Advances in Complex Systems
Vol. 23, No. 8 (2020) 2050020 (28 pages)

DOI: 10.1142/S0219525920500204

A SEMI-MARKOVIAN APPROACH TO DRAWDOWN-BASED MEASURES

GUGLIELMO D'AMICO* and BICE DI BASILIO†

Department of Economics
University 'G.D'Annunzio' of Chieti-Pescara
Pescara 65127, Italy
*guglielmo.damico@unich.it
†bice.dibasilio@unich.it

FILIPPO PETRONI

Department of Management
Marche Polytechnic University
Ancona 60121, Italy
f.petroni@univpm.it

Received 28 December 2020
Revised 21 April 2021
Accepted 26 April 2021
Published 23 June 2021

In this paper we assess the suitability of weighted-indexed semi-Markov chains (WISMC) to study risk measures as applied to high-frequency financial data. The considered measures are the drawdown of fixed level, the time to crash, the speed of crash, the recovery time and the speed of recovery; they provide valuable information in portfolio management and in the selection of investments. The results obtained by implementing the WISMC model are compared with those based on the real data and also with those achieved by GARCH and EGARCH models. Globally, the WISMC model performs much better than the other econometric models for all the considered measures unless in the cases when the percentage of censored units is more than 30% where the models behave similarly.

Keywords: Drawdown-based measures; high-frequency data; right censoring; maximum likelihood estimate.

Mathematics Subject Classification 2020: 60K15, 62P05

1. Introduction

Financial markets are characterized by continuous upward or downward fluctuations in prices, caused by the vast amount of information they receive. A strong price instability, historically and cyclically, often caused strong market collapses. Exactly

†Corresponding author.

in this context, with the aim of quantifying the possible losses connected to an investment or an investment portfolio, risk measures were born.

Over the years, a vast and varied literature on this topic, which aims to analyze the risk with different approaches, has developed. Among the most known risk measures, we mention value-at-risk and expected shortfall whose fame is chiefly linked to the Basel regulatory agreements [3, 4]. Their main weakness is that, being based on quantiles, they neglect the temporal evolution of data, which is a crucial aspect in the analysis of financial time series. Nevertheless, drawdown-based measures overcome this issue; among these, we point out the drawdown of fixed level, the time to crash, the speed of crash, the recovery time and the speed of recovery.

In detail, the drawdown process is the distance of the price process from its running maximum. The drawdown of fixed level, the time to crash and the recovery time scan the time in which the drawdown of an asset attains a selected K-level for the first time, the time taken to have the first K-variation and the time necessary to have the first K'-descendent after having the first K-ascent, respectively [17, 18]. Accordingly, the speed of crash and the speed of recovery describe the rate at which the first K and $(K - K')$ changes occur.

To investigate these risk indicators, it is essential to use stochastic models for financial returns. In literature, various models which shape returns directly or indirectly, have been suggested. As a matter of fact, the most widespread are the econometric models [9, 13, 16]. and the diffusive models [2, 14]. Recently effective alternatives, based on semi-Markovian models, have also been advanced [5–8].

In this paper, we analyze the drawdown-based measures listed above by means of weighted-indexed semi-Markov chain (WISMC) model which is applied to tick-by-tick data of Fiat stock price. WISMC model is an improvement of the general semi-Markov chain (SMC) model which allows to correctly represent the long dependence structure in the squared returns through the addition of an index process. To test the validity of our model we realize comparisons with the GARCH and EGARCH models. In particular, we reproduce synthetic series for each considered model and then, we compute the analyzed risk measures on both real and simulated data. First, we examine the behavior of the drawdown of fixed level as K varies by calculating its main descriptive statistics. Secondly, we explore the time to crash evaluating its best parametric law among the lognormal, the Weibull and the exponential distributions based on the AIC and BIC criterion. In order to do this, we consider right censored data in the maximum likelihood estimates of the parameters because data are affected by the censorship due to the choice of the observation period (i.e. the trading day). Furthermore, we compute the Kullback–Leibler divergence as a function of K to get the closest model to real data. As regard the speed of crash, we analytically derive its density starting from the time to crash's density and then, we scan its behaviour for some values of K. Finally, we explore the recovery time and the speed of recovery for several values of K and K', using the approach just described. Overall, the WISMC model replicates better than the elected GARCH and EGARCH models the considered risk measures unless the percentage of censored data exceeds 30% of observations. In this last situation, the performance of WISMC starts to decrease probably due to an amount of record insufficient for estimating the kernel of the process. Thus, the use of WISMC is strongly recommended for use in future studies when a consistent dataset is considered.

The rest of the paper is organized as follows. In Sec. 2, we provide a formal description of the analyzed drawdown-based measures. In Sec. 3, we present the WISMC model, its application procedures and the maximum likelihood estimate with right censored data. In Sec. 4, we display the results of our analysis. In Sec. 5, we report concluding remarks and future goals. In the appendix, we provide a list of the exploited financial symbols.

2. Risk Measures

The literature on risk measures is really huge and expanding so much that there are financial indicators for every need. The most common and simplest risk measures are the value-at-risk and the expected shortfall which, however, share a significant mutual weakness. In fact, being based on quantiles, they disregard the temporal order of data which is important in financial time series. On the contrary, there are measures that overcome this trouble, such as those based on the drawdown, which take into account the historical evolution of data.

To define the drawdown of an asset, we introduce the discrete-time varying asset price process $X(t)$ and its running maximum process, determined as

$$Y(t) = \max_{s \in \{0,1,\ldots,t\}} \{X(s)\}. \tag{1}$$

The drawdown process, denoted by $D(t)$, is determined as the difference between the running maximum process $Y(t)$ and the price process $X(t)$

$$D(t) = Y(t) - X(t), \quad t \geq 0. \tag{2}$$

It expresses the correction of the asset price with respect to a previous relative maximum.

To understand concretely the definitions above, in Fig. 1 we offer a graphical illustration. In detail, the red, the blue and the black lines stand for the price process, the running maximum process and the drawdown process of Fiat asset during a trading day, respectively. The drawdown process (black line) is obtained by subtracting the asset price process (red line) from the maximum process (blue line).

In this paper, we deal with risk measures based on drawdown and connected to market crashes: the drawdown of fixed level, the time to crash, the speed of crash, the recovery time and the speed of recovery.

The drawdown of fixed level is the first time that the drawdown process attains or overcomes a certain threshold K.

Formally, it is defined as

$$\tau(K) := \min\{t \geq 0 | D(t) \geq K\} \quad \text{where } K \geq 0. \tag{3}$$

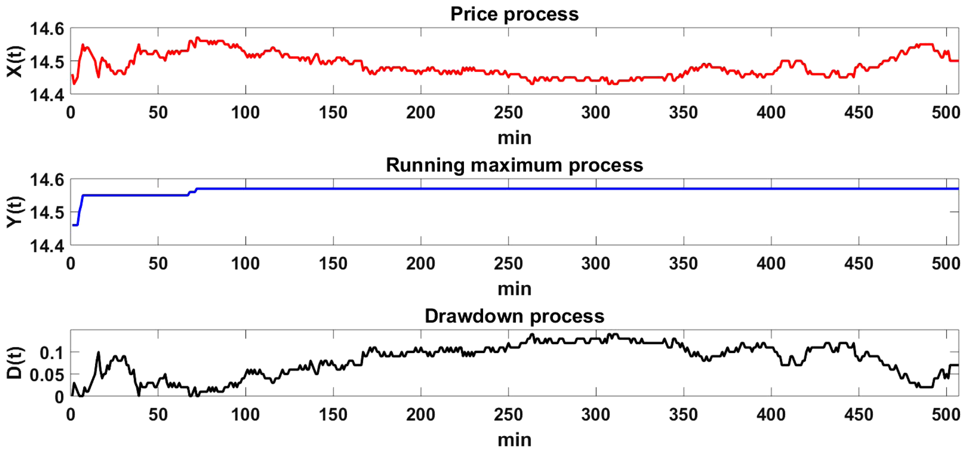

Fig. 1. Price process (red line), running maximum process (blue line) and drawdown process (black line) of Fiat asset during a trading day.

To qualify the time to crash we need to introduce the last visit time of the maximum before the stopping time $\tau(K)$; formally it is defined as

$$\rho(K) := \max\{t \in [0, \tau(K)] | Y(t) = X(t)\}. \tag{4}$$

Using the definition of $\tau(K)$ and $\rho(K)$ we define the time to crash as

$$T_c(K) := \tau(K) - \rho(K). \tag{5}$$

It is the time taken between $\rho(K)$ and $\tau(K)$, i.e. the time the drawdown process employees to have the first drop of level K, the latter identifies the crash.

Consequently, the speed of crash, that is the velocity at which the first K-change occurs, is expressed as

$$S_c(K) := \frac{K}{\tau(K) - \rho(K)} = \frac{K}{T_c(K)}. \tag{6}$$

Our definition of speed of crash is more general than that reported in [17]. In particular, the authors of [17] define the speed of crash as the reciprocal of the time to crash.

In order to introduce the recovery time and the speed of recovery, denoted by $R_t(K, K')$ and $S_r(K, K')$, it is necessary to previously define the quantity $\gamma(K, K')$ as follows

$$\gamma(K, K') := \min\{t > \tau(K) | D(t) \leq K'\} \quad \text{with } K > K'. \tag{7}$$

It identifies the first moment in which the drawdown process drops below the threshold K' after crossing the threshold K for the first time, i.e. the instant in which a $(K - K')$-drop occurs. Therefore it indicates an improvement in terms of risk since the asset passes from a more risky situation, reached when the threshold K is exceeded, to a less risky one achieved when the threshold K' is attained.

Exploiting the definition of $\gamma(K, K')$ and $\tau(K)$, the recovery time is formalized as

$$R_t(K, K') := \gamma(K, K') - \tau(K). \tag{8}$$

It is the time it takes to have the first K'-descent in the drawdown process following the first K-ascent.

Therefore, the speed of recovery, that is the velocity at which a $(K - K')$-variation occurs, is defined as

$$S_r(K, K') := \frac{K - K'}{\gamma(K, K') - \tau(K)} = \frac{K - K'}{R_t(K, K')}. \tag{9}$$

Clearly, all these risk measures provide stakeholders a useful tool with which to assess the riskiness of an investment or investment portfolio. In particular, $\tau(K)$ informs about the riskiness of an asset and depends on the selected K-value (small K indicates low risk events while large K refers to very risky events). Accordingly, $T_c(K)$ and $S_c(K)$ quantify how long it takes and how quickly such risky events occur.

Unlike the previous measures, $R_t(K, K')$ and $S_r(K, K')$ analyze the behavior of a security following the achievement of a certain K-level in its drawdown and, therefore, they show a dependence from both the K and the K'-thresholds. In detail, the recovery time and the speed of recovery describe how long the drawdown process holds the first K-change before having a $(K - K')$ drop and the rate at which this decline occurs.

To graphically display the risk measures just described, in Fig. 2 we provide an illustration. The black line represents the drawdown process of Fiat asset during a trading day while the orange and the yellow dashed lines stand for the thresholds K and K', respectively. After fixing the threshold K we determine both the measures $\tau(K)$ and $T_c(K)$. Consequently, the speed at which the K-intensity market crash occurs, known as $S_c(K)$, is the rate at which we traverse the space K in the time range $[\rho(K), \tau(K)]$.

Conversely, to identify the measure $R_t(K, K')$ it is also necessary to consider the threshold K' in addition to the threshold K. Accordingly, the speed of recovery $S_r(K, K')$ is the rate at which we cross the space $(K - K')$ in the time interval $[\tau(K), \gamma(K, K')]$.

It is important to observe that in this paper the crash should be intended as a drop in the price process whose entity is denoted by the threshold K. If the level K is very large then we are focusing on a huge price fall. Moreover, it should be pointed out that the methodology we are going to advance in the next sections, included the mathematical model, could also be used to investigate a market crash when the price process is substituted with a market index (e.g. the Dow Jones index) representing a segment of the financial market.

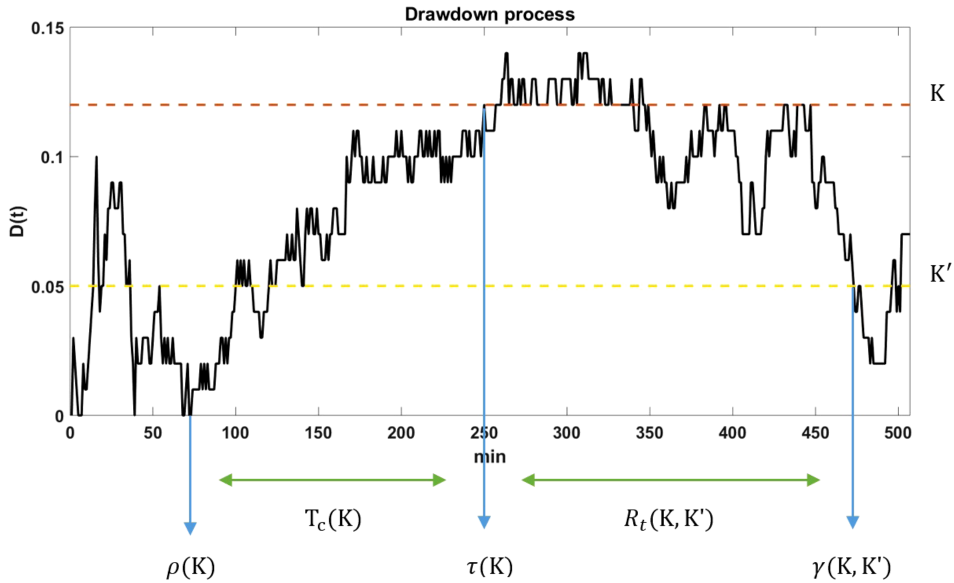

Fig. 2. Drawdown process of Fiat asset during a trading day. The orange and the yellow dashed lines act for the thresholds K and K', respectively.

3. Mathematical Models and Methodology of Application

The content of this section is divided in two parts. In the first part, we shortly present the WISMC model as related to the financial problem. The presentation is only intuitive and it serves the purpose of introducing the considered variables, their probabilistic interrelations and the main features of the model. Additional details and mathematical properties can be retrieved in the literature. In the second part, we explain the methodology of application to assess the suitability of the WISMC model in comparison with popular econometric models to describe the drawdown-based risk measures discussed in the previous section.

3.1. *The weighted-indexed Semi-Markov chain model*

Semi-Markovian-based models are increasingly being used in financial literature to reproduce the main stylized facts related to returns. Unlike discrete-time Markovian models, in which the time until the next transition is geometrically distributed, semi-Markov models assume that the time between transitions can be modeled using any distribution, no memoryless distributions included. The consequence is that, a semi-Markovian process considers not only the current state but also how long it has been in the current state. However, they are unable to capture the persistent autocorrelation in the squared returns. One way to solve this problem is to use an improvement of the semi-Markov chain (SMC) model, called weighted-indexed semi-Markov chain

(WISMC) model. It allows the reproduction of the long-term dependence in the squared returns by means of an index process. The role of the index is exactly to increase the memory of the process by adding information contained in the past trajectory of the returns process. A detailed model definition is in [7, 8].

Basically, the model is described by three stochastic processes: $\{J_n\}_{n\in\mathbb{N}}$, $\{T_n\}_{n\in\mathbb{N}}$ and $\{V_n^\lambda\}_{n\in\mathbb{N}}$, which in our framework are the return process, the corresponding jump time process and the index process, respectively.

Assuming that $X(t)$ is the time varying price of a financial asset, we define the time varying log-returns as

$$R(t) = \log(X(t)/X(t-1)), \quad t \in \mathbb{N}. \tag{10}$$

The series of returns $R(t)$ is converted into a series of discrete returns, denoted by $R_d(t)$, using the following map:

$$\mathcal{M} : \mathbb{R} \to E = \{-i_{\min}\Delta, \dots, -\Delta, 0, \Delta, \dots, i_{\max}\Delta\}, \tag{11}$$

where E is the finite state space of the discrete returns and Δ is the grid amplitude of E. Precisely, $R_d(t) = i\Delta$ whenever $R(t) \in ((i-\frac{1}{2})\Delta, (i+\frac{1}{2})\Delta]$. The lowest discrete return, $R_d(t) = -i_{\min}\Delta$, and the highest discrete return, $R_d(t) = i_{\max}\Delta$, are attained whereas $R(t) \leq (-i_{\min} + \frac{1}{2})\Delta$ and $R(t) > (i_{\max} - \frac{1}{2})\Delta$, respectively.

Subsequently, the sequence of discrete returns $\{R_d(t)\}_{t\in\mathbb{N}}$ is transformed into a series of returns $\{J_n\}_{n\in\mathbb{N}}$ with values in E, describing the value of the return process at the nth change, and into a series of corresponding jump times $\{T_n\}_{n\in\mathbb{N}}$ with values in $\mathbb{N}$, portraying the time in which the nth change in the return process occurred. In order to do this, we set $T_0 = 0$, $J_0 = R_d(0)$ and for $n \geq 1$

$$T_n = \inf\{t \in \mathbb{N}, t > T_{n-1} : R_d(t) \neq R_d(T_{n-1})\}, \tag{12}$$

$$J_n = R_d(T_n). \tag{13}$$

The real novelty, compared to a classic semi-Markov model, is the introduction of the stochastic process $\{V_n^\lambda\}_{n\in\mathbb{N}}$, with values in $\mathbb{R}$. The random variable V_n^λ describes the value of the index process at the nth transition, i.e. it synthesizes the information contained in the past trajectory of the return process up to the n-th transition. It is defined as

$$V_n^\lambda = \sum_{k=0}^{n-1} \sum_{a=T_{n-1-k}}^{T_{n-k}-1} f^\lambda(J_{n-1-k}, T_n, a) + f^\lambda(J_n, T_n, T_n), \tag{14}$$

where f can be any function. It depends on both the past values of returns J_{n-1-k} occurred at times a, the current time T_n and the parameter λ, that has the task of weighing past information. In the next subsection, we provide a specific functional form of f and we describe the calibration of the parameter λ on real data.

In order to build the WISMC model it is necessary to explicit the dependency structure between the random variables J_n, T_n and V_n^λ. Toward this end we adopt the following assumption:

$$\begin{aligned}\mathbb{P}[J_{n+1} = j, T_{n+1} - T_n \leq t|\sigma(J_h, T_h, V_h^\lambda)_{h=0}^n, J_n = i, V_n^\lambda = v] \\ = \mathbb{P}[J_{n+1} = j, T_{n+1} - T_n \leq t|J_n = i, V_n^\lambda = v] =: Q_{ij}^\lambda(v;t),\end{aligned} \tag{15}$$

where $\sigma(J_h, T_h, V_h^\lambda)$ is the natural filtration of the three-variate process $\{J_n, T_n, V_n^\lambda\}$.

The matrix of functions $\mathbf{Q}^\lambda(v;t) = (Q_{ij}^\lambda(v;t))_{i,j\in E}$ plays a fundamental role in this model and it is referred to as the weighted-indexed semi-Markov kernel. It is important to note that if $\mathbf{Q}^\lambda(v;t)$ is constant in v then, the weighted-indexed semi-Markov kernel degenerates into an ordinary semi-Markov kernel, i.e.

$$\begin{aligned}\mathbb{P}[J_{n+1} = j, T_{n+1} - T_n \leq t|\sigma(J_h, T_h, V_h^\lambda)_{h=0}^n, J_n = i, V_n^\lambda = v] \\ = \mathbb{P}[J_{n+1} = j, T_{n+1} - T_n \leq t|J_n = i] =: Q_{ij}(t).\end{aligned} \tag{16}$$

Furthermore, from an ordinary semi-Markov kernel it is recovered the case of a Markov kernel if the probability distributions of the sojourn times $T_{n+1} - T_n$ in the states of the system are geometrically distributed.

The probabilities of the weighted-indexed semi-Markov kernel can be expressed according to

$$\begin{aligned}Q_{ij}^\lambda(v;t) = \mathbb{P}[T_{n-1} - T_n \leq t|J_{n+1} = j, J_n = i, V_n^\lambda = v] \\ \cdot \mathbb{P}[J_{n+1} = j|J_n = i, V_n^\lambda = v] =: G_{ij}(v;t) \cdot p_{ij}(v),\end{aligned} \tag{17}$$

where $p_{ij}^\lambda(\nu)$ and $G_{ij}^\lambda(\nu;t)$ are the indexed transition matrix and the conditional waiting time distribution in the states, respectively. A detailed description of the estimation procedures of each quantity involved in the model and in particular of the weighted-indexed semi-Markov kernel is in [8]. It is worth noting that the index process has the scope of identifying changes in the state of our dynamical system through the identification of different regimes characterized by different semi-Markovian dynamics. Nevertheless, the WISMC model is different from the frequently used Markov regime switching models [12] because in the latter the new dynamics are described by changes in the model parameters that are modulated by a Markov chain that is unobservable. Instead, in the WISMC model the different regimes are connected only to the observable variable (returns in finance) and are optimally included in the model in order to increase its memory.

3.2. *Methodology of application*

In order to apply the model and to assess its suitability to reproduce real data features also in comparison with other econometric models, we develop an application methodology that is composed of several successive steps we are going to describe (see Fig. 3 for a graphical illustration).

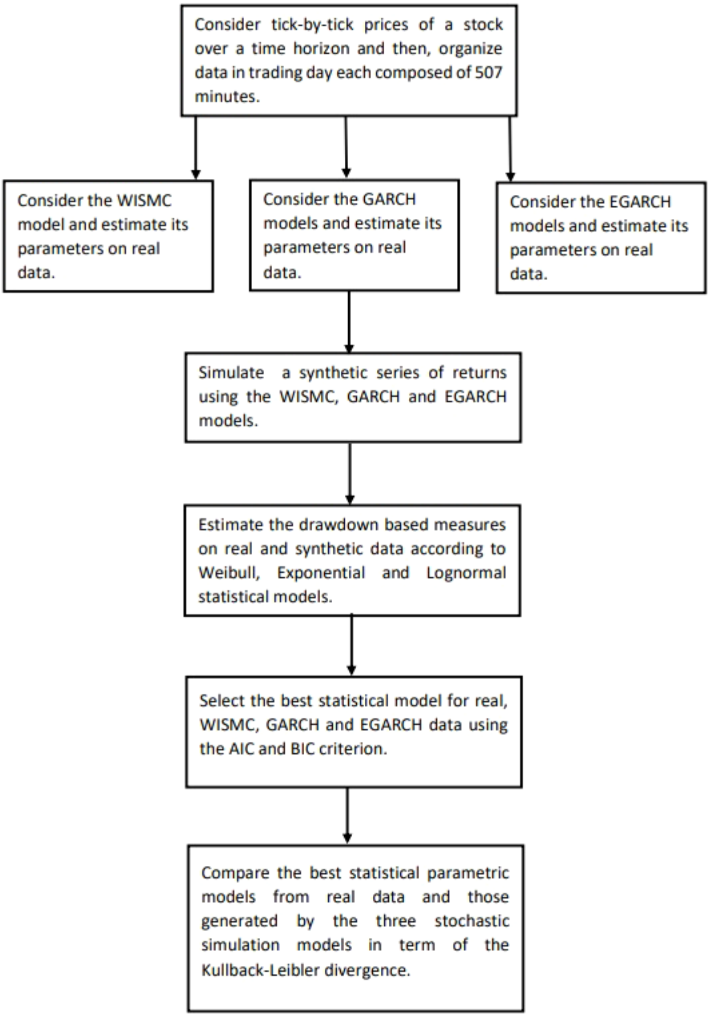

Fig. 3. Flow chart of the steps followed in the analysis.

- Step 1 *Estimation of the model*
  We consider the intraday prices of a stock over a certain time horizon. To explore the behavior of the drawdown-based measures within trading day, we organize the dataset into days each consisting of 507 min or different time grid according to the data availability. Next, we consider the WISMC, GARCH, EGARCH models and then, we estimate their parameters from real data. Specifically, the estimation procedures of the WISMC model are realized according to [8] while, those of the GARCH and EGARCH models are carried out through the Matlab function "*estimate*", following the literature on these topics [9, 13].

- Step 2 *Simulation from different models*
  Using the following algorithm, which was discussed in [8], we simulate a synthetic series of returns from the WISMC model:
  (1) Set $n = 0, J_0 = i, T_0 = 0, V_0^\lambda = v$, horizon time $= T$;
  (2) Sample J from $p^\lambda_{J_n,\cdot}(V_n^\lambda)$ and set $J_{n+1} = J(\omega)$;
  (3) Sample W from $G^\lambda_{J_n,J_{n+1}}(V_n^\lambda, \cdot)$ and set $T_{n+1} = T_n + W(\omega)$;
  (4) Set $V^\lambda_{n+1} = \sum_{k=0}^{n} \sum_{a=T_{n-k}}^{T_{n+1-k}-1} f^\lambda(J_{n-k}, T_{n+1}, a) - f^\lambda(J_{n+1}, T_{n+1}, T_{n+1})$;
  (5) If $T_{n+1} \geq T$ stop
  or else set $n = n + 1$ and go to 2.
  Similarly, by means of the Matlab function "*simulate*", we generate synthetic series from the GARCH and EGARCH models. In this way several synthetic series, generated from the considered stochastic models and estimated on real data, are available.

- Step 3 *Statistical parametric models for drawdown-based measures*
  Using both real and synthetic series, we compute the drawdown-based risk measures. Next, we realize their empirical investigation employing the Weibull, the exponential and the lognormal distributions. All these models' parameters are computed using the maximum likelihood estimate and considering the Type-1 right censoring issue which is linked to the choice of the time interval, i.e. the trading day. The Type-1 right censoring issue is a problem that occurs when an experiment, where a certain number of subjects or objects are observed, is arrested after a specified observation time limit is reached. Subjects still alive at the time limit are censored on the right. Specifically, our selection of the daily time horizon generates Type-1 right censorship on the drawdown of fixed level $\tau(K)$, the time to crash T_c and the recovery time R_t as the interest events may not occur within the 507 minutes of the trading day. To correctly estimate these parameters, we use the Matlab functions "*wblfit*", "*expfit*", and "*lognfit*" which are based on the search for the maximum through an optimization algorithm.
- Step 4 *Comparisons of results on real data and those obtained from the different models.*
  In order to identify the best parametric model on each series, we compute the AIC and BIC criterion using the Matlab function "*aicbic*". In detail, it is possible to directly detect the best parametric model for the real and WISMC data, choosing the one with the lowest AIC/BIC. As for the GARCH and EGARCH families, it is required a further step. In fact, after having identified the best parametric model for each considered GARCH/EGARCH, it is selected the model with the lowest AIC/BIC among the best models. Finally, we quantify the distance between the best parametric model for each family (WISMC, GARCH and EGARCH) and the best parametric model for real data by mean of the Kullback–Leibler divergence. It permits to gauge the semi-distance between real and simulated distributions in order to assess the closest model to real data.

4. Results

Data used in this analysis consist of intra-day prices of Fiat asset, denoted by the symbol F. They have been downloaded from "Borsa Italiana" ("www.borsaitaliana.it") for the period January 2007–December 2010 (4 full years) and then, have been re-sampled by means of the Matlab function "*resample*" to have the frequency of 1 min. Globally, we analyze 1001 trading days each consisting of 507 minutes, for a total of 506506 returns. In Table 1 we provide the main descriptive statistics of Fiat stock.

In order to model asset returns as a WISMC process, the state space has to be discretized according to the map (11). In this work, we discretize asset returns into 5 states, chosen to be symmetrical with respect to returns equal to zero. Actually, each stock exchange fixes a diverse discretization in stock prices which depends on the stock value. Just to give an example, in the Italian stock market the minimum variation for assets with values between 5.0001 and 10 Euro is set to 0.005 Euro. We essay to remain as faithful as possible to this discretization.

The WISMC model requires the specification of the function f in the definition of the index process V_n^λ, see relation (14). Our choice of f is driven by an important empirical evidence, known as volatility clustering. Typically, in real stock market we observe that large (small) changes in prices are followed by large (small) changes, which tend to persist for a certain period. According to [7], we reproduce this empirical feature using, as specification of the function f, an exponentially weighted moving average (EWMA) of the squared returns (volatility), which is expressed as follows:

$$f^\lambda(J_{n-1-k}, T_n, a) = \frac{\lambda^{T_n-a} J_{n-1-k}^2}{\sum_{k=0}^{n-1} \sum_{a=T_{n-1-k}}^{T_{n-k}-1} \lambda^{T_n-a}} = \frac{\lambda^{T_{n-a}} J_{n-1-k}^2}{\sum_{a=1}^{T_n} \lambda^a}. \tag{18}$$

We also discretize the index V_n^λ in 5 states to repeat different levels of volatility, i.e. low, medium low, medium, medium high and high volatility.

Fundamentally, our model depends on two parameters that have to be calibrated: the number of states for the return process and the value of the parameter λ. Both are optimized by minimizing the mean percentage error (MPE) between real and synthetic autocorrelation functions, according to the recommendations given in [8]. The basic idea is to set a number of states and a value of λ to build a trajectory and then, estimate the weighted-indexed semi-Markov kernel and run a Monte Carlo simulation to produce a synthetic series. The next step is to compute the autocorrelation functions for both real and simulated data and compare them by calculating

Table 1. Mean, median, standard deviation, skewness and kurtosis of Fiat stock.

Stock	Mean	Median	SD	Skewness	Kurtosis
F	8.237e-07	0	8.353e-04	−0.017	3.485

Table 2. Descriptive statistics of τ (first quartile, second quartile-median, third quartile, mean, standard deviation, asymmetry index) and related censored units computed on real data as a function of K.

Descriptive statistics of $\tau(K)$							
K	Q1	Q2	Q3	Mean	SD	AI	Censoring rate
0.5%	5	11	27	31.466	72.501	0.847	0.6%
0.8%	11	31	104	97.929	148.267	1.354	6.0%
0.9%	15	41	160.750	121.524	164.290	1.470	9.0%
1.0%	19	53	250.500	147.045	179.973	1.568	12.0%
1.1%	22	66	316	163.796	186.790	1.571	15.0%

the MPE. This procedure is repeated setting different values of states and λ. In the end, we choose the optimal number of states and the optimal value of λ which best represents data by minimizing the MPE.

In Table 2, we display both the main descriptive statistics and the number of censored units related to the drawdown of fixed level τ for different K-values ($K = 0.5\%, K = 0.8\%, K = 0.9\%, K = 1.0\%, K = 1.1\%$). All these quantities are computed on real data.

Using the Kaplan–Meier estimator [10], which provides a non-parametric estimate of the survival function S, the mean and the standard deviation are computed. Focusing on the mean values of the drawdown of fixed level τ, we observe that if the threshold K increases also τ rises on average. This implies that more extreme events take longer to occur. Just to give an example, a drawdown variation of 0.5% is detected on average at the thirty-first minute while to have a major change, such as a 1.0%, we have to wait more, about 147 min on average. Furthermore, as can be noted from the asymmetry index (AI) values, the risk indicator τ has a positive asymmetric distribution that rises as the threshold K grows.

In Fig. 4, we show the real survival curves of the drawdown of fixed level τ as a function of K ($K = 0.5\%, K = 0.8\%, K = 0.9\%, K = 1.0\%, K = 1.1\%$). Analyzing their behavior, it can be observed that they move upwards as the threshold increases. This means that if K rises, i.e. if we look for a riskier event, the probability that it will not occur within the trading day also increases.

To explore the time to crash T_c, the recovery time R_t, the speed of crash S_c and the speed of recovery S_r, we simulate asset returns by means of the WISMC model. Moreover, to test the validity of our model we carry out comparisons with three GARCH models and four EGARCH models. Next, for all the considered models, we generate a synthetic series of returns having the same length as the real one and then, we compute the analyzed risk indicators for both the real and each simulated series.

Given the real and simulated measures of the time to crash T_c and the recovery time R_t, we assess their best parametric law among the lognormal, the Weibull and the exponential distributions based on the AIC and BIC criterion [1, 15]. ll estimation procedures involving these parametric models are accomplished in Matlab software by considering the censorship issue.

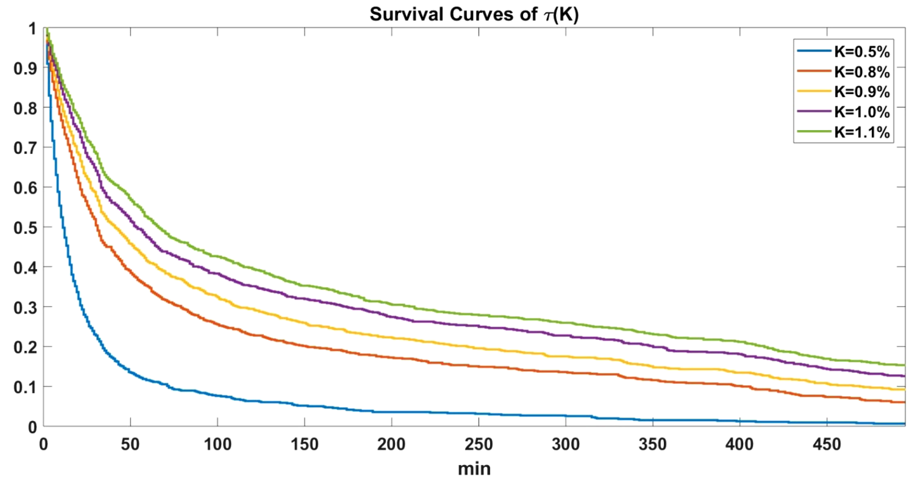

Fig. 4. Survival curves of τ computed on real data in function of K ($K = 0.5\%$, $K = 0.8\%$, $K = 0.9\%$, $K = 1.0\%$, $K = 1.1\%$).

In Table 3, we display the model selection for the time to crash T_c, considering several values of K ($K = 0.5\%, K = 0.9\%, K = 1.1\%$). The AIC and BIC values of the best statistical selected models are in bold. In detail, for the best model choice of real and WISMC data we operate a single step which consists of selecting directly the parametric model with the lowest AIC/BIC. As for the GARCH and EGARCH models, we apply a further step. In fact, after detecting the best parametric law for each considered GARCH and EGARCH model, we choose the one with the smallest AIC/BIC among the best selected models. Just to give an example, looking at $K = 0.5\%$ we have that the GARCH (1, 1)-lognormal, the GARCH (1, 2)-Weibull and the GARCH (2, 1)-lognormal are the best, as a first step. Subsequently, applying the second step, we choose the GARCH (1, 2)-Weibull as it has AIC/BIC lower than the other two.

In Table 4, we show the parameters' point estimate of the best selected models for the time to crash T_c, considering different values of the threshold K ($K = 0.5\%, K = 0.9\%, K = 1.1\%$). The best statistical parametric model is almost always the lognormal one, except in one case where the Weibull model appears to perform better. Furthermore, GARCH models are not stable because their parameters depend on the value of K and thus, there isn't a GARCH that works uniformly better in K than others. This is an undesirable property because it is usually interesting to understand the behavior of T_c for different values of K. This requires the use of a specific GARCH model for each level of the drawdown or choosing a single GARCH model which, consequently, will not be optimal for most of the K values chosen. In contrast, the approach based on the WISMC model does not experience this problem. Indeed, the WISMC is a non-parametric model which is settled once its kernel is estimated. The estimation procedure is independent of the

Table 3. Selection of the best parametric model for the measure T_c on both real and simulated data by mean of AIC and BIC criterion, considering several K-values ($K = 0.5\%$, $K = 0.9\%$, $K = 1.1\%$). AIC and BIC relating to the best models are in bold.

Model selection for $T_c(K)$	Weibull		Lognormal		Exponential	
	AIC	BIC	AIC	BIC	AIC	BIC
$K = 0.5\%$						
Real Data	7.091E+03	7.101E+03	**6.626E+03**	**6.636E+03**	7.792E+03	7.797E+03
WISMC	8.162E+03	8.172E+03	**7.926E+03**	**7.935E+03**	8.246E+03	8.251E+03
GARCH(1,1)	8.029E+03	8.039E+03	7.994E+03	8.003E+03	8.255E+03	8.260E+03
GARCH(1,2)	**7.923E+03**	**7.933E+03**	7.944E+03	7.954E+03	8.199E+03	8.204E+03
GARCH(2,1)	7.924E+03	7.934E+03	7.928E+03	7.938E+03	8.184E+03	8.188E+03
EGARCH(1,1)	7.467E+03	7.476E+03	7.346E+03	7.356E+03	7.696E+03	7.700E+03
EGARCH(1,2)	6.870E+03	6.880E+03	6.757E+03	6.767E+03	7.135E+03	7.130E+03
EGARCH(2,1)	7.384E+03	7.394E+03	7.233E+03	7.243E+03	7.625E+03	7.630E+03
EGARCH(2,2)	6.850E+03	6.860E+03	**6.667E+03**	**6.677E+03**	7.024E+03	7.029E+03
$K = 0.9\%$						
Real Data	9.620E+03	9.629E+03	**9.286E+03**	**9.296E+03**	1.088E+04	1.089E+04
WISMC	9.917E+03	9.927E+03	**9.532E+03**	**9.542E+03**	1.033E+04	1.034E+04
GARCH(1,1)	1.071E+04	1.072E+04	1.018E+04	1.019E+04	1.076E+04	1.077E+04
GARCH(1,2)	1.076E+04	1.077E+04	1.022E+04	1.023E+04	1.088E+04	1.089E+04
GARCH(2,1)	1.067E+04	1.068E+04	**1.015E+04**	**1.016E+04**	1.071E+04	1.072E+04
EGARCH(1,1)	1.004E+04	1.005E+04	9.457E+03	9.467E+03	1.007E+04	1.007E+04
EGARCH(1,2)	9.086E+03	9.095E+03	8.765E+03	8.775E+03	9.140E+03	9.145E+03
EGARCH(2,1)	9.813E+03	9.823E+03	9.339E+03	9.349E+03	9.811E+03	9.816E+03
EGARCH(2,2)	8.664E+03	8.674E+03	**8.532E+03**	**8.542E+03**	8.875E+03	8.880E+03
$K = 1.1\%$						
Real Data	9.869E+03	9.878E+03	**9.585E+03**	**9.595E+03**	1.101E+04	1.102E+04
WISMC	1.039E+04	1.040E+04	**1.003E+04**	**1.004E+04**	1.091E+04	1.092E+04
GARCH(1,1)	1.107E+04	1.108E+04	**1.062E+04**	**1.063E+04**	1.127E+04	1.127E+04
GARCH(1,2)	1.108E+04	1.109E+04	1.065E+04	1.066E+04	1.128E+04	1.129E+04
GARCH(2,1)	1.108E+04	1.109E+04	1.062E+04	1.063E+04	1.127E+04	1.127E+04
EGARCH(1,1)	1.076E+04	1.077E+04	1.022E+04	1.023E+04	1.097E+04	1.097E+04
EGARCH(1,2)	1.006E+04	1.007E+04	9.567E+03	9.576E+03	1.006E+04	1.006E+04
EGARCH(2,1)	1.070E+04	1.071E+04	1.018E+04	1.019E+04	1.080E+04	1.081E+04
EGARCH(2,2)	9.886E+03	9.896E+03	**9.396E+03**	**9.406E+03**	9.886E+03	9.890E+03

levels K and K', thus the result is a unique process that describes correctly the results for all choices of K and K'.

In Table 5, we portray both the descriptive statistics and the number of censored units for the time to crash T_c. They are computed considering the best statistical parametric model for real data as a function of K. Observing Table 5, we note that both the censoring rate and the average values increase as K increases. This indicates that it takes more minutes to cross a large threshold than a small one and consequently, that large risky events tend not to always occur in the trading day. For instance, a 0.5%-change in the drawdown occurs in 12 min on average with 6 censored units while a 1%-variation is achieved in 467 min on average with 148 censored units.

Table 4. Parameters of the best models for the measure T_c, considering several K-values ($K = 0.5\%$, $K = 0.9\%$, $K = 1.1\%$).

Summary of the best statistical parametric model for $T_c(K)$		
	Best Model	Parameters
$K = 0.5\%$		
Real Data	Lognormal	1.7021–1.2561
WISMC	Lognormal	2.4455–1.1193
GARCH(1,2)	Weibull	24.6719–1.5621
EGARCH(2,2)	Lognormal	2.2488–0.7124
$K = 0.9\%$		
Real Data	Lognormal	3.3108–1.9340
WISMC	Lognormal	3.3470–1.2940
GARCH(2,1)	Lognormal	3.9644–0.9129
EGARCH(2,2)	Lognormal	3.1848–0.7092
$K = 1.1\%$		
Real Data	Lognormal	3.9218–2.1097
WISMC	Lognormal	3.7696–1.4432
GARCH(1,1)	Lognormal	4.5716–1.1830
EGARCH(2,2)	Lognormal	3.5949–0.7663

Table 5. Descriptive statistics of T_c (first quartile, second quartile-median, third quartile, mean, standard deviation, asymmetry index) and related censored units as a function of K.

Descriptive statistics of $T_c(K)$							
K	Q1	Q2	Q3	Mean	SD	AI	Censoring rate
0.5%	2.351	5.485	12.798	12.070	23.670	0.835	0.6%
0.9%	7.436	27.407	101.014	177.861	1.140e+03	0.396	9.0%
1.1%	12.168	50.491	209.510	467.415	4.302e+03	0.291	15.0%

In Figs. 5–7, we display the density plots of the time to crash T_c for the best selected parametric models, regarding $K = 0.5\%$, $K = 0.9\%$ and $K = 1.0\%$. Matching the graphs, it is immediate to notice that the densities related to WISMC data are the closest to the real ones for every value of K.

To quantify the distance between the simulated and real distributions of the time to crash T_c, we calculate the Kullback–Leibler divergence [11]. It is sometimes referred to as information divergence, relative entropy or the Kullback–Leibler information criteria (KLIC). We briefly remember that the Kullback–Leibler divergence of the distribution Q from the distribution P, denoted by $D_{KL}(P||Q)$, is the measure of the information lost when Q is used to approximate P. It is defined as follows:

$$D_{KL}(P||Q) = \int_{-\infty}^{+\infty} p(x) \log_2 \left(\frac{p(x)}{q(x)} \right) dx, \tag{19}$$

where p and q denote the probability densities of P and Q. In our framework p and q stand for the synthetic and the real distributions of the time to crash T_c for a selected K-value, respectively. In Table 6 we show the Kullback–Leibler divergences for the

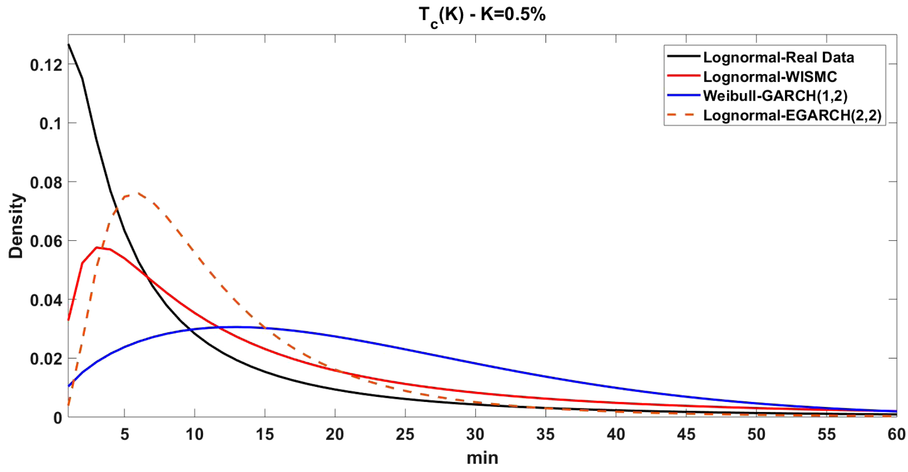

Fig. 5. Density plots of T_c for $K = 0.5\%$.

time to crash T_c, considering three levels of the threshold K ($K = 0.5\%, K = 0.9\%, K = 1.0\%$). It is possible to observe that the WISMC model performs better than the other models, regardless of K.

Once the cdf and the density of the time to crash T_c for the best statistical parametric model have been estimated, it is possible to translate the results to the speed of crash S_c which is a nonlinear transformation of the time to crash T_c. First the cdf can be obtained

$$F_{S_c}(x) = P(S_c \leq x) = P\left(\frac{K}{T_c} \leq x\right) = P\left(\frac{K}{x} \leq T_c\right) = 1 - F_{T_c}\left(\frac{K}{x}\right),$$

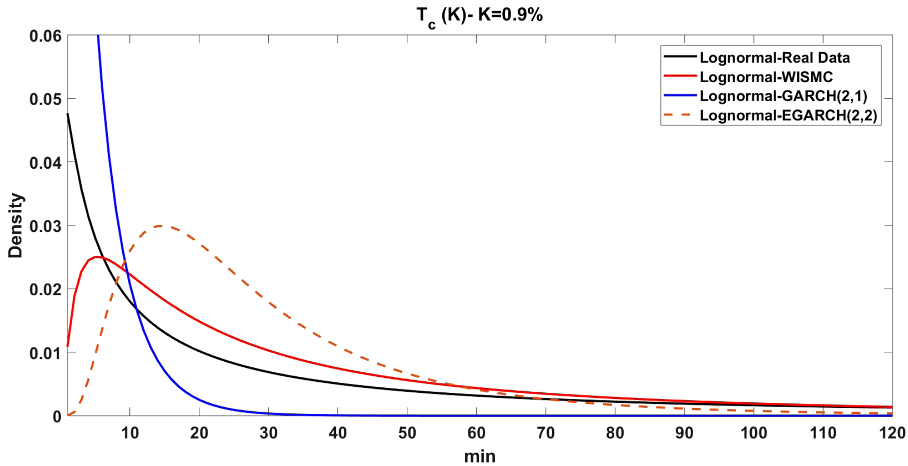

Fig. 6. Density plots of T_c for $K = 0.9\%$.

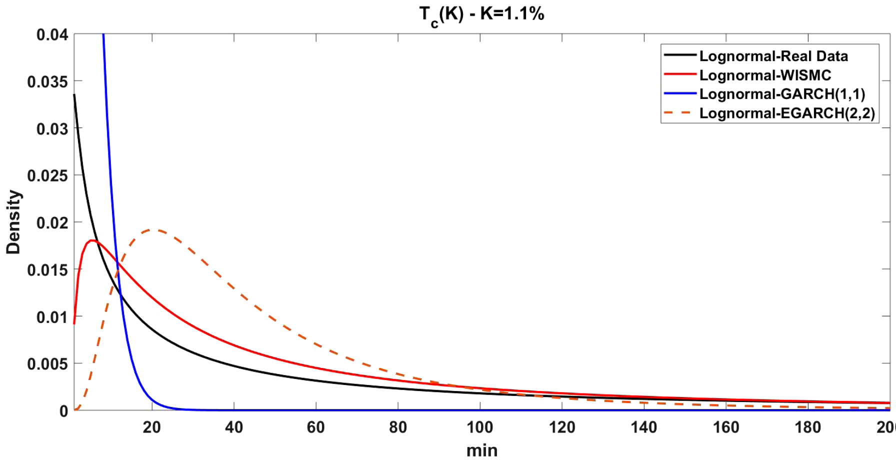

Fig. 7. Density plots of T_c for $K = 1.1\%$.

and by derivation it is possible to obtain the density function:

$$f_{S_c}(x) = \frac{d}{dx}[F_{S_c}(x)] = f_{T_c}\left(\frac{K}{x}\right)\frac{K}{x^2}. \tag{20}$$

In Table 7, we provide the statistics of the speed of crash S_c for different K-values ($K = 0.5\%, K = 0.9\%, K = 1.1\%$). It can be observed that the average speed decreases as K increases. This simply involves that the stock reaches high thresholds more slowly than low thresholds. As a matter of fact, a 0.5% -change in the

Table 6. Kullback–Leibler divergence computed for the risk measure T_c, considering different levels of $K (K = 0.5\%, K = 0.9\%, K = 1.1\%)$. The smaller distances are in bold.

Kullback–Leibler divergence for $T_c(K)$		
	Best Model	KL
$K = 0.5\%$		
WISMC	Lognormal	**0.2705**
GARCH(1,2)	Weibull	0.9189
EGARCH(2,2)	Lognormal	0.4655
$K = 0.9\%$		
WISMC	Lognormal	**0.1816**
GARCH(2,1)	Lognormal	0.6048
EGARCH(2,2)	Lognormal	0.8260
$K = 1.1\%$		
WISMC	Lognormal	**0.1677**
GARCH(1,1)	Lognormal	0.4085
EGARCH(2,2)	Lognormal	0.8522

Table 7. Descriptive statistics of S_c (first quartile, second quartile-median, third quartile, mean, standard deviation, asymmetry index) and related censored units as a function of K.

Descriptive statistics of $S_c(K)$							
K	Q1	Q2	Q3	Mean	SD	AI	Censoring rate
0.5%	4.006e-04	9.167e-04	0.002	0.001	0.001	0.525	0.6%
0.9%	8.911e-05	3.333e-04	0.001	8.934e-04	0.001	1.156	9.0%
1.1%	5.263e-05	2.200e-04	9.167e-04	8.229e-04	0.002	1.154	15.0%

drawdown is achieved with an average velocity of $0.001\,\text{min}^{-1}$ while a bigger variation, such as 1.1%, is attained more slowly, with an average speed of $8.229\text{E-}04\,\text{min}^{-1}$.

Using formula (20), we construct and plot the best density of the speed of crash S_c in function of K as shown in Figures 8–10. Moreover, we don't compute the Kullback–Leibler divergence of S_c as it is invariant under parameter transformations. Therefore, the Kullback–Leibler divergences' values of S_c coincide with those already shown for T_c in Table 6. The results obtained on the measures time to crash and speed of market crash establish a definite superiority of the WISMC model as compared to the GARCH and EGARCH models uniformly in K.

In Tables 8, 13 and 18, we display results on the best statistical model selection for the recovery time R_t, considering several combinations of K and K'. The AIC and BIC values of the best selected models are in bold. In all the analyzed cases, the lognormal law is more suitable than the Weibull and the exponential laws therefore, it represents the best statistical parametric model among those considered in the analysis.

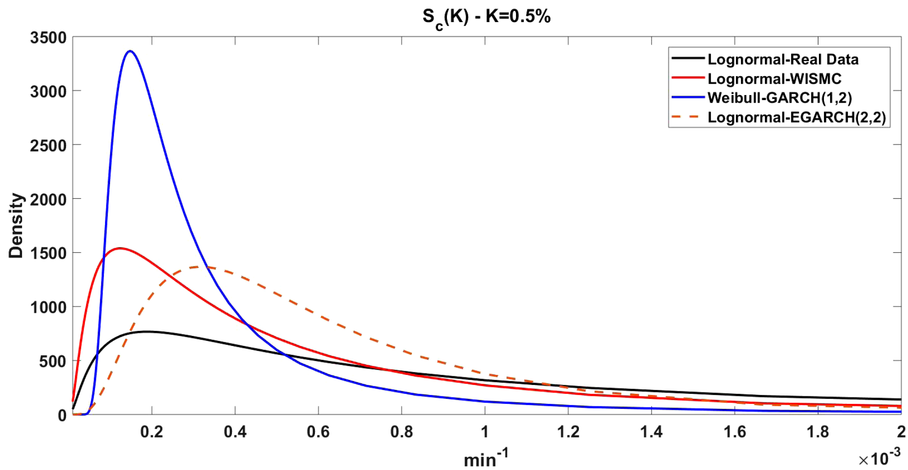

Fig. 8. Density plots of S_c for $K = 0.5\%$.

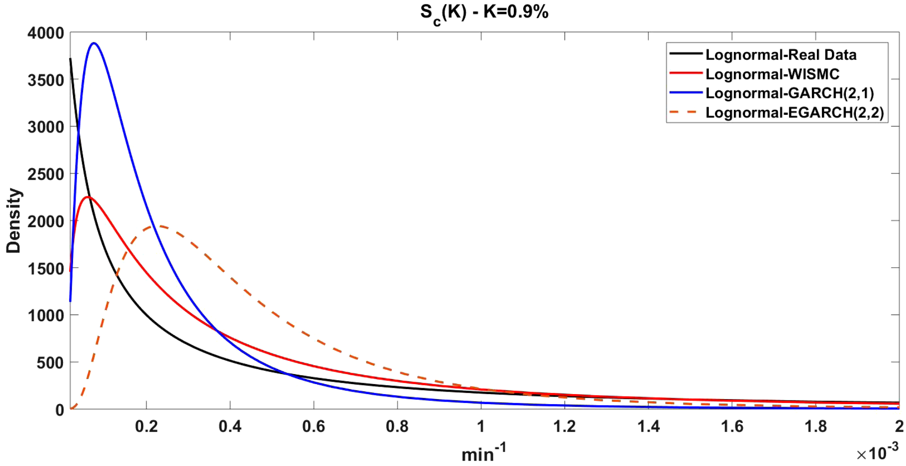

Fig. 9. Density plots of S_c for $K = 0.9\%$.

The parameters' point estimate of the best selected models for the recovery time R_t are shown in Tables 9, 14 and 19. It can be observed that both GARCH and EGARCH models are unstable as each couple of thresholds K and K' identifies different GARCH and EGARCH models.

Tables 10, 15 and 20 display both the main descriptive statistics and the number of censored units as a function of the thresholds K and K'. All these quantities are computed considering the best parametric law for real data as a function of K and K'. In general, the recovery time R_t shows more censored units than the time to crash T_c since it requires two thresholds to be exceeded, K and K'. It is observed that

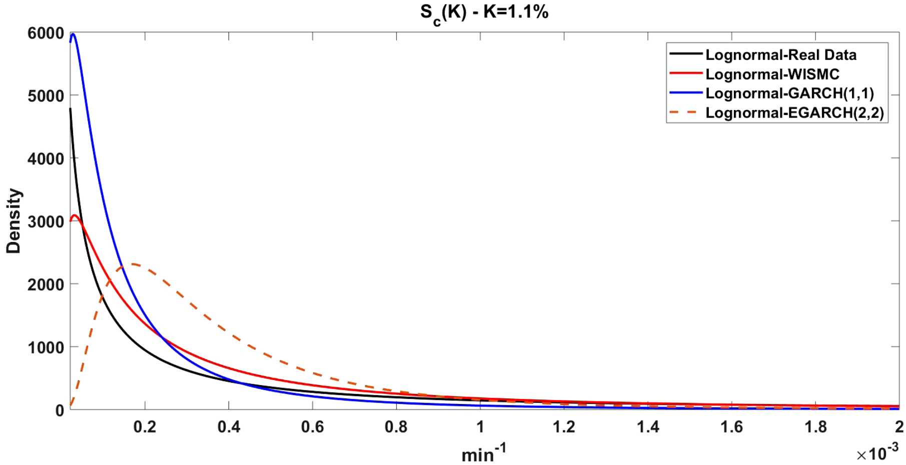

Fig. 10. Density plots of S_c for $K = 1.1\%$.

Table 8. Selection of the best parametric model for the measure R_t on both real and simulated data by mean of AIC and BIC criterion, fixing $K = 0.5\%$ and considering $K' = 0.2\%$ and $K' = 0.3\%$. AIC and BIC relating to the best models are in bold.

Model selection for $R_t(K, K') - K = 0.5\%$						
	Weibull		Lognormal		Exponential	
	AIC	BIC	AIC	BIC	AIC	BIC
$K' = 0.2\%$						
Real Data	8.913E+03	8.923E+03	**8.691E+03**	**8.701E+03**	1.045E+04	1.046E+04
WISMC	9.806E+03	9.816E+03	**9.554E+03**	**9.563E+03**	1.088E+04	1.089E+04
GARCH(1,1)	1.000E+04	1.001E+04	9.743E+03	9.753E+03	1.084E+04	1.085E+04
GARCH(1,2)	1.026E+04	1.027E+04	9.987E+03	9.996E+03	1.103E+04	1.104E+04
GARCH(2,1)	9.842E+03	9.852E+03	**9.609E+03**	**9.619E+03**	1.065E+04	1.066E+04
EGARCH(1,1)	9.921E+03	9.931E+03	9.662E+03	9.671E+03	1.087E+04	1.088E+04
EGARCH(1,2)	9.931E+03	9.941E+03	9.625E+03	9.635E+03	1.100E+04	1.101E+04
EGARCH(2,1)	9.888E+03	9.898E+03	**9.604E+03**	**9.614E+03**	1.091E+04	1.092E+04
EGARCH(2,2)	9.930E+03	9.939E+03	9.635E+03	9.645E+03	1.103E+04	1.104E+04
$K' = 0.3\%$						
Real Data	8.811E+03	8.821E+03	**8.542E+03**	**8.552E+03**	1.073E+04	1.074E+04
WISMC	9.562E+03	9.571E+03	**9.272E+03**	**9.282E+03**	1.094E+04	1.095E+04
GARCH(1,1)	9.680E+03	9.689E+03	**9.386E+03**	**9.396E+03**	1.091E+04	1.091E+04
GARCH(1,2)	1.004E+04	1.005E+04	9.744E+03	9.754E+03	1.108E+04	1.109E+04
GARCH(2,1)	9.781E+03	9.791E+03	9.514E+03	9.523E+03	1.092E+04	1.092E+04
EGARCH(1,1)	9.744E+03	9.754E+03	9.459E+03	9.469E+03	1.101E+04	1.101E+04
EGARCH(1,2)	9.592E+03	9.602E+03	9.261E+03	9.271E+03	1.097E+04	1.097E+04
EGARCH(2,1)	9.613E+03	9.623E+03	9.300E+03	9.310E+03	1.096E+04	1.097E+04
EGARCH(2,2)	9.548E+03	9.558E+03	**9.231E+03**	**9.241E+03**	1.098E+04	1.098E+04

the censorship increases as K increases and it decreases as K' increases. Specifically, they exceed 30% of the total observations for $K = 0.7\% - K' = 0.2\%$, $K' = 0.3\%$, $K' = 0.4\%$. In addition, paying attention to the average values it is possible to note that they decrease as K' increases for each fixed K.

Table 9. Parameters of the best models for R_t, fixing $K = 0.5\%$ and considering $K' = 0.2\%$ and $K' = 0.3\%$.

Summary of the best statistical parametric model for $R_t(K, K') - K = 0.5\%$		
	Best Model	Parameters
$K' = 0.2\%$		
Real Data	Lognormal	4.4397–2.8453
WISMC	Lognormal	4.2893–2.2421
GARCH(2,1)	Lognormal	4.7938–2.1882
EGARCH(2,1)	Lognormal	4.2484–2.1134
$K' = 0.3\%$		
Real Data	Lognormal	3.8138–2.7859
WISMC	Lognormal	3.8102–2.2637
GARCH(1,1)	Lognormal	4.0361–2.2020
EGARCH(2,2)	Lognormal	3.4432–2.1206

Table 10. Descriptive statistics of R_t (first quartile, second quartile-median, third quartile, mean, standard deviation, asymmetry index) and related censored units as a function of K and K'.

Descriptive statistics for $R_t(K, K') - K = 0.5\%$							
K'	Q1	Q2	Q3	Mean	SD	AI	Censoring rate
0.2%	12.436	84.749	577.568	4.854e+03	2.780e+05	0.017	26.0%
0.3%	6.922	45.322	296.741	2.196e+03	1.064e+05	0.020	21.0%

Table 11. Kullback–Leibler divergence computed for the risk measures R_t, fixing $K = 0.5\%$ and considering $K' = 0.2\%$ and $K' = 0.3\%$. The smallest distances are in bold.

Kullback–Leibler for $R_t(K, K') - K = 0.5\%$		
	Best Model	KL
$K' = 0.2\%$		
WISMC	Lognormal	**0.0723**
GARCH(2,1)	Lognormal	0.0953
EGARCH(2,1)	Lognormal	0.1089
$K' = 0.3\%$		
WISMC	Lognormal	**0.0544**
GARCH(1,1)	Lognormal	0.0732
EGARCH(2,2)	Lognormal	0.1030

Moreover, the recovery time R_t is more easily predictable than the time to crash T_c as for every combination of K and K' all the models have similar performances. This evidence is shown in Figure 11 which reports the density plot of the recovery time R_t for $K = 0.5\%$ and $K' = 0.2\%$.

Tables 11, 16 and 21 report the Kullback–Leibler divergences for the measure R_t. As we can observe, if the censored data is less than 30% the WISMC model is the closest to real data, except for the case $K = 0.9\% - K' = 0.7\%$ where GARCH is slightly better than WISMC. Conversely, if the percentage of censored data exceeds 30% the GARCH models is preferable to the WISMC and EGARCH models.

Table 12. Descriptive statistics of S_r (first quartile, second quartile-median, third quartile, mean, standard deviation, asymmetry index) and related censored units as a function of K and K'.

Descriptive statistics of $S_r(K, K') - K = 0.5\%$							
K'	Q1	Q2	Q3	Mean	SD	AI	Censoring rate
0.2%	5.499e-06	3.704e-05	2.614e-04	2.037e-04	4.447e-04	1.124	26.0%
0.3%	6.908e-06	4.495e-05	3.095e-04	1.689e-04	3.304e-04	1.125	21.0%

Table 13. Selection of the best parametric model for the measure R_t on both real and simulated data by mean of AIC and BIC criterion, fixing $K = 0.7\%$ and considering $K' = 0.2\%$ and $K' = 0.3\%$. AIC and BIC relating to the best models are in bold.

Model selection for $R_t(K, K') - K = 0.7\%$						
	Weibull		Lognormal		Exponential	
	AIC	BIC	AIC	BIC	AIC	BIC
$K' = 0.2\%$						
Real Data	8.165e+03	8.174e+03	**8.017e+03**	**8.027e+03**	9.092e+03	9.102e+03
WISMC	9.439e+03	9.449e+03	**9.229e+03**	**9.239e+03**	1.022e+04	1.023e+04
GARCH(1,1)	9.374e+03	9.384e+03	9.182e+03	9.192e+03	9.892e+03	9.902e+03
GARCH(1,2)	9.514e+03	9.523e+03	9.307e+03	9.317e+03	1.005e+04	1.006e+04
GARCH(2,1)	9.101e+03	9.111e+03	**8.922e+03**	**8.932e+03**	9.600e+03	9.610e+03
EGARCH(1,1)	9.763e+03	9.772e+03	9.538e+03	9.548e+03	1.037e+04	1.038e+04
EGARCH(1,2)	9.965e+03	9.975e+03	9.702e+03	9.711e+03	1.074e+04	1.075e+04
EGARCH(2,1)	9.733e+03	9.743e+03	**9.494e+03**	**9.504e+03**	1.039e+04	1.040e+04
EGARCH(2,2)	1.010e+04	1.011e+04	9.843e+03	9.853e+03	1.081e+04	1.082e+04
$K' = 0.3\%$	AIC	BIC	AIC	BIC	AIC	BIC
Real Data	8.382e+03	8.392e+03	**8.217e+03**	**8.227e+03**	9.500e+03	9.501e+03
WISMC	9.534e+03	9.544e+03	**9.305e+03**	**9.315e+03**	1.047e+04	1.047e+04
GARCH(1,1)	9.420e+03	9.430e+03	9.210e+03	9.220e+03	1.014e+04	1.014e+04
GARCH(1,2)	9.845e+03	9.855e+03	9.615e+03	9.625e+03	1.050e+04	1.050e+04
GARCH(2,1)	9.359e+03	9.368e+03	**9.158e+03**	**9.168e+03**	1.002e+04	1.003e+04
EGARCH(1,1)	9.974e+03	9.984e+03	9.724e+03	9.734e+03	1.071e+04	1.072e+04
EGARCH(1,2)	1.001e+04	1.002e+04	9.719e+03	9.729e+03	1.091e+04	1.092e+04
EGARCH(2,1)	9.789e+03	9.799e+03	**9.524e+03**	**9.534e+03**	1.065e+04	1.065e+04
EGARCH(2,2)	1.008e+04	1.009e+04	9.810e+03	9.820e+03	1.095e+04	1.096e+04

Table 14. Parameters of the best models for the measure R_t, fixing $K = 0.7\%$ and considering $K' = 0.2\%$ and $K' = 0.3\%$.

Summary of the best statistical model selection for $R_t(K, K') - K = 0.7\%$		
	Best Model	Parameters
$K' = 0.2\%$		
Real Data	Lognormal	5.8196–3.0246
WISMC	Lognormal	5.1304–2.3359
GARCH(2,1)	Lognormal	5.7300–2.2500
EGARCH(2,1)	Lognormal	5.0986–2.1024
$K' = 0.3\%$		
Real Data	Lognormal	5.4377–3.0407
WISMC	Lognormal	4.8123–2.3468
GARCH(2,1)	Lognormal	5.3411–2.2941
EGARCH(2,1)	Lognormal	4.6954–2.1369

Table 15. Descriptive statistics of R_t (first quartile, second quartile-median, third quartile, mean, standard deviation, asymmetry index) and related censored units as a function of K and K'.

Descriptive statistics of $R_t(K, K') - K = 0.7\%$							
K'	Q1	Q2	Q3	Mean	SD	AI	Censoring rate
0.2%	43.796	336.837	2.591e+03	3.265e+04	3.165e+06	0.010	41.0%
0.3%	29.571	229.913	1.788e+03	2.340e+04	2.382e+06	0.010	37.0%

Table 16. Kullback–Leibler divergence computed for the risk measures R_t, fixing $K = 0.7\%$ and considering $K' = 0.2\%$ and $K' = 0.3\%$. The smallest distances are in bold.

Kullback–Leibler for $R_t(K, K') - K = 0.7\%$		
	Best Model	KL
$K' = 0.2\%$		
WISMC	Lognormal	0.1191
GARCH(2,1)	Lognormal	**0.1053**
EGARCH(2,1)	Lognormal	0.1929
$K' = 0.3\%$		
WISMC	Lognormal	0.1126
GARCH(2,1)	Lognormal	**0.0965**
EGARCH(2,1)	Lognormal	0.1868

Since the speed of recovery S_r is a nonlinear transformation of the recovery time R_t, we gain its cdf analytically as follows

$$\begin{aligned} F_{S_r}(x) &= P(S_r \leq x) = P\left(\frac{(K-K')}{R_t} \leq x\right) = P\left(\frac{(K-K')}{x} \leq R_t\right) \\ &= 1 - F_{R_t}\left(\frac{(K-K')}{x}\right). \end{aligned}$$

Consequently, its density is given by

$$f_{S_r}(x) = \frac{d}{dx}\left[F_{S_r}(x)\right] = f_{R_t}\left(\frac{(K-K')}{x}\right)\frac{(K-K')}{x^2}. \tag{21}$$

Table 17. Descriptive statistics of S_r (first quartile, second quartile-median, third quartile, mean, standard deviation, asymmetry index) and related censored units as a function of K and K'.

Descriptive statistics of $S_r(K, K') - K = 0.7\%$							
K'	Q1	Q2	Q3	Mean	SD	AI	Censoring rate
0.2%	2.366e-06	1.751e-05	1.409e-04	1.976e-04	5.623e-04	0.961	41.0%
0.3%	2.614e-06	1.975e-05	1.569e-04	1.886e-04	4.961e-04	1.021	37.0%

Table 18. Selection of the best parametric model for the measure R_t on both real and simulated data by mean of AIC and BIC criterion, fixing $K = 0.7\%$ and considering $K' = 0.4\%$ and $K' = 0.5\%$. AIC and BIC relating to the best models are in bold.

Model selection for $R_t(K, K') - K = 0.7\%$						
	Weibull		Lognormal		Exponential	
	AIC	BIC	AIC	BIC	AIC	BIC
$K' = 0.4\%$						
Real Data	8.628e+03	8.638e+03	**8.429e+03**	**8.439e+03**	1.002e+04	1.002e+04
WISMC	9.538e+03	9.548e+03	**9.289e+03**	**9.299e+03**	1.071e+04	1.071e+04
GARCH(1,1)	9.621e+03	9.631e+03	**9.386e+03**	**9.396e+03**	1.055e+04	1.056e+04
GARCH(1,2)	9.867e+03	9.877e+03	9.619e+03	9.629e+03	1.075e+04	1.076e+04
GARCH(2,1)	9.625e+03	9.635e+03	9.394e+03	9.403e+03	1.052e+04	1.052e+04
EGARCH(1,1)	9.910e+03	9.920e+03	9.632e+03	9.642e+03	1.090e+04	1.091e+04
EGARCH(1,2)	9.828e+03	9.838e+03	**9.514e+03**	**9.524e+03**	1.087e+04	1.087e+04
EGARCH(2,1)	9.811e+03	9.821e+03	9.514e+03	9.524e+03	1.103e+04	1.104e+04
EGARCH(2,2)	9.928e+03	9.938e+03	9.638e+03	9.648e+03	1.177e+04	1.177e+04
$K' = 0.5\%$						
Real Data	8.602e+03	8.611e+03	**8.352e+03**	**8.361e+03**	1.045e+04	1.046e+04
WISMC	9.297e+03	9.307e+03	**9.018e+03**	**9.028e+03**	1.083e+04	1.084e+04
GARCH(1,1)	9.500e+03	9.510e+03	**9.223e+03**	**9.233e+03**	1.080e+04	1.080e+04
GARCH(1,2)	9.64e+03	9.650e+03	9.345e+03	9.354e+03	1.090e+04	1.091e+04
GARCH(2,1)	9.539e+03	9.549e+03	9.272e+03	9.282e+03	1.078e+04	1.079e+04
EGARCH(1,1)	9.663e+03	9.673e+03	9.345e+03	9.355e+03	1.098e+04	1.098e+04
EGARCH(1,2)	9.501e+03	9.511e+03	**9.156e+03**	**9.165e+03**	1.092e+04	1.092e+04
EGARCH(2,1)	9.627e+03	9.637e+03	9.295e+03	9.305e+03	1.096e+04	1.096e+04
EGARCH(2,2)	9.627e+03	9.637e+03	9.309e+03	9.319e+03	1.099e+04	1.100e+04

In Tables 12, 17 and 22 we show the main descriptive statistics and the number of censored units for the speed of recovery S_r, considering different values of K and K'. Focusing on the average speeds, we note that they decrease as K' increases for each fixed K. For instance, in Table 12 we see that after recording the first 0.5%-change in the drawdown, there is a faster drop of 0.2% rather than 0.3%.

Table 19. Parameters of the best models for the measure R_t, fixing $K = 0.7\%$ and considering $K' = 0.4\%$ and $K' = 0.5\%$.

Summary of the best statistical model selection for $R_t(K, K') - K = 0.7\%$		
	Best Model	Parameters
$K' = 0.4\%$		
Real Data	Lognormal	4.8758–2.9769
WISMC	Lognormal	4.3804–2.3778
GARCH(1,1)	Lognormal	4.7477–2.3065
EGARCH(1,1)	Lognormal	3.9622–2.0625
$K' = 0.5\%$		
Real Data	Lognormal	4.1419–2.9328
WISMC	Lognormal	3.8590–2.4380
GARCH(1,1)	Lognormal	4.1426–2.3417
EGARCH(1,2)	Lognormal	3.4267–2.0523

Table 20. Descriptive statistics of R_t (first quartile, second quartile-median, third quartile, mean, standard deviation, asymmetry index) and related censored units as a function of K and K'.

Descriptive statistics of $R_t(K,K') - K = 0.7\%$							
K'	Q1	Q2	Q3	Mean	SD	AI	Censoring rate
0.4%	17.600	131.079	976.221	1.101e+04	9.251e+05	0.012	31.0%
0.5%	8.7048	62.922	454.884	4.640e+03	3.422e+05	0.013	25.0%

Table 21. Kullback–Leibler divergence computed for the risk measures R_t, fixing $K = 0.7\%$ and considering $K' = 0.4\%$ and $K' = 0.5\%$. The smallest distances are in bold.

Kullback–Leibler for $R_t(K,K') - K = 0.7\%$		
	Best Model	KL
$K' = 0.4\%$		
WISMC	Lognormal	0.0830
GARCH(2,1)	Lognormal	**0.0811**
EGARCH(2,1)	Lognormal	0.2223
$K' = 0.5\%$		
WISMC	Lognormal	**0.0504**
GARCH(2,1)	Lognormal	0.0633
EGARCH(2,1)	Lognormal	0.1898

Table 22. Descriptive statistics of S_r (first quartile, second quartile-median, third quartile, mean, standard deviation, asymmetry index) and related censored units as a function of K and K'.

Descriptive statistics of $S_r(K,K') - K = 0.7\%$							
K'	Q1	Q2	Q3	Mean	SD	AI	Censoring rate
0.4%	3.341e-06	2.449e-05	1.820e-04	1.755e-04	4.157e-04	1.090	31.0%
0.5%	4.561e-06	3.252e-05	2.361e-04	1.519e-04	3.165e-04	1.131	25.0%

Figure 12 depicts the best densities obtained using formula (21) for $K = 0.5\%$ and $K' = 0.2\%$. As we can observe, also the speed of recovery S_r is more foreseeable than the speed of crash S_c since all the models have the same performance. Furthermore, we don't report the Kullback–Leibler divergence of S_r because it is invariant with respect to parameter transformations. Consequently, the Kullback–Leibler divergence's values of the speed of recovery S_r coincide with those already shown in Tables 11, 16 and 21 for the recovery time R_t.

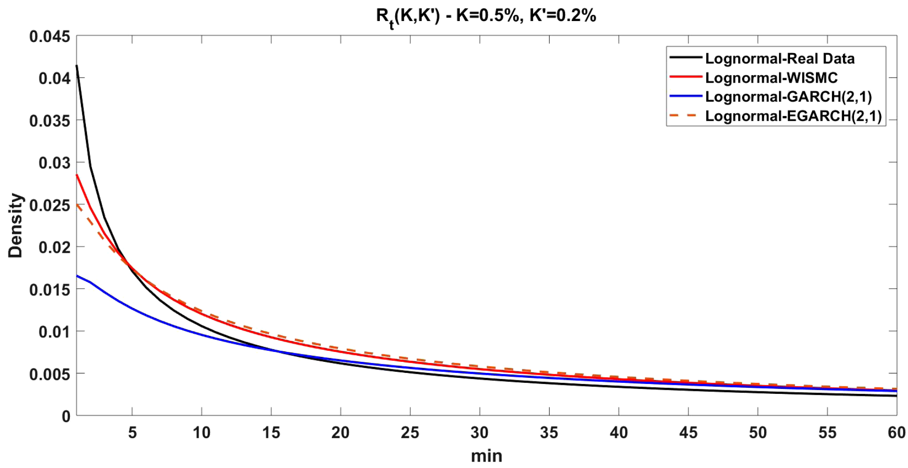

Fig. 11. Density plots of R_t for $K = 0.5\%$ and $K' = 0.2\%$.

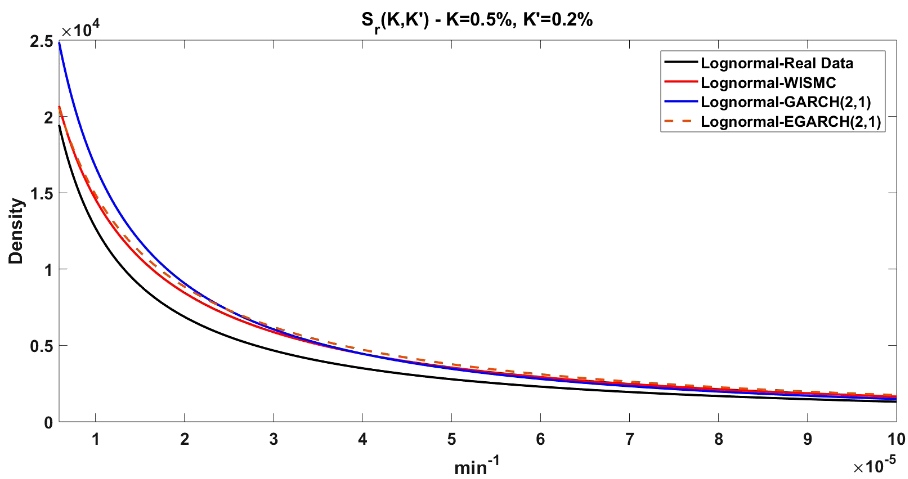

Fig. 12. Density plots of S_r for $K = 0.5\%$ and $K' = 0.2\%$.

5. Concluding Remarks

We analyze several risk measures related to market crises: the drawdown of fixed level, the time to crash, the speed of crash, the recovery time and the speed of recovery. In detail, we study these drawdown-based risk indicators using high-frequency data of Fiat stock, listed on the Italian Stock Exchange. By applying a variant of the classic semi-Markov chain (SMC) model, named weighted-indexed semi-Markov (WISMC) model, we generate a synthetic series of returns. To test the potency of our model we make comparisons with the GARCH and EGARCH models, simulating a synthetic series of returns for each selected econometric model. Next, we compute the drawdown-based risk measures on both real and simulated data and we explore them through parametric models whose estimation procedures are carried out considering the right censorship.

Globally, the WISMC model provides better results than the chosen GARCH and EGARCH models for the measure time to crash, regardless of the chosen K. On the contrary, for the measure recovery time the WISMC model is more efficient only when the number of censored units are less than 30%. In this last situation, the performance of the WISMC model probably decreases because the number of data is insufficient to estimate the kernel of the process. Therefore, our future goal is to use the WISMC model but with an adequate dataset.

Acknowledgments

The authors thank two anonymous referees and the editor for their valuable comments on the earlier versions of this paper.

6. Appendix

Table A.1 provides a list of the financial symbols involved in the analysis.

Table A.1. List of financial symbols

$X(t)$	Price process
$Y(t)$	Running maximum process
$D(t)$	Drawdown process
$\tau(K)$	Drawdown of fixed level
$\rho(K)$	last visit time of the maximum price
$T_c(K)$	Time to crash
$S_c(K)$	Speed of crash
$\gamma(K, K')$	first time drawdown is below K' after crossing level K
$R_t(K, K')$	Recovery time
$S_r(K, K')$	Speed of recovery

References

[1] Akaike, H., A new look at the statistical model identification, *IEEE Trans. Automat. Contr.* **19** (1974) 716–723.

[2] Andersen, T. G., Bollerslev, T., Frederiksen, P. and Nielsen, M. O., Continuous-time Models, realized volatilities, and testable distribution implications for daily stock returns, *J. Appl. Econometrics* **25** (2010) 233–261.

[3] Chen, C. W. S. and Watanabe, T., Bayesian modeling and forecasting of Value-at-Risk via threshold realized volatility, *Appl. Stoch. Model. Bus. Ind.* **35** (2019) 747–765.

[4] Cont, R., Deguest, R. and Scandolo, G., Robustness and sensitivity analysis of risk measurement procedures, *Quant. Finance.* **10** (2010) 593–606.
[5] D'Amico G. and Petroni, F., A semi-Markov model with memory for price changes, *J. Stat. Mech. Theory Exp.* **2011** (2011) P12009.
[6] D'Amico G. and Petroni, F., A semi-Markov model for price returns, *J. Phys A.* **391** (2012) 4867–4876.
[7] D'Amico G. and Petroni, F., Weighted-indexed semi-Markov model for modeling financial returns, *J. Stat. Mech. Theory Exp.* **2012** (2012) P07015.
[8] D'Amico G. and Petroni, F., Copula based multivariate semi-Markov models with application in high-frequency finance, *European J. Oper. Res.* **267** (2018) 765–777.
[9] Engle, R. F., Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations, *Econometrica* **50** (1982) 987–1007.
[10] Kaplan, E. L. and Meier, P., Nonparametric estimation from incomplete observations, *J. Am. Stat. Assoc.* **53** (1958) 457–481.
[11] Kullback, S. and Leibler, R. A., On Information and Sufficiency, *Ann. Math. Statist.* **22** (1951) 79–86.
[12] Liu, R., Di Matteo, T. and Lux, T., Multifractality and long-range dependence of asset returns: the scaling behavior of the Markov-switching multifractal model with lognormal volatility components, *Adv. Complex Syst.* **11** (2008) 669–684.
[13] Nelson, D. B., Conditional heteroskedasticity in asset return: A new approach, *Econometrica* **59** (1991) 347–370.
[14] Pospisil, L., Vecer, J. and Hadjiliadis, O., Formulas for stopped diffusion processes with stopping times based on drawdowns and drawups, *Stoch. Process. Their Appl.* **119** (2009) 2563–2578.
[15] Schwarz, G. E., Estimating the dimension of a model, *Ann. Stat.* **6** (1978) 461–464.
[16] Vrontos, I. D., Giakoumatos, S. G., Dellaportas P. and Politis, D. N., An application of three bivariate time-varying volatility models, *Appl. Stoch. Model. Bus. Ind.* **17** (2001) 121–133.
[17] Zhang, H. and Hadjiliadis, O., Drawdown and the Speed of Market Crash, *Methodol. Comput. Appl. Probab.* **14** (2012) 739–752.
[18] Zhang, H., *Stochastic Drawdowns* (World Scientific, Singapore, 2018).