

RESEARCH

Open Access



# An evaluation of the adequacy of Lévy and extreme value tail risk estimates

Sharif Mozumder<sup>1</sup>, M. Kabir Hassan<sup>2\*</sup>  and M. Humayun Kabir<sup>3</sup>

\*Correspondence:  
mhassan@uno.edu

<sup>1</sup> University of Dhaka, Dhaka,  
Bangladesh

<sup>2</sup> University of New Orleans,  
Kenner, LA, USA

<sup>3</sup> Massey University, Palmerston  
North, New Zealand

## Abstract

This study investigates the simplicity and adequacy of tail-based risk measures—value-at-risk (VaR) and expected shortfall (ES)—when applied to tail targeting of the extreme value (EV) model. We implement Lévy–VaR and ES risk measures as full density-based alternatives to the generalized Pareto VaR and the generalized Pareto ES of the tail-targeting EV model. Using data on futures contracts of S&P500, FTSE100, DAX, Hang Seng, and Nikkei 225 during the Global Financial Crisis of 2007–2008, we find that the simplicity of tail-based risk management with a tail-targeting EV model is more attractive. However, the performance of EV risk estimates is not necessarily superior to that of full density-based relatively complex Lévy risk estimates, which may not always give us more robust VaR and ES results, making the model inadequate from a practical perspective. There is randomness in the estimation performances under both approaches for different data ranges and coverage levels. Such mixed results imply that banks, financial institutions, and policymakers should find a way to compromise or trade-off between “simplicity” and user-defined “adequacy”.

**Keywords:** Lévy–Kintchine-formula, Value-at-risk, Expected shortfall, Generalized extreme value

**JEL Classification:** C52, G13

*It can scarcely be denied that the supreme goal of theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience.*

*-Albert Einstein<sup>1</sup>*

## Introduction

Value-at-risk (VaR) is an intuitively simple tail-based risk measure that is popular among practitioners and academics. Recent applications of VaR were highlighted by Perignon and Smith (2010a), Frésard et al. (2011), and Perignon and Smith (2010b). However, the VaR measure has some limitations. It fails to satisfy the requirement of sub-additivity, implying that it does not fulfill the requirement of coherence. Further, VaR fixes the tail events corresponding to a specific confidence level. Although it considers the likelihood

<sup>1</sup> The quote is from Albert Einstein's Herbert Spencer Lecture delivered at Oxford (June 10, 1933) (On the Method of Theoretical Physics, published in *Philosophy of Science*, Vol. 1, No. 2 (April 1934), pp. 163–169., p. 165).

of conditional tail events, it ignores the magnitude of the catastrophe after the occurrence of a tail event. In a nutshell, VaR provides a snapshot of unsystematic losses while failing to consider the actual size of unsystematic losses that exceed the cut-off points. To offset such limitations and ensure that the coherence (sub-additivity) requirements are met, the expected shortfall (ES) measure has been introduced. ES estimates the unsystematic loss by weighing all the possible losses in the tail of the distribution, thus circumventing the limitation of VaR.

Many studies (e.g., Longin 1996; McNeil and Frey 2000; Jondeau and Rockinger 2003; Gençay and Selçuk 2004; Tolikas and Gettinby 2009; So and Wong 2012; Cheng et al. 2015; Du and Escanciano 2017; Bayer and Dimitriadis 2022; Otto and Breitung 2022) have used simple to adequately complex methodologies and assumed different distributional properties in data-generating process for estimating and backtesting VaR and ES models. Some recent studies have also identified VaR forecasting breakdown due to structural change and a break in the data-generating process of returns (Chavez-Demoulin et al. 2014; Quintos et al. 2001). Through joint modeling, time-varying conditional VaR, and ES, Taylor (2019) produced forecasts with generalized autoregressive conditional heteroskedasticity (GARCH) (1,1) and Glosten–Jagannathan–Runkle GARCH (1,1) models using maximum likelihood based on a Student  $t$ -distribution, and the asymmetric Laplace likelihood was used to evaluate post-sample VaR and ES forecasts. Patton et al. (2019) also jointly modeled VaR and ES in a new dynamic framework, which is semiparametric and agnostic about the conditional distribution of returns, and confirmed via simulation that the proposed new ES-VaR models outperform forecasts based on GARCH or rolling window models. The forecasting process of VaR and ES requires sophisticated and complicated models. Lazar and Zhang (2019) examined whether the inadequacy of modeling leads to the model risk of such risk measures and found that ES estimates using GARCH models require more minor corrections for model risk than VaR.

However, there is some trade-off between simplicity and adequacy when deciding on models and the underlying data-generating processes. Under the Basel framework, the Bank for International Settlements (BIS 2019) requires banks to establish “an *adequate* system for monitoring and reporting risk exposures” to assess risk profiles. With this “adequacy” in mind, Hoga and Demetrescu (2022) developed a sequential procedure that can directly and continuously determine risk assessments based on VaR and ES forecasts with controlled size based on the  $t$ -GARCH model. Kourouma et al. (2010) found an underestimation of the risk of loss for unconditional VaR models (historical and extreme values (EVs) theory VaR model) compared with conditional models. The conditional EVT model is more accurate for predicting risk losses during the 2008 Global Financial Crisis. Despite their accuracy, banks are reluctant to use conditional EVT models as the Basel II agreement penalizes banks for using such models.

Therefore, a relevant question arises as to why banks or financial institutions are reluctant to use some models that are adequate to satisfy the BIS framework. Why is the simplicity of tail-targeting EV models, which are much easier to implement, not attractive and adequate compared with models chosen by banks and financial institutions, or do they define “adequacy” from a different perspective? For our study, the heuristic adequacy of the measure is the naive “closest to the empirical estimate.” The

EV models are based on the distribution of extreme returns instead of all returns (Bali 2007). The simplicity of the tail, characterized by the EV model, leads to the analytic formulation of tail-focused risk measures of VaR and the tail-aggregate risk measure of ES.

On the other hand, the Lévy heavy-tailed generalized hyperbolic (GH) family models are based on full-density distributions and are not easy to implement. We consider the Lévy family models and purely extreme tail-based extreme value (EV) models from an adequacy perspective to quantify investors' risks. It is the adequacy perspective that should guide us in the choice of models in risk measures but not the simplicity. However, if simplicity is advocated, then how should the adequacy of models be determined? Thus, whether the performance of EV risk estimates is superior to that of Lévy risk estimates is an empirical question.

In this study, we conduct a heuristic study to determine the adequacy of VaR and ES estimates of investment risks in leading indices during a period when markets were falling and recovering from the global financial or subprime crisis (2007–2008). The standard VaR measures provide inaccurate estimates of losses during highly volatile periods as they require an explicit functional form (normal or lognormal) on the distribution. EV models have seen many applications in modeling extremities of weather, reserves, and financial extremities (Pidgeon 2012; Monier and Gao 2015; Cheng et al. 2014). However, unlike the spectral risk measure, VaR and ES are purely tail-based risk measures. As EV is also purely a tail-based method, it might cause an uncanny impulse that VaR and ES with the EV method might be a good alternative.

Lévy-based heavy-tailed models belong to another class of models that has also been investigated in modeling extreme fluctuations. Lévy models use full data to estimate the parameters. In contrast, an EV model uses only the partial data remaining in the extreme tail of a distribution concerning a certain cut-off point.

EV models use only a few extremely large returns on the tail in calibration when one believes that the extreme tail data follow a generalized EV distribution. This presumption makes us not worry about the true distribution of returns smaller than the threshold that are not in the tail. This idea may be sufficient to get reasonable numbers for the risk measures of VaR and ES defined on the tail. Such an approach applies fewer data points than models that calibrate historical data. It is accepted that the relevance of systematic fluctuations (small return values) not on the tail can be ignored when modeling. Nevertheless, we do not possess an axiomatic justification to assume this will necessarily be true.

This prompts us to investigate the risk of investment in world markets falling and recovering during the Global Financial Crisis of 2007–2008. We adopt a range of models with moderate time-varying volatility and incorporate a stochastic diffusive perturbation of the markets with time as our benchmark. We consider models of the GH family that include stochastic volatility through stochastic time changing without an explicit dynamic for volatility. We assess the comparative performance of tail modeling using both Lévy (both systematics and unexpected returns) and EV models (only unexpected returns), followed by a comparison of the performances of the respective tail-based measures of risk VaR and ES based on both approaches. Thus, we follow a procedure of fixing the tail as applied in standard EV calibration that only uses unexpected returns.

We then obtain Lévy-tails with calibrations that use both systematics and unexpected returns.

The mathematically elegant Lévy approach has a significant limitation, i.e., except for a few trivial cases, there is no analytic formula for the risk measure VaR, without even mentioning the case of ES. Therefore, the VaR calculation is relatively difficult to implement. Complicacies and huge time requirements in implementation have deterred practitioners from using Lévy models to forecast VaR and ES, as observed in a VaR back-testing study. To the best of our knowledge, this is the first study to compare the performance of tail-based VaR and ES estimated for the EV and Lévy models by contrasting the adequacy of tail-based risk measures while considering simplicity in estimating with this tail-targeting model.

The superior method is unclear at the outset. The presumably advantageous use of the returns in the EV model might not be advantageous in practice when the evaluator is obliged to consider small return values that influence the fits, which must determine the shape of the tail. Moreover, there are concerns about whether applying extreme return observations on the tail can be sufficient to model extreme fluctuations, even when market movements are not extreme (such as the decade following the 2007–2008 Global Financial Crisis). In this study, we seek to shed empirical light by estimating the tail-based VaR and ES following both approaches in the existing theoretical framework using data from major indices. The sample period is when markets suffered and recovered from the Global Financial Crisis.

The contribution of this study is that we assess the relative importance of the adequacy and simplicity of EV and Lévy models in estimating VaR and ES. We try to answer the following questions: “Is the simplicity of EV models adequate to guarantee that they would perform robustly in describing extremely unexpected return phenomena?” “Should the adequacy of the Lévy models be more important in the backdrop of the simplicity of EV models?” We find that the simplicity of the tail-characterized EV model leads to the analytic formulation of the tail-focused risk measures of VaR and the tail-aggregate risk measure of ES. However, the performance of EV risk estimates is not necessarily superior to that of Lévy risk estimates. On the other hand, VaR estimates based on Lévy models are more stable than those based on the EV model. However, it is not possible to establish an indisputable rule for a particular Lévy model. The performance of models with only a few unexpected extreme returns (EV model) and modeling with both numerous smaller expected and few extreme unexpected returns (Lévy models) are mixed.

Our model testing with heuristic adequacy measure reinforces the theoretical fundamental that only extreme observations of the EV model (discarding smaller systematic returns) and smaller and extreme observations of Lévy models are different approaches with the common goal of a meaningful simplification of reality. Their relative performance for a particular time window may fail to constitute any guarantee. There is some randomness of estimation performances under both approaches. The explanatory power of the approaches is hardly distinguishable across our data.

In this study, we determine whether a model is solely tail-targeting or has little effect on VaR and ES forecasts at least for the most crucial periods, comprising both sharp market downturns and smooth recoveries of the 2007–08 Global Financial Crisis.

Here, the choice of model can be based on a compromise or trade-off between simplicity and user-defined adequacy. As many banks and financial institutions do not follow adequacy requirements in risk measures as stipulated in the Basel agreement, our findings shed empirical light on such complexity. Our results imply that when the results are mixed, banks and financial institutions should find a way to compromise simplicity and adequacy.

The remaining parts of the paper are structured as follows. Sections “[Characterization in Lévy framework](#)” and “[Initial data analysis](#)” discuss the characterization of a Lévy framework and initial data analysis. Sections “[Estimation of risk measures](#)” and “[GOF: EV versus Lévy](#)” discuss the estimation of VaR and ES under Lévy, its contender EV approaches, and the goodness of fits (GOFs) under contending approaches. Section “[Comparison of Risk Measures](#)” discusses the Comparison of Risk Measures. Section “[VaR and ES Backtesting](#)” compares the forecasts of VaR and ES for the approaches. Section “[Discussion](#)” provides a discussion, and we conclude in Sect. “[Conclusion](#)”.

### Characterization in Lévy framework

Lévy models have recently been applied in modeling extreme behavior analysis (German 2002; Fajardo 2015; Fajardo and Mordecki 2006, 2014; Kim et al. 2008; Fuse and Meucci 2008; Wong and Guan 2011, De Oliver et al. 2018; Farkas and Mathys 2022). The characteristic function of a stochastically continuous process that starts at zero and possesses stationary independent increments has the following general form (see Sato (1999) and Schouten (2003)):

$$\Phi_{X_t}(s) = E \left[ e^{isX_t} \right] = \exp \left\{ t \left[ ias - \frac{s^2 b^2}{2} \right] + \int_{\mathfrak{R} \setminus \{0\}} \left[ e^{isx} - 1 - isxI_{(-1,1)}(x) \right] \nu(dx) \right\} \tag{1}$$

for  $s \in \mathfrak{R}, t \geq 0$  and constants  $a \in \mathfrak{R}, b \in \mathfrak{R}^+$ , where  $\nu$  is the so-called Lévy measure defined on  $\mathfrak{R} \setminus \{0\}$  that satisfies square inerrability of tiny ( $< 1$ ) jumps:

$$\int_{\mathfrak{R} \setminus \{0\}} \left( |x|^2 \wedge 1 \right) \nu(dx) < \infty \tag{2}$$

Equation (1) is the so-called Lévy–Kinchine representation of a Lévy process,<sup>2</sup> which is closely aligned with the concept of infinitely divisible distribution:

$$\Phi_{X_t}(s) = \left[ \Phi_{X_1}(s) \right]^t \tag{3}$$

Thus, the inverse Fourier transform can be applied to obtain the numerical transition density from the characteristic function (1) with the Lévy measure  $\nu$  of a particular Lévy process, which always exists. The numerical transition densities can then estimate the risk measure VaR under different model assumptions. However, in this study, our interest is mainly on the primitive members of Lévy processes belonging to the GH class,

<sup>2</sup> The theory of Lévy processes is well documented, among others, in Bertoin (1996), Sato (1999), Schoutens (2003), and Kyprianou (2006).

which have been widely used in financial modeling (Barndorff-Nielson 1977, 1978, 1995; Eberlein and Prause 1998; Prause 1999; Eberlien and Keller 1995; Bingham and Kiesel 2001; Eberlien and Hammerstein 2002) due to the availability of closed-form densities.

We focus on the GH subclass<sup>3</sup> of Lévy processes (variance gamma (VG), normal-inverse Gaussian (NIG), hyperbolic distribution (HYP), and GH) and estimate the risk measures—VaR and ES—to investigate the relative adequacy of a purely tail-based simple analytic EV risk model compared with full density-based Lévy risk models. Among others, these measures have been investigated by Cotter and Dowd (2006) in the context of future contracts and by Sorwar and Dowd (2010) in the context of options contracts.

Let  $X_1 = \log(S_{t+1}/S_t)$  for nonnegative integer  $t$  and is characterized by Eq. (1) (the Lévy-Kintchine Formula). For the models we consider, the equivalent processes are given more effectively by their densities (see Schouten 2003):

$$f_{X_1}^{vg}(x) = \frac{2e^{\frac{\theta x}{\sigma^2}}}{\sigma\sqrt{2\pi}v^{\frac{1}{v}}\Gamma(\frac{1}{v})} \left(\frac{|x|}{\sqrt{\frac{2\sigma^2}{v} + \theta^2}}\right)^{\frac{1}{v}-\frac{1}{2}} K_{\frac{1}{v}-\frac{1}{2}}\left(\frac{|x|\sqrt{\frac{2\sigma^2}{v} + \theta^2}}{\sigma^2}\right) \tag{4}$$

where  $K_I$  is the modified Bessel function of the third type with index  $I$ ;  $\theta$  is the skewness parameter; and  $v$  is the percentage excess kurtosis in the distribution for the VG model.

$$f_{X_1}^{nig}(x) = \frac{\alpha\delta}{\pi} e^{(\delta\sqrt{\alpha^2-\beta^2}+\beta x)} \frac{K_1(\alpha\sqrt{\delta^2+x^2})}{\sqrt{\delta^2+x^2}} \tag{5}$$

$$f_{X_1}^{hyp}(x) = \frac{\sqrt{\alpha^2-\beta^2}}{2\delta\alpha K_1(\delta\sqrt{\alpha^2-\beta^2})} e^{(-\alpha\sqrt{\delta^2+x^2}+\beta x)} \tag{6}$$

$$f_{X_1}^{gh}(x) = \frac{(\alpha^2-\beta^2)^{\frac{v}{2}}}{\sqrt{2\pi}\alpha^{(v-\frac{1}{2})}\delta^v K_v(\delta\sqrt{\alpha^2-\beta^2})} (\delta^2+x^2)^{\frac{v-1}{2}} K_{v-\frac{1}{2}}(\alpha\sqrt{\beta^2+x^2}) e^{\beta x} \tag{7}$$

Due to these closed-form densities, obtaining standard errors (SEs) of each parameter becomes easier by easily computing Fisher’s information matrix.

In our context, the competing idea to the Lévy approach assumes that only the extreme returns characterize the performance of the risk measure of VaR and ES. As in the studies by Dowd (2005), Cotter and Dowd (2006), and Mozumder et al. (2017), perhaps the most elegant tool to utilize in such a context is the peaks-over-threshold, which lies in the fact that as threshold  $u$  becomes larger, the distribution of exceedances converges to generalized Pareto (GP) distribution, having the following two-parametric characterization:

<sup>3</sup> VG, NIG, and HYP are versions of a GH model with some or other parameters restricted. Thus, to observe the effect of full flexibility and selected restrictions for this family of processes, in addition to the restricted models, we include the GH model.

$$GP_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\beta}\right)^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\beta}\right) & \text{if } \xi = 0 \end{cases} \tag{8}$$

where

$$x \in \begin{cases} [0, \infty) & \text{if } \xi \geq 0 \\ \left[0, -\frac{\beta}{\xi}\right] & \text{if } \xi < 0 \end{cases} \tag{9}$$

$\xi$  and  $\beta > 0$  are shape and scale parameters, respectively, contingent on each choice of threshold  $u$ .

### Initial data analysis

We employ future contracts return data; our empirical analysis is based on the returns of the S&P500, FTSE100, DAX, Hang Seng, and Nikkei 225 indices. We choose futures contracts because there is a lack of studies on EV and Levy that employ futures contracts data. The data are about futures contracts from January 1, 2007 to December 31, 2017, which expire in the following trading months. The rollover from one expiring contract to the next occurs at the start of each trading month. Datastream considers padding the dataset and considers bank holidays’ end-of-day price as the previous trading day’s end-of-day price—a technique accepted by practitioners and ensures we have the same daily returns for all indices (2762). Our sample period comprises the period of the 2008 Global Financial Crisis and beyond. This helps us to check the robustness of the adequacy versus simplicity of the competing approaches (and methods) in terms of the tail risk measures—VaR and ES.

Table 1 presents the summary statistics of the returns of all index futures. For our analysis, we identify the cut-off point in each extreme tail according to the EV theory discussed in a recent study on the VaR backtest (Mozumder et al. 2017). We note that the extents of extremity in return series corresponding to various indices are not similar.

Table 2 presents a good fit to the data for both long and short positions obtained with GP distribution (GPD); the tail indices are positive except for the Nikkei225; and the

**Table 1** Summary statistics of the futures returns

	SPX	FTSE100	DAX	Hang Seng	Nikkei225
Mean	0.022	0.065	−0.055	−0.0138	−0.0564
Median	0.065	0.0496	−0.083	−0.0343	−0.0653
Maximum	13.202	9.5803	6.858	7.6758	7.8446
Minimum	−10.400	−9.6992	−5.816	−5.6327	−7.4204
Std. dev	01.265	1.2098	1.252	1.1762	1.4419
Skewness	−0.1521	−0.212	0.249	0.336	0.321
Kurtosis	17.254	10.838	5.639	5.909	7.065
Jarque–Bera	23,908.05	7193.87	686.07	766.20	836.40

We report the summary statistics of returns on futures for the world indexes S&P500, FTSE100, DAX, Hang Seng, and Nikkei 225. From January 1, 2007, to December 31, 2017, Futures contracts expire in the following trading months; rollover from one expiring contract to the next occurs at the start of each trading month



**Table 2** Parameter estimates for Generalized Pareto Distribution (GDP)

Index	Position	Threshold ( $u$ )	Prob	$N_u$	Scale ( $\beta$ )	Tail ( $\xi$ )
S&P500	Long	2	0.041	131	0.602 (0.070)	0.173 (0.102)
	Short	2	0.033	117	0.743 (0.121)	0.107 (0.096)
FTSE100	Long	1.5	0.069	249	0.622 (0.063)	0.084 (0.091)
	Short	1.5	0.083	278	0.741 (0.071)	0.033 (0.072)
DAX	Long	2	0.076	234	1.214 (0.102)	0.016 (0.061)
	Short	2	0.074	231	1.020 (0.118)	0.051 (0.083)
Hang Seng	Long	2	0.102	349	1.169 (0.110)	0.114 (0.058)
	Short	2	0.108	365	1.139 (0.092)	0.145 (0.064)
Nikkei 225	Long	2	0.084	271	0.846 (0.082)	-0.018 (0.053)
	Short	2	0.079	272	1.059 (0.080)	-0.059 (0.057)

Maximum likelihood estimates of the GPD parameters for long and short positions are based on daily % returns of futures contracts from January 1, 2007, to December 31, 2017.  $u$  is the threshold (selected using the threshold selection procedure with mean residual plot),  $N_u$  is the number of exceedances in excess of  $u$ , Prob is the probability of observation in excess of  $u$ ,  $\xi$  is the tale, and  $\beta$  is the scale parameter, respectively. In parenthesis, estimated standard errors of the parameters are reported

estimated scale parameters fluctuate around 1. Table 2 also provides assumed thresholds  $u$ , the number of exceedances ( $N_u$ ) contingent on the choice of thresholds  $u$ , and the observed exceedance probabilities (Prob) contingent on the choice of thresholds  $u$ . Table 2 also presents the asymmetry of long and short positions of tail definition ( $u$ ) choices in an EV model. The same cut-off point results in a different number of left-alone observations on the tails for long and short positions. Parameter estimates are expected to differ with respect to different positions under EV models. In the case of full-density-based Lévy models, it results in only a sign alteration of the skewness characterizing parameter<sup>4</sup> corresponding to long and short positions.<sup>5</sup>

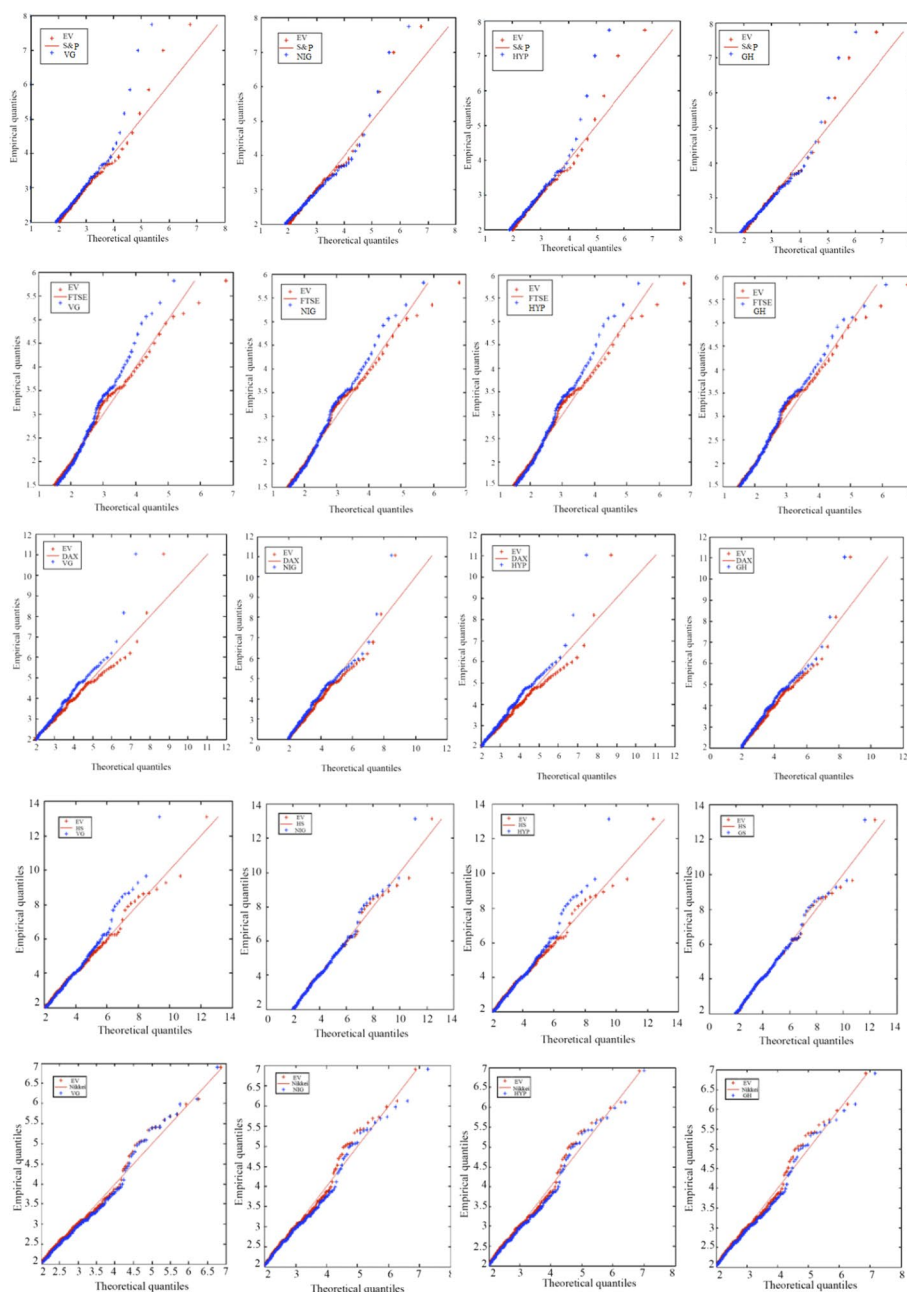
To visualize the differences in the models’ fit in the tails, for all indices under consideration, we separately present the GP EV tail alongside the tails for each of our considered Lévy models. Our strategy is to obtain the EV quantiles above thresholds and the corresponding quantiles from the Lévy models. Thus, instead of fixing tail mass, we set the thresholds. Figure 1 depicts the QQ-plot of EV with each of the Lévy models for all indices. There is clear evidence of deviation between EV and Lévy quantiles at the extreme, although the EV models reveal smaller deviations in most cases. Such differences may be attributed to how different Lévy models feed information from observations outside the tails in fitting the tails.

Table 3 presents the maximum likelihood estimate of the parameters for all five indices and all four Lévy models. The threshold value selection is an important factor as it has the strongest effect on the results. While larger thresholds produce few EVs and lead to large variances, smaller thresholds generate a sample that approximates the models poorly. We select the smallest threshold value among those that produce EVs, following the limit exceedance model. We use the mean residual plot to determine thresholds  $u$  and the probabilities of exceedances. Both the EV and the Lévy

<sup>4</sup> Every member of the GH family has sign-of-skewness explicitly characterized by the sign of a single parameter of the model. That is why we consider GH family in our context.

<sup>5</sup> The difference in tail masses for GH members, with respect to long and short positions under the same cut-off point, is simply due to the sign change of the skewness parameter.





**Fig. 1** Q-Q plot in excess of threshold: long position in S&P500, FTSE, DAX, HangSeng, Nikkei225

models provide a similar data description. However, as expected of long and short positions, the tail-based EV parameters are significantly different—a sharp contrast to Lévy parameters, which use the entire data set. As Lévy models on the complete data of short and long positions flip the densities along the y-axis, long and short positions alter the sign of the parameter that characterizes the skewness in the model. Thus, it is sufficient to report the estimates corresponding to long positions alone for which the risk measures VaR and ES under tail targeting and full density-based models will be investigated.

**Table 3** Maximum likelihood estimates for Generalized Hyperbolic Lévy models

Index	u	Prob	$\sigma$	$\theta$	$\nu$	
<i>Panel A: Variance Gamma (VG)</i>						
S&P500	2	0.021	1.109 (0.030)	-0.027 (0.039)	0.734 (0.025)	
FTSE100	1.5	0.074	1.064 (0.049)	-0.052 (0.040)	0.674 (0.008)	
DAX	2	0.056	1.305 (0.046)	-0.012 (0.006)	0.536 (0.013)	
Hang Seng	2	0.093	1.705 (0.024)	-0.066 (0.054)	0.768 (0.023)	
Nikkei 225	2	0.078	1.329 (0.014)	0.016 (0.067)	0.428 (0.088)	
	u	Prob	$\alpha$	$\beta$	$\delta$	
<i>Panel B: Normal Inverse Gaussian (NIG)</i>						
S&P500	2	0.025	0.821 (0.044)	-0.010 (0.056)	0.901 (0.071)	
FTSE100	1.5	0.076	0.971 (0.062)	-0.036 (0.025)	1.147 (0.064)	
DAX	2	0.061	0.437 (0.051)	-0.054 (0.061)	1.154 (0.053)	
Hang Seng	2	0.079	0.538 (0.071)	-0.022 (0.019)	1.708 (0.091)	
Nikkei 225	2	0.069	0.837 (0.024)	0.062 (0.042)	1.937 (0.232)	
<i>Panel C: Hyperbolic (HYP)</i>						
S&P500	2	0.029	1.028 (0.065)	-0.022 (0.032)	0.704 (0.009)	
FTSE100	1.5	0.081	0.978 (0.043)	-0.039 (0.021)	0.731 (0.157)	
DAX	2	0.059	1.095 (0.039)	-0.027 (0.075)	0.602 (0.107)	
Hang Seng	2	0.098	0.507 (0.098)	-0.051 (0.014)	0.989 (0.018)	
Nikkei 225	2	0.063	1.047 (0.025)	0.039 (0.056)	1.036 (0.062)	
<b>Panel A: Generalized Hyperbolic (GH)</b>						
	u	Prob	$\alpha$	$\beta$	$\delta$	$\mu$
S&P500	2	0.028	0.896 (0.082)	-0.041 (0.031)	0.821 (0.147)	0.013 (0.087)
FTSE100	1.5	0.069	0.768 (0.127)	-0.038 (0.057)	1.504 (0.516)	-0.096 (0.128)
DAX	2	0.093	0.907 (0.091)	-0.058 (0.072)	1.196 (0.236)	-0.017 (0.018)
Hang Seng	2	0.095	0.428 (0.105)	-0.047 (0.013)	1.464 (0.539)	-0.053 (0.602)
Nikkei 225	2	0.074	1.101 (0.263)	0.009 (0.049)	1.926 (0.618)	0.076 (0.721)

Based on daily % returns for long futures positions on daily % returns of futures contracts, from January 1, 2007, to December 31, 2017. We report the estimates of the VG, NIG, HYP, and GH parameters. Prob is the probability of observation in excess of u (the same thresholds selected for the EV model in Table 2). In parenthesis, estimated standard errors of the parameters are reported.

**Estimation of risk measures**

Except for a few specific cases, VaR is obtained by solving the following quantile-integral equation:

$$\int_{x_{\min}}^{VaR} f(u)du - \alpha = 0 \tag{10}$$

where  $\alpha$  is the coverage level. The VaR for different Lévy models can be obtained by solving Eq. (10) with corresponding Lévy density.

The major problem with VaR is that it indicates the magnitude of loss to a certain level but ignores the magnitude of losses that exceed the pre-fixed confidence level. Thus, VaR identifies the tail to a given level but has no answer regarding how concerting that tail is with respect to the pre-fixed confidence level. In addition to identifying the tail to a given pre-fixed level, ES as a measure provides the average of losses belonging to the identified tail contingent on the pre-fixed level.

As in Eq. (9), the high  $\alpha$ th quantile (i.e., VaR at a very high confidence level  $\alpha$ ) is given as follows:

$$VaR(\alpha) = u + \frac{\beta}{\xi} \left\{ \left( \frac{n}{N_u} \alpha \right)^{-\xi} - 1 \right\} \tag{11}$$

and the ES of the same confidence level of  $\alpha$  is as follows:

$$ES^{gp}(\alpha) = \frac{VaR(\alpha)}{1 - \xi} + \frac{\beta - \xi u}{1 - \xi} \tag{12}$$

where  $\beta$  and  $\xi$  are scale and shape parameters contingent on threshold  $u$ , respectively.

In Eq. (11),  $n$  is the total number of observations, and  $N_u$  is the number of observations that exceed threshold  $u$ . The ES yields from the fundamental equation are as follows:

$$ES(\alpha) = \frac{1}{1 - \alpha} \int_{\alpha}^1 VaR(u) du \tag{13}$$

$$= \frac{1}{1 - \alpha} \int_{\alpha}^1 \left\{ S \left| \int_{x_{\min}}^S f(x) dx = u \right. \right\} du \tag{14}$$

For a VG model, ES is then obtained with VG density:

$$ES^{vg}(\alpha) = \frac{1}{1 - \alpha} \left\{ \int_{\alpha}^1 S \left| \int_{x_{\min}}^S \left[ \frac{2e^{-\frac{\theta x}{\sigma^2}}}{\sigma \sqrt{2\pi} v^{\frac{1}{v}} \Gamma(\frac{1}{v})} \left( \frac{|x|}{\sqrt{\frac{2\sigma^2}{v} + \theta^2}} \right)^{\frac{1}{v} - \frac{1}{2}} K_{\frac{1}{v} - \frac{1}{2}} \left( \frac{|x| \sqrt{\frac{2\sigma^2}{v} + \theta^2}}{\sigma^2} \right) \right] dx = u \right. \right\} du \tag{15}$$

The approach is similar to obtaining the ES for other Lévy models but incorporates different densities in Eq. (14). We apply the parametric bootstrap to get the SEs and confidence intervals (CIs) of risk measures, following Cotter and Dowd (2006). However, as Lévy models have no closed-form expressions for risk measures, it is a computational challenge that we overcome using a machine with a powerful configuration.

**GOF: EV versus Lévy**

Among various GOF tests, one that is particularly suitable for tail-based risk management studies<sup>6</sup> is the Anderson–Darling (AD) test. It is about a weighing rule introduced by Anderson and Darling (1952, 1954) in the Kolmogorov–Smirnov test that emphasizes the observations in the tail. Anna et al. (2005) provided a formula for an AD test statistic when the distribution of the complete sample is unknown and observations are only available at the extreme tail, referred to as left-truncated data adapted AD test. This adaptation fits the test of the EV model. For the AD test version applied with complete distributions such as our Lévy models having closed-form densities, the  $p$ -values are analytically available. However, for the AD test adapted to left truncated data,  $p$ -values need to be calculated through bootstrapping or Monte Carlo simulation. In this study, we consider 1,000 resampling and calculate the  $p$ -value for the EV model using bootstrap. We use VaR for our Lévy models and VaR for the EV model as critical values for

<sup>6</sup> Other popular GOF tests such as a Chi-square test are not comparable for Lévy models on complete data and left truncated version for EV model with incomplete data.

the tests (the left truncated version of the test of VaR for the EV model remains consistent as it is computed from a left-truncated density):

$$AD^2 = -N - \sum_{i=1}^N \frac{(2i-1)}{N} [\log(F(x_i)) + \log(1 - F(x_{N+1-i}))] \tag{16}$$

$$AD_{ev}^2 = -n + 2n \log(1 - F(u)) - \frac{1}{n} \sum_{i=1}^n (1 + 2(n-i)) \log(1 - F(x_j)) + \dots \tag{17}$$

$$+ \frac{1}{n} \sum_{i=1}^n (1 - 2i) \log(F(x_j) - F(u))$$

where  $u$  is the truncation level;  $x_j$  is the  $j$ th observed value of the order statistic  $X_1 \leq X_2 \leq \dots \leq X_n$ ; and  $n$  represents observations in the tail (total number).

Table 4 presents the statistics for applying both the GOF tests (AD and its left truncated version  $AD_{ev}$ ). By the very nature of the tests, AD and  $AD_{ev}$  are differently informative about the tail fits. Based on Table 4, EV and full density Lévy models perform statistically almost similarly on the tail. However, as AD and  $AD_{ev}$  values exhibit

**Table 4** Anderson Darling (Lévy) and left truncated Anderson Darling (EV) Goodness-of-fit

Index	Model	AD-stat	1%CV	5%CV	10%CV	p-value
S&P500	EV	0.241	3.343*	2.008*	1.501*	0.826
	VG	1.715	3.877*	1.818*	1.637	0.050
	NIG	1.521	3.141*	1.777*	1.311	0.071
	HYP	1.458	3.449*	1.638*	1.124	0.072
	GH	1.419	2.898*	1.710*	1.102	0.066
FTSE100	EV	0.691	3.416*	2.912*	1.421*	0.727
	VG	0.651	3.211*	1.906*	1.202*	0.202
	NIG	0.410	4.153*	1.920*	1.410*	0.481
	HYP	0.408	3.003*	1.883*	1.400*	0.336
	GH	0.313	3.233*	1.847*	1.382*	0.400
DAC	EV	1.124	3.997*	3.043*	1.677*	0.710
	VG	1.503	4.211*	2.732*	1.651*	0.142
	NIG	1.310	3.639*	2.421*	1.597*	0.163
	HYP	1.151	3.909*	2.466*	1.715*	0.172
	GH	1.147	3.333*	2.411*	1.774*	0.255
Hang Seng	EV	1.417	5.462*	3.265*	2.118*	0.718
	VG	1.008	4.938*	3.119*	2.061*	0.300
	NIG	0.522	4.697*	2.528*	2.199*	0.381
	HYP	0.625	5.110*	3.121*	2.152*	0.449
	GH	0.505	5.033*	3.077*	2.136*	0.371
Nikkei225	EV	0.819	4.000*	3.011*	1.772*	0.793
	VG	0.825	4.189*	2.637*	1.835*	0.473
	NIG	0.723	4.167*	2.501*	1.810*	0.341
	HYP	0.633	3.993*	2.477*	1.730*	0.303
	GH	0.729	4.033*	2.710*	1.689*	0.443

In the case of a left-truncated Anderson Darling test, 1000 resampling is considered to obtain the  $p$ -values by bootstrapping. (\*) denotes that models survive the test to the given significance level. AD-stat for the EV model is obtained with  $AD_{ev}$  of Eq. (17). There is no clear preference between the approaches (EV vs. Lévy)

observations outside the tails, they influence the test results even when the weights attached to such observations are much less than those attached to the tail.

### Comparison of risk measures

This section analyzes the Lévy and EV estimates of VaR and ES. The estimates of VaR are based on GPD and the four Lévy models. We report the parameters of each model calibrated under both approaches in Tables 5, 6, 7, 8 and 9. As VaR and ES are based on higher coverage levels, they account for trading losses at a very high level due to extreme (unexpected) events. Surprisingly, VaR estimates across the empirical values are approximately of the same order of magnitude for all indices, but that is not true for ES estimates. For the ES model, the estimates depend on the entire tail shape of the model but not only on a specific quantile of the tail. However, between the approaches, it is difficult to claim with certainty that, based on the estimates of VaR and ES, any particular approach is better than another. The EV model occasionally provides VaR and/or ES estimates that deviate less from their empirical counterparts. However, Lévy–VaR and/or Lévy–ES estimates have less deviation from their empirical counterparts on different occasions.

Looking into the precision of VaR estimations may help us ascertain some preference between the approaches. Overall, the SEs of the Lévy model VaR estimates are much

**Table 5** Estimates of VaR and ES risk measures for S&P500 futures position: EV versus Lévy

Model	Risk Measure	$\alpha = 0.99$	$\alpha = 0.995$	$\alpha = 0.999$
Empirical	VaR	2.953	3.435	5.647
	ES	4.415	4.529	6.824
GP	VaR	2.918 (0.127) [0.935–1.075]	3.489 (0.204) [0.908–1.089]	5.125 (0.659) [0.797–1.189]
	ES	4.012 (0.173) [0.981–1.124]	4.411 (0.252) [0.931–1.104]	<b>6.610 (0.911)</b> <b>[0.825–1.161]</b>
VG	VaR	2.889 (0.121) [0.935–1.072]	3.384 (0.165) [0.926–1.094]	4.528 (0.374) [0.870–1.145]
	ES	3.736 (0.137) [0.951–1.129]	4.153 (0.171) [0.951–1.183]	5.527 (0.304) [0.915–1.310]
NIG	VaR	3.001 (0.146) [0.925–1.085]	3.617 (0.203) [0.883–1.086]	<i>5.139 (0.547)</i> <i>[0.840–1.180]</i>
	ES	<b>4.003 (0.162)</b> <b>[0.951–1.133]</b>	<b>4.533 (0.220)</b> <b>[0.941–1.177]</b>	6.099 (0.414) [0.910–1.091]
HYP	VaR	2.902 (0.123) [0.934–1.073]	<i>3.409 (0.164)</i> <i>[0.898–1.073]</i>	4.587 (0.371) [0.889–1.172]
	ES	3.922 (0.133) [0.991–1.120]	4.225 (0.160) [0.961–1.166]	6.111 (0.401) [0.924–1.107]
GH	VaR	<i>2.989 (0.139)</i> <i>[0.928–1.081]</i>	3.569 (0.189) [0.888–1.081]	4.967 (0.454) [0.869–1.179]
	ES	3.903 (0.140) [0.951–1.006]	4.344 (0.201) [0.942–1.112]	5.490 (0.402) [0.972–1.210]

The estimates are based on the parameter values in Tables 2 and 3 using daily % return. Here,  $\alpha$  is the coverage level of VaR and ES when estimated under the EV and Lévy approaches, corresponding to a holding period of 1 day. Next to each estimate, SEs are reported, and the 90% confidence intervals are immediately below (normalized by bootstrapped estimates). The adequate ES estimates are depicted in bold, and the adequate VaR are shown in bold and Italics. Again there is no clear pattern of preference between the approaches (EV vs. Lévy)

**Table 6** Estimates of VaR and ES risk measures for FTSE100 futures position: EV versus Lévy

Model	Risk measure	$\alpha = 0.99$	$\alpha = 0.995$	$\alpha = 0.999$
Empirical	VaR	3.306	3.739	5.118
	ES	4.141	4.614	5.812
GP	VaR	3.058 (0.147) [0.924–1.086]	3.674 (0.243) [0.908–1.119]	5.273 (0.556) [0.837–1.190]
	ES	<b>3.883 (0.174)</b> <b>[0.951–1.191]</b>	4.704 (0.241) [0.941–1.200]	5.923 (0.681) [0.901–1.200]
VG	VaR	2.969 (0.111) [0.942–1.064]	3.424 (0.145) [0.938–1.078]	4.453 (0.324) [0.882–1.129]
	ES	3.702 (0.120) [0.936–1.080]	4.400 (0.155) [0.951–1.079]	5.101 (0.331) [0.901–1.077]
NIG	VaR	3.026 (0.124) [0.937–1.071]	3.542 (0.168) [0.931–1.088]	4.772 (0.452) [0.866–1.191]
	ES	3.822 (0.138) [0.953–1.200]	4.412 (0.160) [0.920–1.099]	<b>5.492 (0.422)</b> <b>[0.914–1.121]</b>
HYP	VaR	2.993 (0.116) [0.939–1.067]	3.470 (0.153) [0.935–1.082]	4.569 (0.300) [0.887–1.123]
	ES	3.690 (0.120) [0.943–1.105]	4.225 (0.140) [0.945–1.128]	5.311 (0.336) [0.914–1.134]
GH	VaR	3.027 (0.131) [0.933–1.075]	3.579 (0.184) [0.926–1.096]	4.980 (0.412) [0.863–1.156]
	ES	3.809 (0.155) [0.951–1.110]	<b>4.520 (0.200)</b> <b>[0.913–1.181]</b>	5.928 (0.545) [0.891–1.206]

The estimates are based on the parameter values in Tables 2 and 3 using daily % return. Here,  $\alpha$  is the coverage level of VaR and ES when estimated under the EV and Lévy approaches, corresponding to a holding period of 1 day. Next to each estimate, SEs are reported, and the 90% confidence intervals are immediately below (normalized by bootstrapped estimates). The most adequate ES estimate is depicted in bold, and the most adequate VaR is shown in bold and italics. Again there is no clear pattern of preference between the approaches (EV vs. Lévy)

lower than those of the EV model VaR estimates. The rise in coverage levels reinforces this observation. The coefficient of variations (estimated risk measure value divided by corresponding SE) helps us to double-check this observation. Thus, the VaR estimates based on the Lévy models are more stable than those based on the EV model. However, this may be partly due to a few observations under the EV model.

Tables 5, 6, 7, 8 and 9 report a 90% CI for VaR estimates obtained with bootstrapped estimates. We find that at low coverage levels, the estimated CIs for both EV and Lévy models are symmetric. However, at higher coverage levels, confidence levels are asymmetric, with the upper bound moving further away from the mean of the estimates (bootstrapped). In contrast to the GP, it is difficult to establish clear-cut results for the four Lévy models through a comparison of the SEs of VaR estimates. Unlike GP, often at ultra-high coverage (0.999), the CIs exhibit ultra-spread, which indicates unstable forecasts at ultra-high coverage, presumably because the estimations are based on a few extreme observations above the threshold<sup>7</sup> alone.

Overall, the ES CIs are narrower than those of VaR, indicating that ES measures are more precisely estimated than VaRs. VaR and ES bootstrapped statistics (SE and CI) are informative regarding some differences and similarities between Lévy and EV approaches. VaR and ES bootstrapped statistics are narrower for Lévy models than for EV models, which indicates that the estimation with a Lévy approach is more stable than

<sup>7</sup> We also conducted VaR and ES estimation for sample periods with or without the 2008 Global Financial Crisis period. The results do not change significantly. For brevity, we do not report the results here.

**Table 7** Estimates of VaR and ES risk measures for DAX futures position: EV versus Lévy

Model	Risk measure	$\alpha = 0.99$	$\alpha = 0.995$	$\alpha = 0.999$
Empirical	VaR	4.397	4.940	6.594
	ES	6.091	6.511	9.910
GP	VaR	4.330 (0.195) [0.926–1.074]	5.178 (0.326) [0.914–1.115]	4.330 (0.195) [0.860–1.151]
	ES	<b>5.822 (0.184)</b> <b>[0.950–1.110]</b>	<b>6.300 (0.253)</b> <b>[0.931–1.139]</b>	<b>9.032 (0.651)</b> <b>[0.890–1.206]</b>
VG	VaR	3.971 (0.165) [0.919–1.063]	4.634 (0.257) [0.925–1.102]	6.157 (0.556) [0.872–1.165]
	ES	5.043 (0.177) [0.951–1.166]	5.911 (0.192) [0.940–1.009]	8.128 (0.420) [0.922–1.113]
NIG	VaR	4.119 (0.206) [0.931–1.079]	4.932 (0.282) [0.909–1.085]	6.928 (0.709) [0.828–1.149]
	ES	5.512 (0.232) [0.952–1.104]	6.410 (0.248) [0.921–1.124]	8.401 (0.553) [0.918–1.203]
HYP	VaR	4.003 (0.171) [0.917–1.065]	4.690 (0.240) [0.915–1.085]	6.284 (0.548) [0.851–1.126]
	ES	5.029 (0.220) [0.952–1.137]	5.922 (0.240) [0.920–1.120]	7.729 (0.428) [0.925–1.202]
GH	VaR	4.115 (0.203) [0.932–1.078]	4.916 (0.277) [0.910–1.084]	6.865 (0.689) [0.831–1.146]
	ES	5.442 (0.228) [0.954–1.123]	6.200 (0.248) [0.914–1.150]	8.200 (0.472) [0.913–1.202]

The estimates are based on the parameter values in Tables 2 and 3 using daily % return. Here,  $\alpha$  is the coverage level of VaR and ES when estimated under the EV and Lévy approaches, corresponding to a holding period of 1 day. Next to each estimate, SEs are reported, and the 90% confidence intervals are immediately below (normalized by bootstrapped estimates). The most adequate ES estimate is depicted in bold, and the most adequate VaR is shown in bold and italics. Again there is no clear pattern of preference between the approaches (EV vs. Lévy)

**Table 8** Estimates of VaR and ES risk measures for Hang Seng futures position: EV versus Lévy

Model	Risk measure	$\alpha = 0.99$	$\alpha = 0.995$	$\alpha = 0.999$
Empirical	VaR	5.269	6.273	9.169
	ES	7.324	7.939	11.548
GP	VaR	5.240 (0.269) [0.912–1.099]	6.385 (0.367) [0.907–1.092]	9.494 (1.083) [0.817–1.198]
	ES	<b>7.188 (0.303)</b> <b>[0.953–1.094]</b>	<b>7.389 (0.463)</b> <b>[0.929–1.201]</b>	<b>11.622 (1.131)</b> <b>[0.908–1.127]</b>
VG	VaR	5.057 (0.209) [0.936–1.071]	5.917 (0.278) [0.878–1.129]	7.901 (0.642) [0.878–1.129]
	ES	7.003 (0.211) [0.946–1.103]	7.204 (0.312) [0.922–1.135]	10.250 (0.707) [0.902–1.105]
NIG	VaR	5.288 (0.258) [0.925–1.084]	6.372 (0.356) [0.883–1.085]	9.052 (0.901) [0.853–1.159]
	ES	7.100 (0.235) [0.961–1.126]	8.135 (0.371) [0.916–1.139]	11.005 (0.812) [0.861–1.091]
HYP	VaR	5.084 (0.215) [0.935–1.073]	5.968 (0.286) [0.899–1.073]	8.019 (0.666) [0.876–1.132]
	ES	6.626 (0.236) [0.955–1.071]	7.444 (0.301) [0.934–1.114]	9.396 (0.670) [0.914–1.202]
GH	VaR	5.285 (0.266) [0.923–1.087]	6.413 (0.374) [0.878–1.089]	9.281 (0.980) [0.845–1.169]
	ES	7.093 (0.320) [0.942–1.150]	8.306 (0.440) [0.911–1.118]	11.401 (1.078) [0.852–1.168]

The estimates are based on the parameter values in Table 2 and 3 using daily % return. Here,  $\alpha$  is the coverage level of VaR and ES when estimated under the EV and Lévy approaches, corresponding to a holding period of 1 day. Next to each estimate, SEs are reported, and the 90% confidence intervals are immediately below (normalized by bootstrapped estimates). The most adequate ES estimate is depicted in bold, and the most adequate VaR is shown in bold and italics. Again there is no clear pattern of preference between the approaches (EV vs. Lévy)



**Table 9** Estimates of VaR and ES risk measures for Nikkei225 futures position: EV versus Lévy

Model	Risk measure	$\alpha = 0.99$	$\alpha = 0.995$	$\alpha = 0.999$
Empirical	VaR	3.758	4.770	5.891
	ES	5.102	5.610	6.735
GP	VaR	<i>3.846 (0.143)</i> [0.945–1.061]	4.447 (0.223) [0.927–1.090]	5.821 (0.472) [0.882–1.134]
	ES	4.815 (0.171) [0.911–1.115]	5.310 (0.224) [0.944–1.104]	6.328 (0.501) [0.913–1.145]
VG	VaR	3.946 (0.158) [0.942–1.063]	4.526 (0.189) [0.937–1.069]	<i>5.833 (0.382)</i> [0.887–1.108]
	ES	4.814 (0.151) [0.971–1.105]	5.615 (0.205) [0.941–1.079]	<b>6.496 (0.310)</b> <b>[0.904–1.108]</b>
NIG	VaR	3.996 (0.163) [0.928–1.064]	<i>4.636 (0.206)</i> [0.936–1.083]	6.142 (0.541) [0.879–1.159]
	ES	4.699 (0.169) [0.962–1.183]	5.444 (0.225) [0.942–1.094]	7.138 (0.478) [0.890–1.107]
HYP	VaR	3.980 (0.148) [0.931–1.057]	4.592 (0.180) [0.934–1.059]	5.995 (0.434) [0.877–1.135]
	ES	<b>4.920 (0.153)</b> <b>[0.944–1.113]</b>	5.590 (0.201) [0.953–1.110]	6.703 (0.510) [0.879–1.109]
GH	VaR	3.993 (0.160) [0.928–1.063]	4.622 (0.201) [0.937–1.081]	6.083 (0.520) [0.882–1.155]
	ES	4.811 (0.202) [0.943–1.104]	<b>5.611 (0.220)</b> <b>[0.934–1.121]</b>	6.809 (0.520) [0.909–1.092]

*Note:* The estimates are based on the parameter values in Tables 2 and 3 using daily % return. Here,  $\alpha$  is the coverage level of VaR and ES when estimated under the EV and Lévy approaches, corresponding to a holding period of 1 day. Next to each estimate, SEs are reported, and the 90% confidence intervals are immediately below (normalized by bootstrapped estimates). The most adequate ES estimate is depicted in bold, and the most adequate VaR is shown in bold and italics. Again there is no clear pattern of preference between the approaches (EV vs. Lévy)

that with an EV approach. Moreover, the estimation performance deteriorates under both approaches with increased coverage. For NIG and GH models, estimation instability seems to propagate much faster, especially at higher coverage levels.

### VaR and ES backtesting

We conduct dynamic calibration on a rolling window for backtesting. As daily  $\text{VaR}(\alpha)$  is estimated on daily return, the loss for the one-day holding of an asset can only be violated for 100%  $\alpha$  of the time, allowing all possible extremities. We use an indicator variable describing the hit sequence that identifies the day of VaR violation in the following  $T$  trading days. The hit sequence is Bernoulli distributed with a probability  $\alpha$  of assuming 1. We implement three VaR tests—unconditional, independence, and conditional coverage. We use two tests without distributional assumption for ES backtests—unconditional-normal and unconditional  $t$ -test.

#### VaR backtesting

##### Unconditional test

The unconditional hypothesis of backtesting does not hold any assumption regarding today's violation status when it provides statistical evidence as to whether the observed proportion of violation (PV) of a VaR model tomorrow is significantly different from its

promised fraction  $\alpha$ . However, the evidence is provided through an asymptotical test statistic following  $\chi^2$  with one degree of freedom:

$$LR_{uc} = -2\log \left[ \frac{(1 - \alpha)^{T_0} \alpha^{T_1}}{\left(1 - \frac{T_1}{T}\right)^{T_0} \left(\frac{T_1}{T}\right)^{T_1}} \right] \sim \chi^2(1) \tag{18}$$

Here,  $T = T_1 + T_0$  is assumed to be significantly large, and  $T_0$  and  $T_1$  are the number of days with no violation and days with a violation, respectively. We use the Monte Carlo simulation to calculate the  $p$ -value. We compute the Monte Carlo  $p$ -value to simulate 999 test values,  $\{LR(i)\}_{i=1}^{999}$ , each of which is based on a Bernoulli ( $\alpha$ ) sample of hit sequences having the same size as the original sample:

$$P^{mc} = \frac{1}{1000} \left\{ 1 + \sum_{i=1}^{999} I_{[LR(i)] > LR_{uc}} \right\} \tag{19}$$

The simulated  $p$ -value is the proportion of simulated test values given that a simulated test value is more significant than the test value to roughly the number of simulations.

**Independence test**

The independence test checks whether VaR violations are truly random and not clustered over time. As assets with volatility clustering yield VaR exhibiting clustering, we can predict that if there is a violation today, then we will most likely find a violation tomorrow, which is more than 100%  $\alpha$  likely. VaR adjusts to the predictions of high volatility as useful information, ensuring the risk model is correctly specified, and the violation of VaR remains unpredictable. The test statistic of the independence test is given as follows (Christoffersen 2003):

$$LR_{ind} = -2\log \left[ \frac{L(p)}{L(\hat{\Pi}_1)} \right] \sim \chi^2 \tag{20}$$

where the matrix of transitional probabilities of conditional violations is given as follows:

$$\hat{\Pi}_1 = \begin{bmatrix} \frac{T_{00}}{T_{00}+T_{01}} & \frac{T_{01}}{T_{00}+T_{01}} \\ \frac{T_{10}}{T_{10}+T_{11}} & \frac{T_{11}}{T_{10}+T_{11}} \end{bmatrix} \tag{21}$$

Thus, we can write

$$L(\hat{\Pi}_1) = \left( \frac{T_{00}}{T_{00} + T_{01}} \right)^{T_{00}} \left( \frac{T_{01}}{T_{00} + T_{01}} \right)^{T_{01}} \left( \frac{T_{10}}{T_{10} + T_{11}} \right)^{T_{10}} \left( \frac{T_{11}}{T_{10} + T_{11}} \right)^{T_{11}} \tag{22}$$

where  $p$  characterizes the matrix of transitional probabilities of violation, ensuring no dependence between 0 and 1 in the hit sequence:

$$\hat{\Pi} = \begin{bmatrix} 1 - p & p \\ 1 - p & p \end{bmatrix} = \begin{bmatrix} \frac{T_0}{T_0+T_1} & \frac{T_1}{T_0+T_1} \\ \frac{T_0}{T_0+T_1} & \frac{T_1}{T_0+T_1} \end{bmatrix} \tag{23}$$

$L(p)$  is similar to the unconditional hypothesis.  $LR_{ind}$  provides the statistical significance of the likelihood of independence in the hit sequence over the likelihood of dependence.

**Conditional coverage test**

The conditional coverage test checks whether the average number of violations changes with the level of a risk model. The conditional coverage test statistic has a similar expression as the independence test statistic with  $p = \frac{T_1}{T}$  of the independence statistic replaced by the coverage level  $\alpha$  of the risk model (Christoffersen 2003; Dowd 2005):

$$LR_{cc} = -2\log \left[ \frac{LR(\alpha)}{LR(\hat{\Pi}_1)} \right] = LR_{uc} + LR_{ind} \sim \chi^2 \tag{24}$$

**ES backtesting**

The unconditional coverage test statistic proposed by Acerbi et al. (2014) for ES backtesting is as follows:

$$Z_{unconditional} = \frac{1}{N(1 - \alpha)} \sum_{t=1}^N \frac{X_t I_t}{ES_{\alpha,t}} + 1 \tag{25}$$

where  $X_t$  represents profit and loss distribution along a real but unknowable distribution and is forecasted by a model predictive distribution conditional to previous information used to compute ES, and  $I_t$  is an indicator function, which is equal to 1 when the forecasted VaR is violated, that is,  $X_t < -VaR_{\alpha,t}$  and 0 otherwise.

We use two tests without distributional assumption—unconditional-normal test and unconditional- $t$  test (see Acerbi et al. 2014; Acerbi et al. 2017). The unconditional-normal test assumes that  $X_t$  follows a standard normal distribution, whereas the unconditional- $t$  test assumes that  $X_t$  follows a  $t$ -distribution. The unconditional test statistic is sensitive to the severity of the VaR failures relative to the ES estimate and the frequency of VaR failures. As a result, a rare but colossal VaR failure(s) relative to the ES may result in the rejection of a model over a particular timeframe.

However, when the ES estimate is large on a violation day, it may not impact the test results as much as a large loss would have if a smaller ES were encountered. Similarly, a model can be rejected due to many VaR failures even if all VaR violations are just slightly higher than the VaR as such failure contributes to making the test statistics negative. Thus, glimpses of asymmetry in several VaR violations among different models and asymmetry in expected and observed severity (severity ratio) are critical for ES backtesting.

**Backtesting results**

We now examine the sensitivity of the risk measure VaR to new observations for dynamic calibration on a rolling window of four business years with a two-year look-back window and continue to increase the window length. To avoid the problem with EV dynamic calibration that considers only extreme observations, we increase the proportion of

extreme observations by adjusting the threshold and expanding the length of the look-back window. We consider the extreme 30% observations in calibration dynamically for EV and the coverage levels of 95% and 99%.

First, we calibrate all the models on the time series of returns for 2007–2010 on December 31, 2010 and use the calibrated parameters to predict the VaR and ES for January 1, 2011. This gives us an additional new observation of returns on January 1, 1995. We remove the oldest observation to accommodate this new observation in our fixed length look-back window and then calibrate the models in a new window to predict the VaR and ES on January 2, 2011. The process continues until the end of 2003. Thus, the dynamic calibration starts on January 1, 2011 and ends on December 31, 2017. The unconditional, independence, and conditional coverage hypotheses are tested with 95% VaRs. The backtesting checks whether unconditional and conditional distributions influence conditional and unconditional coverage hypotheses tests. The backtesting results for long positions in all indices and the PVs are presented in Table 10.<sup>8</sup>

Table 10 reveals that the EV–VaR is not distinguishable from that of the full density-based Lévy–VaR given the observed PV performance of backtesting. The PVs corresponding to an EV model are closer to the promised fraction of violations for all indices, except for Nikkei225. The PVs corresponding to EV models for Nikkei225 deviate more from the promised fraction of violations than those for Lévy models. Thus, the tail-based risk measure of VaR obtained for the tail-based model of EV and full density-based Lévy models are almost similar.

For the remaining indices, the results from the hypotheses testing are mixed. As the VaR violations are clustered at 95%, the independence test fails in most cases. However, the test passes the 99% coverage. On the other hand, the conditional coverage hypothesis is supported in most cases, although both unconditional and independence hypotheses are not supported. A significant deviation of the observed PV from the promised PV for unconditional coverage may have contributed to the rejection of the conditional coverage hypothesis. We also report the Chi-square and the Monte Carlo simulated  $p$ -values, which test the effectiveness of the test statistics. We find that both the Chi-square and  $p$ -values are close to each other, implying that our tests are relevant.

The last two columns in each table report the results of the ES backtesting. We report two ES backtesting results without distributional assumption—the unconditional-normal test (unconditional- $N$ ) and the unconditional- $t$  test (unconditional- $t$ ). The  $p$ -values of the tests, which represent the success rate when multiplied by 100, are reported in parentheses. We identify each test as a “pass” (P) or a “fail” (F) based on the  $p$ -values in the table. All tests are conducted at a 95% confidence level. None of the unconditional- $N$  tests passed. However, the unconditional- $t$  tests passed in only a few instances. Thus, our results do not suggest any preference for the EV or Lévy models through ES backtesting.

## Discussion

We have investigated four full-density Lévy models and estimated the tail-focused risk measure VaR and its coherent version ES, in addition to estimating VaR and ES for an EV model (a tail targeting approach). The parameters calibrated for all five models under

<sup>8</sup> The unconditional, independence, and conditional coverage hypotheses are also tested with 99% VaRs. The results are similar to the 95% VaR.

**Table 10** Backtesting results for conditional and unconditional models (95% Coverage)

Index	Model	VaR backtest										ES backtest											
		Unconditional coverage (95%)					Independence test (95%)					Conditional coverage (95%)				Unconditional-N		Unconditional-t					
		Stat	1% CV	5% CV	P	MC	Stat	1% CV	5% CV	P	MC	PV	P	MC	Stat	1% CV	5% CV	P	MC	PV	Pass (P) / Fail (F) (p-values)	Pass (P) / Fail (F) (p-values)	
S&P 500	GP	5.31	5.22	3.48	0.00	1e-3	6.88	1.03	5.22	3.48	0.72	1.01	8.15	9.35	8.35	5.34	2e-4	1e-4	6.88	F (0.00)	F (0.00)	F (0.00)	F (0.00)
	VG	24.26	5.22	3.48	0.00	1e-3	6.29	20.05	5.22	3.48	1e-6	4.30	21.51	7.54	5.34	0.25	0.29	6.29	F (0.00)	F (0.00)	F (0.00)	F (0.00)	
	NIG	20.02	5.22	3.48	0.00	1e-3	6.45	2.06	5.22	3.48	0.26	0.31	4.48	20.65	8.24	5.34	2e-7	1e-4	6.45	F (0.00)	F (0.00)	F (0.00)	F (0.00)
	HYP	23.24	5.22	3.48	0.00	1e-3	7.84	1.86	5.22	3.48	0.30	0.55	6.21	29.02	8.35	5.34	2e-7	2e-3	7.84	F (0.00)	F (0.00)	F (0.00)	F (0.00)
	GH	30.06	5.22	3.48	0.00	1e-3	7.22	2.35	5.22	3.48	0.21	1.03	7.24	30.13	8.22	5.34	4e-4	2e-4	7.22	F (0.00)	F (0.00)	F (0.00)	F (0.00)
TSE 100	GP	5.11	5.22	3.48	0.00	0.02	6.42	9.21	5.22	3.48	0.00	0.00	6.15	12.24	8.35	5.34	0.02	3e-4	6.42	F (0.00)	P (0.050)	F (0.00)	P (0.050)
	VG	4.69	5.22	3.48	0.02	0.01	5.59	8.66	5.22	3.48	0.00	0.00	5.20	10.65	8.35	5.34	4e-4	4e-4	5.59	F (0.00)	F (0.00)	F (0.00)	F (0.00)
	NIG	6.32	5.22	3.48	0.01	0.02	5.37	9.53	5.22	3.48	0.00	0.00	5.34	10.24	8.35	5.34	4e-3	3e-4	5.37	F (0.00)	F (0.00)	F (0.00)	F (0.00)
	HYP	5.05	5.22	3.48	0.00	0.00	6.20	8.36	5.22	3.48	0.00	0.02	6.18	9.69	8.12	5.34	4e-3	2e-3	6.20	F (0.00)	F (0.00)	F (0.00)	F (0.00)
	GH	4.88	5.22	3.48	0.01	0.00	6.49	8.47	5.22	3.48	0.00	0.02	6.42	10.88	8.35	5.34	4e-4	3e-4	6.49	F (0.00)	F (0.00)	F (0.00)	F (0.00)
DAX	GP	20.15	5.22	3.48	4e-4	3e-3	5.33	14.51	5.22	3.48	3e-5	2e-3	6.14	34.15	8.35	5.34	3e-7	2e-3	5.33	F (0.00)	F (0.00)	F (0.00)	F (0.00)
	VG	20.47	5.22	3.48	3e-5	2e-3	6.48	15.47	5.22	3.48	3e-5	2e-3	6.47	30.49	8.35	5.34	4e-8	2e-4	6.48	F (0.00)	F (0.00)	F (0.00)	F (0.00)
	NIG	19.36	5.22	3.48	2e-6	1e-3	7.00	15.63	5.22	3.48	5e-5	3e-3	7.08	38.19	8.35	5.34	4e-7	4e-3	7.00	F (0.00)	F (0.00)	F (0.00)	F (0.00)
	HYP	24.01	5.22	3.48	2e-4	1e-3	6.91	15.38	5.22	3.48	5e-5	3e-3	6.09	35.57	7.89	5.34	4e-6	2e-3	6.91	F (0.00)	F (0.00)	F (0.00)	F (0.00)
	GH	22.98	5.22	3.48	5e-3	3e-3	6.24	16.14	5.22	3.48	8e-5	4e-3	7.44	34.05	8.35	5.34	3e-9	3e-4	6.24	F (0.00)	F (0.00)	F (0.00)	F (0.00)

**Table 10** (continued)

Index	Model	VaR backtest										ES backtest																						
		Unconditional coverage (95%)					Independence test (95%)					Conditional coverage (95%)					Unconditional-N					Unconditional-t												
		Stat	1% CV	5% CV	P	MC	Stat	1% CV	5% CV	P	MC	PV	Stat	1% CV	5% CV	P	MC	P	MC	PV	Stat	1% CV	5% CV	P	MC	Stat	1% CV	5% CV	P	MC	PV			
Hang Seng	GP	3.04	5.22	3.48	0.13	0.14	4.19	12.66	5.22	3.48	2e-6	1e-3	4.78	12.22	8.35	5.34	5e-5	4e-4	4.19	F (0.00)	12.22	8.35	5.34	5e-5	4e-4	4.19	F (0.00)	12.22	8.35	5.34	5e-5	4e-4	4.19	F (0.00)
	VG	2.17	5.22	3.48	0.12	0.18	3.89	11.32	5.22	3.48	1e-6	2e-3	3.99	15.24	8.35	5.34	5e-5	2e-3	3.89	F (0.00)	15.24	8.35	5.34	5e-5	2e-3	3.89	F (0.00)	15.24	8.35	5.34	5e-5	2e-3	3.89	F (0.00)
	NIG	5.22	5.22	3.48	0.20	0.31	3.78	14.49	5.22	3.48	2e-6	2e-3	4.18	12.89	8.35	5.34	6e-4	2e-3	3.78	F (0.00)	12.89	8.35	5.34	6e-4	2e-3	3.78	F (0.00)	12.89	8.35	5.34	6e-4	2e-3	3.78	F (0.00)
	HYP	4.48	5.22	3.48	0.14	0.14	4.04	12.79	5.22	3.48	2e-6	2e-3	4.91	16.15	8.35	5.34	6e-5	3e-4	4.04	F (0.00)	16.15	8.35	5.34	6e-5	3e-4	4.04	F (0.00)	16.15	8.35	5.34	6e-5	3e-4	4.04	F (0.00)
	GH	4.28	5.22	3.48	0.31	0.26	4.62	14.00	5.22	3.48	1e-6	2e-3	3.52	11.20	8.22	5.34	6e-4	4e-4	4.62	F (0.00)	11.20	8.22	5.34	6e-4	4e-4	4.62	F (0.00)	11.20	8.22	5.34	6e-4	4e-4	4.62	F (0.00)
Nikkei 225	GP	0.94	5.22	3.48	0.22	0.31	6.20	1.04	5.22	3.48	0.33	0.25	4.28	1.08	8.35	5.34	0.29	0.30	6.20	F (0.00)	1.08	8.35	5.34	0.29	0.30	6.20	F (0.00)	1.08	8.35	5.34	0.29	0.30	6.20	F (0.00)
	VG	1.00	5.22	3.48	0.35	0.36	5.20	1.61	5.22	3.48	0.26	0.36	5.99	1.68	8.35	5.34	0.27	0.37	5.20	F (0.00)	1.68	8.35	5.34	0.27	0.37	5.20	F (0.00)	1.68	8.35	5.34	0.27	0.37	5.20	F (0.00)
	NIG	0.95	5.22	3.48	0.42	0.25	6.14	0.55	5.22	3.48	0.31	0.34	4.89	1.54	8.35	5.34	0.26	0.39	6.14	F (0.00)	1.54	8.35	5.34	0.26	0.39	6.14	F (0.00)	1.54	8.35	5.34	0.26	0.39	6.14	F (0.00)
	HYP	1.32	5.22	3.48	0.30	0.37	5.69	1.01	5.22	3.48	0.31	0.28	5.08	1.65	8.35	5.34	0.30	0.45	5.69	F (0.00)	1.65	8.35	5.34	0.30	0.45	5.69	F (0.00)	1.65	8.35	5.34	0.30	0.45	5.69	F (0.00)
	GH	0.42	5.22	3.48	0.41	0.25	6.10	0.88	5.22	3.48	0.33	0.24	5.04	2.45	8.35	5.34	0.28	0.41	6.10	F (0.00)	2.45	8.35	5.34	0.28	0.41	6.10	F (0.00)	2.45	8.35	5.34	0.28	0.41	6.10	F (0.00)

Backtesting results for conditional and unconditional models: S&P500, FTSE100, DAX, HangSeng, and Nikkei225. PV stands for the proportion of VaR violations. P-values from both Chi-square and Monte Carlo (MC) simulations are reported. The last two columns show the unconditional-normal and unconditional-t test results for ES backtesting. We report whether the tests Pass (P) or Fail (F) and the p-values in the parentheses

both approaches are presented in Table 2 and 3. VaR and ES are based on high coverage levels, accounting for extreme events governing high trading losses. We analyze the performance of VaR and ES risk measures, utilizing full-density Lévy models of VG, NIG, HYP, and GH, and compare them with the VaR and ES estimates obtained with the tail density-based EV model. The results reveal that it is very difficult to ascertain any comprehensive superiority of one approach over the other.

Table 11 presents the frequency distribution of significant estimates reported in Table 5, 6, 7, 8 and 9. We have 15 estimates of VaR and 15 estimates of ES risk under the risk model EV and its contender Lévy estimated across all five indices and three coverage levels. Generally (11 out of 15 estimates), we find that the Lévy–VaR forecasts are closer to respective empirical estimates than their EV counterparts. Nevertheless, such observation does not allow us to declare inadequacy to disfavor the EV approach. However, in the Lévy category, the NIG (4 out of 11) and the GH (6 out of 11) models provide much more appreciable forecasts (in the sense of having minimum absolute deviation with empirical estimates) of VaR compared with the VG and the HYP Lévy models. Among the remaining Lévy models, VaR forecasts favor the HYP model, supporting the derivative pricing concept (Schouten 2003). This implies that a fully flexible GH model forecasts the quantiles more befittingly than its restricted versions.

Regarding the restricted versions, NIG characterization seems to have minimal effect on forecasts due to restriction. Regarding 15 ES forecasts, the EV model counts for the 8 most favorable forecasts, which is not sufficient to be deemed adequate. Looking into odds for Lévy-ES forecasts, we find 3 out of 7 for NIG, 2 out of 7 for GH, 1 out of 7 for the VG, and 1 out of 7 for the HYP model. Thus, the presumed myth that a tail-targeting model is more likely to provide superior, consistent forecasts for tail-focused risk measures VaR and ES is empirically confronted. The simplicity of EV–VaR and EV-ES is attractive, but a comparison with empirical values often disputes their adequacy.

We find some randomness in classifying the superiority of an approach over another given a specific timeframe. The frequency distribution of adequacy between the simple tail targeting EV analytic risk model and the relatively complex full density driven Lévy risk models is presented in Table 11. While it is difficult to claim that a particular full-density Lévy model is superior irrespective of data ranges, it is impossible to claim that the tail-targeting EV model has any sense of adequacy irrespective of all data ranges.

**Table 11** Frequency distribution of tail risk estimates of analytic EV and root search-based Lévy models

	Extreme value model		Total	Lévy models		Total
	VaR	ES		VaR	ES	
GP	4 (15)	8 (15)	12(30)	–	–	–
VG	–	–	–	0 (15)	1 (15)	1 (30)
NIG	–	–	–	4 (15)	2 (15)	6 (30)
HYP	–	–	–	1 (15)	1 (15)	2 (30)
GH	–	–	–	6 (15)	3 (15)	9 (30)
Total	4	8	–	11	7	18

The numbers are of the adequate estimates found in Tables 5, 6, 7, 8 and 9 corresponding to EV and Lévy models. The total number of estimates is reported in parentheses. The approaches have no discernible preference pattern (EV vs. Lévy)



Given such similarity of forecast performances in both approaches, the choice is likely to be determined by a compromise between user-define-simplicity and user-perceived-adequacy. This suggests that the performances of risk measure VaR and its coherent version ES are different and fail to adequately identify the risk profiles of assets. When the VaR and ES models identify the risk profiles of assets similarly, both VaR and ES would be adequate for tail targeting either the EV model or some full-density Lévy model. However, the performance of VaR and ES are mixed across EV and Lévy models. This should not be interpreted that the test statistics results of model fitting performances are contradictory.

It is well known that the AD test is tail-emphasized. Therefore, quantile mismatch outside the tail is barely detected by the  $AD_{ev}$  test, which is applied to the tail-targeting EV model. Based on the AD test statistics in Table 3, the EV model has some preference over the Lévy models. However, this is only based on the tail quantile match of EV. This means that it hardly bears information on quantile matches far outside the tail. This is why an AD test value of a solely tail-based EV model might turn deceptive when compared with the AD test value of an entire distribution-based Lévy model. This deception can hardly be adequately detected by applying the GOF test emphasized on the tail. Thus, it is not surprising that the seemingly preferable EV model turns elusive and does not yield the most adequate forecasts of risk measures, i.e., VaR and ES. Our backtesting results about VaR and ES also confirm such findings. The EV–VaR or EV–ES results are not significantly different from VaR and ES based on Lévy models. In most cases, the results are mixed.

## Conclusion

We investigate and compare the simplicity and adequacy of tail-focused VaR and ES risk measures for tail-targeting EV models with the full density-focused Lévy–VaR and Lévy–ES using data on futures contracts of S&P500, FTSE100, DAX, Hang Seng, and Nikkei 225 indices from January 1, 2007 to December 31, 2017, covering the 2007–2008 Global Financial Crisis that led to the subprime mortgage debacle in the US. We find that returns discarded by the EV model (as they do not characterize the extreme unexpected market losses) and incorporated by the Lévy models do have some effect on the performance of tail-focused risk measures VaR and its coherent version ES. Thus, without an immutable law justifying any preference between “tail-alone” and “full-density-based” models for tail-focused risk management, this study provides a heuristic analysis illustrating the potential effects of the observations discarded by an EV model on risk estimates when considered under Lévy models.

The tail-based EV models are simpler for the analytic formulation of the tail-focused risk measures of VaR and tail-aggregate risk measure of ES compared with the Lévy-based measure. Moreover, the EV models are simpler to implement in risk measure calculations. However, we find that the EV models are inadequate as the performance of EV risk estimates is not necessarily superior to that of Lévy risk estimates. On the other hand, we cannot assure that a full density-based model based on Lévy distribution adequately assesses risk measures. Thus, the adequacy of a simpler model with a more straightforward implementation becomes a relative consideration. Our model testing reinforces the theoretical fundamental that the extreme observations in the tails of the

EV model (discarding smaller systematic returns) and all observations, including smaller and extreme ones in Lévy models, are different approaches with the common goal of meaningful simplification of reality. Their relative performance for a particular time window may fail to offer any guarantee. Given such randomness of estimation performances under both approaches (for different ranges of data and coverage levels), the choice should be determined by a compromise or trade-off between simplicity and user-defined adequacy. Our study period covers the 2007–2008 Global Financial Crisis. The analysis can be extended to other financial crisis periods, e.g., the Russian financial crisis, the default crisis related to the Long-Term Capital Management of 1998,<sup>9</sup> and more recent the coronavirus-related turmoil in the financial market during 2020–2021.

Our analysis is based on only a selection of EV and Levy models. The study can be extended using other types of models. Practitioners should not rely on one set of desired models and ignore others when implementing VaR and ES estimation. The simplicity of a model does not guarantee its adequacy. However, the adequacy of a model based on “full-density,” e.g., the Lévy-based model, may not always be the best when a simpler tail-based model will provide more robust VaR and ES results. As many banks and financial institutions do not follow adequacy requirements in risk measures following the Basel agreement, our findings shed empirical light on such complexity. Therefore, when the results are mixed, banks and financial institutions as well as policymakers should find a way to compromise simplicity and adequacy.

#### **Acknowledgements**

None

#### **Author contributions**

All authors contributed to the study conception and design. All authors read and approved the final manuscript. SM was involved in concept development. MKH developed the idea further and was involved in editing. SM wrote the first draft. MKH and HK further revised the draft.

#### **Funding**

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

#### **Availability of data and materials**

The data will be available from the authors upon request.

#### **Declarations**

##### **Competing interests**

The authors have no relevant financial or non-financial interests to disclose.

Received: 20 July 2022 Accepted: 31 December 2023

Published online: 03 May 2024

#### **References**

- Acerbi C, Szekely B (2014) Backtesting expected shortfall. MSCI Inc., New York
- Acerbi C, Szekely B (2017) General properties of backtestable statistics. SSRN Electron J. <https://doi.org/10.2139/ssrn.2905109>
- Anderson TW, Darling DA (1952) Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes. *Ann Math Stat* 193–212
- Anderson TW, Darling DA (1954) A test of goodness of fit. *J Amer Stat Assoc* 49(268):765–769
- Anna C, Rachev S, Fabozzi F (2005) Composite goodness-of-fit tests for left-truncated loss samples. Working Paper, Department of Statistics and Applied Probability. University of California, Santa Barbara
- Bali TG (2007) A generalized extreme value approach to financial risk measurement. *J Money Credit Bank* 39(7):1613–1649

---

<sup>9</sup> Kabir and Hasan (2005, 2009) examine the implications of these two crisis.

- Bank for International Settlements (2019) Basel Framework (Basel), Accessed September 2021, [http://www.bis.org/basel\\_framework/index.htm?export=pdf](http://www.bis.org/basel_framework/index.htm?export=pdf)
- Barndorff-Nielsen OE (1995) Normal Inverse Gaussian distributions and the modeling of stock returns. Research report no 300, Department of Theoretical Statistics, Aarhus University 401–419
- Barndorff-Nielsen O (1977) Exponentially decreasing distributions for the logarithm of particle size. *Proc R Soc Lond A* 353:401–419
- Barndorff-Nielsen O (1978) Hyperbolic distributions and distributions on hyperbolae. *Scand J Stat* 5:151–157
- Bayer S, Dimitriadis T (2022) Regression-based expected shortfall backtesting. *J Financ Econ* 20(3):437–471
- Bertoin J (1996) *Levy processes*. Cambridge University Press, Cambridge
- Bingham N H, Kiesel R (2001) Modeling asset returns with hyperbolic distributions. *Return Distributions on finance* 1–20. Butterworth-Heinemann.
- Chavez-Demoulin V, Paul E, Sylvain S (2014) Extreme-quantile tracking for financial time series. *J Econ* 181(1):44–52
- Cheng L, AghaKouchak A, Gilleland E, Katz RW (2014) Non-stationary extreme value analysis in a changing climate. *Clim Change* 127:353–369
- Cheng J, Hong Y, Tao J (2015) How do risk attitudes of clearing firms matter for managing default exposure in futures markets? *Eur J Financ* 22(10):909–940
- Christoffersen P (2003) *Element of financial risk management*. Academic Press, Cambridge
- Cotter J, Dowd K (2006) Extreme spectral risk measures: an application to futures clearinghouse margin requirements. *J Bank Finance* 30:3469–3485
- De Olivera F, Fajardo J, Mordecki E (2018) Skewed Lévy models and implied volatility skew. *Int J Theor Appl Finance* 21(02):1850003
- Dowd K (2005) *Measuring market risk*. John Wiley & Sons Ltd, Hoboken
- Du Z, Escanciano JC (2017) Backtesting expected shortfall: accounting for tail risk. *Manage Sci* 63(4):940–958
- Eberlein E, Hammerstein EA (2002) The generalized hyperbolic and inverse Gaussian distributions: limiting cases and approximation of processes. FDM preprint 80, University of Freiburg.
- Eberlein E, Prause K (1998) The generalized hyperbolic model: financial derivatives and risk measures. FDM preprint 56, University of Freiburg
- Eberlein E, Keller U (1995) Hyperbolic distributions in finance. *Bernoulli* 1(3):281–299
- Fajardo J (2015) Barrier style contracts under Lévy processes: an alternative approach. *J Bank Finance* 53:179–187
- Fajardo J, Mordecki E (2006) Symmetry and duality in Lévy markets. *Quant Finance* 6(3):219–227
- Fajardo J, Mordecki E (2014) Skewness premium with Lévy processes. *Quant Finance* 14(9):1619–1626
- Farkas W, Mathys L (2022) Geometric step options and Lévy models: duality, PIDEs, and semi-analytical pricing. *Front Math Finance* 1(1):1–51
- Frésard L, Perignon C, Wilhelmsson A (2011) The pernicious effects of contaminated data in risk management. *J Bank Finance* 35(10):2569–2583
- Fuse G, Meucci A (2008) Pricing discretely monitored Asian options under Lévy processes. *J Bank Finance* 32(10):2076–2088
- Gençay R, Selçuk F (2004) Extreme value theory and value-at-risk: relative performance in emerging markets. *Int J Forecast* 20(2):287–303
- German H (2002) Pure jump Lévy processes for asset price modeling. *J Bank Finance* 26(7):1297–1316
- Hoga Y, Demetrescu M (2022) Monitoring value-at-risk and expected shortfall forecasts. *Manag Sci*. <https://doi.org/10.1287/mnsc.2022.4460>
- Jondeau E, Rockinger M (2003) The tail behavior of stock returns: emerging versus mature markets. *J Empir Financ* 10:559–581
- Kabir MH, Hassan MK (2005) The near-collapse of LTCM, US financial stock returns, and the fed. *J Bank Finance* 29(2):441–460
- Kabir MH, Hassan MK (2009) Russian financial crisis, US financial stock returns and the IMF. *Appl Financ Econ* 19(5):409–426
- Kim YS, Rachev ST, Bianchi ML, Fabozzi FJ (2008) Financial market models with Lévy processes and time-varying volatility. *J Bank Finance* 32(7):1363–1378
- Kourouma L, Dupre D, Sanfilippo G, Taramasco O (2010) Extreme value at risk and expected shortfall during the financial crisis. Available at SSRN 1744091
- Kyprianou A (2006) *Introductory lectures on fluctuations of Levy processes with applications*. Springer, Berlin
- Lazar E, Zhang N (2019) Model risk of expected shortfall. *J Bank Finance* 1(105):74–93
- Longin FM (1996) The asymptotic distribution of extreme stock market returns. *J Bus* 69(3):383–408
- McNeil AJ, Frey R (2000) Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *J Empir Financ* 7:271–300
- Monier E, Gao X (2015) Climate change impacts on extreme events in the United States: an uncertainty analysis. *Clim Change* 131(1):67–81
- Mozumder S, Dempsey M, Kabir MH (2017) Backtesting extreme value and Lévy value-at-risk models: evidence from international futures markets. *J Risk Finance* 18(1):88–118
- Otto S, Breitung J (2022) Backward CUSUM for testing and monitoring structural change with an application to COVID-19 pandemic data. *Econom Theory*. <https://doi.org/10.1017/S0266466622000159>
- Patton AJ, Ziegel JF, Chen R (2019) Dynamic semiparametric models for expected shortfall (and value-at-risk). *J Econom* 211(2):388–413
- Perignon C, Smith DR (2010a) The level and quality of value-at-risk disclosure by commercial banks. *J Bank Finance* 34(2):362–377
- Perignon C, Smith DR (2010b) Diversification and value-at-risk. *J Bank Finance* 34(1):55–66
- Pidgeon N (2012) Climate change risk perception and communication: addressing a critical moment? *Risk Anal* 32:951–956

- Prause K (1999) The Generalized Hyperbolic Model: Estimation, financial derivatives and risk measures. Ph.D. Thesis, University of Freiburg
- Quintos C, Zhenhong F, Peter CBP (2001) Structural change tests in tail behaviour and the Asian crisis. *Rev Econ Studies* 68(3):633–663
- Sato K (1999) *Lévy processes and infinitely divisible distributions*. Cambridge University Press, Cambridge
- Schoutens W (2003) *Lévy processes in finance: pricing financial derivatives*. John Wiley & Sons Ltd, Hoboken
- So MKP, Wong CM (2012) Estimation of multiple period expected shortfall and median shortfall for risk management. *Quant Finance* 12(5):739–754
- Sorwar G, Dowd K (2010) Estimating financial risk measures for options. *J Bank Finance* 34(8):1982–1992
- Taylor JW (2019) Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. *J Bus Econ Stat* 37(1):121–133
- Tolikas K, Gettinby GD (2009) Modelling the distribution of the extreme share returns in Singapore. *J Empir Financ* 16(2):254–263
- Wong HY, Guan P (2011) An FFT-network for Lévy option pricing. *J Bank Finance* 35(4):988–999

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.