When to Be Discrete: The Importance of Time Formulation in the Modeling of Extreme Events in Finance

Katarzyna Bień-Barkowska[a], Rodrigo Herrera[b]

[a]*Institute of Econometrics, Warsaw School of Economics, Madalinskiego 6/8, Warsaw, 02-513, Poland*
[b]*Facultad de Economía y Negocios, Universidad de Talca, Av. Lircay s/n, Talca, Chile*

Abstract

We propose a novel extension of the score-driven POT (SPOT) model within a discrete-time framework. This adaptation is motivated by the fact that financial returns and, consequently, extreme events are typically observed at discrete time intervals. Our primary objective is to assess whether this discrete-time SPOT model provides a more precise representation and superior fit for tail risk forecasting. The study reveals several important findings. First, we demonstrate that continuous-time approaches can result in inaccurate value-at-risk and expected shortfall forecasts. In contrast, discrete-time models provide a more accurate description of the dynamics of extreme losses. Empirical evidence supports the superiority of discrete-duration models, outperforming various continuous-time SPOT specifications and GARCH models. Overall, our study has substantial implications for the modeling and forecasting of extreme financial events, offering a more accurate and efficient approach than traditional approaches.

Keywords: Extreme events, Tail risk, Discrete time, Value at Risk, Score-driven models

1. Introduction

Over the last decade, the dynamics of extreme events, and specifically, tail risk, has been a highly researched topic in financial econometrics. One of the main debates in this field is whether volatility models or some of their extensions that incorporate high-frequency information are sufficient to capture tail risk or if alternative methods are needed. For instance, Trapin (2018) shows that models such as the heterogeneous autoregressive (HAR) model by Corsi (2009) do not generate clustering of extreme events, a commonly observed stylized fact of stock market returns. However, high-frequency volatility models that exhibit extreme dependence and leverage effects are the only ones that can explain the dynamics of extreme financial events. On the other hand, Hoga (2021) shows that while the use of covariate information in volatility models can improve the accuracy of extreme risk forecasts, it also increases uncertainty. In particular, additional covariates can lead to overfitting and increased forecast uncertainty, particularly for extreme events.

Several authors have recently proposed alternative risk forecasting methods that directly exploit the dynamics of extreme events, i.e., losses exceeding a high threshold. There are two leading methodologies in the literature. The first relies on a censored time series model for the time-varying probability of an extreme event (i.e., exceedance probability), while the dynamics associated with the magnitude of extreme event is captured using a time-varying generalized Pareto distribution. Here, the exceedance probability is defined as the probability that a loss will exceed a high threshold for a predefined future time period. Thus, the occurrence or non-occurrence of an extreme event is characterized by a dichotomous variable, which is subsequently modelled using a dynamic specification. For instance, Massacci (2017) and Bee et al. (2019) proposed two different dynamic specifications to model the exceedance probability. Massacci (2017) harness this probability utilizing a dynamic power law function, with parameters equipped with a score-based updating mechanism, while Bee et al. (2019) utilize a dynamic logit model with realized measures of volatility as covariates.

The second methodology uses the marked point process (MPP) theory, which defines the

timing and severity of extreme events through a dynamic conditional intensity function and a time-varying generalized Pareto distribution. Unlike the first methodology, the second one accounts for the dynamics of inter-exceedance times, which are defined as the time intervals between any consecutive extreme events, by offering a wide range of specifications. Some of the most recent empirical applications include the Hawkes-POT model introduced by Chavez-Demoulin and McGill (2012) in the context of high-frequency financial time series to estimate high-quantile-based risk measures, the multivariate intensity POT model proposed by Hautsch and Herrera (2020) to jointly model the occurrence of extreme observations and the temporal clustering of their magnitudes using an autoregressive multiplicative error model, and the SPOT approach introduced by Fuentes et al. (2023) whose main feature is that the dynamics of the parameters are functions of the inter-exceedance times and magnitudes through a score-driven specification.

In this study, we opt for the second methodology and propose a novel extension of the SPOT model in a discrete-time framework. This approach is motivated by the fact that returns in finance are typically observed at discrete intervals, such as daily or weekly, and thus extreme events are also observed over a grid of discrete time points. While continuous-time models offer an adequate representation of the dynamics of extreme events (Herrera and Schipp, 2013; Fuentes et al., 2018; Hautsch and Herrera, 2020), the aim of this article is to determine whether a discrete-time model offers a more accurate representation and a better fit for tail risk forecasting.

We conducted a comprehensive forecasting study to evaluate the overall accuracy of the proposed discrete-time approach. The key findings are as follows. First, a Monte Carlo study shows that continuous-time models tend to deliver biased estimates of risk because they assume that inter-exceedance times range from zero to infinity and not over a grid of discrete values from one to infinity, as it is observed in daily returns. Second, the empirical analysis shows that discrete-time SPOT models offer a better in-sample fit than their continuous counterparts do. Third, the discrete SPOT specifications show superior performance in terms

of the Value at Risk (VaR) and Expected Shortfall (ES) risk measures during the out-of-sample period. In particular, for higher tail risk levels, the discrete-time SPOT specifications rank among the top prediction models, including comparisons with GARCH volatility models. This result is most evident during financial turmoil periods, such as the recent COVID19 pandemic.

The remainder of this paper is organized as follows. In Section 2, we motivate the study of some stylized factors associated with extreme events in important stock market indices. Section 3 introduces the discrete-time SPOT framework, and highlights its flexibility and possible specifications. Section 4 presents a Monte Carlo study that characterizes the primary differences between employing discrete and continuous hazard functions in the SPOT models. Section 5 presents our empirical results. Finally, we conclude.

2. Motivation

Figure 1 presents the sample path of daily negative log returns, shown as percentages, i.e., $y_t = -100 \cdot \ln(p_t/p_{t-1})$, where p_t represents the closing price for the Dow Jones index from January 2, 1981, to December 28, 2022. We observe that volatility clustering in the data corresponds to clustering of large negative returns. For this initial illustration, we define an extreme event as a negative return that surpasses the 95th percentile threshold. In this case, both the occurrence and magnitude of these large losses exhibit a high degree of serial dependence. First, the inter-exceedance times tend to be positively autocorrelated so that they follow a wave-like pattern. When a negative return exceeds this threshold, there is a heightened likelihood that another return of similar magnitude will occur in the near future. Second, if an exceptionally large negative return above this threshold is recorded, the likelihood that the subsequent negative extreme return will also be substantial increases.

In addition to the abovementioned clustering effects, we aim to shed light on a second observation. An important aspect of extreme returns in financial markets is that the arrival of such extreme events occurs in discrete time because prices are commonly observed at regular

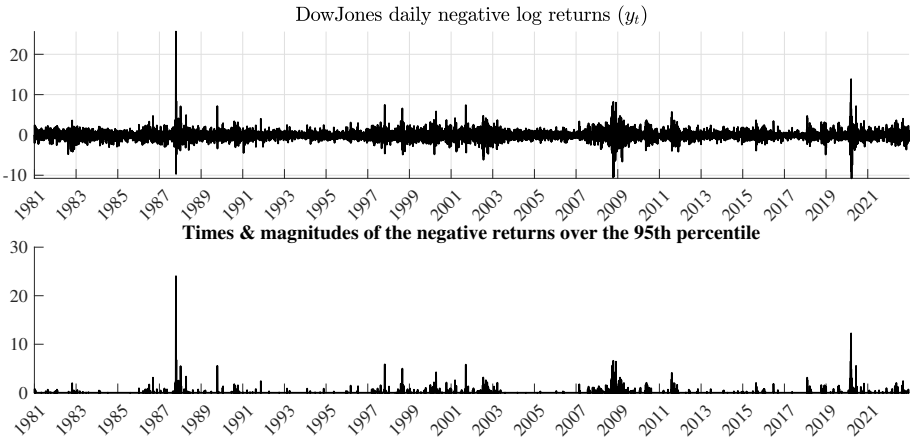

Figure 1: Daily negative log return time series of the Dow Jones Index from January 3, 1981, to December 28, 2022. Highlighted within this series are the specific dates and magnitudes on which negative returns exceed the 95th percentile.

time intervals, such as weekly, daily, every 5 min, or 1 s. Additionally, the corresponding inter-exceedance times exhibit stylized features, such as positive serial correlation, that have been widely documented in literature using continuous-valued time series (Chavez-Demoulin and McGill, 2012; Herrera and Schipp, 2013; Hautsch and Herrera, 2020). However, current literature adopting a discrete-time framework is limited (see Bień-Barkowska, 2020; Bień-Barkowska, 2023).

The left panel of Figure 2 illustrates a prominent feature of extreme stock market returns for a set of well-known US stock market indexes, including Dow Jones, Nasdaq, S&P 500, and Wilshire 5000. Assuming thresholds that consider 10%, 7%, and 5% of the largest losses in these markets, we observe that most of the inter-exceedance times are distributed among discrete values ranging from one to three days. Additionally, as the threshold increases, so does the frequency of inter-exceedance times within the one to three day range. Furthermore, the tail distribution of the inter-exceedance times exhibits heavy-tailed behavior.

These stylized facts indicate that extreme events tend to occur close to each other, which can be described by means of the clustering coefficient $\theta \in [0, 1]$ called the extremal index. The extremal index measures the degree of short-term dependence in extreme events, whereas the inverse of the extremal index is approximately the mean cluster size (see Ferro and Segers,

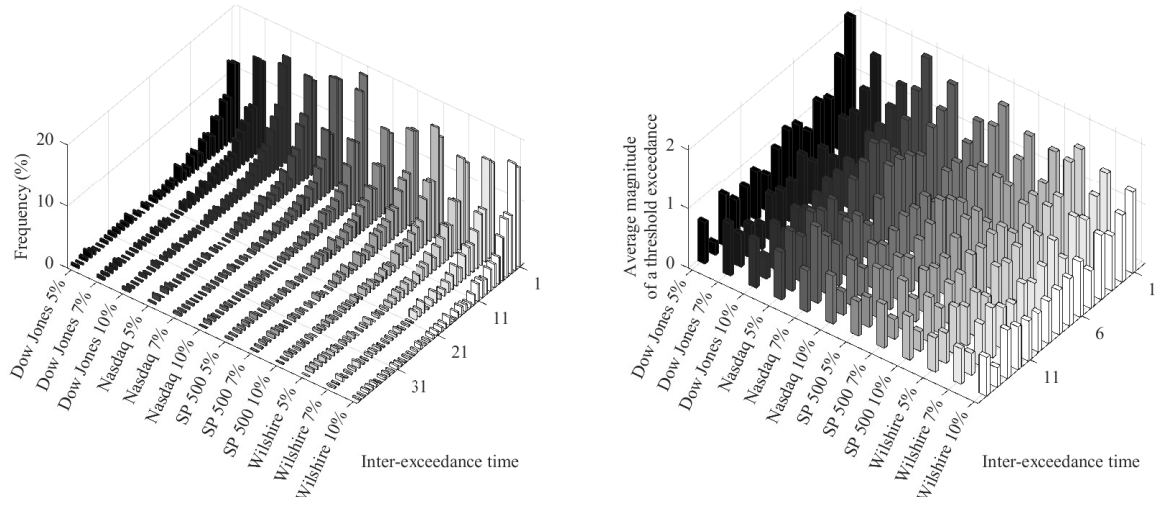

Figure 2: Left panel: Frequency histogram for the inter-exceedance times of four major stock market returns from January 3, 1981, to December 28, 2022, for different threshold values. Right panel: Average magnitude of extreme events associated with the threshold exceedance time. The thresholds cut off 10%, 7%, and 5% of the most negative extreme returns in these markets.

2003; Ferreira and Ferreira, 2018; Holešovský and Fusek, 2020).

Table 1 illustrates the relationship between the extremal index and mean cluster size. For stock market indexes, we have observed extremal indices falling between 0.22 to 0.44 for various thresholds. This corresponds to mean cluster sizes ranging from 4.5 to 2.2 days and suggests a tendency for a short (long) inter-exceedance time to be followed by another short (long) inter-exceedance time. Hence, we can employ heavy-tailed autoregressive models to capture this dependence for integer-valued time series. Additionally, it is important to note that short inter-exceedance times are typically associated with extreme events of larger average magnitudes. This relationship becomes more prominent at higher thresholds, as shown in the right-hand panel of Figure 2.

In the next section, we introduce a score-based approach that employs well-known discrete distributions and their continuous counterparts to capture the clustering behavior of the inter-exceedance times. Simultaneously, we harness the dynamics of extreme event magnitudes by incorporating a time-varying scale parameter within the generalized Pareto distribution function.

Thresholds	Dow Jones	Nasdaq	S&P 500	Wilshire
10%	0.44	0.40	0.46	0.44
7%	0.44	0.31	0.37	0.40
5%	0.35	0.22	0.28	0.24

Table 1: Estimates of the extremal index using the interval estimator proposed by Ferro and Segers (2003), which relies on the asymptotic results of the limiting distribution of inter-exceedance times. The results consider 10%, 7%, and 5% of the most negative extreme events in these stock market indices.

3. The discrete-time SPOT framework

Let $Y_t \in \mathbb{R}$ be a random variable for a negative log-return observed on day t. Thus, a large positive value of Y_t represents the realization of a large loss on day t. Under the POT approach, an *extreme loss* is defined as such Y_t that exceeds a *sufficiently* high threshold $u > 0$ (Smith, 1984). Let $T_i \in \mathbb{Z}_{>0}$ (with $i \in \mathbb{Z}_{>0}$) denote a discrete random variable, representing time at which the i-th extreme loss occurs, i.e., $Y_{t_i} > u$, while $\tilde{Y}_i := Y_{t_i} - u$ denotes the (random) magnitude of the i-th threshold exceedance. In particular, the sequence $\{t_i, \tilde{y}_i\}_{i\in\{1,2,...,n\}}$ represents the sample path of MPP, where t_i is the realized *occurrence time* of the i-th extreme event and $\tilde{y}_i$ denotes the associated magnitude or *mark*.

A simple MPP can be described by a sequence of event times t_i, or a sequence of corresponding durations (i.e., inter-exceedance times), $x_i = t_i - t_{i-1}$ for $i = 2, 3, ...$ (with $x_1 = t_1$). Letting $\mathcal{F}_{t_{i-1}}$ be the σ-field generated by $\{t_1, ..., t_{i-2}, t_{i-1}, \tilde{y}_1, ..., \tilde{y}_{i-2}, \tilde{y}_{i-1}\}$, the evolution of MPP can be expressed in terms of a joint conditional probability density function of the random ith inter-exceedance time $X_i \in \mathbb{Z}_{>0}$, and the associated mark $\tilde{Y}_i$:

$$f\left(x_i, \tilde{y}_i \mid \mathcal{F}_{t_{i-1}}\right) \;=\; g\left(x_i \mid \mathcal{F}_{t_{i-1}}\right) q\left(\tilde{y}_i \mid x_i, \mathcal{F}_{t_{i-1}}\right),$$

where $g\left(x_i \mid \mathcal{F}_{t_{i-1}}\right)$ is the probability function of the inter-exceedance time X_i conditional on the bivariate filtration of threshold exceedance times and their magnitudes $\mathcal{F}_{t_{i-1}}$, and $q\left(\tilde{y}_i \mid x_i, \mathcal{F}_{t_{i-1}}\right)$ is the probability density function of the threshold exceedance $\tilde{Y}_i$ conditional on the contemporaneous inter-exceedance time and $\mathcal{F}_{t_{i-1}}$. The latter density function is parameterized using the conditional generalized Pareto distribution, which will be described in

Section 3.2. An adequate parametrization of $g\left(x_i \mid \mathcal{F}_{t_{i-1}}\right)$ and $q\left(\tilde{y}_i \mid x_i, \mathcal{F}_{t_{i-1}}\right)$ is very important, since it allows us to investigate the time-varying daily probability of an extreme event arrival, the associated expected magnitude of the threshold exceedance, as well as their joint dynamics.

The likelihood function associated with the distribution of n points (i.e., threshold exceedances) over the period $(0, \mathcal{T}]$ can be expressed in terms of a joint density $f(x_i, \tilde{y}_i \mid \mathcal{F}_{t_{i-1}})$, also provided by Chavez-Demoulin and McGill (2012), as follows

$$\begin{aligned}\mathcal{L} &= \left[\prod_{i=1}^{n} f\left(x_i, \tilde{y}_i \mid \mathcal{F}_{t_{i-1}}\right)\right] P\left(T_{n+1} > \mathcal{T} \mid \mathcal{F}_{t_n}\right) \\ &= \left[\prod_{i=1}^{n} g\left(x_i \mid \mathcal{F}_{t_{i-1}}\right)\right] P\left(X_{n+1} > \mathcal{T} - T_n \mid \mathcal{F}_{t_n}\right) \left[\prod_{i=1}^{n} q\left(\tilde{y}_i \mid t_i, \mathcal{F}_{t_{i-1}}\right)\right],\end{aligned}$$

3.1. Modeling the hazard function

Within our proposed framework, we characterize the time-varying probability of extreme event occurrence, $P\left(Y_t > u \mid \mathcal{F}_{t-1}\right)$, using a conditional hazard function. Among the competing hazard functions, we consider these associated with the Weibull, Burr, and generalized gamma distributions in both the continuous and discrete versions. Additionally, we incorporate a novel discrete distribution recently introduced by Gorgi (2020), which to the best of our knowledge has no counterpart in the continuous case[1]. This is important to note that under such (i.e., Weibull, Burr, and generalized gamma) discrete-time SPOT models, for any time t, where $t_{i-1} < t \leq t_i$, the conditional hazard function $h(t \mid \mathcal{F}_{t_{i-1}})$ directly represents the probability of an extreme loss at time t (i.e., for example, on day t) – given that no extreme loss event was observed since t_{i-1}:

$$h\left(t \mid \mathcal{F}_{t_{i-1}}\right) \;=\; \Pr\left(T_i = t \mid T_i \geq t; \mathcal{F}_{t_{i-1}}\right) = \frac{g\left(t - t_{i-1} \mid \mathcal{F}_{t_{i-1}}\right)}{S\left(t - t_{i-1} - 1 \mid \mathcal{F}_{t_{i-1}}\right)}, \tag{1}$$

[1]An interesting feature of discrete distributions is that they retain many of the reliability characteristics of their continuous counterparts. See for more details Nakagawa and Osaki (1975); Krishna and Pundir (2009); Chakraborty (2015).

where

$$g\left(t - t_{i-1} \mid \mathcal{F}_{t_{i-1}}\right) \;=\; S\left(t - t_{i-1} - 1 \mid \mathcal{F}_{t_{i-1}}\right) - S\left(t - t_{i-1} \mid \mathcal{F}_{t_{i-1}}\right)$$

corresponds to the discrete probability function associated with the survival function $S(\cdot \mid \mathcal{F}_{t_{i-1}})$ of a corresponding right-shifted continuous distribution. This shifting of statistical distribution to the right by 1 is necessary for the strictly-positive discrete domain of $g\left(t - t_{i-1} \mid \mathcal{F}_{t_{i-1}}\right)$.

In addition, to characterize the dynamics of the unobserved scale parameter, denoted as Ψ_i, we utilize the score-driven framework introduced by Creal et al. (2013) and Harvey (2013). This approach employs a first-order autoregressive dynamic conditional score model for the logarithm of Ψ_i

$$\ln(\Psi_i) = \omega_h + \beta_h \ln(\Psi_{i-1}) + \alpha_h s_{i-1} + \eta_h \tilde{y}_{i-1}, \tag{2}$$

where $\omega_h \in \mathbb{R}$, $\beta_h \in (0,1) \subset \mathbb{R}$, $\alpha_h \in \mathbb{R}_{>0}$, $\eta_h \in \mathbb{R}$ are unknown scalar parameters, and $s_i = \frac{\partial \ln g\left(x_i|\mathcal{F}_{t_{i-1}}\right)}{\partial \ln(\Psi_i)}$ corresponds to the score. Note that Ψ_i remains constant between times t_i and t_{i-1}.

3.1.1. The Weibull specification

We begin with the continuous case of the Weibull ($\mathcal{W}$) distribution which has been widely used in financial data modeling (see Engle and Russell, 1998; Herrera and Schipp, 2013; Bień-Barkowska, 2020; Bień-Barkowska, 2023). Under this assumption, the inter-exceedance time X_i follows a conditional $\mathcal{W}$ distribution with (continuous) density function given by

$$g^{[\mathcal{W}]}\left(x_i \mid \mathcal{F}_{t_{i-1}}\right) = \frac{\gamma}{x_i}\left(\frac{x_i}{\Psi_i}\right)^{\gamma} \exp\left[-\left(\frac{x_i}{\Psi_i}\right)^{\gamma}\right], \tag{3}$$

for $x_i \in \mathbb{Z}_{>0}$, where $\gamma \in \mathbb{R}_{>0}$ denotes the shape parameter, and $\Psi_i \in \mathbb{R}_{>0}$ denotes the time-varying scale parameter with score-driven dynamics proposed in Eq. (2). The conditional

survival function is given by

$$S^{[\mathcal{W}]}\left(x_i \mid \mathcal{F}_{t_{i-1}}\right) = \exp\left[-\left(\frac{x_i}{\Psi_i}\right)^{\gamma}\right], \tag{4}$$

while the corresponding conditional hazard function has the following form

$$h^{[\mathcal{W}]}\left(t \mid \mathcal{F}_{t_{i-1}}\right) = \gamma\ \Psi_i^{-\gamma}\ (t - t_{i-1})^{\gamma-1} \qquad \text{for} \qquad t_{i-1} < t \leq t_i.$$

Alternatively, plugging Eq.(4) in Eq.(1), we can obtain the discrete Weibull ($\mathcal{DW}$) hazard function

$$h^{[\mathcal{DW}]}\left(t \mid \mathcal{F}_{t_{i-1}}\right) = \frac{\exp\left(-\left(\frac{t-t_{i-1}-1}{\Psi_i}\right)^{\gamma}\right) - \exp\left(-\left(\frac{t-t_{i-1}}{\Psi_i}\right)^{\gamma}\right)}{\exp\left(-\left(\frac{t-t_{i-1}-1}{\Psi_i}\right)^{\gamma}\right)} \qquad \text{for} \qquad t_{i-1} < t \leq t_i.$$

Notice that we adopt a discretization approach based on the survival functions of the right-shifted continuous distributions. This is because the inter-exceedance times are always greater than or equal to 1.

The score s_i utilized for updating of the scale parameter in Eq. (2) is obtained by computing the first derivative of the i-th contribution to the log-likelihood function (with respect to $\ln(\Psi_i)$), while using the density function in Eq.(3)

$$s_i^{[\mathcal{W}]} = \gamma\left(\frac{x_i}{\Psi_i}\right)^{\gamma} - \gamma.$$

In the discrete case the corresponding score is given by

$$\begin{aligned} s_i^{[\mathcal{DW}]} &= \left(\exp\left(-\left(\frac{x_i-1}{\Psi_i}\right)^{\gamma}\right) - \exp\left(-\left(\frac{x_i}{\Psi_i}\right)^{\gamma}\right)\right)^{-1} \\ &\cdot\ \gamma\left(\exp\left(-\left(\frac{x_i-1}{\Psi_i}\right)^{\gamma}\right)\left(\frac{x_i-1}{\Psi_i}\right)^{\gamma} - \exp\left(-\left(\frac{x_i}{\Psi_i}\right)^{\gamma}\right)\left(\frac{x_i}{\Psi_i}\right)^{\gamma}\right). \end{aligned}$$

In Appendix A we present the general formula for the scores associated with the discrete distributions based on the parametric functional forms for the conditional density and survival function associated with the underlying continuous distributions. For further information on the discrete Weibull distribution we refer to Nakagawa and Osaki (1975), while for additional details regarding the score in the continuous case we refer to Fuentes et al. (2023).

3.1.2. The Burr specification

The Burr ($\mathcal{B}$) distribution offers a broad spectrum of skewness and kurtosis values, making it highly applicable for modeling financial durations in various contexts (see Grammig and Maurer, 2000; Hallin and Vecchia, 2020; Jacobi and Tzur, 2021). When the inter-exceedance times follow a conditional $\mathcal{B}$ distribution, the conditional density function of X_i is expressed as

$$g^{[\mathcal{B}]}\left(x_i \mid \mathcal{F}_{t_{i-1}}\right) = \frac{\kappa\, \zeta\, \Psi_i^{-\kappa}\, x_i^{\kappa-1}}{\left(1+\Psi_i^{-\kappa}\, x_i^{\kappa}\right)^{\zeta+1}}, \tag{5}$$

where $\zeta \in \mathbb{R}_{>0}$, $\kappa \in \mathbb{R}_{>0}$ denote shape parameters of the $\mathcal{B}$ distribution, and Ψ_i is the time varying scale parameter. The associated conditional survival and hazard functions are given by

$$S_{X_i}^{[\mathcal{B}]}\left(x_i \mid \mathcal{F}_{t_{i-1}}\right) = (1+\Psi_i^{-\kappa}\, x_i^{\kappa})^{-\zeta}. \tag{6}$$

and

$$h^{[\mathcal{B}]}\left(t \mid \mathcal{F}_{t_{i-1}}\right) = \frac{\kappa\, \zeta\, \Psi_i^{-\kappa}\, (t-t_{i-1})^{\kappa-1}}{1+\Psi_i^{-\kappa}\, (t-t_{i-1})^{\kappa}} \qquad \text{for} \qquad t_{i-1} < t \leq t_i,$$

respectively. The distributional properties and reliability characteristics of the discrete Burr ($\mathcal{DB}$) distribution were developed by Krishna and Pundir (2009). Observe that using the

results of Eq.(6) in Eq.(1) we can obtain the discrete hazard function

$$h^{[\mathcal{DB}]}\left(t \mid \mathcal{F}_{t_{i-1}}\right) = \frac{\left(1+\Psi_i^{-\kappa}\left(t-t_{i-1}-1\right)^{\kappa}\right)^{-\zeta}-\left(1+\Psi_i^{-\kappa}\left(t-t_{i-1}\right)^{\kappa}\right)^{-\zeta}}{\left(1+\Psi_i^{-\kappa}\left(t-t_{i-1}-1\right)^{\kappa}\right)^{-\zeta}}$$

$$\text{for} \qquad t_{i-1} < t \leq t_i.$$

For the $\mathcal{B}$ distribution, the score corresponding to the log scale parameter can be calculated as follows

$$s_i^{[\mathcal{B}]} = \frac{\kappa\zeta x_i^{\kappa}\Psi_i^{-\kappa}-\kappa}{x_i^{\kappa}\Psi_i^{-\kappa}+1},$$

while for the discrete case it is given by

$$s_i^{[\mathcal{DB}]} = \frac{\kappa\zeta\Psi_i^{-\kappa}\left(\left(1+(x_i-1)^{\kappa}\Psi_i^{-\kappa}\right)^{-\zeta-1}(x_i-1)^{\kappa}-\left(1+x_i^{\kappa}\Psi_i^{-\kappa}\right)^{-\zeta-1}x_i^{\kappa}\right)}{\left(1+\Psi_i^{-\kappa}\left(x_i-1\right)^{\kappa}\right)^{-\zeta}-\left(1+\Psi_i^{-\kappa}x_i^{\kappa}\right)^{-\zeta}}.$$

3.1.3. The generalized gamma specification

In the Generalized Gamma ($\mathcal{GG}$) SPOT model the conditional distribution function of the inter-exceedance time is a tree-parameter $\mathcal{GG}$ distribution

$$g^{[\mathcal{GG}]}\left(x_i \mid \mathcal{F}_{t_{i-1}}\right) = \frac{\gamma}{\Gamma(\nu)}\frac{1}{\Psi_i}\left(\frac{x_i}{\Psi_i}\right)^{\gamma\nu-1}\exp\left[-\left(\frac{x_i}{\Psi_i}\right)^{\gamma}\right], \tag{7}$$

where $\Gamma(\cdot)$ denotes the gamma function, $\nu \in \mathbb{R}_{>0}$ and $\gamma \in \mathbb{R}_{>0}$ are the two shape parameters, and Ψ_i is the score-driven scale parameter. If $\gamma = 1$, the $\mathcal{GG}$ distribution becomes the gamma distribution, and for $\nu = 1$, it nests the Weibull distribution with a shape parameter γ. The conditional survival function for the $\mathcal{GG}$ distribution is:

$$S^{[\mathcal{GG}]}\left(x_i \mid \mathcal{F}_{t_{i-1}}\right) = 1 - \boldsymbol{\gamma}\left(\nu, \left(\frac{x_i}{\Psi_i}\right)^{\gamma}\right), \tag{8}$$

where

$$\boldsymbol{\gamma}\left(\nu,\left(\frac{x_i}{\Psi_i}\right)^{\gamma}\right)=\int_0^{\left(\frac{x_i}{\Psi_i}\right)^{\gamma}}\frac{\exp(-s)s^{\nu-1}}{\Gamma(\nu)}\mathtt{d}s$$

is the regularized lower incomplete gamma function. The hazard function in this case is expressed as follows

$$h^{[\mathcal{GG}]}\left(t\mid\mathcal{F}_{t_{i-1}}\right) \;=\; \frac{\frac{\gamma}{\Gamma(\nu)\Psi_i}\left(\frac{t-t_{i-1}}{\Psi_i}\right)^{\gamma\nu-1}\exp\left[-\left(\frac{t-t_{i-1}}{\Psi_i}\right)^{\gamma}\right]}{1-\boldsymbol{\gamma}\left(\nu,\left(\frac{t-t_{i-1}}{\Psi_i}\right)^{\gamma}\right)} \qquad \text{for} \qquad t_{i-1}<t\leq t_i.$$

By applying the results from Eq.(8) to Eq.(1), the corresponding discrete generalized gamma ($\mathcal{DGG}$) hazard function proposed by Chakraborty (2015) can be expressed as

$$h^{[\mathcal{DGG}]}\left(t\mid\mathcal{F}_{t_{i-1}}\right) \;=\; \frac{-\boldsymbol{\gamma}\left(\nu,\left(\frac{t-t_{i-1}-1}{\Psi_i}\right)^{\gamma}\right)+\boldsymbol{\gamma}\left(\nu,\left(\frac{t-t_{i-1}}{\Psi_i}\right)^{\gamma}\right)}{1-\boldsymbol{\gamma}\left(\nu,\left(\frac{t-t_{i-1}-1}{\Psi_i}\right)^{\gamma}\right)} \qquad \text{for} \qquad t_{i-1}<t\leq t_i.$$

The corresponding score s_i, can be derived as

$$s_i^{[\mathcal{GG}]} \;=\; \gamma\left(\left(\frac{x_i}{\Psi_i}\right)^{\gamma}-\nu\right).$$

in the continuous-time case, while in the discrete-time it is given by

$$\begin{aligned} s_i^{[\mathcal{DGG}]} \;&=\; \left(-\boldsymbol{\gamma}\left(\nu,\left(\frac{x_i-1}{\Psi_i}\right)^{\gamma}\right)+\boldsymbol{\gamma}\left(\nu,\left(\frac{x_i}{\Psi_i}\right)^{\gamma}\right)\right)^{-1} \\ &\cdot\; \frac{\gamma}{\Gamma(\nu)}\left(\left(\frac{x_i-1}{\Psi_i}\right)^{\gamma\nu}\exp\left(-\left(\frac{x_i-1}{\Psi_i}\right)^{\gamma}\right)\right. \\ &-\; \left.\left(\frac{x_i}{\Psi_i}\right)^{\gamma\nu}\exp\left(-\left(\frac{x_i}{\Psi_i}\right)^{\gamma}\right)\right), \end{aligned}$$

3.1.4. The Beta-Negative Binomial specification

Gorgi (2020) has recently introduced the beta–negative binomial ($\mathcal{BNB}$) distribution in the context of integer-valued time series of variables with heavy-tailed distributions. Thus,

this distribution enables us to account for the dynamics observed in Section 2. In this scenario, we treat the inter-exceedance time as a positive count variable with a unit right-shifted $\mathcal{BNB}$ distribution. The conditional probability density function of X_i under this distribution is given by:

$$g^{[\mathcal{BNB}]}\left(x_i \mid \mathcal{F}_{t_{i-1}}\right) = \frac{\Gamma(x_i - 1 + r)}{\Gamma(x_i)\Gamma(r)} \frac{\boldsymbol{B}(\tau + r, (\tau - 1)\Psi_i/r + x_i - 1)}{\boldsymbol{B}(\tau, (\tau - 1)\Psi_i/r)},$$

where $\boldsymbol{B}(\cdot, \cdot)$ denotes the beta function, $\Psi_i \in \mathbb{R}_{>0}$ is the dynamic parameter representing the conditional expectation of $X_i - 1$, $r \in \mathbb{R}_{>0}$ and $\tau \in \mathbb{R}_{>1}$ are the dispersion parameter and the tail parameter, respectively. The $\mathcal{BNB}$ hazard function does not have a closed analytic form, however, for $t_{i-1} < t \leq t_i$ it can be derived as

$$\begin{aligned} h^{[\mathcal{BNB}]}\left(t \mid \mathcal{F}_{t_{i-1}}\right) &= g^{[\mathcal{BNB}]}\left(t - t_{i-1} \mid \mathcal{F}_{t_{i-1}}\right) \\ &\cdot \left(1 - \sum_{k=1}^{k=t-t_{i-1}-1} g^{[\mathcal{BNB}]}\left(k \mid \mathcal{F}_{t_{i-1}}\right)\right)^{-1} \qquad \text{for} \qquad t - t_{i-1} \geq 2, \end{aligned}$$

and $h^{[\mathcal{BNB}]}\left(t \mid \mathcal{F}_{t_{i-1}}\right) = g^{[\mathcal{BNB}]}\left(t - t_{i-1} \mid \mathcal{F}_{t_{i-1}}\right)$ for $t - t_{i-1} = 1$. Finally, the score s_i under the assumption of a $\mathcal{BNB}$ distribution is computed as

$$\begin{aligned} s_i^{[\mathcal{BNB}]} &= \gamma_s \Psi_i [\psi(\gamma_s \Psi_i + x_i - 1) + \psi(\gamma_s \Psi_i + \tau) \\ &- \psi(\gamma_s \Psi_i + x_i - 1 + \tau + r) - \psi(\gamma_s \Psi_i)], \end{aligned} \tag{9}$$

where $\gamma_s = (\tau - 1)/r$ and $\psi(\cdot)$ denotes the digamma function.

3.2. Modeling the marks

Following the Pickands–Balkema–De Haan theorem (Pickands III, 1975; Balkema and de Haan, 1974), the associated density function of $\tilde{Y}_i$ can be well described by a conditional

generalized Pareto probability density function

$$q(\tilde{y}_i \mid x_i, \mathcal{F}_{t_{i-1}}) = \frac{1}{\sigma_i}\left(1 + \xi\frac{\tilde{y}_i}{\sigma_i}\right)_+^{-1/\xi-1}, \tag{10}$$

where for ease of the exposition the shape ξ and scale σ_i parameters are assumed strictly positive. In this framework, the scale parameter of the generalized Pareto distribution will also be updated following a score-driven specification that includes a Box-Cox transformation of the contemporaneous inter-exceedance time x_i, as an additional explanatory variable

$$\ln \sigma_i = \omega_s + \beta_s \ln \sigma_{i-1} + \alpha_s s^*_{i-1} + \eta_{1,s}(x_i^{\eta_{2,s}} - 1)/\eta_{2,s}, \tag{11}$$

where $\omega_s \in \mathbb{R}$, $\beta_s \in (0,1) \subset \mathbb{R}$, $\alpha_s \in \mathbb{R}_{>0}$, $\eta_{1,s} \in \mathbb{R}$, $\eta_{2,s} \in \mathbb{R}\backslash\{0\}$, are the parameters to be estimated, and $s^*_i = \frac{\partial \ln q\left(\tilde{y}_i|\mathcal{F}_{t_{i-1}}\right)}{\partial \ln(\sigma_i)}$ corresponds to the conditional score with respect to the logarithm of σ_i

$$s^*_i = \frac{\tilde{y}_i - \sigma_i}{\xi\tilde{y}_i + \sigma_i}. \tag{12}$$

3.3. Risk Measures

In the past few years, financial risk managers have incorporated alternative metrics like ES, alongside conventional measures such as VaR, to better capture extreme risk (Bayer and Dimitriadis, 2022; Deng and Qiu, 2021; Patton et al., 2019).

To calculate these risk measures within the SPOT framework, we set the probability that the daily return Y_t exceeds the VaR at the α confidence level at time t – denoted by υ_t^α,

$$1 - \alpha = P\left(Y_t > \upsilon_t^\alpha \mid \mathcal{F}_{t-1}\right) = P\left(Y_t > u \mid \mathcal{F}_{t-1}\right) P\left(Y_t - u > \upsilon_t^\alpha - u \mid Y_t > u, \mathcal{F}_{t_{i-1}}, x(t)\right),$$

where $x(t) = t - t_{i-1}$, for $t_{i-1} < t \leq t_i$. The solution is obtained as follows:

$$\upsilon_t^{\alpha} = u + \frac{\sigma_t}{\xi}\left\{\left(\frac{h(t \mid \mathcal{F}_{t_{i-1}})}{1-\alpha}\right)^{\xi} - 1\right\},$$

where $\ln \sigma_t = \omega_s + \beta_s \ln \sigma_{i-1} + \alpha_s s_{i-1}^* + \eta_{1,s}((t - t_{i-1})^{\eta_{2,s}} - 1)/\eta_{2,s}$. Similarly, to calculate the ES denoted by e_t^{α}, we need to determine the expectation of losses beyond a certain threshold given by υ_t^{α}:

$$e_t^{\alpha} = \frac{1}{1-\alpha}\int_{\alpha}^{1} \upsilon_t^s ds = \frac{\upsilon_t^{\alpha}}{1-\xi} + \frac{\sigma_t - u\xi}{1-\xi}.$$

For a more detailed derivation of both risk measures, see Hautsch and Herrera (2020).

4. Competing hazard functions: A Monte Carlo study

In this section, we use a Monte Carlo study to characterize the primary differences between discrete and continuous SPOT specifications. It is crucial to note that, for any time t, the conditional hazard function derived from SPOT models is used either (1) to approximate (for continuous-time SPOT models) or (2) to directly represent (for discrete-time SPOT models) the *failure probability*, which is the probability of an extreme loss at that time unit (for instance, on day t).

4.1. Relationship between exceedance time and score

In Figure we examine the relationship between the i-th observation of exceedance time x_i and the resulting score s_i for all specifications[2]. No considerable discrepancies are observed between the scores related to the continuous distributions and their discrete counterparts. However, an important observation is that x_i and s_i exhibit a nonlinear relationship. The degree of curvature of these relationships depends on the tail parameter value.

Specifically, the sensitivity of s_i to x_i can be reduced for a large x_i, and the degree of attenuation increases as the tail distribution becomes heavier. For the $\mathcal{W}$, $\mathcal{DW}$, $\mathcal{GG}$, and

[2]We display this relation with calibrated scores, such that $E\left(X_i \mid \mathcal{F}_{t_{i-1}}\right) \approx 10$

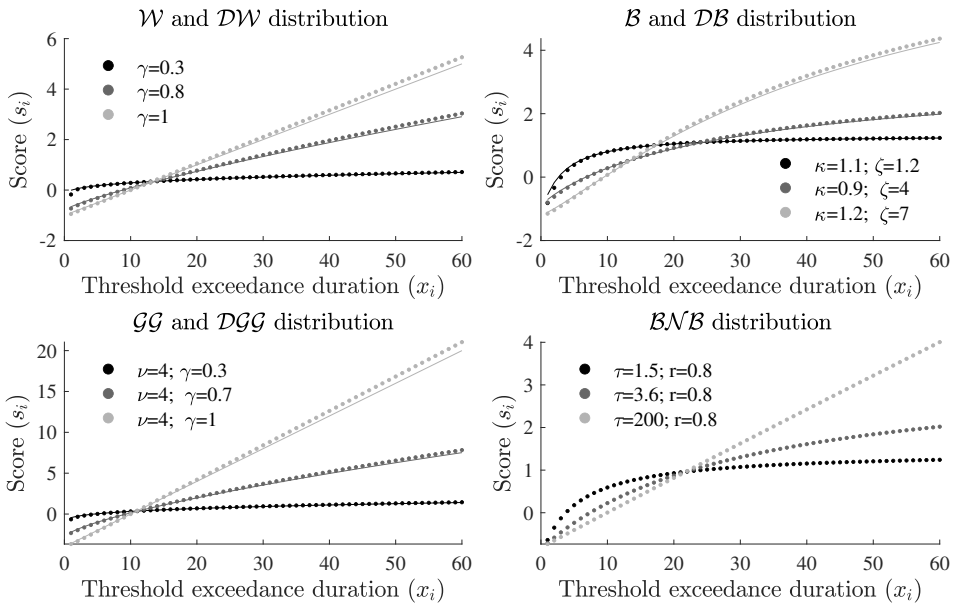

Figure 3: Sensitivity of the score (s_i) to the inter-exceedance time x_i in different SPOT models: $\mathcal{W}$, $\mathcal{DW}$, $\mathcal{GG}$, $\mathcal{B}$, $\mathcal{DB}$ and $\mathcal{BNB}$ distribution. The scores are calibrated such that $E\left(X_i \mid \mathcal{F}_{t_{i-1}}\right) \approx 10$.

$\mathcal{DGG}$ distributions, a lower value of the tail parameter γ results in a higher attenuation degree of an impact produced by large x_i. The tail index of the $\mathcal{B}$ and $\mathcal{DB}$ distributions equals $\kappa\zeta$, with a smaller product of these shape parameters leading to a lower sensitivity of s_i to a very large x_i. This occurs because 'small' tail indexes characterize heavy-tailed distributions with numerous extreme observations in the tail. The SPOT model accurately alleviates the potentially significant impact of a very large x_i on the scale parameter Ψ_i, leaving this effect to be described by the tail parameters. A similar observation was made for the $\mathcal{BNB}$ distribution in Gorgi (2020). In this case, the smaller the tail parameter τ, the more robust s_i is towards outliers in the inter-exceedance times.

4.2. Model misspecification

We now explore the implications of potentially misusing continuous distributions when modeling discrete exceedance times, particularly for forecasting the failure probability, that is, the conditional probability of an extreme loss at time t. We focus on two data generating processes (DGPs): (1) the $\mathcal{DB}$ and (2) the $\mathcal{DGG}$ score-driven specification. The parameters

are chosen to closely resemble the parameter estimates obtained in the empirical study. For the conditional $\mathcal{DB}$ distribution, we assume $\omega = 0.3$, $\beta = 0.9$, $\alpha = 0.4$, $\kappa = 0.95$, and $\zeta = 3$. For the conditional $\mathcal{DGG}$ distribution, we assume $\omega = 0$, $\beta = 0.9$, $\alpha = 0.4$, $\nu = 2.5$, and $\gamma = 0.5$. The simulation study uses 1000 replications and sample sizes of 250, 500, and 2000 sampled values. For DGP I, the kernel density plots for the estimates of shape parameters $\hat{\kappa}$ and $\hat{\zeta}$ are shown in Figure 4 (upper panel). As the sample size increases, the parameter estimates for the $\mathcal{DB}$ distribution converge toward the true parameter values. Moreover, their distributions are symmetrical (except for the slightly skewed $\hat{\zeta}$ distribution in the small samples). However, after estimating the score-driven continuous $\mathcal{B}$ distribution using this artificial data, the obtained estimates are strongly biased, and this bias does not vanish as the sample size increases.

The κ estimates for $\mathcal{B}$ distribution are strongly upward biased and converge to a value of approximately 1.5, whereas the ζ estimates are downward-biased with an average value of approximately 1.1. These results have a significant impact on the feasible shapes of the conditional hazard functions. According to DGP I, where $\kappa < 1$, the 'true' conditional hazard is always a decreasing function, but according to the mean parameter estimates obtained using the $\mathcal{B}$ distribution, the mean value of the κ estimate is larger than 1 (i.e., $\bar{\hat{\kappa}} > 1$); thus, the conditional hazard function can be hump-shaped. However, if we focus only on the integer-valued arguments of the continuous $\mathcal{B}$ hazard function, its resulting shape depends not only on the shape parameters but also on the value of the dynamic scale parameter Ψ_i, and hence, on the conditional expectation of the inter-exceedance time[3].

Some plausible shapes of the hazard are illustrated in Figure 4. For example, in extremely volatile periods, when the expected inter-exceedance times are very small, the $\mathcal{B}$ hazard tends to overestimate the $\mathcal{DB}$ failure probability immediately after the arrival of an extreme event.

[3]Observe that the conditional expectation of X_i based on the $\mathcal{B}$ specification is given by $E^{[\mathcal{B}]}\left(X_i \mid \mathcal{F}_{t_{i-1}}\right) = \Psi_i \zeta \boldsymbol{B}(\zeta - 1/\kappa, 1 + 1/\kappa)$ where $\boldsymbol{B}(\cdot,\cdot)$ denotes the beta function, while under the $\mathcal{GG}$ specification it can be calculated as $E^{[\mathcal{GG}]}\left(X_i \mid \mathcal{F}_{t_{i-1}}\right) = \Psi_i \frac{\Gamma(\nu+1/\gamma)}{\Gamma(\nu)}$. For discrete distributions, the conditional expectation of X_i can easily be computed as a finite approximation of $E\left(X_i \mid \mathcal{F}_{t_{i-1}}\right) = \sum_{k=1}^{\infty} g_{X_i}\left(X_i = k \mid \mathcal{F}_{t_{i-1}}\right) \cdot k$.

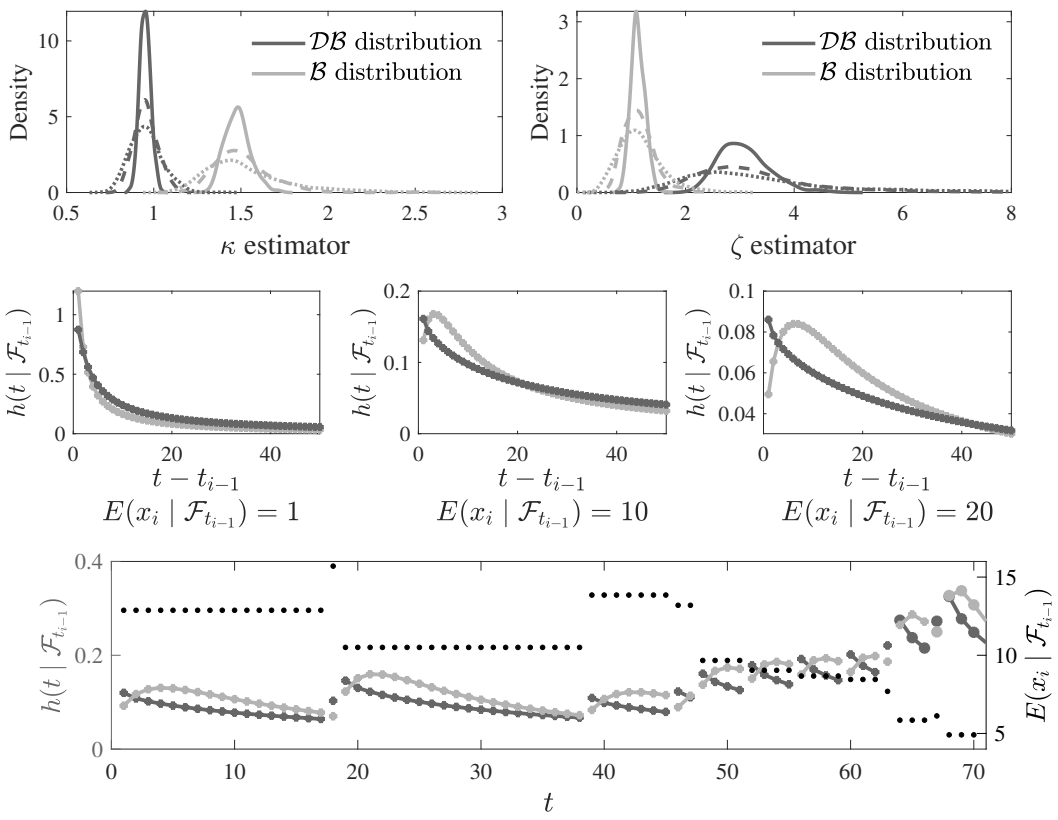

Figure 4: Upper Panel: Kernel plots illustrating shape parameter estimates, specifically $\hat{\kappa}$ and $\hat{\zeta}$, for the conditional $\mathcal{DB}$ (DGP I) and the conditional $\mathcal{B}$ distributions at sample sizes n=250, 500, and 2000. Middle Panel: Visual representation of potential shapes for the conditional hazard function, derived for the average parameter estimates, $\bar{\hat{\kappa}}$ and $\bar{\hat{\zeta}}$, for a sample size of n=2000. Lower Panel: Time-varying conditional probability of an extreme loss event according to the DGP I (in dark gray), and the corresponding conditional probability according to the continuous $\mathcal{B}$ distribution (in light gray) for the mean values of the parameter estimates.

However, more often, after an extreme event occurs, the $\mathcal{B}$ hazard tends to systematically underestimate the probability of the next extreme event initially and overestimate this probability a few time units (days) later.

Similar findings emerge from the simulation based on DGP II. In Figure 5, we observe that after estimating the $\mathcal{DGG}$ hazard specification, the kernel density plots for $\hat{\nu}$ and $\hat{\gamma}$ are symmetrical and resemble bell-shaped curves, and the estimates converge to the true parameter values as the sample size increases. However, when fitting the conditional continuous $\mathcal{G}$ distribution to the integer-valued inter-exceedance times, we obtain parameter estimates that are severely upward biased. Consequently, the resulting shapes of the $\mathcal{GG}$ conditional

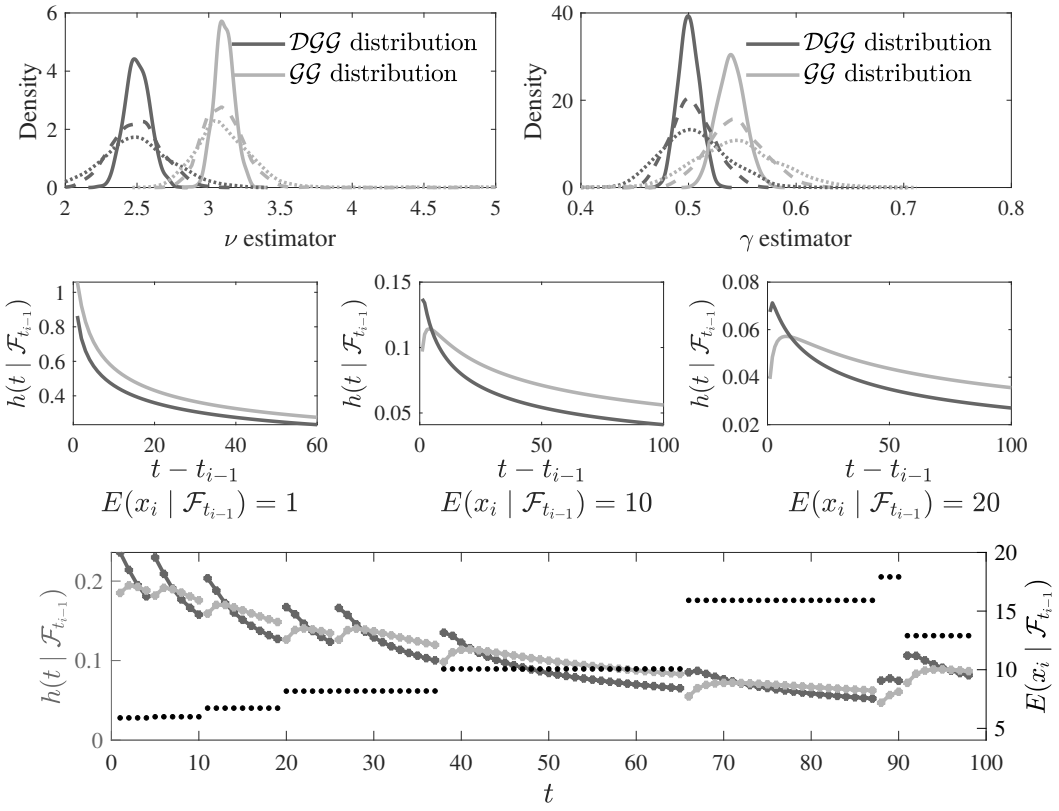

Figure 5: Upper Panel: Kernel plots illustrating shape parameter estimates, specifically $\hat{\nu}$ and $\hat{\gamma}$, of the conditional $\mathcal{DGG}$ (DGP I) and the conditional $\mathcal{GG}$ distributions at sample sizes n=250, 500, and 2000. Middle Panel: Visual representation of potential shapes for the conditional hazard function, derived for the average parameter estimates, $\bar{\hat{\nu}}$ and $\bar{\hat{\gamma}}$, for a sample size of n=2000. Lower Panel: Time-varying conditional probability of an extreme loss event according to the DGP II (in dark gray), and the corresponding conditional probability according to the continuous $\mathcal{GG}$ distribution (in light gray) for the mean values of the parameter estimates.

hazard function differ significantly from those of true $\mathcal{DGG}$. In summary, we observe that the conditional hazard of a continuous distribution tends to either overestimate or underestimate the conditional probability of an extreme loss event.

This result may arise because continuous hazard functions tend to assign a non-zero probability to inter-exceedance times with values less than one, which, in empirical applications, have a zero probability of occurrence. However, even though the results of this experiment seem to favor discrete specifications, it is not possible to directly extrapolate their implications to empirical applications. To better understand these aspects, we will analyze them in the following section.

5. Empirical Results

In this study, we analyze the time series of daily negative log returns, expressed as $y_t = -100 \cdot \ln(p_t/p_{t-1})$, where p_t represents the stock market index price. The study spans from January 2, 1981, to December 28, 2022, and focuses on four major U.S. equity indices: (i) the Dow Jones Industrial Average, which tracks the performance of 30 leading blue-chip companies; (ii) the NASDAQ Composite, which measures the performance of more than 3,000 stocks listed on the NASDAQ stock exchange; (iii) the S&P 500; and (iv) Wilshire 5000, both representing the broad U.S. equity market.

5.1. Threshold selection and estimation procedure

Several approaches have been proposed to address the threshold selection problem in EVT (Scarrott and MacDonald, 2012; Danielsson et al., 2016; Schneider et al., 2021). Each approach has its advantages and disadvantages, and the choice depends on the specific application and the available data.

In this study, we determine the optimal threshold using a data-driven method, following the approach outlined in Hoga (2019). We use a scoring rule to identify observations that are considered extreme events. In particular, we apply a quantile-weighted continuous-ranked probability score (S^u_{QCRPS}). A detailed description of this method can be found in Appendix B. In short, we select an optimal threshold from an extensive range of values, aiming to minimize the scoring function S^u_{QCRPS}, and hence, to achieve the best in-sample goodness-of-fit of the SPOT models at the tail region of distribution.[4]

We also employ a multi-step approach for our analysis. First, we estimated all SPOT specifications using in-sample data from January 2, 1981, to December 30, 2010. We then selected the optimal threshold using the score rule S^u_{QCRPS} and generated VaR and ES forecasts at different α levels of risk for the subsequent two years (from January 2, 2011, to December

[4]Specifically, we use a dense grid of possible u_j values defined as in-sample j-percentiles of Y_t for $j \in \{88, 88.1, 88.2, ..., 96.5\}$.

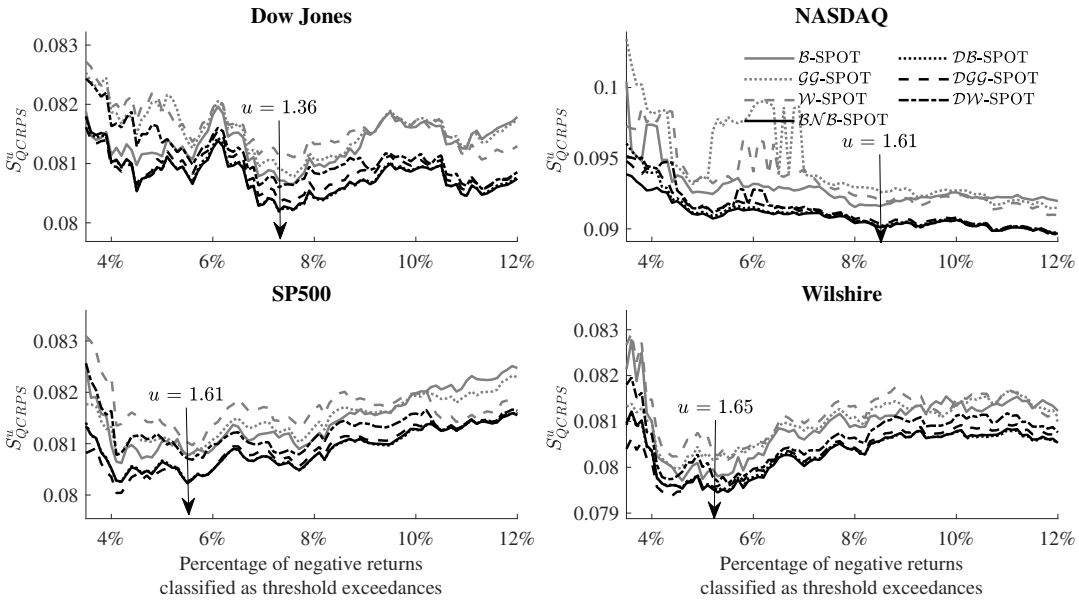

Figure 6: In-sample quantile-weighted continuous-ranked probability score (S^u_{QCRPS}) estimates for SPOT models across a range of thresholds for each stock market index.

31, 2012). We then re-estimated the SPOT models after adding the last 2-years to the in-sample period (so that it spans from January 2, 1981, to December 31, 2012) and forecasted the risk measures for the following two years (from January 2, 2013, to December 31, 2014). This procedure was repeated four times. Finally, we used the 12-year out-of-sample period (from January 2, 2011, to December 28, 2022) of daily VaR and ES forecasts to evaluate the accuracy of all SPOT specifications.

Figure 6 illustrates the estimates of S^u_{QCRPS} for the in-sample period across the range of examined threshold levels. Several important observations can be made. First, discrete-time SPOT models exhibit superior goodness-of-fit compared with their continuous-time counterparts. Regardless of the threshold level, the discrete-time SPOT models (black lines) consistently yield lower values of the scoring function than the continuous-time SPOT models (grey lines). Although the differences in performance among various discrete-time SPOT specifications are negligible, the $\mathcal{DW}$-SPOT model appears slightly inferior in most cases. Second, the relationship between the proportion of threshold exceedances and S^u_{QCRPS} estimates re-

Est.	$\mathcal{W}$	$\mathcal{B}$	$\mathcal{GG}$	$\mathcal{DW}$	$\mathcal{DB}$	$\mathcal{DGG}$	$\mathcal{BNB}$
			Probability of threshold exceedance				
ω_h	0.336 (0.075)	0.286 (0.082)	0.000 (-)	0.332 (0.079)	0.476 (0.134)	0.001(0.183)	0.357(0.089)
β_h	0.882 (0.027)	0.878 (0.033)	0.878 (0.027)	0.877 (0.031)	0.872 (0.032)	0.870(0.032)	0.870(0.032)
α_h	0.210 (0.034)	0.258 (0.041)	0.248 (0.039)	0.294 (0.051)	0.354 (0.058)	0.345(0.060)	0.395(0.067)
η_h	-0.064 (0.016)	-0.038 (0.017)	-0.059 (0.017)	-0.067 (0.019)	-0.055 (0.020)	-0.058(0.022)	-0.059(0.027)
γ	0.902 (0.024)		0.482 (0.012)	0.783 (0.023)		0.451 (0.092)	
κ		1.411(0.082)			0.947 (0.055)		
ζ		1.284(0.218)			3.644 (1.149)		
r							0.860 (0.079)
τ							3.522 (0.858)
ν			3.458 (0.185)			2.872 (1.102)	
			Threshold exceedance magnitude				
ω_s				0.347 (0.074)			
β_s				0.787 (0.046)			
α_s				0.136 (0.033)			
$\eta_{1,s}$				-0.780 (0.290)			
$\eta_{2,s}$				-1.439 (0.483)			
ξ				0.077 (0.033)			
			Log Likelihood (Log$\mathcal{L}$) and Information Criteria				
Log$\mathcal{L}$	-3058.330	-3018.453	-3022.461	-2981.932	-2974.3297	-2974.6342	-2973.2351
AIC	8.363	8.257	8.268	8.155	8.137	8.138	8.134
BIC	8.432	8.332	8.343	8.224	8.212	8.213	8.209

Table 2: Maximum likelihood estimates for the SPOT models applied to the Dow Jones from January 2, 1981, to December 30, 2020. Standard errors are provided in parentheses. *Log*$\mathcal{L}$ represents the maximum log-likelihood value. *AIC* and *BIC* refer to the Akaike Information Criterion and the Schwarz Information Criterion, respectively.

sembles either a U-or L-shaped pattern, clearly indicating that excessively high thresholds expose too few exceedances. In contrast, thresholds that are too low also negatively impact in-sample goodness-of-fit. Based on the visual inspection of Figure 6, we determine the best threshold values as $u = 1.36$ for Dow Jones (7.3% of the data), $u = 1.61$ for NASDAQ (8.4%), $u = 1.61$ (5.5%), and $u = 1.65$ (5.3%) for the S&P 500 and the Wilshire, respectively.

5.2. In-sample Analysis

Table 2 presents the estimation results for the Dow Jones index, providing more insight into the differences between the continuous-time and discrete-time SPOT models (the estimation results for the remaining stock indices can be found in Appendix D). For all specifications, the parameter estimates of α_h and β_h, harnessing the dynamics of the scale parameter Ψ_i, are highly significant and indicate high persistence ($\hat{\beta}_h > 0.87$). The negative impact of the last observed mark on the inter-exceedance time is confirmed ($\hat{\eta}_h < 0$), suggesting that the size of a previous extreme loss event increases the (daily) probability of an

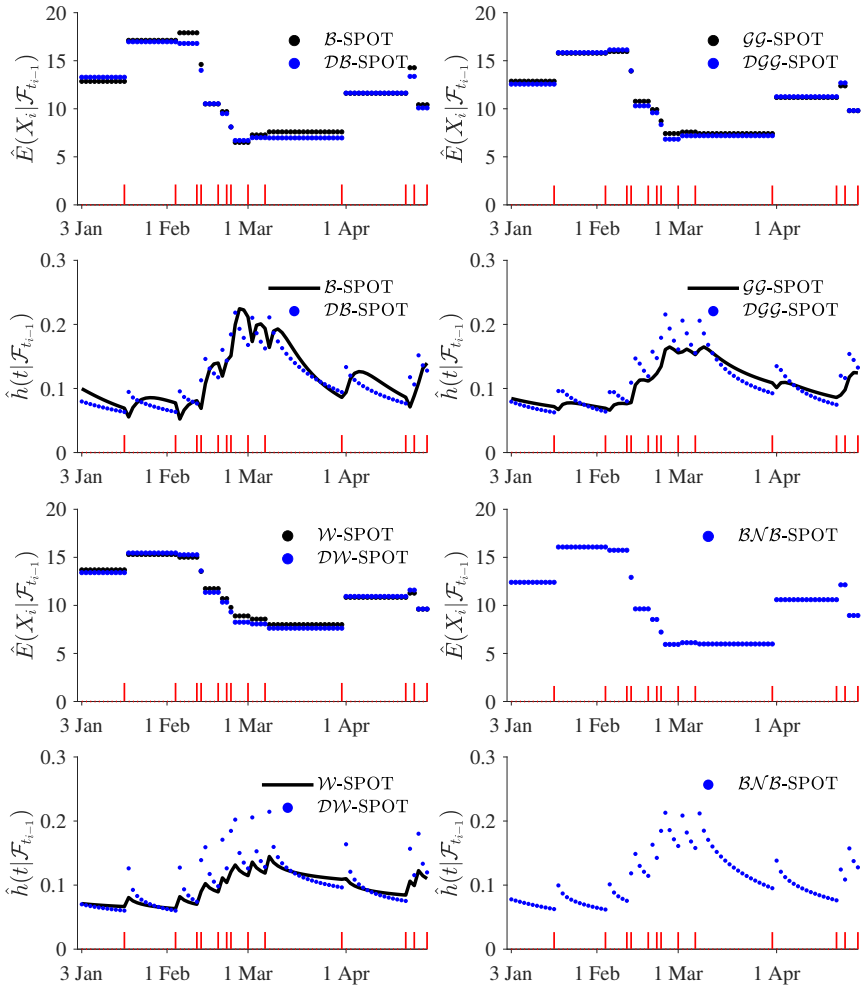

Figure 7: Conditional expectations of the inter-exceedance times $\hat{E}\left(X_i \mid \mathcal{F}_{t_{i-1}}\right)$ and the daily probability of an extreme loss $\hat{h}\left(t \mid \mathcal{F}_{t_{i-1}}\right)$ for Dow Jones over the period from January 3, 2022 to May 27, 2022 according to the SPOT models (red lines indicate times of extreme events).

upcoming extreme loss event. Interestingly, the estimates of the shape parameters for the $\mathcal{B}$ distribution, i.e., $\hat{\kappa} > 1$, and the $\mathcal{GG}$ distribution, i.e., $\hat{\nu}\hat{\gamma} > 1$, can display a conditional hazard function with an inverted U shape. However, this outcome is not confirmed by the $\mathcal{DB}$-SPOT model, which yields only a decreasing conditional hazard for $\hat{\kappa} < 1$. Although the $\mathcal{DGG}$-SPOT model can also produce a non-monotonic conditional hazard, our simulations show that it holds only for very large values of the scale parameter Ψ_i.

Figure 7 offers a more comprehensive perspective on the dynamics underlying the daily probability of an extreme event. It depicts the evolution paths for the conditional expecta-

tions of the inter-exceedance time, along with their associated conditional hazard functions, over a four-month span from January 3, 2022, to April 29, 2022, for the Dow Jones index. Notably, 13 threshold exceedances were observed during this volatile period.

The dynamics of the conditional expectation follows a step function, and its values are updated immediately after an extreme event occurs. The magnitude of each upward/downward movement depends on the entire history of exceedance times and the size of the previously observed threshold exceedance, which is captured by the score-driven equation of $\ln(\Psi_i)$. There are no significant discrepancies between the sample paths of conditional expectations for all the SPOT specifications under examination. For instance, immediately after the first extreme loss on January 18, 2022, the expected inter-exceedance time for the next extreme event was forecasted to be between 15 and 18 days using all SPOT specifications. Furthermore, the differences between the conditional expectation levels corresponding to the $\mathcal{B}$-SPOT ($\mathcal{GG}$-SPOT, $\mathcal{W}$-SPOT) model and its discrete-time counterpart are negligible. However, this observation does not hold for the time paths of conditional hazards, because their shapes vary significantly among specifications. The form of the hazard function relies on the estimated shape parameters.

The results indicate that discrete-time SPOT models forecast a higher failure probability (i.e., the probability of extreme loss) for a day or a few days directly following an extreme loss event. By contrast, continuous-time SPOT models, owing to their hump-shaped hazard ($\mathcal{B}$-SPOT, $\mathcal{GG}$-SPOT) or a hazard with a smaller slope (for the $\mathcal{W}$-SPOT model), underestimate this probability and react to an extreme event with a noticeable delay in time. This empirical example strongly supports our simulation results, where continuous-time models lead to distorted shapes of conditional hazard functions when fitted to integer value data.

Regarding the dynamics of the threshold exceedance magnitudes, we obtained highly significant estimates of α_s and β_s, harnessing the score-driven evolution of the scale parameter associated with the generalized Pareto distribution (see Eq. 11). There is also a significant negative impact of the inter-exceedance times on the threshold exceedance sizes ($\hat{\eta}_{1,s} < 0$,

$\hat{\eta}_{2,s} < 0$). This effect can be observed in Figure 8 (upper panel), where the trajectory of $\hat{\sigma}_t$ is depicted. Accordingly, the anticipated magnitude of the threshold exceedance decreases with increasing time distance to a previously recorded extreme loss event. In the lower panel of Figure 8 we present how the time variation in the daily probability of extreme loss (exhibited by the estimated conditional hazard function) and the time evolution of $\hat{\sigma}_t$ impacts the daily VaR forecast according to the $\mathcal{B}$-SPOT or $\mathcal{DB}$-SPOT models. VaR dynamics are characterized by abrupt jumps immediately following extreme loss events; however, the magnitude of these adjustments depends on the SPOT specification. In accordance with the discrepancies between the shapes of continuous-time and discrete-time hazards, in the 4-month period under study, the discrete-duration SPOT models tend to forecast higher VaR on days directly after extreme losses in comparison to the continuous-time SPOT models.

5.3. Sensitivity of risk measure forecasting to threshold selection

We examine the sensitivity of the out-of-sample accuracy of the SPOT models to the choice of the threshold u. To this end, the risk measures generated by a broad range of competing models are compared using the concept of strictly consistent scoring functions introduced in Fissler and Ziegel (2016). In particular, evaluating the average value of these scoring functions at different risk levels of VaR and ES allows us to rank SPOT specifications based on their actual scores, as suggested by Patton et al. (2019).

In our study, we utilize the $FZ0L$ score function proposed by Patton et al. (2019) (for details, see Appendix C). Figure 9 shows the average of the $FZ0L$ scores (denoted as $\overline{FZGL}^{\alpha} = 1/(\check{\mathcal{T}} - \mathcal{T}) \sum_{t=\mathcal{T}+1}^{\check{\mathcal{T}}} FZGL_t^{\alpha}$, with risk level $\alpha \in 0.95, 0.975, 0.99$, where $\check{\mathcal{T}} - \mathcal{T}$ represents the number of days in the 12-year out-of-sample period) across different thresholds. The findings indicate that the lowest average for this score function is associated with the discrete-time SPOT specifications, and this result is not threshold-dependent for either risk measure. Furthermore, the choice of discrete distribution is not as crucial as ensuring that the $\mathcal{DB}$, $\mathcal{DGG}$, $\mathcal{DW}$, and $\mathcal{BNB}$ distributions are appropriately designed to capture the discreteness of inter-exceedance times. Interestingly, within the examined range of discrete-

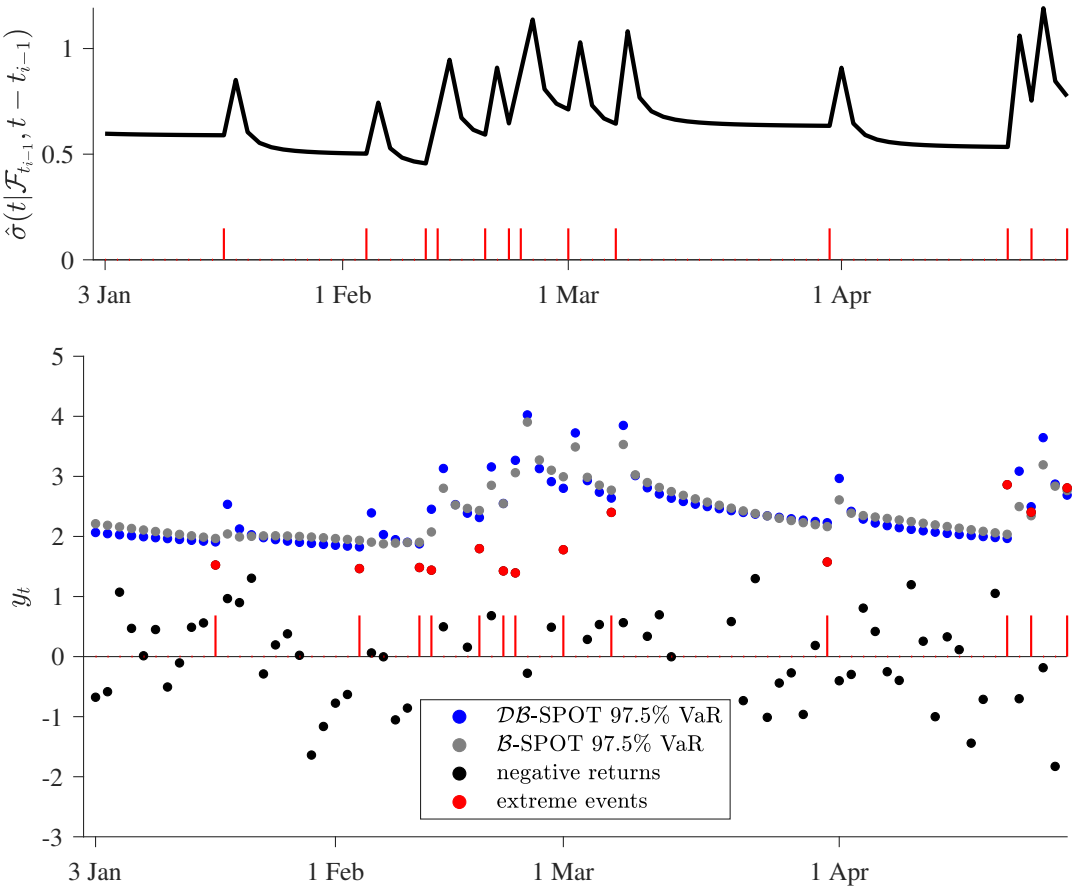

Figure 8: Estimates of the time-varying scale parameter σ_t for the magnitudes of threshold exceedances in Dow Jones over the period from January 3, 2022, to May 27, 2022 (upper panel) and the corresponding 97.5% VaR associated with the $\mathcal{B}$-SPOT and $\mathcal{DB}$-SPOT model (red lines indicate times of extreme events).

time SPOT models, the $\mathcal{DW}$ distribution (which has only one shape parameter) tends to outperform more complex distributions such as $\mathcal{DB}$, $\mathcal{DGG}$, or $\mathcal{BNB}$. It is difficult to rank continuous-time SPOT models based on their performance because of the significant variability that arises when different threshold values are selected.

5.4. Comparative Analysis

Figure 10 exhibits the performance of the alternative SPOT specifications based on the out-of-sample average of the $FZ0L$ scores, denoted by $\overline{FZ0L}^{\alpha}$, with $\alpha \in \{0.95, 0.975, 0.99\}$. We expand the range of competing models by including restricted specifications of the SPOT (RSPOT). These versions omit the explanatory variables $\tilde{Y}_{i-1}$ and X_i in Eq. (2) and Eq.

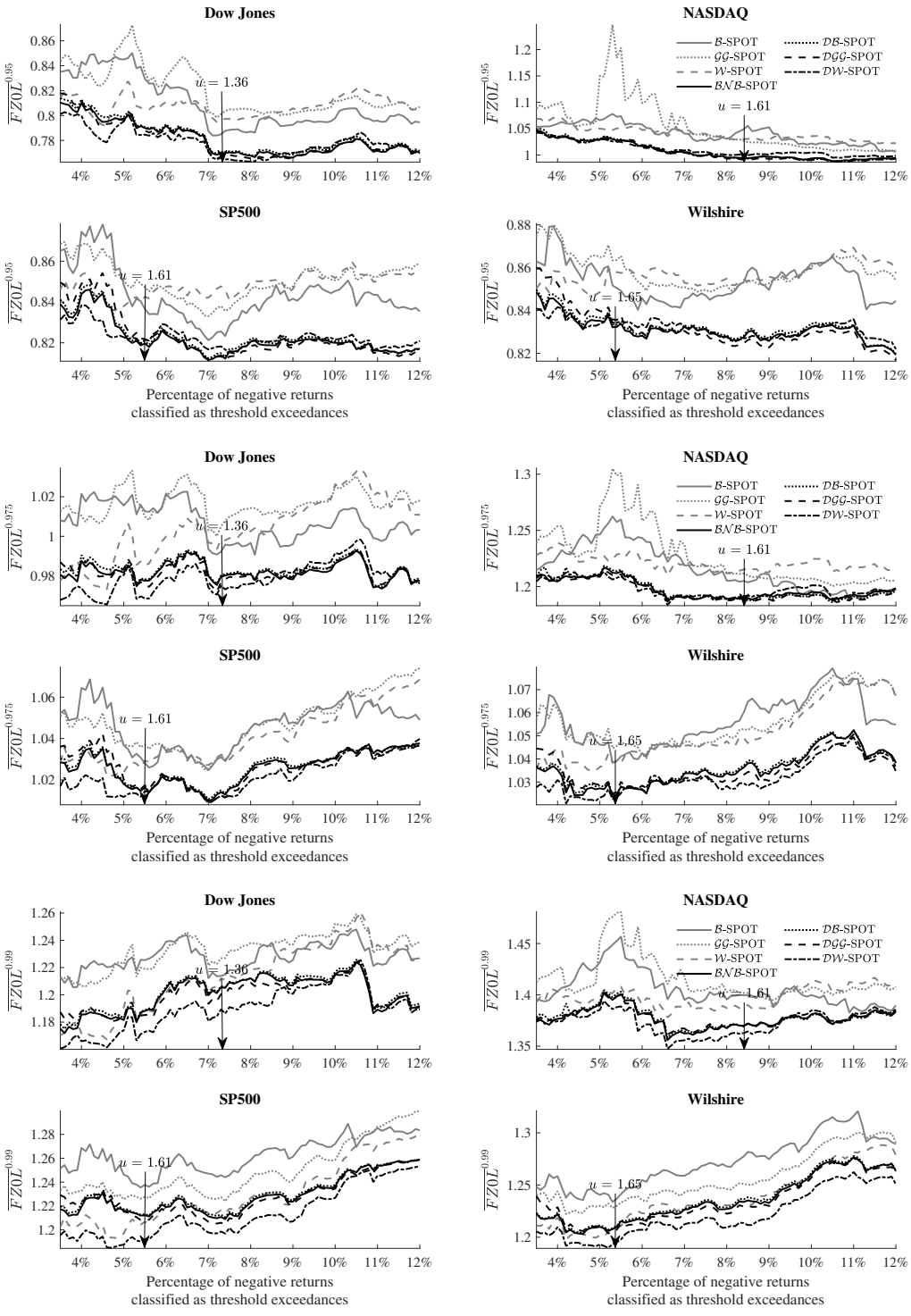

Figure 9: Average scores utilizing the FZ0L loss for various SPOT model specifications across a range of u values. This 12-year out-of-sample analysis spans from January 2, 2011, to December 28, 2022. The risk levels are $\alpha = 0.95$ (top panel), $\alpha = 0.975$ (central panel), and $\alpha = 0.99$ (lower panel).

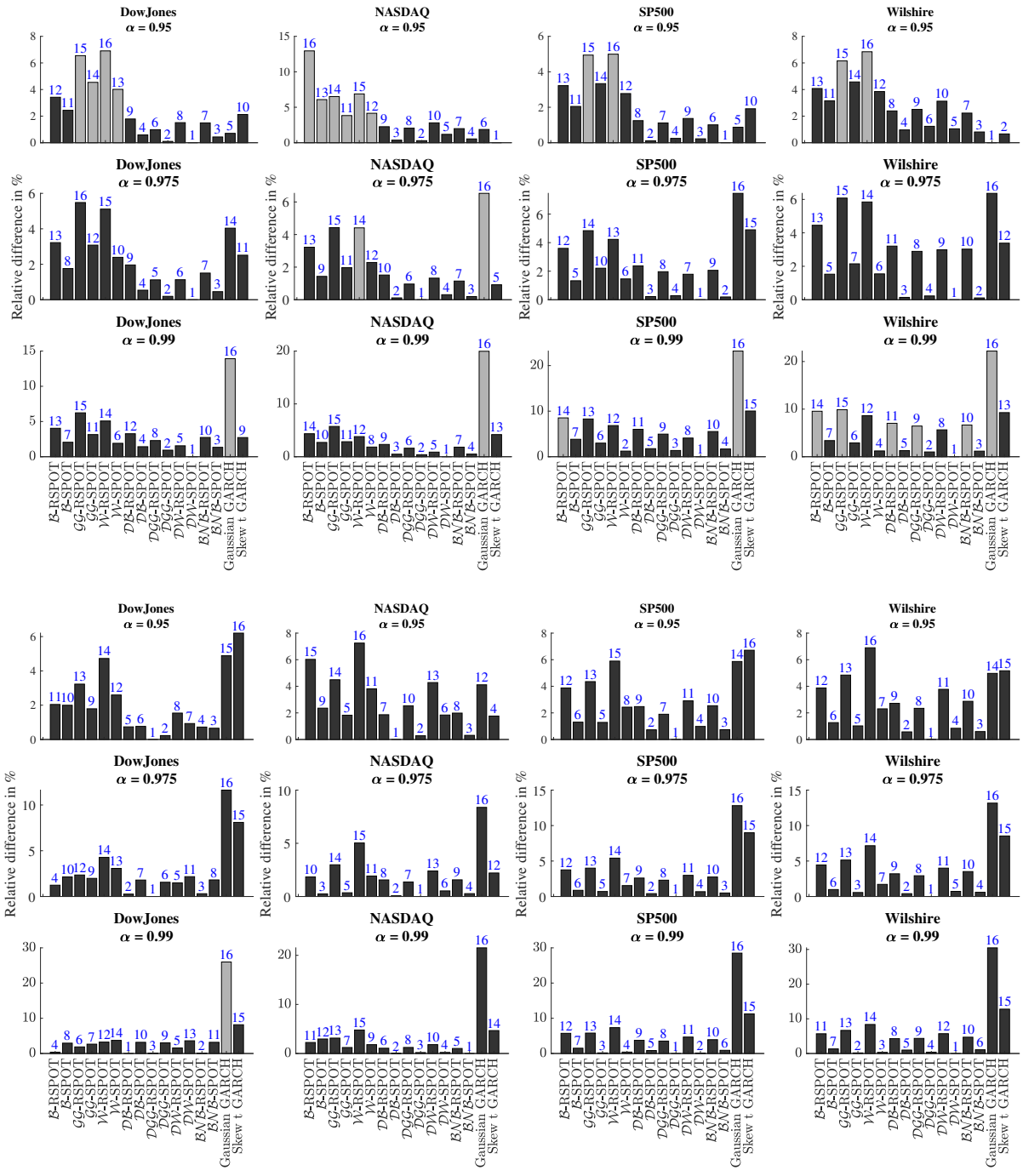

Figure 10: Relative differences in performance between individual models and the top-performing model based on $\overline{FZ0L}^{\alpha}$, where α ranges from $\{0.95, 0.0975, 0.99\}$. This analysis encompasses a 12-year period from January 2, 2011, to December 28, 2022 (upper panel), and a focused 4-year subperiod from January 2, 2019, to December 28, 2022 (lower panel). Model rankings are highlighted in blue, from the highest performer (1) to the lowest performer (18). Light grey bars denote models outside the 90% confidence level for the Model Confidence Set (MCS).

	$\mathcal{B}$-RSPOT	$\mathcal{B}$-SPOT	$\mathcal{GG}$-RSPOT	$\mathcal{GG}$-SPOT	$\mathcal{W}$-RSPOT	$\mathcal{W}$-SPOT	$\mathcal{B}$-RSPOT	$\mathcal{B}$-SPOT	$\mathcal{GG}$-RSPOT	$\mathcal{GG}$-SPOT	$\mathcal{W}$-RSPOT	$\mathcal{W}$-SPOT	$\mathcal{B}$-RSPOT	$\mathcal{B}$-SPOT	$\mathcal{GG}$-RSPOT	$\mathcal{GG}$-SPOT	$\mathcal{W}$-RSPOT	$\mathcal{W}$-SPOT
	$\alpha = 0.95$						$\alpha = 0.975$						$\alpha = 0.99$					
Dow Jones																		
$\mathcal{DB}$-**RSPOT**	**0.00**	**0.02**	**0.00**	0.10	**0.00**	0.05	0.12	0.56	**0.01**	0.18	**0.03**	0.38	0.25	0.73	0.07	0.53	0.19	0.77
$\mathcal{DB}$-**SPOT**	**0.00**	**0.00**	**0.01**	**0.04**	**0.00**	**0.01**	**0.05**	0.12	**0.02**	**0.04**	**0.03**	0.10	0.11	0.28	*0.07*	0.15	0.12	0.39
$\mathcal{DGG}$-**RSPOT**	**0.01**	**0.02**	**0.00**	**0.02**	**0.00**	**0.01**	0.07	0.35	**0.00**	**0.02**	**0.00**	0.15	0.14	0.54	*0.00*	0.24	0.05	0.59
$\mathcal{DGG}$-**SPOT**	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	**0.04**	0.12	**0.01**	**0.01**	**0.01**	**0.03**	*0.08*	0.22	**0.04**	**0.04**	*0.07*	0.24
$\mathcal{DW}$-**RSPOT**	**0.04**	0.25	**0.03**	0.31	**0.00**	*0.09*	0.15	0.38	**0.01**	**0.03**	**0.00**	*0.09*	0.14	0.43	**0.01**	0.16	**0.00**	0.39
$\mathcal{DW}$-**SPOT**	**0.04**	**0.05**	**0.01**	*0.07*	**0.00**	**0.00**	*0.07*	0.17	**0.01**	0.10	**0.01**	**0.01**	*0.07*	0.15	**0.03**	**0.03**	**0.04**	**0.04**
$\mathcal{BNB}$-**RSPOT**	**0.00**	**0.02**	**0.00**	*0.08*	**0.00**	**0.03**	*0.07*	0.42	**0.01**	**0.03**	**0.01**	**0.01**	0.15	0.64	**0.04**	0.38	0.12	0.68
$\mathcal{BNB}$-**SPOT**	**0.00**	**0.00**	**0.01**	**0.04**	**0.00**	**0.01**	**0.05**	0.12	**0.02**	**0.04**	**0.02**	*0.09*	0.11	0.27	*0.07*	0.13	0.12	0.37
NASDAQ																		
$\mathcal{DB}$-**RSPOT**	**0.00**	**0.00**	**0.01**	0.14	**0.01**	0.16	0.15	0.51	**0.03**	0.41	**0.03**	0.30	0.13	0.44	**0.03**	0.44	0.17	0.60
$\mathcal{DB}$-**SPOT**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	*0.06*	0.21	**0.01**	*0.08*	**0.01**	0.05	*0.05*	*0.10*	**0.03**	*0.07*	0.10	0.18
$\mathcal{DGG}$-**RSPOT**	**0.00**	**0.00**	**0.00**	*0.03*	**0.00**	**0.03**	0.13	0.41	**0.00**	0.22	**0.00**	0.11	*0.09*	0.34	**0.00**	0.28	**0.03**	0.45
$\mathcal{DGG}$-**SPOT**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	*0.07*	0.22	**0.01**	**0.03**	**0.01**	**0.01**	*0.06*	0.11	**0.02**	**0.03**	*0.09*	0.11
$\mathcal{DW}$-**RSPOT**	**0.00**	**0.02**	**0.01**	0.2	**0.00**	0.10	0.21	0.48	**0.02**	0.37	**0.00**	0.17	*0.09*	0.27	**0.01**	0.20	**0.00**	0.28
$\mathcal{DW}$-**SPOT**	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	0.13	0.32	**0.02**	0.12	**0.01**	**0.00**	*0.07*	0.14	**0.02**	**0.05**	*0.06*	**0.03**
$\mathcal{BNB}$-**RSPOT**	**0.00**	**0.00**	**0.01**	*0.08*	**0.00**	0.10	0.12	0.43	**0.02**	0.31	**0.01**	0.21	*0.09*	0.36	**0.02**	0.34	*0.09*	0.50
$\mathcal{BNB}$-**SPOT**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	*0.07*	0.24	**0.02**	*0.09*	**0.01**	**0.05**	*0.06*	*0.11*	**0.03**	**0.07**	0.10	0.17
SP500																		
$\mathcal{DB}$-**RSPOT**	**0.00**	0.53	**0.03**	0.29	**0.03**	0.41	*0.08*	0.83	*0.04*	0.55	0.11	0.74	**0.01**	0.92	0.11	0.96	0.33	0.99
$\mathcal{DB}$-**SPOT**	**0.00**	**0.01**	**0.00**	**0.02**	**0.00**	**0.02**	**0.02**	0.11	**0.01**	*0.08*	**0.03**	0.16	**0.00**	**0.02**	**0.02**	0.23	**0.04**	0.64
$\mathcal{DGG}$-**RSPOT**	**0.02**	0.41	**0.00**	0.17	**0.00**	0.25	0.11	0.67	**0.00**	0.42	*0.02*	0.68	**0.01**	0.71	**0.01**	0.91	*0.08*	0.99
$\mathcal{DGG}$-**SPOT**	**0.00**	**0.02**	**0.00**	**0.00**	**0.00**	**0.01**	**0.02**	0.15	**0.01**	**0.03**	**0.02**	0.15	**0.00**	**0.02**	**0.01**	*0.09*	**0.02**	0.54
$\mathcal{DW}$-**RSPOT**	0.11	0.51	**0.04**	0.32	**0.00**	0.38	0.18	0.59	**0.04**	0.41	**0.00**	0.63	**0.03**	0.54	**0.03**	0.69	**0.01**	0.98
$\mathcal{DW}$-**SPOT**	**0.01**	0.12	**0.04**	**0.00**	**0.00**	**0.00**	*0.06*	0.24	**0.02**	0.10	**0.01**	**0.03**	**0.00**	**0.04**	**0.01**	*0.08*	**0.01**	0.10
$\mathcal{BNB}$-**RSPOT**	*0.06*	0.74	**0.03**	0.46	*0.07*	0.67	*0.06*	0.74	**0.03**	0.47	*0.08*	0.68	**0.00**	0.86	*0.07*	0.93	0.23	0.99
$\mathcal{BNB}$-**SPOT**	**0.02**	0.13	**0.00**	*0.08*	**0.03**	0.15	**0.02**	0.13	**0.01**	*0.08*	**0.03**	0.16	**0.00**	**0.03**	**0.02**	0.22	**0.04**	0.62
Wilshire																		
$\mathcal{DB}$-**RSPOT**	**0.03**	0.51	**0.01**	0.25	**0.00**	0.35	0.10	0.91	**0.02**	0.76	**0.04**	0.90	**0.02**	0.97	*0.07*	0.99	0.19	0.99
$\mathcal{DB}$-**SPOT**	**0.00**	**0.02**	**0.00**	**0.02**	**0.00**	**0.01**	**0.01**	0.10	**0.00**	0.10	**0.01**	0.11	**0.00**	**0.04**	**0.01**	0.18	**0.01**	0.52
$\mathcal{DGG}$-**RSPOT**	*0.08*	0.55	**0.00**	0.26	**0.00**	0.38	*0.10*	0.82	**0.00**	0.71	**0.01**	0.88	**0.01**	0.92	**0.01**	0.98	*0.07*	0.99
$\mathcal{DGG}$-**SPOT**	**0.00**	**0.03**	**0.00**	**0.01**	**0.00**	**0.01**	**0.01**	0.10	**0.00**	*0.06*	**0.00**	0.91	**0.00**	**0.02**	**0.00**	*0.09*	**0.01**	0.43
$\mathcal{DW}$-**RSPOT**	0.33	0.68	0.10	0.45	**0.00**	0.66	0.23	0.75	**0.04**	0.66	**0.00**	0.91	*0.05*	0.80	**0.04**	0.87	**0.01**	0.99
$\mathcal{DW}$-**SPOT**	**0.02**	0.12	**0.00**	**0.05**	**0.00**	**0.00**	**0.03**	0.20	**0.01**	0.13	**0.00**	**0.04**	**0.00**	**0.00**	**0.04**	*0.09*	**0.00**	0.11
$\mathcal{BNB}$-**RSPOT**	**0.03**	0.50	**0.01**	0.25	**0.00**	0.35	0.10	0.87	**0.02**	0.71	**0.03**	0.87	**0.01**	0.96	0.05	0.98	0.14	0.99
$\mathcal{BNB}$-**SPOT**	**0.00**	**0.02**	**0.00**	**0.02**	**0.00**	**0.01**	**0.01**	0.12	**0.01**	0.11	**0.01**	0.12	**0.00**	**0.04**	**0.01**	0.18	**0.01**	0.49

Table 3: The results of the Diebold-Mariano test for the superior forecasting performance of the discrete-time SPOT model specifications over the continuous-time SPOT models using the loss function $FZ0_t^{\alpha}$ with $\alpha \in \{0.95, 0.0975, 0.99\}$, based on the out-of-sample period from January 3, 2011, to December 28, 2022. Bolded values denote statistically significance at the 0.05 level, while italicized values highlight significance at the 0.1 level.

(11).

To provide a more comprehensive risk assessment, we also include ARMA-GARCH specifications, both with a Gaussian or skewed Student's t-distribution, for the error term. The bars in Figure 10 show the percentage differences in $\overline{FZ0L}^{\alpha}$ values between each model and the best-performing model (i.e., the model with the lowest $\overline{FZ0L}^{\alpha}$ value for a given α). Because forecasting performance is inversely proportional to $\overline{FZ0L}^{\alpha}$, a higher bar indicates that the corresponding competing model is less effective than the best one. The rankings are

indicated by blue numbers, with the top-performing model ranked 1.

We chose two sub-periods for the out-of-sample data: (1) the 12-year-long period from January 2, 2011, to December 28, 2022, and (2) a significantly more volatile 4-year-long sub-period from January 2, 2019, to December 28, 2022, encompassing the COVID-19 pandemic and the Russia-Ukraine war. Several important observations can be made from this analysis. First, the model rankings depend on the level of risk α and volatility in the examined period. For NASDAQ and Wilshire, using a 12-year-long out-of-sample period and $\alpha = 0.95$, the discrete-time SPOT models cannot outperform the skew-t ARMA-GARCH models. However, when moving deeper into the tails, specifically for $\alpha = 0.975$ and $\alpha = 0.99$, one of the discrete-time SPOT variants consistently ranked as the best performing risk model. Good results were achieved by the discrete $\mathcal{DW}$-SPOT model, which ranked first at $\alpha = 0.99$ for all equity indices. The superiority of the SPOT models over the ARMA-GARCH specifications becomes particularly evident for $\alpha = 0.99$ and for the second period, which covers the COVID-19-induced financial turmoil.

In these cases, the discrete-time SPOT specifications rank first for all the three α levels. It is important to note that continuous-time SPOT models consistently exhibit inferior performance compared to their discrete-time counterparts. In addition, RSPOT models generally underperform in unrestricted SPOT specifications, indicating that the appropriate inclusion of explanatory variables $\tilde{Y}_{i-1}$ and X_i is recommended for models developed in either continuous or discrete time.

Although the visual illustration of the out-of-sample model performance suggests the superiority of the discrete-time SPOT specifications, it does not prove that the loss differences are statistically significant. To address this, we present the p-values for the Diebold-Mariano (DM) t-statistics of these loss differentials in Table 3. The tests are conducted such that the null hypothesis for the DM statistics states that the 'row model' in Table 3 has equal predictive accuracy as the 'column model,' and the alternative hypothesis states that the 'row model' outperforms the 'column model.' The results show that at $\alpha = 0.95$, and for all equity

	$\mathcal{B}$-SPOT	$\mathcal{DB}$ SPOT	$\mathcal{GG}$-SPOT	$\mathcal{DGG}$-SPOT	$\mathcal{W}$- SPOT	$\mathcal{DW}$-SPOT	$\mathcal{BNB}$-SPOT	Gaussian GARCH	Skew-t GARCH
Strict ESR test p-value (period A)									
$\alpha = 0.95$	0.818	0.682	0.868	0.647	0.730	0.634	0.699	**0.000**	**0.016**
$\alpha = 0.975$	0.974	0.927	0.956	0.957	0.876	0.853	0.929	**0.000**	**0.043**
$\alpha = 0.99$	0.994	0.994	0.975	0.984	0.995	0.945	0.989	**0.000**	*0.097*
Auxiliary ESR test p-value (period A)									
$\alpha = 0.95$	0.873	0.740	0.931	0.727	0.690	0.647	0.757	**0.001**	**0.014**
$\alpha = 0.975$	0.951	0.933	0.999	0.908	0.843	0.863	0.919	**0.000**	**0.034**
$\alpha = 0.99$	0.956	0.999	0.971	0.993	0.993	0.944	0.999	**0.001**	*0.097*
Strict ESR test p-value (period B)									
$\alpha = 0.95$	0.766	0.745	0.770	0.750	0.665	0.698	0.759	**0.001**	**0.038**
$\alpha = 0.975$	0.880	0.830	0.853	0.853	0.705	0.770	0.830	**0.001**	0.266
$\alpha = 0.99$	0.868	0.858	0.919	0.863	0.872	0.841	0.861	**0.001**	0.722
Auxiliary ESR test p-value (period B)									
$\alpha = 0.95$	0.729	0.738	0.778	0.764	0.592	0.733	0.747	**0.003**	**0.035**
$\alpha = 0.975$	0.889	0.845	0.878	0.899	0.693	0.790	0.852	**0.001**	0.236
$\alpha = 0.99$	0.899	0.909	0.952	0.900	0.908	0.874	0.891	**0.019**	0.696

Table 4: ES backtesting results for the Dow Jones index. Panel A: the out-of-sample period from January 2, 2011, to December 28, 2022. Panel B: the out-of-sample period from January 2, 2019, to December 28, 2022.

indices, at least two (for S&P 500) or three specifications (for the other indices) of discrete-time SPOT models significantly outperform all continuous-time SPOT models. For example, in the case of the Dow Jones index, the $\mathcal{DB}$-SPOT, $\mathcal{DGG}$-SPOT, $\mathcal{DGG}$-RSPOT, and $\mathcal{BNB}$-SPOT models delivered statistically significant forecasting gains over all continuous-time approaches. As we move deeper into the tails, the superiority of discrete-time specifications becomes less clear. However, for the Dow Jones index and $\alpha = 0.975$, the $\mathcal{DGG}$-SPOT model still significantly outperforms five out of the six continuous-time SPOT specifications.

Interestingly, for VaR and ES at $\alpha = 0.99$, the best forecasting accuracy is achieved by the parsimonious $\mathcal{DW}$-SPOT specification. At the 10% significance level, this model significantly outperformed five of the six continuous-time variants of the SPOT models across all equity indices.

5.5. ES forecasts

To assess and compare the quality of ES forecasts, we first used the method introduced by Emmer et al. (2015) and subsequently applied to multivariate intensity POT models by

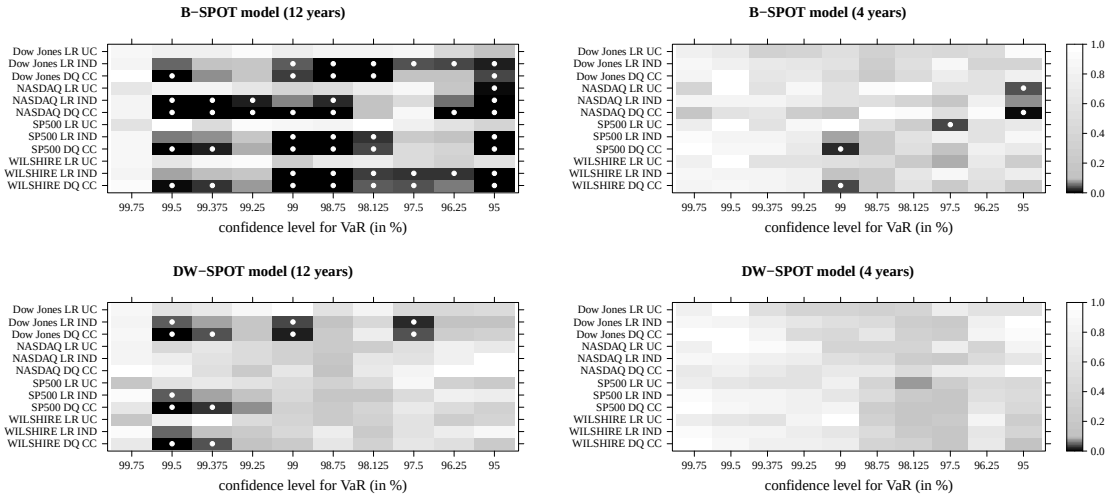

Figure 11: From top to bottom: Out-of-sample heatmap charts for the p-values of the VaR backtesting procedures for the risk models: $\mathcal{B}$-SPOT, and $\mathcal{DW}$-SPOT. White dots indicate p-value < 0.05. Left side: The 12-year out-of-sample period from January 2, 2011, to December 28, 2022. Right side: The 4-year out-of-sample period from January 2, 2019, to December 28, 2022.

Hautsch and Herrera (2020). This approach hinges on approximating ES, denoted as e_t^α, using the following formula: $e_t^\alpha \approx 1/4(\upsilon_t^\alpha + \upsilon_t^{0.75\alpha+0.25} + \upsilon_t^{0.5\alpha+0.5} + \upsilon_t^{0.25\alpha+0.75})$. This approximation technique enables the utilization of the established VaR backtesting methodologies. Specifically, we implement the Likelihood-Ratio-Based Test of Unconditional Coverage (LR UC) and the Test of Independence in VaR Exceedances (LR Ind), both of which were introduced by Christoffersen (1998). We also applied the dynamic quantile (DQ) test of conditional coverage developed by Engle and Manganelli (2004), where, based on auxiliary regressions, the demeaned VaR exceedances were regressed on their lags and contemporaneous VaR.

Figure 11 presents heatmaps that display the p-values corresponding to these backtesting procedures, conducted for ES at risk levels $\alpha = 0.99$, 0.975, and 0.95[5]. Figure 11 confronts the forecasting performance of the continuous-time $\mathcal{B}$-SPOT model and the discrete-time $\mathcal{DW}$-SPOT model of risk which, on average, rendered the best $\overline{FZ0L}^\alpha$ score (the results for the remaining models are enclosed in Appendix D). Interestingly, the SPOT specifications

[5]As a result, the LR UC, LR Ind, and DQ CC tests were executed across the range of ten VaR risk levels: 0.95, 0.9625, 0.975, 0.98125, 0.9875, 0.99, 0.9925, 0.99375, 0.995, 0.9975

perform much better in the more turbulent period from January 3, 2019, to December 28, 2022, when the threshold exceedances occurred more frequently and their magnitudes were larger. However, it is evident that the $\mathcal{DW}$-SPOT leads to the most accurate forecasts of the tail risk and that the gain in the ES forecast accuracy of discrete-time SPOT models over continuous-time SPOT models is substantial, favoring the latter specifications. For all equity indices, out of 120 instances of VaR backtests, the null hypotheses of the LR UC, LR Ind, or DQ CC tests were rejected in 47 cases for the $\mathcal{B}$-SPOT VaR model and only in 12 cases for $\mathcal{DW}$-SPOT VaR specification. In the volatile period from January 3, 2019, to December 28, 2022, for all ten risk levels, none of the null hypotheses were rejected for the $\mathcal{DW}$-SPOT model.

To evaluate the accuracy of ES forecasts directly, we use two variants of the backtesting procedures recently developed by Bayer and Dimitriadis (2022), namely the Auxiliary ESR (ES regression) and the Strict ESR test[6]. These tests rely on a system of two auxiliary regressions, which for the negative returns Y_t are given as: $-Y_t = \beta_0 + \beta_1 m_t + u_t^q$ and $-Y_t = \gamma_0 + \gamma_1 e_t^\alpha + u_t^e$, where $m_t = -v_t^\alpha$ (for the Auxiliary ESR test) or $m_t = -e_t^\alpha$ (for the Strict ESR test), and $Q_{1-\alpha}(u_t^q|\mathcal{F}_{t-1}) = 0$, $ES_{1-\alpha}(u_t^e|\mathcal{F}_{t-1}) = 0$ (with $Q(\cdot)$ denoting the conditional quantile). The regression system is estimated jointly to test $H_0: \ (\gamma_0, \gamma_1) = (0, 1)$ against $H_1 \ \ (\gamma_0, \gamma_1) \neq (0, 1)$.

Table 4 presents the p-values associated with these tests for the Dow Jones index (for other indices, see Appendix D). The SPOT specifications are not rejected uniformly across all ES risk levels and model specifications. This finding is particularly intriguing because, unlike the approach proposed by Emmer et al. (2015), this methodology does not differentiate between the performances of individual models. However, ES forecasts based on ARMA GARCH models, most notably those incorporating Gaussian innovations, are rejected.

[6]We used the R library esback of Bayer and Dimitriadis (2022), available at https://cran.r-project.org/web/packages/esback/index.html)

6. Conclusions

In this study, we introduce a novel extension of dynamic POT models for extreme events in financial markets. Our approach is motivated by the fact that extreme daily returns can only occur on a grid of discrete points in time (e.g., distinct days), with a probability mass predominantly concentrated within intervals ranging from to 1 to 2-5 days. Accordingly, we propose parametric specifications for discrete-time score-driven POT models. In our framework, we treat the time intervals between extreme returns, the so-called inter-exceedance times, as integer-value positive variables.

These inter-exceedance times are dynamically modeled using conditional discrete probability distributions, such as the discrete Weibull, discrete Burr, discrete Generalized Gamma, or Beta Negative Binomial. In particular, to capture the temporal clustering of extreme events, we employ a score-driven approach to model the scale parameter of the associated distribution. Furthermore, we employ a time-varying Generalized Pareto Distribution, guided by a score-driven scale parameter, to effectively model the magnitudes of extreme returns. This approach allows us to capture the serial correlation inherent in the sizes of extreme events efficiently.

Our study yields several key insights that have significant implications for the modeling and forecasting of extreme events. First, based on a simulation example, we show that the discreteness of the inter-exceedance times might lead to a non-monotonic hump shape of a hazard function when using a continuous-time approach. This is because the continuous density function, and consequently the likelihood function, is fitted to discrete positive data and attempts to assign a zero value to data points less than one. In the context of the continuous-time point-process approach to POT models, the hazard function associated with the inter-exceedance times approximates the daily probability of an extreme event. Hence, any distortion in its shape can lead to inaccurate VaR and ES forecasts.

In contrast, discrete-time SPOT models offer greater accuracy because they deliver a direct probability of extreme loss on a given day rather than an approximation. Our empirical

results, which span a 12-year forecasting period, clearly indicate that discrete duration models generally yield more accurate VaR and ES forecasts. These findings are further substantiated by the application of consistent scoring functions and backtesting methods.

In a empirical application, we find that the risk measures derived from SPOT models, whether implemented in continuous or discrete time, outperform those based on GARCH models with either a normal or skew-t distribution for the error term. While identifying the optimal parametric specification for a discrete-time SPOT model warrants further research, our preliminary findings, based on four U.S. stock indices, suggest that the discrete-Weibull SPOT model is the most parsimonious and promising specification.

References

Balkema, A.A., de Haan, L., 1974. Residual life time at great age. The Annals of Probability 2.

Bayer, S., Dimitriadis, T., 2022. Regression-based expected shortfall backtesting. Journal of Financial Econometrics 20, 437–471.

Bee, M., Dupuis, D.J., Trapin, L., 2019. Realized peaks over threshold: A time-varying extreme value approach with high-frequency-based measures. Journal of Financial Econometrics 17, 254–283.

Bień-Barkowska, K., 2020. Forecasting extreme returns in financial markets: A discrete duration framework. Acta Phys. Pol. 138, 48–59.

Bień-Barkowska, K., 2023. Forecasting extreme negative returns in gold and silver: A discrete-duration approach to POT models. Applied Stochastic Models in Business and Industry .

Chakraborty, S., 2015. A new discrete distribution related to generalized gamma distribution and its properties. Communications in Statistics - Theory and Methods 44, 1691–1705.

Chavez-Demoulin, V., McGill, J., 2012. High-frequency financial data modeling using Hawkes processes. J. Bank. Financ. 36, 3415–3426.

Christoffersen, P., 1998. Evaluating interval forecasts. International Economic Review 39, 841–862.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 7, 174–196.

Creal, D., Koopman, S.J., Lucas, A., 2013. Generalized autoregressive score models with applications. Journal of Applied Econometrics 28, 777–795.

Danielsson, J., Ergun, L.M., de Haan, L., de Vries, C.G., 2016. Tail index estimation: Quantile driven threshold selection. Available at SSRN 2717478 .

Deng, K., Qiu, J., 2021. Backtesting expected shortfall and beyond. Quantitative Finance 21, 1109–1125.

Emmer, S., Kratz, M., Tasche, D., 2015. What is the best risk measure in practice? A comparison of standard measures. Journal of Risk 18, 31–60.

Engle, R., Manganelli, S., 2004. CAViaR: Conditional autoregressive value at risk by regression quantiles. Journal of Business & Economic Statistics 22, 367–381.

Engle, R., Russell, J., 1998. Autoregressive conditional duration: A new model for irregularly spaced transaction data. Econometrica 66, 1127–1162.

Ferreira, H., Ferreira, M., 2018. Estimating the extremal index through local dependence. Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 54.

Ferro, C.A.T., Segers, J., 2003. Inference for clusters of extreme values. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65, 545–556.

Fissler, T., Ziegel, J.F., 2016. Higher order elicitability and Osband's principle. The Annals of Statistics 44, 1680–1707.

Fuentes, F., Herrera, R., Clements, A., 2018. Modeling extreme risks in commodities and commodity currencies. Pacific-Basin Finance Journal 51, 108–120.

Fuentes, F., Herrera, R., Clements, A., 2023. Forecasting extreme financial risk: A score-driven approach. International Journal of Forecasting 39, 720–735.

Gorgi, P., 2020. Beta–negative binomial auto-regressions for modelling integer-valued time series with extreme observations. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 82, 1325–1347.

Grammig, J., Maurer, K.O., 2000. Non-monotonic hazard functions and the autoregressive conditional duration model. Economet. J. 3, 16–38.

Hallin, M., Vecchia, D.L., 2020. A simple r-estimation method for semiparametric duration models. Journal of Econometrics 218, 736–749.

Harvey, A.C., 2013. Dynamic models for volatility and heavy tails: with applications to financial and economic time series. volume 52. Cambridge University Press.

Hautsch, N., Herrera, R., 2020. Multivariate dynamic intensity peaks-over-threshold models. J. Appl. Economet. 35, 248–272.

Herrera, R., Schipp, B., 2013. Value at risk forecasts by extreme value models in a conditional duration framework. J. Empir. Finance 23, 33–47.

Hoga, Y., 2019. Where does the tail begin? an approach based on scoring rules. Econometric Reviews 39, 579–601.

Hoga, Y., 2021. The uncertainty in extreme risk forecasts from covariate-augmented volatility models. International Journal of Forecasting 37, 675–686.

Holešovský, J., Fusek, M., 2020. Estimation of the extremal index using censored distributions. Extremes 23, 197–213.

Jacobi, A., Tzur, J., 2021. Wealth distribution across countries: Quality of weibull, dagum and burr XII in estimating wealth over time. Finance Research Letters 43, 102023.

Krishna, H., Pundir, P.S., 2009. Discrete Burr and discrete Pareto distributions. Stat. Methodol. 6, 177–188.

Massacci, D., 2017. Tail risk dynamics in stock returns: Links to the macroeconomy and global markets connectedness. Management Science 63, 3072–3089.

Nakagawa, T., Osaki, S., 1975. The discrete Weibull distribution. IEEE T. Reliab. 24, 300–301.

Patton, A.J., Ziegel, J.F., Chen, R., 2019. Dynamic semiparametric models for expected shortfall (and value-at-risk). Journal of Econometrics 211, 388–413.

Pickands III, J., 1975. Statistical inference using extreme order statistics. The Annals of Statistics 3.

Scarrott, C., MacDonald, A., 2012. A review of extreme value threshold es-timation and uncertainty quantification. REVSTAT–Statistical Journal 10, 33–60.

Schneider, L.F., Krajina, A., Krivobokova, T., 2021. Threshold selection in univariate extreme value analysis. Extremes 24, 881–913.

Smith, R.L., 1984. Threshold methods for sample extremes, in: Statistical Extremes and Applications. Springer Netherlands, pp. 621–638.

Trapin, L., 2018. Can volatility models explain extreme events? Journal of Financial Econometrics 16, 297–315.

Appendix A. Scores for the discretized distributions

Let C denote a type of a continuous distribution, where $C \in \{\mathcal{W}, \mathcal{B}, \mathcal{GG}\}$), and D a type of the corresponding discrete distribution for the inter-exceedance times: $D \in \{\mathcal{DW}, \mathcal{DB}, \mathcal{DGG}\}$.

Let ϵ_i be defined as $\epsilon_i = x_i/\Psi_i$; so that $g^{[\mathcal{C}]}\big(x_i \mid \mathcal{F}_{t_{i-1}}\big) = \frac{1}{\Psi_i} g_{\epsilon}^{[\mathcal{C}]}\big(\epsilon_i \mid \mathcal{F}_{t_{i-1}}\big)$. In the continuous-time case, the score can be derived as:

$$s_i^{[\mathcal{C}]} = \frac{\Psi_i}{g^{[\mathcal{C}]}\left(x_i \mid \mathcal{F}_{t_{i-1}}\right)} \frac{\partial g^{[\mathcal{C}]}\big(x_i \mid \mathcal{F}_{t_{i-1}}\big)}{\partial \Psi_i}. \tag{A.1}$$

In the discrete-time case, the score can be obtained as:

$$\begin{aligned} s_i^{[\mathcal{D}]} &= \frac{\Psi_i}{S^{[\mathcal{C}]}\left(x_i - 1 \mid \mathcal{F}_{t_{i-1}}\right) - S^{[\mathcal{C}]}\left(x_i \mid \mathcal{F}_{t_{i-1}}\right)} \\ &\cdot \frac{\partial\left(S^{[\mathcal{C}]}\left(x_i - 1 \mid \mathcal{F}_{t_{i-1}}\right) - S^{[\mathcal{C}]}\left(x_i \mid \mathcal{F}_{t_{i-1}}\right)\right)}{\partial \Psi_i}. \end{aligned} \tag{A.2}$$

Using the equivalences:
$\frac{\partial S^{[\mathcal{C}]}\left(x_i \mid \mathcal{F}_{t_{i-1}}\right)}{\partial \Psi_i} = -g_{\epsilon}^{[\mathcal{C}]}\left(\epsilon_i \mid \mathcal{F}_{t_{i-1}}\right) \frac{\partial \epsilon_i}{\partial \Psi_i} = g_{\epsilon}^{[\mathcal{C}]}\left(x_i \mid \mathcal{F}_{t_{i-1}}\right) \frac{x_i}{\Psi_i^2} = g^{[\mathcal{C}]}\left(x_i \mid \mathcal{F}_{t_{i-1}}\right) \frac{x_i}{\Psi_i}$, we can provide the general parametric form for the score in Eq. (A.2) as:

$$s_i^{[\mathcal{D}]} = \frac{\left(g^{[\mathcal{C}]}\left(x_i - 1 \mid \mathcal{F}_{t_{i-1}}\right)(x_i - 1) - g^{[\mathcal{C}]}\left(x_i \mid \mathcal{F}_{t_{i-1}}\right) x_i\right)}{S^{[\mathcal{C}]}\left(x_i - 1 \mid \mathcal{F}_{t_{i-1}}\right) - S^{[\mathcal{C}]}\left(x_i \mid \mathcal{F}_{t_{i-1}}\right)}. \tag{A.3}$$

Observe, that the Eq. (A.3) delivers the score for discrete-time SPOT models based on the parametric functional forms of the conditional density and survival function associated with the corresponding continuous distributions.

The score for the continuous-time duration model (A.1) and the discrete-time duration model (A.2) are closely related, since

$$S^{[\mathcal{C}]}\left(x_i - 1 \mid \mathcal{F}_{t_{i-1}}\right) - S^{[\mathcal{C}]}\left(x_i \mid \mathcal{F}_{t_{i-1}}\right) = \int_{x_i - 1}^{x_i} g^{[\mathcal{C}]}(s \mid \mathcal{F}_{t_{i-1}}) ds \approx 1 \cdot g^{[\mathcal{C}]}(x_i \mid \mathcal{F}_{t_{i-1}}).$$

Appendix B. Threshold selection based on scoring rules

In this study, we chose the threshold u in a data-driven manner following the method in Hoga (2019). We use a quantile-weighted continuous-ranked probability score to identify the point at which the distribution tail begins. A dense grid of possible thresholds u was used to evaluate and compare the in-sample goodness-of-fit of $v_t^{\alpha}(u)$ estimates (i.e., VaR) derived using threshold u for a predefined range of risk levels, that is, α values: $\alpha \in \langle \alpha^*, \infty)$. Although Hoga (2019) originally used the Weissmann estimator for extreme quantiles, to make the threshold u perfectly compatible with our approach, we use the conditional quantiles of Y_t derived from the SPOT models instead. Specifically, we search for a u value that minimizes the scoring function v_t^{α} given as

$$S_{QCRPS}^{u} = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{1-\alpha^*}\int_{\alpha^*}^{1} 2\left(\boldsymbol{I}_{\{y_t \leq v_t^{\alpha}(u)\},} - \alpha\right)\left(\hat{v}_t^{\alpha}(u) - y_t\right)\, \mathrm{d}\alpha, \tag{B.1}$$

where T is the number of in-sample observations, $\boldsymbol{I}_{\{.\}}$ denotes the indicator function, and $\hat{v}_t^{\alpha}(u)$ is a t-th day $\alpha \times 100$ % VaR derived from a SPOT model which was estimated using a threshold u. We set $\alpha^* = 0.975$.

To calculate S_{QCRPS}^{u} one performs a discrete approximation of the integral in Eq. (B.1):

$$\hat{S}_{QCRPS}^{u} = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{T^*}\sum_{\alpha=1-1/T}^{1-T^*/T} 2\left(\boldsymbol{I}_{\{y_t \leq \hat{v}_t^{\alpha}(u)\}} - \alpha\right)\left(\hat{v}_t^{\alpha}(u) - y_t\right), \tag{B.2}$$

where $T^* = (1-\alpha^*)T$.

Appendix C. Strictly consistent scoring functions

The forecasts of risk measures generated by a broad range of competing models are compared using the concept of strictly consistent scoring functions introduced in Fissler and Ziegel (2016). A pair of statistical functionals, such as the VaR and ES, is called jointly elicitable using a scoring function (i.e., a loss function), if the correct (true) forecast of these functionals is the unique minimizer of the expected loss. Comparing the mean value of the scoring function at the forecasted levels of VaR and ES makes it possible to rank risk models in terms of their realized scores (Patton et al., 2019). Using our notation for negative returns y_t, which imposes that the traditional 'left-tail risk' corresponds to positive values of ϵ_t^α and υ_t^α, with $\epsilon_t^\alpha > \upsilon_t^\alpha$ (unlike in Fissler and Ziegel (2016), where $\epsilon_t^\alpha < \upsilon_t^\alpha$), a class of strictly consistent scoring functions for VaR and ES will given as:

$$\begin{aligned} FZL_t(\upsilon_t^\alpha, e_t^\alpha) &= (\boldsymbol{I}_{\{y_t \geq \hat{\upsilon}_t^\alpha\}} - (1-\alpha))G_1(-\upsilon_t^\alpha) - \boldsymbol{I}_{\{y_t \geq \hat{\upsilon}_t^\alpha\}} G_1(-y_t) \\ &+ G_2(-e_t^\alpha)\frac{\boldsymbol{I}_{\{y_t \geq \hat{\upsilon}_t^\alpha\}}(y_t - \upsilon_t^\alpha)}{1-\alpha} + G_2(-\epsilon_t^\alpha)(\upsilon_t^\alpha - \epsilon_t^\alpha) - \mathcal{G}_2(-\epsilon_t^\alpha), \end{aligned} \tag{C.1}$$

where G_1 is a weakly increasing function, G_2 is a strictly increasing positive function, and $\mathcal{G}_2' = G_2$. In our empirical applications, we use the scoring function introduced in Patton et al. (2019) as the *FZ0* loss. It delivers loss differentials that are homogeneous of degree 0 and this property is advocated in Patton and Sheppard (2009) and Patton et al. (2019), because it results in higher power of the Diebold and Mariano (1995) tests, used–at a later stage– to investigate the significant differences in forecast gains. Patton et al. (2019) set $G_1(z) = 0$ and $G_2(z) = -1/z$, which, after substituting to Eq. (C.1), yields:

$$FZ0_t^\alpha = \frac{1}{\epsilon_t^\alpha(1-\alpha)}\boldsymbol{I}_{y_t \geq \upsilon_t^\alpha}(y_t - \upsilon_t^\alpha) + \frac{\upsilon_t^\alpha}{\epsilon_t^\alpha} + \ln(\epsilon_t^\alpha) - 1. \tag{C.2}$$

Appendix D. Tables and Figures

Parameter	$\mathcal{W}$	$\mathcal{B}$	$\mathcal{GG}$	$\mathcal{DW}$	$\mathcal{DB}$	$\mathcal{DGG}$	$\mathcal{BNB}$
			Nasdaq				
			Probability of threshold exceedance				
ω_h	0.237 (0.044)	0.214 (0.059)	0.000	0.227 (0.056)	0.343 (0.094)	0.000	0.278(0.063)
β_h	0.925 (0.014)	0.870 (0.035)	0.926 (0.016)	0.924 (0.020)	0.907 (0.026)	0.905(0.019)	0.907(0.021)
α_h	0.187 (0.019)	0.185 (0.033)	0.201 (0.024)	0.260 (0.042)	0.288 (0.048)	0.287(0.041)	0.335(0.051)
η_h	-0.066 (0.014)	-0.055 (0.018)	-0.065 (0.016)	-0.068 (0.018)	-0.068 (0.020)	-0.069(0.019)	-0.077(0.021)
γ	0.916 (0.023)		0.464 (0.072)	0.772 (0.021)		0.433 (0.077)	
κ		1.926(0.157)			0.968 (0.059)		
ζ		0.644(0.107)			3.069 (0.904)		
r							0.877 (0.089)
τ							3.165 (0.622)
ν			3.969 (1.289)			3.110 (1.091)	
			Threshold exceedance magnitude				
ω_s				0.263 (0.059)			
β_s				0.787 (0.053)			
α_s				0.127 (0.029)			
$\eta_{1,s}$				-0.351 (0.108)			
$\eta_{2,s}$				-0.801 (0.264)			
ξ				0.0001			
			Log Likelihood ($Log\mathcal{L}$) and Information Criteria				
$Log\mathcal{L}$	-3543.9456	-3472.389	-3491.8775	-3428.2200	-3417.8158	-3416.2292	-3416.7075
AIC	8.394	8.228	8.274	8.121	8.099	8.095	8.096
BIC	8.395	8.295	8.341	8.183	8.166	8.162	8.163

Parameter	$\mathcal{W}$	$\mathcal{B}$	$\mathcal{GG}$	$\mathcal{DW}$	$\mathcal{DB}$	$\mathcal{DGG}$	$\mathcal{BNB}$
			S&P 500				
			Probability of threshold exceedance				
ω_h	0.338 (0.078)	0.278 (0.083)	0.000	0.336 (0.081)	0.448(0.129)	-0.293(0.435)	0.384(0.113)
β_h	0.892 (0.025)	0.885 (0.031)	0.887 (0.030)	0.887 (0.029)	0.876(0.033)	0.879(0.037)	0.874(0.036)
α_h	0.228 (0.037)	0.278 (0.049)	0.266 (0.046)	0.302 (0.055)	0.377(0.068)	0.361(0.072)	0.416(0.078)
η_h	-0.066 (0.023)	-0.039 (0.019)	-0.062 (0.025)	-0.070 (0.020)	-0.055(0.022)	-0.057(0.030)	-0.066(0.032)
γ	0.833 (0.031)		0.445 (0.013)	0.733 (0.024)		0.325 (0.023)	
κ		1.376(0.094)			0.942 (0.060)		
ζ		1.157(0.225)			2.811 (0.773)		
r							0.854 (0.096)
τ							2.670 (0.625)
ν			3.544 (0.238)			4.896 (3.008)	
			Threshold exceedance magnitude				
ω_s				0.378 (0.095)			
β_s				0.720 (0.053)			
α_s				0.166 (0.040)			
$\eta_{1,s}$				-0.582 (0.220)			
$\eta_{2,s}$				-0.867 (0.303)			
ξ				0.106 (0.041)			
			Log Likelihood ($Log\mathcal{L}$) and Information Criteria				
$Log\mathcal{L}$	-2448.6303	-2409.6493	-2416.5765	-2387.6591	-2376.8292	-2377.7164	-2375.4811
AIC	8.879	8.742	8.767	8.659	8.624	8.627	8.619
BIC	8.965	8.836	8.861	8.745	8.717	8.721	8.713

Parameter	$\mathcal{W}$	$\mathcal{B}$	$\mathcal{GG}$	$\mathcal{DW}$	$\mathcal{DB}$	$\mathcal{DGG}$	$\mathcal{BNB}$
			Wilshire				
			Probability of threshold exceedance				
ω_h	0.311 (0.073)	0.332 (0.098)	0.000	0.318 (0.077)	0.501(0.140)	-0.061(0.269)	0.397(0.116)
β_h	0.906 (0.021)	0.872 (0.035)	0.893 (0.029)	0.898 (0.026)	0.875(0.033)	0.883(0.035)	0.874(0.036)
α_h	0.238 (0.039)	0.332 (0.053)	0.284 (0.051)	0.316 (0.053)	0.399(0.069)	0.380(0.078)	0.436(0.087)
η_h	-0.068 (0.025)	-0.044 (0.022)	-0.064 (0.028)	-0.073 (0.021)	-0.063(0.024)	-0.063(0.033)	-0.075(0.036)
γ	0.846 (0.032)		0.440 (0.014)	0.741 (0.026)		0.389 (0.105)	
κ		1.307(0.094)			0.903 (0.060)		
ζ		1.307(0.284)			3.423 (1.157)		
r							0.798 (0.089)
τ							2.996 (0.805)
ν			3.612 (0.269)			3.405 (1.755)	
			Threshold exceedance magnitude				
ω_s				0.445 (0.097)			
β_s				0.689 (0.061)			
α_s				0.141 (0.041)			
$\eta_{1,s}$				-0.697 (0.237)			
$\eta_{2,s}$				-0.918 (0.307)			
ξ				0.092 (0.045)			
			Log Likelihood ($Log\mathcal{L}$) and Information Criteria				
$Log\mathcal{L}$	-2254.1451	-2226.4534	-2229.1780	-2198.8253	-2192.9443	-2193.4637	-2191.3054
AIC	8.918	8.813	8.824	8.700	8.681	8.683	8.674
BIC	9.009	8.913	8.923	8.792	8.781	8.783	8.774

Table D.5: Maximum likelihood estimates of SPOT models for the NASDAQ, S&P 500, and Wilshire for the period from January 2, 1981, to December 30, 2020. Standard errors are provided in parentheses. $Log\mathcal{L}$ represents the maximum log-likelihood value. AIC and BIC refer to the Akaike Information Criterion and the Schwarz Information Criterion, respectively.

	$\mathcal{B}$-SPOT	$\mathcal{DB}$ SPOT	$\mathcal{GG}$-SPOT	$\mathcal{DGG}$-SPOT	$\mathcal{W}$- SPOT	$\mathcal{DW}$-SPOT	$\mathcal{BNB}$-SPOT	Gaussian GARCH	Skew-t GARCH
Nasdaq									
Strict ESR test p-value (period A)									
$\alpha = 0.95$	0.446	0.925	0.861	0.982	0.723	0.957	0.934	**0.000**	**0.025**
$\alpha = 0.975$	0.346	0.930	0.784	0.882	0.858	0.936	0.910	**0.000**	**0.018**
$\alpha = 0.99$	0.851	0.974	0.856	0.973	0.920	0.953	0.965	**0.000**	*0.080*
Auxiliary ESR test p-value (period A)									
$\alpha = 0.95$	0.556	0.925	0.896	0.977	0.694	0.989	0.919	**0.000**	**0.027**
$\alpha = 0.975$	0.509	0.895	0.827	0.929	0.850	0.966	0.901	**0.000**	**0.019**
$\alpha = 0.99$	0.924	0.961	0.855	0.981	0.911	0.910	0.954	**0.000**	0.100
S&P 500									
Strict ESR test p-value (period A)									
$\alpha = 0.95$	0.777	0.817	0.885	0.763	0.759	0.772	0.806	**0.000**	**0.022**
$\alpha = 0.975$	0.643	0.814	0.810	0.792	0.976	0.920	0.806	**0.000**	**0.016**
$\alpha = 0.99$	0.798	0.663	0.722	0.588	0.531	0.320	0.534	**0.000**	*0.059*
Auxiliary ESR test p-value (period A)									
$\alpha = 0.95$	0.914	0.877	1.000	0.867	0.715	0.767	0.866	**0.000**	**0.016**
$\alpha = 0.975$	0.568	0.819	0.831	0.813	0.976	0.918	0.827	**0.000**	**0.017**
$\alpha = 0.99$	0.840	0.772	0.838	0.708	0.544	0.461	0.792	**0.000**	*0.055*
Wilshire									
Strict ESR test p-value (period A)									
$\alpha = 0.95$	0.958	0.949	0.995	0.947	0.808	0.897	0.954	**0.000**	**0.028**
$\alpha = 0.975$	0.747	0.890	0.899	0.925	0.963	0.987	0.882	**0.000**	**0.012**
$\alpha = 0.99$	0.652	0.732	0.767	0.725	0.721	0.610	0.700	**0.000**	**0.033**
Auxiliary ESR test p-value (period A)									
$\alpha = 0.95$	0.953	0.984	0.993	0.988	0.648	0.854	0.976	**0.000**	**0.024**
$\alpha = 0.975$	0.727	0.926	0.889	0.930	0.924	0.995	0.923	**0.000**	**0.010**
$\alpha = 0.99$	0.752	0.846	0.881	0.803	0.835	0.676	0.822	**0.000**	**0.030**

Table D.6: ES backtesting results for the NASDAQ, S&P 500, and Wilshire index in the out-of-sample period (A): from January 3, 2011, to December 28, 2022.

	$\mathcal{B}$-SPOT	$\mathcal{DB}$ SPOT	$\mathcal{GG}$-SPOT	$\mathcal{DGG}$-SPOT	$\mathcal{W}$- SPOT	$\mathcal{DW}$-SPOT	$\mathcal{BNB}$-SPOT	Gaussian GARCH	Skew-t GARCH
Nasdaq									
Strict ESR test p-value (period B)									
$\alpha = 0.95$	0.850	0.973	0.977	0.938	0.833	0.917	0.970	**0.003**	0.191
$\alpha = 0.975$	0.487	0.754	0.750	0.775	0.804	0.894	0.880	**0.009**	0.303
$\alpha = 0.99$	0.661	0.861	0.870	0.857	0.868	0.836	0.841	**0.047**	0.285
Auxiliary ESR test p-value (period B)									
$\alpha = 0.95$	0.850	0.979	0.959	0.948	0.741	0.890	0.974	**0.009**	0.197
$\alpha = 0.975$	0.568	0.738	0.780	0.804	0.880	0.843	0.756	**0.009**	0.276
$\alpha = 0.99$	0.859	0.910	0.898	0.901	0.874	0.878	0.918	0.100	0.275
S&P 500									
Strict ESR test p-value (period B)									
$\alpha = 0.95$	0.925	0.903	0.918	0.924	0.688	0.883	0.911	**0.000**	**0.048**
$\alpha = 0.975$	0.650	0.884	0.880	0.778	0.923	0.812	0.856	**0.016**	0.109
$\alpha = 0.99$	0.971	0.979	0.806	0.959	0.769	0.941	0.976	**0.012**	0.168
Auxiliary ESR test p-value (period B)									
$\alpha = 0.95$	0.952	0.908	0.899	0.943	0.698	0.860	0.902	**0.012**	**0.03**6
$\alpha = 0.975$	0.913	0.765	0.928	0.810	0.953	0.926	0.760	**0.016**	0.105
$\alpha = 0.99$	0.969	0.986	0.840	0.965	0.939	0.961	0.987	*0.052*	0.107
Wilshire									
Strict ESR test p-value (period B)									
$\alpha = 0.95$	0.917	0.938	0.914	0.944	0.673	0.889	0.942	**0.002**	0.121
$\alpha = 0.975$	0.948	0.965	0.969	0.953	0.960	0.943	0.955	**0.010**	*0.099*
$\alpha = 0.99$	0.997	0.995	0.978	0.998	0.976	0.972	0.994	**0.012**	0.276
Auxiliary ESR test p-value (period B)									
$\alpha = 0.95$	0.941	0.917	0.908	0.922	0.560	0.839	0.906	**0.035**	0.134
$\alpha = 0.975$	0.924	0.950	0.981	0.956	0.923	0.955	0.955	**0.010**	*0.096*
$\alpha = 0.99$	0.988	0.994	0.963	1.000	0.987	0.995	0.992	**0.028**	0.262

Table D.7: ES backtesting results for the Nasdaq, S&P 500, and Wilshire index in the out-of-sample period (B): from January 3, 2019, to December 28, 2022.

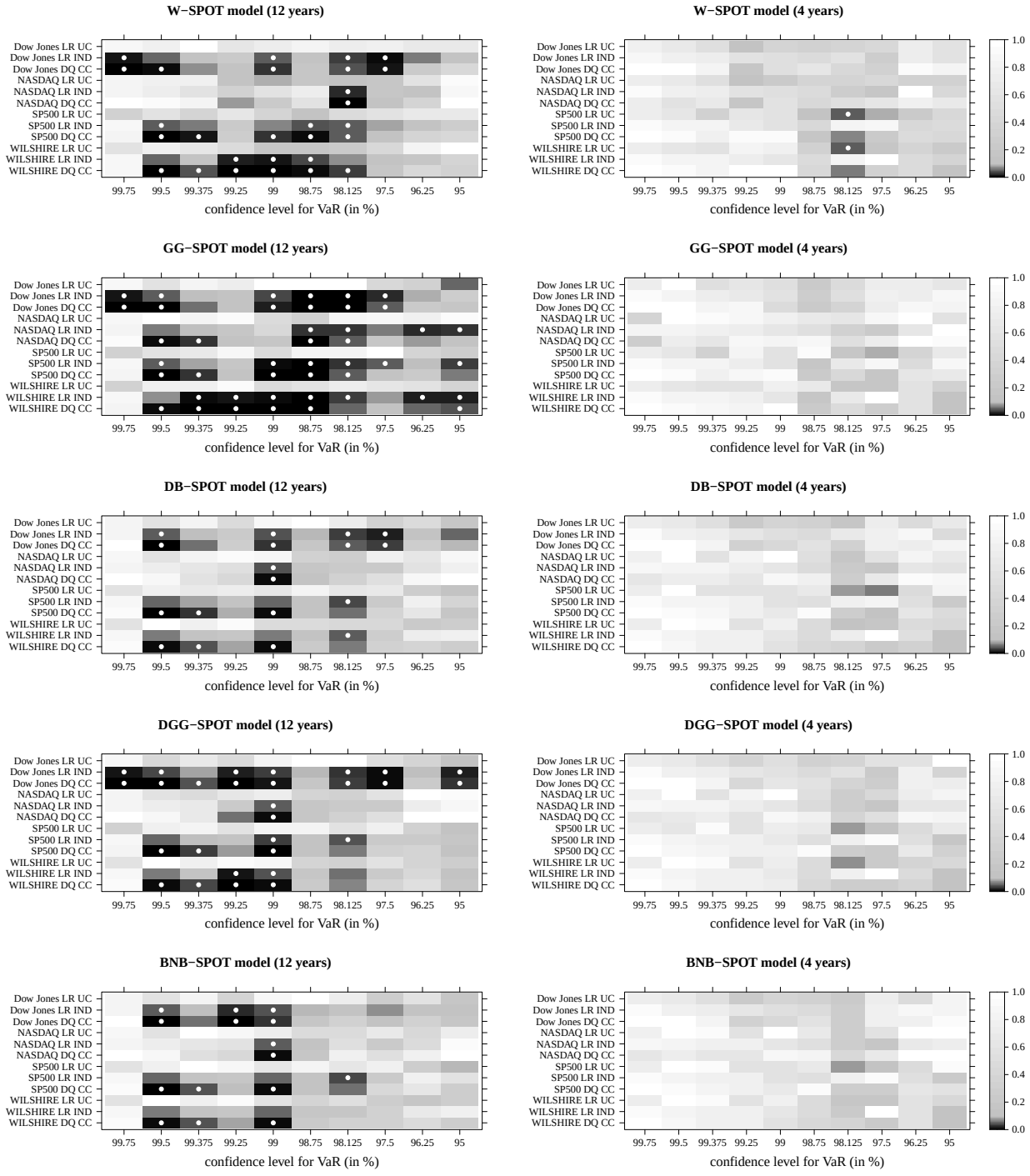

Figure D.12: From top to bottom: Out-of-sample heatmap charts for the p-values of the VaR backtesting procedures for the risk models: $\mathcal{W}$-SPOT, $\mathcal{GG}$-SPOT, $\mathcal{DB}$-SPOT, $\mathcal{DGG}$-SPOT, and $\mathcal{BNB}$-SPOT. White dots indicate p-value < 0.05. Left side: The 12-year out-of-sample period from January 2, 2011 to December 28, 2022. Right side: The 4-year out-of-sample period from January 2, 2019, to December 28, 2022.

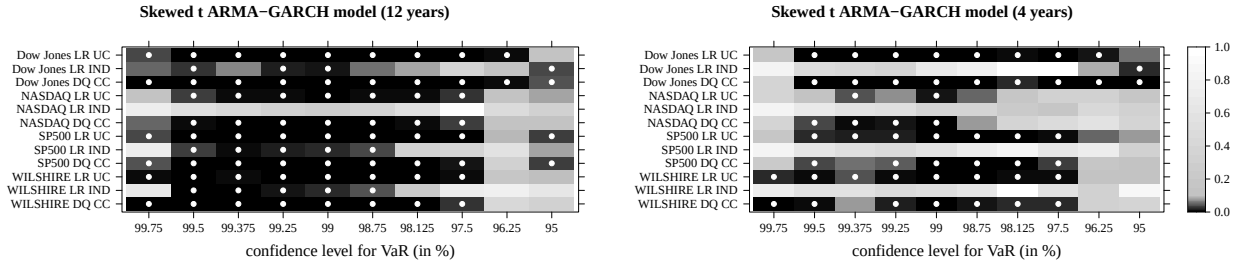

Figure D.13: Out-of-sample heatmap charts for the p-values of the VaR backtesting procedures for the skewed-t ARMA-GARCH risk model. White dots indicate p-value < 0.05. Left side: The 12-year out-of-sample period from January 2, 2011 to December 28, 2022. Right side: The 4-year out-of-sample period from January 2, 2019, to December 28, 2022.