pp. 281-299.
6. Bibby, B. M. and Sørensen, M., 1996. A hyperbolic diffusion model for stock prices. *Finance and Stochastics*, 1(1), pp. 25-41. doi: 10.1007/s007800050015
7. Bradley, R. C., 2005. Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*, 2, pp. 107-144.
8. Cont, R., 2001. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1, pp. 223-236. doi: 10.1080/713665670
9. Embrechts, P., Klüppelberg, C. and Mikosch, T., 1997. *Modelling Extremal Events: for Insurance and Finance*. Berlin: Springer. doi: 10.1007/978-3-642-33483-2
10. Forman, J. L., 2005. Least squares estimation for autocorrelation parameters with applications to sums of Ornstein-Uhlenbeck type of processes. *Working paper*. Available at: <http://www.math.ku.dk/~julief/forsk/paper1.pdf>.
11. Fornari, F. and Mele, A., 2000. *Stochastic Volatility in Financial Markets: Crossing the Bridge to Continuous Time*. New York: Springer. doi: 10.1007/978-1-4615-4533-0
12. Grahovac, D. [et al.], 2015. Asymptotic properties of the partition function and applications in tail index inference of heavy-tailed data. *Statistics*, 49(6), pp. 1221-1242. doi: 10.1080/02331888.2014.969267
13. Gray, B. J. and French, D. W., 1990. Empirical comparisons of distributional models for stock index returns. *Journal of Business Finance & Accounting*, 17(3), pp. 451-459. doi: 10.1111/j.1468-5957.1990.tb01197.x
14. Heyde, C. C. and Leonenko, N. N., 2005. Student processes. *Advances in Applied Probability*, 37(2), pp. 342-365. doi: 10.1239/aap/1118858629
15. Heyde, C. C. and Liu, S., 2001. Empirical realities for a minimal description risky asset model. The need for fractal features. *Journal of the Korean Mathematical Society*, 38(5), pp. 1047-1059.
16. Hurst, S. R. and Platen, E., 1997. *The marginal distributions of returns and volatility*. Lecture Notes – Monograph Series, 31, pp. 301-314. Hayward: Institute of Mathematical Statistics. doi: 10.1214/lnms/1215454146

17. Iacus, S. M., 2009. *Simulation and Inference for Stochastic Differential Equations: with R Examples*. New York: Springer.
18. Jondeau, E. and Rockinger, M., 2003. Testing for differences in the tails of stock-market returns. *Journal of Empirical Finance*, 10(5), pp. 559-581. doi: 10.1016/S0927-5398(03)00005-7
19. Leonenko, N. N. and Šuvak, N., 2010. Statistical inference for Student diffusion process. *Stochastic Analysis and Applications*, 28(6), pp. 972-1002. doi: 10.1080/07362994.2010.515476
20. Mandelbrot, B., 1963. The variation of certain speculative prices. *The Journal of Business*, 36(4), pp. 394-419. doi: 10.1086/294632
21. Miletić, M. and Miletić, S., 2015. Performance of Value at Risk models in the midst of the global financial crisis in selected CEE emerging capital markets. *Economic Research-Ekonomska istraživanja*, 28(1), pp. 132-166. doi: 10.1080/1331677X.2015.1028243
22. Rydberg, T. H., 1999. Generalized hyperbolic diffusion processes with applications in finance. *Mathematical Finance*, 9(2), pp. 183-201. doi: 10.1111/1467-9965.00067
23. Serfling, R. J., 1980. *Approximation Theorems of Mathematical Statistics*. New York: John Wiley & Sons. doi: 10.1002/9780470316481
24. Tsay, R. S., 2010. *Analysis of Financial Time Series*. New York: John Wiley & Sons. doi: 10.1002/9780470644560
25. Žiković, S. and Pečarić, M., 2010. Modelling extreme events: Application to Zagreb Stock Exchange. *Ekonomski pregled*, 61(1-2), pp. 19-37.