Erasmus University Rotterdam

Erasmus School of Economics

Bachelor Thesis BSc2 Econometrics and Economics

Decoupling tail index and extrapolation location for conditional tail risk forecasts in ARMA-TGARCH models

Author:	*Student ID Number:*	*Supervisor:*	*Second Assessor:*
Lennart Rooden	469404	Prof. dr. Chen Zhou	Msc. Bram van Os

July 4, 2021

Abstract

The Extreme Value Theory (EVT) approach to tail risk estimation has gained immense traction in view of risk management regulation and policy directives becoming increasingly quantitative in nature. We validate and improve this oft relied-upon method by means of verification of past findings, entertaining a comparison with Filtered Historical Simulation (FHS), the treatment of misspecification issues, and a generalisation of the Weissman quantile estimator with new asymptotic theory for confidence intervals. The generalisation introduces an extrapolation location (l), within the tail corresponding to the extreme value index estimator, to separately optimise the uncertainty tradeoffs in the quantile and tail index estimators. For economic applications, we find that proper specification of temporal dependencies is essential, particularly concerning asymmetries as we evaluate the left tail only, whereas several choices for l yield comparable results for moderate confidence levels (0.5%-2.5%). Although FHS performance is not significantly different from the more intricate EVT approach, outperformance occurs for two out of six indices in favour of our proposed data-driven heuristic for the optimal extrapolation location (l^*). Even more promising is performance for $\alpha = 0.1\%$ ($\alpha = 0.01\%$) with $l = l^*$ significantly outperforming contemporary literature for five (six) out of six indices under consideration, which warrants further investigation in applications that model such extreme events.

Keywords: ARMA-TGARCH filter, Conditional Expected Shortfall, Conditional Value-at-Risk, extrapolation location, generalised Weissman estimator

Acknowledgements: My sincere appreciation goes out to Chen Zhou for his assiduous guidance, expertise and critical eye. His feedback helped me improve the contents of this paper and motivated me to explore unconventional ideas that, for a large part, have made writing this paper so interesting and rewarding. I would also like to express my gratitude towards Yannick Hoga, and by extension Laura Spierdijk, for providing their R code.

Contents

1 Introduction

> *'If you don't invest in risk management, it doesn't matter what business you're in, it's a risky business,'*
>
> — Gary Cohn*, 25 September 2011*

More than ten years after the Global Financial Crisis, the COVID-19 pandemic serves as a stark reminder of the unpredictable nature and tremendous impact brought about by the recurring distortions in financial markets. The unprecedented turmoil that ensued in financial markets caused volatility, as measured by the Cboe Volatility Index, to peak at 89.53 on October 24th, 2008, and close at a record high (82.69) on March 16th, 2020 (Kawa, 2020). Such events re-emphasise the need for proper risk management in order to establish policies and regulations that ensure the stability of the financial system, particularly in case of momentous unforeseen adverse events. The risks involved in these extreme events, commonly referred to as tail risk, are quantified using measures of tail risk. Tail risk measures are loosely defined as random variables that quantify a certain risk beyond some threshold.

Such risk measures have been conceptualised in abundance, but with their widespread use in regulatory frameworks such as Basel III and Solvency II, Conditional Value-at-Risk (CVaR) and Conditional Expected Shortfall (CES) are the two most established, and hence most important to investigate, tail-based risk measures (Liu & Wang, 2020). Left-tailed CVaR is defined as the value for which, conditional on past returns, there is some small pre-specified probability that losses exceed CVaR. Correspondingly, CES describes the expected return given that the loss exceeds CVaR. Specifically, Basel III and Solvency II capital requirement guidelines for credit risk stipulate that financial institutions ought to be able to incur losses amounting to the 99%-CVaR on a one year holding period, whereas this is 99.5% for insurance companies. Moreover, the shift from CVaR to CES has already been initiated for credit risk, along with communicating the intention to implement similar changes for market risk regulation (Laas & Siegel, 2017). The primary motivation behind this shift concerns the desire to identify the expected magnitude of extreme losses, rather than merely a quantile that signals the expected minimal extreme loss for some confidence level chosen to represent the likelihood of extreme losses. Additionally, in contrast to CES, CVaR does not qualify as a coherent risk measure as defined by the four criteria formulated by Artzner et al. (1997) and Artzner et al. (1999) to set out favourable properties of such measures.

Most methods for tail risk estimation, such as Filtered Historical Simulation (FHS), suffer from

the fact that quantiles must be contained within the sample size for the approach to work. For instance, if weekly data is available and time-varying behaviour is expected, a once-in-a-thousand scenario can not be predicted adequately. To alleviate this shortcoming, researchers and practitioners oftentimes resort to the principles of Extreme Value Theory (EVT), which was first introduced to risk modelling by McNeil & Frey (2000). By means of semiparametric assumptions regarding the shape of the tail of the distribution, extrapolation beyond sample size becomes possible. Chan et al. (2007) and Hoga (2019) derive asymptotic theory for EVT-based CVaR and CES estimates.

The aim of this paper is to present and evaluate generalised CVaR and CES estimators and their confidence intervals by means of a simulation study and empirical application in the methodological framework of Hoga (2019). To this end, we first replicate the methods in Hoga (2019) , before applying them to a proposed generalised estimator that extrapolates from some optimal point within the tail, aiming at point estimates with lower loss scores and confidence intervals based on new asymptotic theory that exhibit superior coverage vis-à-vis extrapolation from the boundary of the tail. Moreover, we refine the framework by using an ARMA-TGARCH filter, thereby shedding light upon misspecification issues and the consistency required for correct coverage of the confidence intervals. All findings are placed in an economic context to substantiate their validity and reveal additional insights regarding the consequences of recent extreme events in financial markets.

In a societal context, this research is relevant as it provides an intuitive method of achieving more reliable estimators of tail risk, on which many policies and regulations rely, thereby functioning as an indispensable piece of the puzzle of achieving a more stable financial system. This, in turn, is essential for sustaining optimal employment levels, stimulating investment, and facilitating sustainable economic growth. More importantly, warding off financial instability helps prevent severe pension fund losses, mistrust of financial institutions, hyperinflation, and other detrimental consequences (World Bank, 2012). The scientific relevance is grounded in the lack of research regarding the optimal extrapolation location in the EVT approach to estimating risk measures, which has hitherto been assumed implicitly to be at the interior bound of the tail determined by some other criterion such as the goodness-of-fit with the theoretical asymptotic Pareto shape of the distribution. In this paper, we consider separately the identification of the tail and extreme value index on the one hand, and the extrapolation location within the tail on the other, aiming at more reliable CVaR and CES point and interval estimates. Additionally, we find evidence in favour of the importance of proper ARCH-type specification, which contradicts Jalal & Rockinger (2008).

In order to place this paper into the relevant economic and scientific context, we commence

with a literature review regarding relevant stylised facts in financial econometrics and the issue of tail index threshold selection that closely resembles our extrapolation location selection (Section 2). Then, we consider the model setup and methodological approach in Section 3, before validating the findings by Hoga (2019) in Section 4. Subsequently, we entertain model selection in Section 5, followed by the presentation and interpretation of a forecast comparison for various extrapolation location choices in Section 6. Finally, Section 7 concludes and provides limitations and suggestions for further research. Appendices A, B, and C offer supporting materials regarding replication results, extension results and the slope of the Weissman (1978) estimator extrapolation factor respectively.

2 Literature Review

This paper is concerned with improving the application of Extreme Value Theory (EVT) to the modelling of tail risk. In time series applications, the series must first be filtered appropriately to obtain independent and identically distributed (*iid*) innovation terms.

2.1 Time series models for financial returns

At the foundation of the non-Gaussian ARCH-type specifications employed to model temporal dependencies in the expectation and variance of a series lie two stylised facts of log-returns in financial econometrics. Firstly, after filtering out autoregressive and moving average (ARMA) components from a series, thereby modelling the conditional mean, the estimated residuals exhibit heteroskedasticity (Finkenstadt & Rootzén, 2003). Whereas there is little autocorrelation in log-returns themselves, squared log-returns are characterised by a small but significant and persistent positive autocorrelation. This results in, for example, volatility clustering, an oft-observed phenomenon in financial time series. The principal approach to extracting the *iid* innovation terms underlying these heteroskedastic shocks are the (generalised) autoregressive conditional heteroskedasticity ((G)ARCH) models presented in the seminal works of Engle (1982) and Bollerslev (1986). Failing to adequately model heteroskedasticity invalidates the use of innumerable econometric models, as these generally assume *iid* innovations. Secondly, residuals are leptokurtic, so Gaussian models are inappropriate, which is detrimental to the economics of extremes, where the differences are most noticeable. It is fairly straightforward to illustrate that for Gaussian standardised shocks, the GARCH process yields ARMA shocks with excess kurtosis relative to the normal distribution (p.49 of Francq & Zakoian (2019)). However, this is often insufficient, leading practitioners to employ alternative distributions, the most favoured being the heavy-tailed Student's t distribution (e.g. Harvey & Chakravarty (2008), Wilhelmsson (2006), and Sun & Zhou (2014)).

In addition to heteroskedasticity and excess kurtosis in ARMA residuals, asymmetric impact of shocks is often noted based on the sign of the shock. Black (1976) observed that negative shocks impact the volatility of securities considerably more severely than do positive shocks and coined this the leverage effect. However established this term may have become, academics widely agree that this feature has little to do with financial leverage (e.g. Hasanhodzic & Lo (2011) and Hens & Steude (2009)). Combined with their finding that significant and persistent positive autocorrelation — also termed the long memory property — is present not only for squared residuals, but more generally for many $|\varepsilon_t|^\delta$, where $\delta > 0$, Ding et al. (1993) explicated the Asymmetric Power ARCH (APARCH) model. It allows for asymmetric impact on volatility driven by the sign of the shocks and models volatility by means of specifying autoregressive dependencies for σ_t^δ rather than σ_t^2. In its most general form, the APARCH model includes the power term δ to be estimated based on the series under consideration. However, APARCH nests several prominent asymmetric volatility models for given values of δ, with Threshold GARCH (TGARCH) for $\delta = 1$ (Zakoian, 1994) and GJR-GARCH for $\delta = 2$ (Glosten et al., 1993) receiving most attention from academics and practitioners.

When applying tail risk estimation methods reliant on the *iid* innovations assumption, we must not pretend that any real-life process will be specified entirely correctly, resulting in estimated residuals drawn from a genuinely *iid* process. Instead, we may hope to specify a process sufficiently well such that the approaches based on this assumption work well in practice. In that regard, Jalal & Rockinger (2008) find that, given the appropriate conditional distribution of the filtered innovations, applying a GARCH filter to non-GARCH processes to estimate Value-at-Risk and Expected Shortfall in an Extreme Value Theory framework yields excellent forecasting performance. Surprisingly, even omitting the heteroskedasticity filter altogether still results in satisfactory predictions.

Building on the works of, among others, Embrechts et al. (1997) and Reiss et al. (1997) on an Extreme Value Theory approach to modelling the distribution of asset returns, the EVT method of estimating tail risk proposed by McNeil & Frey (2000), for which McNeil (1999) lays the groundwork, is a relatively recent addition to the financial econometrics literature regarding tail risk prediction. Nonetheless, it has certainly gained traction and crucially allows for more reliable extrapolation beyond the sample size by incorporating more appropriate distributional assumptions than previous approaches such as assuming conditional normality. Moreover, incorporating theoretically derived properties of the tail for large quantiles essentially integrates additional information, which improves estimates without requiring supplemental observations.

In this paper, we validate the findings for and extend the ARMA-GARCH filter used in contem-

porary literature (McNeil & Frey (2000) and Hoga (2019) among others) to capture asymmetries by means of an ARMA-TGARCH formulation. Furthermore, we investigate misspecification issues and contrast our findings with those of Jalal & Rockinger (2008). Lastly, we evaluate the stylised facts put forth in past literature in the context of major global indices with recent data to verify that recent data still exhibits these features.

2.2 Tail index threshold selection

For tail fitting, McNeil & Frey (2000) note that practically all established continuous probability density functions exhibit a Generalised Pareto Distribution shape in the limit. Specifically, the ordinary Pareto distribution serves as a first-order tail approximation (Danielsson & De Vries, 1997). This property lies at the foundation of the Weissman (1978) estimator, which extrapolates from the innermost boundary of the tail by extending the theoretical asymptotic Pareto shape to the desired quantile. Given some extreme value index estimate, this estimator gives rise to an uncertainty tradeoff regarding the extrapolation location, which in previous literature coincides with the tail boundary for the tail index estimator. To wit, when extrapolating to quantiles more extreme than the extrapolation location, the variance of the order statistic increases as the threshold increases, whereas the variance of the extrapolation factor diminishes, which dominates asymptotically (see Section 3.6 and Section 3.3 for further elucidation). To this end, both the tail boundary and the tail index require estimation. This decision boils down to a bias-variance tradeoff (Danielsson et al., 2016), where including more observations reduces the estimator variance at the cost of introducing bias by increasingly deviating from the asymptotic Pareto shape. A common selection method relies on bootstrapping or maximum random fluctuation to minimise the mean squared error of the estimator (recall that the mean squared error is comprised of the variance and the squared bias) as treated in the works of, for instance, Hall (1990), Danielsson et al. (2001), and Drees & Kaufmann (1998). Alternatively, several heuristics have been used; most notably, a fixed fraction of the sample size or inspection of Hill plots (Drees et al., 2000) offer suggestions for threshold selection. Recently, Danielsson et al. (2016) proposed to simplify the tail identification by minimising the maximum deviation from a fitted Pareto tail for a reasonable range of sample size fractions, with favourable finite sample performance.

Whereas research on CVaR and CES point forecasts is abundant, very few papers delve into the limiting distribution and subsequent confidence intervals (CIs) for these estimators. The earliest discussion of CIs of EVT-based risk measure forecasts without strict distributional assumptions or computationally intensive bootstrapping is presented by Chan et al. (2007), who restrict the

analysis of McNeil & Frey (2000) to a nonparametric GARCH model with heavy-tailed innovations and derive CIs for the CVaR estimator proposed by the latter. To construct CIs, Chan et al. (2007) advocate a normal approximation or data tilting. To improve finite sample performance by relying on the observations themselves rather than asymptotic theory, Shao (2010) proposes self-normalisation for stationary time series.

Inspired by these developments, Hoga (2019) generalises the normal approximation from Chan et al. (2007) to an ARMA-GARCH framework and implements self-normalisation to arrive at two methods for computing CVaR and CES confidence intervals. Moreover, the author considers multiple tail index estimators and demonstrates the implementation of the threshold selection method recently proposed by Danielsson et al. (2016). The author validates the theoretical contributions using a simulation study and presents the findings of an empirical study. The forecasts in the simulation study exhibit undercoverage but the best performance in the empirical application for the ARMA-GARCH model (as opposed to GARCH) with k based on minimising the maximum distance from a Pareto distribution in the tail (as compared to a function of sample size, see Chan et al. (2007)), with the Hill (1975) estimator (as compared to MR, see Daníelsson et al. (1996)), with a self-normalised interval (as compared to a normal approximation, see Chan et al. (2007)). Section 4 contains a more elaborate discussion of the findings in Hoga (2019).

For the purposes of this paper, we implement the recently suggested tail index threshold selection of Danielsson et al. (2016) and decouple the extrapolation location from the tail boundary to separately optimise the tradeoffs in threshold selection for the Weissman (1978) estimator on the one hand, and the tail index on the other. To this end, we propose a new heuristic to determine our optimal extrapolation location, thereby deviating from currently proposed methods that are aimed mostly at the extreme value index bias-variance tradeoff. Moreover, we conjecture new asymptotic theory that deviates from the works of Chan et al. (2007) and Hoga (2019) for we employ an ARMA-TGARCH process specification and rely on our proposed generalised Weissman estimator.

3 Methodology

First, the model and estimators used in Hoga (2019) are presented. Then, we present alternative CVaR and CES estimators, provide asymptotic confidence intervals for the point estimators and conjecture new asymptotic results, followed by the details of identification, estimation, and forecasting. Lastly, forecast evaluation and the data used in the simulation study and empirical application are treated. All methods described are implemented in R 4.0.5 (R Core Team, 2020).

3.1 The ARMA-GARCH model

Let y_t, where $t = 1, \ldots, T$, be a univariate stationary time series driven by ARMA dynamics and GARCH innovation terms ε_t. Following most of the notation in Hoga (2019), a general ARMA$(\overline{p}, \overline{q})$-GARCH$(p, q)$ specification follows as

$$y_t = \sum_{i=1}^{\overline{p}} \phi_i y_{t-i} + \varepsilon_t - \sum_{j=1}^{\overline{q}} \theta_j \varepsilon_{t-j}, \text{ where } \varepsilon_t = \sigma_t U_t, \text{ and} \tag{1}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \varphi_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \vartheta_j \varepsilon_{t-j}^2. \tag{2}$$

Here, $p, q, \overline{p}, \overline{q} \in \mathbb{N}$ and standardised shocks U_t are *iid* with a zero mean and unit variance. The interested reader is referred to Hoga (2019) for the parameter restrictions required to derive asymptotic confidence intervals for CVaR and CES under their estimator formulation.

3.2 The ARMA-TGARCH model

As per Section 2.1, financial time series often exhibit an asymmetric effect of shocks on volatility. TGARCH models the variance of log-returns and incorporates asymmetry through an additional term — the absolute value of the lagged shock. In an APARCH-type formulation, the ARMA$(\overline{p}, \overline{q})$-TGARCH$(p, q)$ process is then given by Equation (1) with conditional standard deviation

$$\sigma_t = \omega + \sum_{i=1}^{p} \varphi_i \sigma_{t-i} + \sum_{j=1}^{q} \vartheta_j (|\varepsilon_{t-j}| - \varrho_j \varepsilon_{t-j}). \tag{3}$$

3.3 Estimators for CVaR and CES

Let us first formally define the one-step ahead right-tail CVaR and CES, following Hoga (2019), as

$$\begin{aligned} \text{CVaR}_{\alpha,t} &:= \inf\{y \in \mathbb{R} : P(y_{t+1} \leq y | \mathcal{I}_t) \geq 1 - \alpha\} \text{ and} \\ \text{CES}_{\alpha,t} &:= \mathbb{E}(y_{t+1} | y_{t+1} > \text{CVaR}_{\alpha,t}, \mathcal{I}_t), \end{aligned}$$

where $\mathcal{I}_t = \{y_1, \ldots, y_t\}$ denotes the information set at time t and $\alpha \in (0, 1)$. Hoga (2019) suggests rewriting y_t into $\mu_t + \sigma_t U_t$, with $\mu_t = \sum_{i=1}^{\overline{p}} \phi_i y_{t-i} - \sum_{j=1}^{\overline{q}} \theta_j (y_{t-j} - \mu_{t-j})$ and σ_t given by Equation (3). Let their estimators be denoted $\hat{\mu}_t(\hat{\boldsymbol{\psi}})$ and $\hat{\sigma}_t(\hat{\boldsymbol{\psi}})$ respectively to obtain CVaR and CES estimators

$$\widehat{\text{CVaR}}_{\alpha,t} := \hat{\mu}_{t+1}(\hat{\boldsymbol{\psi}}) + \hat{\sigma}_{t+1}(\hat{\boldsymbol{\psi}}) \hat{Q}^U(1 - \alpha) \quad \text{and} \quad \widehat{\text{CES}}_{\alpha,t} := \hat{\mu}_{t+1}(\hat{\boldsymbol{\psi}}) + \hat{\sigma}_{t+1}(\hat{\boldsymbol{\psi}}) \widehat{\text{ES}}^U(\alpha), \tag{4}$$

where the estimator of generic parameter vector $\boldsymbol{\psi} \in \Theta_{\text{ARMA}} \times \Theta_{\text{GARCH}} \subset \mathbb{R}^{\bar{p}+\bar{q}} \times \mathbb{R}_{>0} \times \mathbb{R}_{\geq 0}^{p+q}$ is often omitted for brevity, $\hat{Q}^U(\cdot)$ denotes the estimator of the quantile function of any U_t, and $\widehat{\text{ES}}^U(\alpha)$ represents the estimator of the Expected Shortfall (ES) for any U_t (i.e. of $\mathbb{E}(U_1|U_1 > Q^U(1-\alpha))$.

In the tail we assume that the quantiles of U_t are of regular variation at infinity with extreme value index $\gamma > 0$, which has estimator $\hat{\gamma}$. That is, for the $(1-1/x)$-quantile given by tail function $U(x) := Q^U(1-1/x)$ and $\gamma > 0$, we assume $\lim_{x\to\infty} U(xy)/U(x) = y^\gamma$ for all $y > 0$. Moreover, let $\hat{\varepsilon}_t$ represent the estimator of the GARCH-driven ARMA residual at time t. Using $\hat{U}_t := \hat{\varepsilon}_t/\hat{\sigma}_t$ and defining order statistics $U_{1:T} \leq \cdots \leq U_{T:T}$ of $\{\hat{U}_1, \ldots, \hat{U}_T\}$, the Weissman (1978) estimator of $Q^U(1-\alpha)$ and the $\text{ES}^U(\alpha)$ estimator presented by Hoga (2019) follow as

$$\hat{Q}^U_{\text{WM}}(1-\alpha) := U_{T-k:T}\left(\frac{\alpha T}{k}\right)^{-\hat{\gamma}} \quad \text{and} \quad \widehat{\text{ES}}^U_{\text{WM}}(\alpha) := \frac{\hat{Q}^U_{\text{WM}}(1-\alpha)}{1-\hat{\gamma}}, \tag{5}$$

where k indicates the boundary of the tail. For tail boundary k_T of a series of length T, $k_T \to \infty$ and $k_T/T \to 0$ are presumed to hold as $T \to \infty$. Henceforth, let $k := k_T$.

The Weissman (1978) estimator extrapolates to an extreme quantile from the interior bound of the tail. However, as α becomes exceedingly small, uncertainty in the estimator exhibits superlinear growth (see Appendix C), which defeats the purpose of employing EVT. Namely, to accurately determine quantiles so far in the tail that order statistics are unreliable or unobserved entirely. To this end, let us consider an alternative estimator of $Q^U(1-\alpha)$ that extrapolates from some location within the tail to optimise the uncertainty tradeoff present. That is, order statistics become less reliable further in the tail, whereas the extrapolation factor becomes less reliable as the target is further from the point of extrapolation. For tuning threshold l, the generalised estimator of $Q^U(1-\alpha)$ is given by

$$\hat{Q}^U_l(1-\alpha) := U_{T-l:T}\left(\frac{l}{\alpha T}\right)^{\hat{\gamma}}, \tag{6}$$

where $l \in \{0, \ldots, k\}$ denotes the extrapolation point and $\hat{\gamma}$ is based on the entire tail defined by k. For optimal extrapolation point $l^*_T(\alpha)$, which may vary with α, of a series of length T, $l^*_T(\alpha) \to \infty$ and $l^*_T(\alpha)/T \to 0$ are presupposed as $T \to \infty$, analogous to k_T. Similarly, let $l^* := l^*_T(\alpha)$ hereinafter. From this quantile estimator, an alternative $\text{ES}^U(\alpha)$ estimator given by $\widehat{\text{ES}}^U_l(\alpha) := \hat{Q}^U_l(1-\alpha)/(1-\hat{\gamma})$ follows akin Equation (5).

3.4 Shape parameter and extrapolation location

Let k^* characterise the tail of the distribution of U_t defined as $\{U_{T-k^*:T}, \ldots, U_{T:T}\}$. As discussed in Section 2.2, the tail theoretically approximates a Pareto distribution, which also underlies the extrapolation factor in the Weissman (1978) estimator and Equation (6). Therefore, k^* is chosen to optimise the Pareto fit of the tail within some reasonable range $\{k_{\min}, \ldots, k_{\max}\}$. More formally,

$$k^* := \underset{k \in \{k_{\min}, \ldots, k_{\max}\}}{\arg\min} \left(\sup_{t \in \{1, \ldots, k_{\max}\}} \left| U_{T-t:T} - U_{T-k:T} \left(\frac{t}{k} \right)^{-\hat{\gamma}} \right| \right), \tag{7}$$

as proposed in Danielsson et al. (2016). The optimal k thus minimises the largest discrepancy between observed and extrapolated order statistics. The last component of Equation (5) left unaddressed is $\hat{\gamma}$. The Hill (1975) estimator is most established among researchers and practitioners. Moroever, Tsourti & Panaretos (2001) demonstrate that for Pareto-type data, it has the best small-sample performance compared to several well-known alternatives, whereas the Moments Ratio (MR) estimator of Daníelsson et al. (1996) outperforms for larger sample sizes. The Hill and MR estimators follow as

$$\hat{\gamma}_H := \frac{1}{k^*} \sum_{t=0}^{k^*} \ln U_{T-t:T} / U_{T-k^*:T} \quad \text{and} \quad \hat{\gamma}_{MR} := \frac{1}{2\hat{\gamma}_H k^*} \sum_{t=0}^{k^*} \left(\ln U_{T-t:T} / U_{T-k:T} \right)^2 .$$

We consider both and assess which is more appropriate for our proposed estimator. Lastly, l^* is estimated by minimising within some reasonable range a CVaR forecast loss function given by

$$\hat{l}^* := \underset{l \in \{l_{\min}, \ldots, k\}}{\arg\min} \sum_{t=t_1}^{t_2} \frac{S_1(\widehat{\text{CVaR}}_{\alpha,t-1}, y_t)}{t_2 - t_1 + 1}, \tag{8}$$

where t_1 and t_2 denote the indices corresponding to the first and last observations respectively of the subseries used to evaluate the forecast performance (the selection set, in short). Moreover, $S_1(\cdot)$ is the realised Quantile score of observation t (Taylor, 2019) to quantify CVaR forecast quality, calculated as

$$S_1(\widehat{\text{CVaR}}_{\alpha,t-1}, y_t) = \sum_{t=t_1}^{t_2} \left(y_t - \widehat{\text{CVaR}}_{\alpha,t-1} \right) \left(\alpha - I_{\{y_t \geq \widehat{\text{CVaR}}_{\alpha,t-1}\}} \right) / (t_2 - t_1 + 1),$$

for conventional indicator function $I_{\{\cdot\}}$.

3.5 Confidence interval estimators for CVaR and CES

The main result of Hoga (2019) is the asymptotic distribution under mild conditions of the CVaR and CES estimators based on Equation (5). The author used a normal approximation and self-normalisation, yielding asymptotic $(1-\tau)$-confidence intervals

$$I_{\text{na}}^{1-\tau} := \left[\hat{z}(1)\exp\left(-\Phi(1-\tau/2)\hat{\sigma}_{\hat{\gamma},\gamma}\frac{\ln(k/(n\alpha))}{\sqrt{k}}\right),\ \hat{z}(1)\exp\left(\Phi(1-\tau/2)\hat{\sigma}_{\hat{\gamma},\gamma}\frac{\ln(k/(n\alpha))}{\sqrt{k}}\right)\right],$$
$$I_{\text{sn}}^{1-\tau} := \left[\hat{z}(1)\exp\left(-\sqrt{V_{t_0,1-\tau}\int_{t_0}^{1} t^2\ln^2\left(\frac{\hat{z}(t)}{\hat{z}(1)}\right)\mathrm{d}t}\right),\ \hat{z}(1)\exp\left(\sqrt{V_{t_0,1-\tau}\int_{t_0}^{1} t^2\ln^2\left(\frac{\hat{z}(t)}{\hat{z}(1)}\right)\mathrm{d}t}\right)\right],$$

where $z \in \{\text{CVaR}_{\alpha,t}, \text{CES}_{\alpha,t}\}$, $t_0 = 0.2$ as in Hoga (2019), $\hat{\sigma}_{\hat{\gamma}_H,\gamma} = \gamma^2$, $\hat{\sigma}_{\hat{\gamma}_{MR},\gamma} = 2\gamma^2$, $\Phi(\cdot)$ denotes the standard Gaussian cumulative distribution function, and $V_{t_0,1-\tau}$ represents the $(1-\tau)$-quantile of V_{t_0} (see Table 4). For standard Brownian motion $W(\cdot)$, $V_{t_0} := W^2(1)/\int_{t_0}^{1}(W(t)-tW(1))^2\,\mathrm{d}t$ is the asymptotic distribution of the self-normalised expression used to derive $I_{\text{sn}}^{1-\tau}$.

Put simply, these intervals are derived from the limiting distribution of the extrapolation factor in the Weissman (1978) estimator, the variance of which dominates that of the order statistic and ARMA-GARCH parameters under certain conditions as further elaborated upon in Section 3.6. For root-T consistent ARMA-TGARCH parameter estimates, $\lim_{T\to\infty} \alpha T/l = 0$, and $\lim_{T\to\infty} \ln(\alpha T/l)/\sqrt{k} = 0$ (*nota bene*, $\sqrt{k}$ corresponds to the Weissman (1978) estimator that still relies on k observations), we may conjecture that the underlying theorem still applies, such that the asymptotic confidence intervals $(1-\tau)$-confidence intervals based on the estimator in Equation (6) follow as

$$I_{\text{na}}^{1-\tau} := \left[\hat{z}(1)\exp\left(-\Phi(1-\tau/2)\hat{\sigma}_{\hat{\gamma},\gamma}\frac{\ln(l/(n\alpha))}{\sqrt{k}}\right),\ \hat{z}(1)\exp\left(\Phi(1-\tau/2)\hat{\sigma}_{\hat{\gamma},\gamma}\frac{\ln(l/(n\alpha))}{\sqrt{k}}\right)\right],$$
$$I_{\text{sn}}^{1-\tau} := \left[\hat{z}(1)\exp\left(-\sqrt{V_{t_0,1-\tau}\int_{t_0}^{1} t^2\ln^2\left(\frac{\hat{z}(t)}{\hat{z}(1)}\right)\mathrm{d}t}\right),\ \hat{z}(1)\exp\left(\sqrt{V_{t_0,1-\tau}\int_{t_0}^{1} t^2\ln^2\left(\frac{\hat{z}(t)}{\hat{z}(1)}\right)\mathrm{d}t}\right)\right],$$

where $\hat{z}(\cdot)$ is based on $Q_l^U(1-\alpha)$ rather than $Q_{\text{WM}}^U(1-\alpha)$. This new asymptotic result allows for more flexibility in the bias-variance tradeoff in the Hill (1975) estimator and the tradeoff between uncertainty in the order statistic and the extrapolation factor of the Weissman (1978) estimator.

3.6 Confidence interval derivation for CVaR and CES

In order to substantiate our asymptotic results, we now provide a condensed overview of the steps in the proof in Appendix A of Hoga (2019) , from which the author's CIs are derived. Of the mild assumptions made, several relate to our TGARCH and extrapolation location extensions. Firstly,

ARMA and GARCH models must be (strictly) stationary and invertible. Secondly, the estimation of their parameters must be root-T consistent. Moreover, the author stresses the importance of consistent parameter estimation for the validity of subsequent interval forecasts, emphasising the importance of proper model specification. Finally, α is an integer sequence that adheres to $\lim_{T\to\infty} \alpha T/k = 0$ and $\lim_{T\to\infty} \ln(\alpha T/k)/\sqrt{k} = 0$. Regarding the asymptotically vanishing expressions, k corresponds to the extrapolation location, whereas $\sqrt{k}$ follows from the convergence rate of the Weissman (1978) estimator. The former limit is required for the Weissman (1978) estimator to be valid, whereas the latter is required for the uncertainty in the extrapolation factor, as opposed to the order statistic, to dominate in the asymptotic distribution.

Starting from the CVaR and CES estimators in Equation (4), root-T consistency implies that uncertainty in the conditional mean and variance vanish, such that only the tail risk of the standardised residuals is required as in Equation (5). Hence, finding the limiting distribution of the CVaR estimator of a process reduces to transforming the quantile forecast of the standardised residuals. This is the case not only for CVaR, but also for CES, as this expression is asymptotically equivalent to a function of CVaR and the tail index (see Equation (5)), where uncertainty in the CVaR estimator dominates. Subsequently, $\lim_{T\to\infty} \ln(\alpha T/k)/\sqrt{k} = 0$ ensures that the extrapolation factor in the Weissman (1978) estimator dominates the order statistic, such that the limiting distribution is determined by that of the extreme value index estimator. Hoga (2019) finds that the limiting distribution of the tail index estimator is a scaled Brownian motion with convergence rate $\sqrt{k}$, which can be approximated by means of a normal approximation or self-normalisation to obtain approximate distributions of which we can deduce asymptotically valid quantile estimates.

Extending these findings to an ARMA-TGARCH framework with our generalised Weissman estimator likely only requires minor adaptations. We can hypothesise that root-T consistency of (strictly) stationary, invertible ARMA and TGARCH processes still ensures that we may restrict our attention towards quantiles of the standardised residuals and that under $\lim_{T\to\infty} \alpha T/l = 0$, and $\lim_{T\to\infty} \ln(\alpha T/l)/\sqrt{k} = 0$, the generalised Weissman estimator remains asymptotically valid and its extrapolation factor dominates in the asymptotic distribution. This, in turn, leaves us with the asymptotic distribution of the tail index, which is unchanged as compared to Hoga (2019).

3.7 Simulation study

As a proof of concept, the CVaR and CES forecasts based on our generalised estimator are compared against the findings in Hoga (2019) based on the Weissman (1978) estimator.

3.7.1 Data

The quantiles of V_{t_0} are determined based on simulating a million instances of a Brownian motion with ten thousand time steps. Then, ten thousand series of length one thousand each are generated based on the three data generating processes given by

$$y_t = \varepsilon_t, \quad \sigma_t^2 = 1.5079 + 0.15 \cdot \varepsilon_{t-1}^2 + 0.8 \cdot \sigma_{t-1}^2, \quad U_t \sim st_3(5), \tag{M1}$$

$$y_t = \varepsilon_t, \quad \sigma_t^2 = 3.2 \cdot 10^{-6} + 0.0349 \cdot \varepsilon_{t-1}^2 + 0.9373 \cdot \sigma_{t-1}^2, \quad U_t \sim st_{4.2}(0), \tag{M2}$$

$$y_t = 0.2714 y_{t-1} + \varepsilon_t, \quad \sigma_t^2 = 3.4 \cdot 10^{-6} + 0.1407 \cdot \varepsilon_{t-1}^2 + 0.7914 \cdot \sigma_{t-1}^2, \quad U_t \sim st_{5.3}(0.8531), \tag{M3}$$

where $\varepsilon_t = \sigma_t U_t$ and $st_\nu(\xi)$ represents the skewed Student's t distribution with skewness parameter ξ and ν degrees of freedom (Azzalini & Capitanio, 2003).

3.7.2 Identification, estimation, and forecasting

Following Hoga (2019), we fit AR(1)-GARCH(1,1) models with the Hill and MR estimators of γ to the empirical log-returns using Laplace Quasi Maximum Likelihood Estimation (Hoga, 2019) with a burn-in period of ten observations for initialisation effects. Moreover, as it may occur that $\hat{\gamma}(t)$ exceeds one, thereby invalidating the ES forecast $\widehat{\mathrm{ES}}_i^U(\alpha) = \hat{Q}_i^U(1-\alpha)/(1-\hat{\gamma})$, we instead use $\widehat{\mathrm{ES}}_i^U(\alpha) = \hat{Q}_i^U(1-\alpha)/\max(0.1, 1-\hat{\gamma})$, where $i \in \{l, \mathrm{WM}\}$. Consequently, γ and k^* are estimated as described in Section 3.3 and Section 3.4 with k^* being chosen in a range of 5% to 20% of observations following Hoga (2019). This forecasting performance evaluation requires one-step-ahead forecasts of the conditional mean, by taking the expectation conditional on information until time T using information set $\mathcal{I}_T$, and conditional variance, which is deterministic for ARCH-type models such as TGARCH, given by

$$\hat{y}_{T+1|T} = \mathbb{E}(y_{T+1}|\mathcal{I}_T) = \sum_{i=1}^{\overline{p}} \hat{\phi}_i \hat{y}_{T+1-i} - \sum_{j=1}^{\overline{q}} \hat{\theta}_j \hat{\varepsilon}_{T+1-j},$$

$$\hat{\sigma}_{T+1|T}^2 = \hat{\omega} + \sum_{i=1}^{p} \hat{\varphi}_i \hat{\sigma}_{T+1-i}^2 + \sum_{j=1}^{q} \hat{\vartheta}_j \hat{\varepsilon}_{T+1-j}^2 \text{ for GARCH, and}$$

$$\hat{\sigma}_{T+1|T} = \hat{\omega} + \sum_{i=1}^{p} \hat{\varphi}_i \hat{\sigma}_{T+1-i} + \sum_{j=1}^{q} \hat{\vartheta}_j (|\hat{\varepsilon}_{T+1-j}| - \hat{\varrho}_j \hat{\varepsilon}_{T+1-j}) \text{ for TGARCH,}$$

where $\hat{\cdot}$ refers to the estimator of that variable in Equations (1), (2), and (3).

For 0.5%, 1%, and 2.5% right-tail quantiles, the CVaR and CES point forecasts are evaluated based on bias and root-mean-square error (RMSE), whereas the 95% confidence intervals are as-

sessed in terms of coverage and interval length. Since the choice of l^* is of particular interest, the accuracy measures are also evaluated for an appropriate range of l in line with the assumptions required on l for the asymptotic confidence intervals to apply.

3.8 Empirical application

Most importantly, the generalised estimator should demonstrate superior performance in real-life data. As such, we employ an empirical investigation with a focus on the left tail of the returns.

3.8.1 Data

The six global indices under consideration are the two largest in terms of daily traded value in The Americas, Europe, and Asia, vis. the Standard & Poor's 500 (S&P 500), NASDAQ 100, Nikkei 225, Hang Seng Index (HSI), FTSE 100, and DAX 30, ranging from 2001 to 2020, covering about five thousand observations. For model and parameter selection, these series are used from 1990 to 2000. The log-returns are used with characteristics as given in Table 1 and findings are contrasted with those in Hoga (2019) concerning 1997–2016 with the Dow Jones Industrial Average (DJIA) and CAC 40 instead of the S&P 500 and FTSE 100. All empirical data is downloaded from *Yahoo Finance* and *The Wall Street Journal*.

3.8.2 Identification, estimation, and forecasting

The approach set out in Section 3.7.2 is also applicable to the empirical data, except that a simple GARCH(1,1) with the Hill (1975) estimator is included and a moving window of a thousand observations is applied in the estimation and forecasting procedure to allow for time-varying parameters. Subsequently, we add a leverage effect to the GARCH component resulting in the TGARCH model to facilitate asymmetric effect of innovations. We implement the ARCH-type misspecification tests proposed in Engle & Ng (1993) that focus on asymmetry. Additionally, verification of the robustness of the selection of l^* is done by visually inspecting plots of this scoring function for varying values of l for the 1990–2000 series to see whether a clear global minimum can be distinguished. Determining and evaluating the optimal l requires splitting the data up into a fitting set to fit the model, a selection set to choose the optimal l, and a test set to assess true out-of-sample performance. To avoid an unfair advantage in the number of observations used, we stick to the moving window of a thousand observations, the first five hundred of which are used for fitting, and the remaining five hundred for forecast evaluation aimed at minimising the Quantile score to obtain l^*. Then, an out-of-sample one-step-ahead forecast is made, based on the optimal extrapolation location and autoregressive parameters fitted to the complete moving window, to be tested jointly with the other

out-of-sample forecasts. Extrapolation location l^* is selected as described in Section 3.4 and is thus determined by finding the minimum of the realised Quantile scores of the resulting point forecasts corresponding to a range of values for l. For the purposes of computational feasibility, l^* is estimated at a lower frequency — that is, every one hundred observations. Moreover, point forecasts are described in Section 3.7.2 and forecast evaluation is directed more towards quantifying the validity of the forecasts. Point forecasts are assessed by means of scoring functions that represent a loss to be minimised for CVaR, given by the argument in Equation (8), and for CVaR and CES jointly, specified in Taylor (2019) as the average between t_1 and t_2 of

$$S_2(\widehat{\text{CVaR}}_{\alpha,t-1}, \widehat{\text{CES}}_{\alpha,t-1}, y_t) := -\ln\left(\frac{\alpha-1}{\widehat{\text{CES}}_{\alpha,t-1}}\right) - \frac{\left(y_t - \widehat{\text{CVaR}}_{\alpha,t-1}\right)\left(\alpha - I_{\{y_t \geq \widehat{\text{CVaR}}_{\alpha,t-1}\}}\right)}{\alpha\widehat{\text{CES}}_{\alpha,t-1}}.$$

CVaR interval forecasts are subjected to statistical tests for correct unconditional and conditional coverage (UC & CC) as described in Christoffersen (1998), where the null-hypothesis of correct (un)conditional coverage may be rejected. Essentially, unconditional coverage tests verify that the total fraction of violations does not differ significantly from the nominal level, whereas conditional coverage tests take this a step further and test whether the fraction of violations, conditional on whether past realisations constituted violations, deviates significantly. As such, the conditional coverage test is comprised of an unconditional coverage element, as well as a test for independence of CVaR violations (Christoffersen, 1998). Importantly, we evaluate these performance measures for the cases of several choices for the extrapolation location with two-sided Diebold-Mariano (DM) tests (Diebold & Mariano, 2002) and a Bonferroni correction (Weisstein, 2004) for the scoring functions to determine whether a statistically significant improvement is achieved with the implementation of the tuning threshold.

4 Principal findings with l = k*

To provide the necessary background, the key results from the simulation study in Hoga (2019) are set out, followed by those of the author's empirical application in the context of the indices and period described in Section 3.8.1. First, however, we briefly expand on the data characteristics to substantiate the methodology. Table 1 contains the summary statistics of the daily log-returns for the six indices used from 2001 to 2020.

As evidenced by Table 1, the daily (log-)returns (recall $\ln(1+r) \approx r$ for small r) are on average positive for most indices with considerable variance. Most indices are negatively skewed,

Table 1: Daily log-return summary statistics for six indices from 2001 to 2020.

	Mean	SD	Skew	Kurt	Low	High	AC	S-AC	JB	Obs
S&P 500	-2.1e-4	1.2e-2	0.42	11.7	-0.11	0.13	-0.12	0.33	2.9e3	5031
NASDAQ	3.4e-4	1.5e-2	-0.14	7.4	-0.13	0.13	-0.07	0.27	1.1e4	5030
Nikkei	1.4e-4	1.5e-2	-0.38	6.5	-0.12	0.13	-0.03	0.28	8.9e3	4896
HSI	1.2e-4	1.4e-2	-0.07	8.5	-0.14	0.13	-0.13	0.37	1.5e4	4923
FTSE 100	-8.8e-6	1.2e-2	0.34	8.4	-0.09	0.13	-0.04	0.22	1.5e4	5145
DAX 30	1.5e-4	1.5e-2	-0.18	6.0	-0.13	0.11	-0.01	0.14	7.7e3	5072

Note. Given are the mean, standard deviation (SD), skewness (skew), kurtosis (kurt), lowest (low) and highest (high) observation, first order autocorrelation (AC), first order autocorrelation in the squared series (S-AC), Jarque-Bera statistic (JB), and the number of observations (obs). The values for JB and S-AC are all significant at 1%, whereas no values for AC are at 10%.

implying frequent moderate positive returns with occasional severe losses. Moreover, the log-returns are fat-tailed (leptokurtic), a well-known stylised fact of log-returns (e.g. Cont et al. (2001)) where extreme quantiles are relatively likely. These metrics highlight the need to model extreme quantiles, particularly losses, which have more probability mass and range up to 13% for the average index. Since the first order autocorrelation is not significant, modelling the conditional mean will likely not prove convincingly beneficial. In contrast, that of the squared returns, which can be used as a proxy for the variance (Patton, 2011), is highly significant, evidencing the stylised facts described in Section 2.1. Moreover, it is positive, hinting at volatility clustering. This heteroskedasticity can be filtered using a GARCH model, which models part of the excess kurtosis compared to a Normal distribution (see Section 2.1), but the degree of excess kurtosis and skewness observed justifies the use of the skewed Student's t distribution.

4.1 Simulation study

Reproducibility of published scientific work is of utmost importance to the validity of its claims. Hence, we note that all our results demonstrate a high degree of agreement with those of Hoga (2019), especially concerning those on which the author's conclusions hinge primarily (Table 5). Except for the bias and RMSE, values are off by at most a few per cent and all statistically significant results remain as such.

In this section, we restrict our attention to the case where $l = k^*$ as tabulated in Table 5, for which Hoga (2019) concludes that bias increases as α becomes more extreme for the Hill estimator, especially for CES estimates. In addition, the RMSE for CES exceeds that of CVaR and both decrease in α, while there is little difference between the RMSE for the Hill and MR estimators, which contradicts the theoretical result that the asymptotic RMSE for the Hill estimator should be half that of the MR estimator (Hoga, 2019). Moreover, by and large, the substantial undercoverage for

$I_{\text{na}}^{0.95}$ is substantially improved upon through self-normalisation, but is still below the nominal 95% coverage. A plausible cause is that the parameter uncertainty vanishes only asymptotically, leading to underestimating the confidence bounds for finite samples. Surprisingly, the normal approximation CI exhibits the best coverage for CVaR estimates, whereas self-normalised CIs perform best for CES, which one would expect to be more complicated to estimate. Lastly, self-normalised confidence intervals are much wider than those of the normal approximation, which improves coverage but also causes issues regarding implementation, where narrow intervals are desirable.

In the end, the validity of $I_{\text{na}}^{0.95}$ and $I_{\text{sn}}^{0.95}$ largely depends on the finite-sample performance of the asymptotic distributions in Hoga (2019) (Eq. (24) & (25)), for which probability-probability (PP) plots are presented in Figure 5 that indicate the superior fit for the self-normalised (Eq. (25)) asymptotic distribution for both the Hill and MR estimators as compared to the normal approximation (Eq. (24)). Based on Figures 6 and 7, Hoga (2019) also concludes that the choice of k^* is crucial in obtaining satisfactory coverage probabilities, bias, and RMSE, particularly for the Hill estimator, and that on average the selection criterion in Equation (7) performs excellently.

These findings suggest the use of self-normalised confidence intervals using the Hill estimator and k^* as in Equation (7) as a benchmark for our extrapolation location generalisation.

4.2 Empirical application

Table 6 contains the performance of the model proposed by Chan et al. (2007), which is a GARCH(1,1) with $k = \lfloor 1.5(\ln T)^2 \rfloor$ and the Hill estimator (abbreviated G-k-H), and of the AR(1)-GARCH(1,1) for both the Hill and MR estimators with k as in Equation (7) (abbreviated AG-k^*-H and AG-k^*-MR respectively) as used by Hoga (2019) . Regarding CVaR interval forecasts, correct unconditional coverage is only rejected twice for AG-k^*-MR, whereas correct conditional coverage is rejected for all but one series for AG-k^*-MR at 1 or 5%. It is rejected at 10% for half of the data sets based on G-k-H and never using AG-k^*-H. In terms of point forecasts, AG-k^*-H significantly outperforms G-k-H the majority of the instances, and it scores lowest ten out of twelve instances, with AG-k^*-MR performing best in the other two. Notably, we see no reason to differentiate between CVaR and joint forecasts in terms of forecast quality in contrast to Hoga (2019). Applying this analysis to data from 2001 till 2020 (see Table 7), rather than 1997–2016, with two alternative indices, we observe that coverage has considerably deteriorated as almost half of the coverage tests are significant. Again, CC is rejected more often and AG-k^*-H performed best. For the scoring functions, AG-k^*-H still dominates — obtaining the lowest score in all twelve combinations of data sets and scoring functions — but statistical outperformance occurs less frequently and only once for the

CVaR scoring function.

Clearly, absolute performance of all methods and relative performance of AG-k^*-H are less convincing for more recent data, which may well be due to the COVID-19 pandemic. In line with the Great Moderation, which some argue continued after the Great Recession (e.g. Gadea et al. (2018) and Perron & Yamamoto (2021)), typically positive returns with relatively little variation preceded crashes of many indices in early 2020, followed predominantly by profits again. As such, the rolling window approach on moderately varying data led to underestimating the extreme quantiles at the start of the pandemic, which in turn prompted overestimating extreme losses in the remaining observations. In Figure 10, we observe that a moving average ($\sim$1 year) of the squared standardised residuals for the NASDAQ exhibit periods of excess and lacking volatility vis-à-vis the much more stable pseudo-random *iid* squared shocks associated with this series (Hoga, 2019) in Figure 8, particularly following periods of turmoil in financial markets. This illustrates the insufficient robustness of the ARMA-GARCH filter employed to black swan events such as a pandemic, which is essential when modelling the economics of extremes governed by the consequences of such extreme events.

In line with simulation results, but less convincingly so, Table 7 points towards AG-k^*-H as the superior model to utilise as an exemplar of satisfactory performance in the forthcoming comparisons.

5 Model selection

Prior to forecast evaluation, the required consistency of the autoregressive process dictates that the formulation of the applied filter be addressed. Additionally, we provide a theoretical and empirical basis for potential extrapolation location choices.

5.1 Autoregressive specification

In light of the focus on one tail only, it is essential to facilitate asymmetry with respect to the sign of innovations in their effect on the conditional volatility. To that end, an established alternative to the GARCH model that does allow for asymmetry is the APARCH model that nests many ARCH-type models, amongst which the most well-known are GJR-GARCH and TGARCH (see Section 2.1). APARCH reduces to the former for a power term equal to two, and to the latter when the power term is one. Estimation of the 1990–2000 series indicates that this power term is quite narrow around and close to one (1.27 average), suggesting that using TGARCH is a sensible simplification and robustification. A set of rather general ARCH-type adequacy tests with an emphasis on asymmetry was proposed by Engle & Ng (1993). The p-values for the three tests and 1990–2000 series are tabulated in Table 8 and illustrate TGARCH superiority. For the AR(1)-GARCH(1,1), eleven out

of eighteen tests reject an adequate fit at 1% significance, whereas none are rejected for AR(1)-TGARCH(1,1). Thus, the forthcoming empirical analyses rely on the TGARCH filter rather than the GARCH filter.

As further inspected in Section 6.2, a comparison between Tables 3 and 7 reveals that incorporating asymmetries by means of a TGARCH formulation for heteroskedasticity brings about substantial improvements in the loss functions for all indices except the NASDAQ. For an explanation, we may turn to an economic perspective and entertain cross-country heterogeneity in risk aversion as a potential mechanism both driving this improved performance and explaining why no such improvement is observed for the NASDAQ. In a recent survey regarding risk attitude towards financial investment and covering most of the Western World, Ferreira (2018) finds that investors in the United States are on average risk-neutral, whereas an incontestable majority of investors in all European countries studied, including the United Kingdom and Germany, are risk-averse. Moreover, a similar inquiry by Gandelman et al. (2015) uncovered that individuals from Asian countries have decidedly more conservative risk appetite than their American counterparts. Under the premise of increased exposure to indices corresponding to the country of residence for the average investor, for example owing to a perceived informational edge or to avoid currency risk, distributional asymmetries are expected in Europe and Asia more so than the US. That is, investors have been found to react more severely to negative news than positive news effectuated by risk-aversion (Kim & Lee, 2008), leading to asymmetries in conditional volatility depending on the sign of past innovations. Hence, European and Asian indices exhibit more noticeable differences in volatility reactions, mostly through sell-offs driven by risk aversion after negative shocks, which underlies the superior TGARCH performance observed. For the NASDAQ, however, this asymmetry is not expected because American investors are distributed approximately symmetrically around risk-neutrality (Ferreira, 2018). Being the world's most traded and international index, this is less applicable to the S&P 500 as it is to a larger extent subject to international investor sentiment, potentially explaining why incorporating asymmetries does improve S&P 500 CVaR and CES forecast performance.

5.2 Extrapolation location selection

A crucial element in extending the tail risk estimators by means of a tuning threshold is the selection of the optimal value thereof. The furthest quantiles are of most interest because these are most likely to improve under alternative choices of l because of the superlinear growth of the uncertainty in the extrapolation factor (see Appendix C). Hence, we focus on the case where $\alpha = 0.005$.

5.2.1 Simulation study

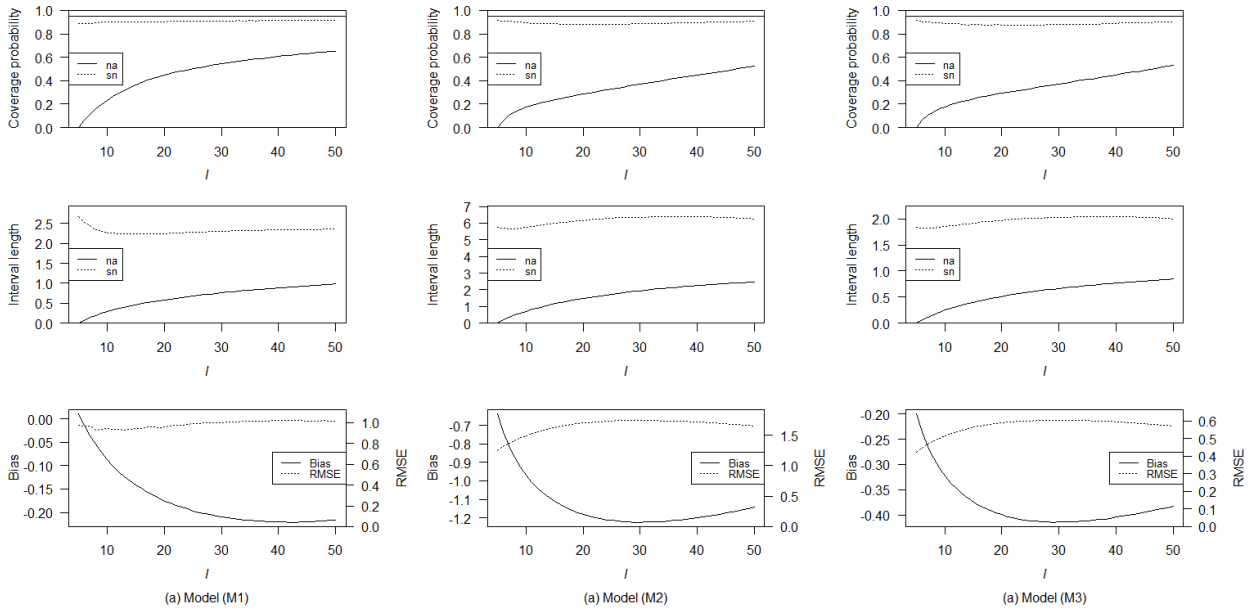


Figure 1: CVaR forecast performance measures for models (M1), (M2), and (M3).

Depicted in Figure 1 (and in Figure 9 of Appendix B for the alternative models described in Appendix B of Hoga (2019)) are the bias and RMSE for the CVaR point forecasts, and interval length and coverage probabilities for the 95% CIs from a normal approximation and self-normalisation as in Section 3.5. Clearly, normal approximation coverage is abysmal for small l but comparable to Hoga (2019) for large l due to the assumption of $\lim_{T\to\infty} \alpha T/l = 0$ being satisfied more adequately as l increases for given α and T. On that account, the remainder of the analysis is restricted to the self-normalised CIs. The effectiveness of the extrapolation location depends largely on the shape of the distribution used and the extent to which it deviates from the theoretical Pareto distribution. As such, we ought to consider a wider range of potential distributions. Appendix B of Hoga (2019) contains the model specifications of processes with a Burr distribution and a heavily skewed Student's t distribution, such that the finite sample Pareto shape is diverged from considerably. Performance measures are depicted in Figure 9 similar to Figure 1 and exhibit a high level of agreement between models (M1), (M2), and (M3) on the one hand, and the alternative processes on the other. However, Figure 9 more strongly suggests using a small l. Throughout the depicted range for l, self-normalised CI coverage is excellent and interval length seems lowest around ten and bias and RMSE decrease in magnitude as l becomes closer to αT or k^*. Concluding the simulation study regarding the location parameter, this suggests considering the ordinary Weissman

(1978) estimator ($l = k^*$), Filtered Historical Simulation ($l = \alpha T$), and extrapolation for relatively small l ($l = 2\alpha T$). However, no choice proves conclusively superior.

5.2.2 Empirical application

Since the ultimate objective of scientific inquiry is generally the application to real-life situations, identifying the optimal extrapolation location in empirics is the principal challenge at hand. To this end, Figure 2 contains the CVaR and joint loss functions for six indices between 1990 and 2000. These are plotted for a range of l up to the lower bound of k^* to locate the optimal tuning threshold and highlight differences between CVaR and joint performance.

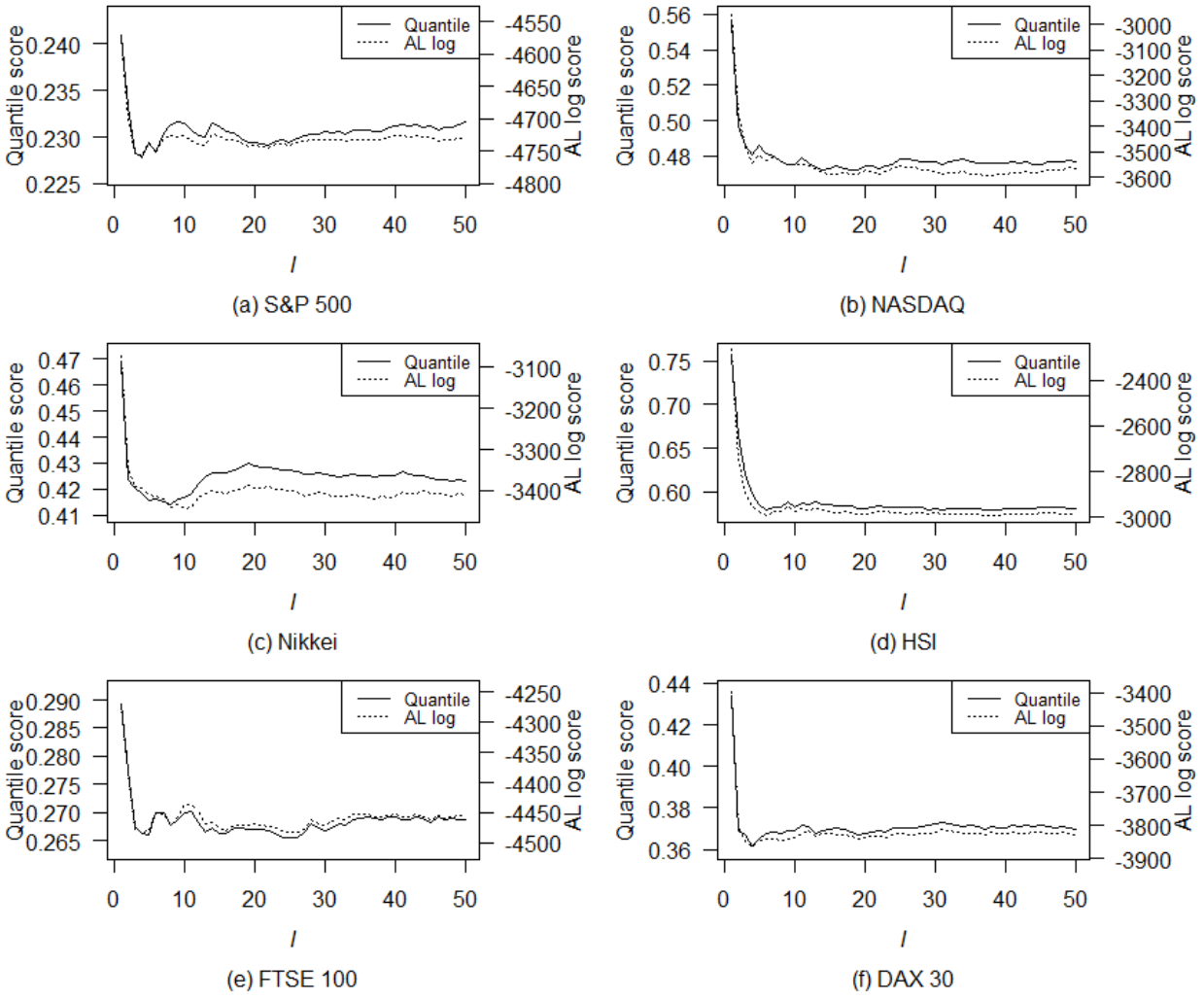


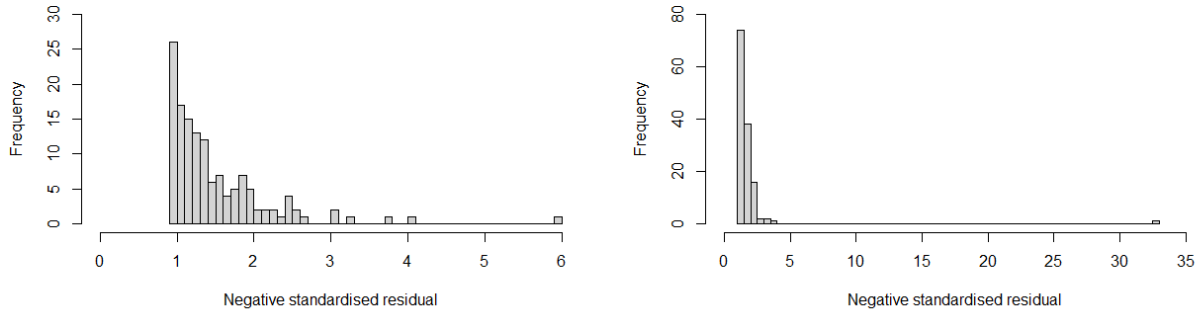
Figure 2: CVaR (solid) and joint (dotted) loss scores for six indices from 1990 to 2000 and a range of fixed extrapolation locations.

Several insightful features arise from Figure 2. The main similarity between performance on the varying indices is that a clear decline takes place below $l = \alpha T$. This behaviour stems from the fact

that, here, the uncertainty in the order statistic dominates that of the extrapolation factor, making corresponding point forecasts considerably less reliable. Subsequently, the S&P 500, HSI, and DAX 30 reach their lowest points, making a case for the consideration of $l = \alpha T$. An additional benefit of this extrapolation location choice is that it is equivalent to Filtered Historical Simulation, the comparison with which is unaddressed in Hoga (2019). Because of its simplicity and solid theoretical foundation, benchmarking against FHS is certainly called for when investigating more intricate methods that for all intents and purposes nest FHS in terms of point forecasts (vis. for $k^* = \alpha T$ or $l^* = \alpha T$). For the Nikkei, the lowest scores are attained around $l = 2\alpha T$, which is therefore chosen as the fixed $l < k^*$ from which extrapolation takes place. Because most indices either reach the lowest scores for a larger l or have several extrapolation locations with similar corresponding loss scores, the final choice of l is data- and time-dependent as described in Section 3.7.2 to accommodate differences between series and over time. Finally, we treat $l = k^*$ not only because performance exhibits low losses as l approaches k^*, but also because generalisations should be implemented exclusively in case they provide superior performance. Except for our data-driven l^*, these location parameter choices are in line the simulation results in Section 5.2.1 and illustrate a relatively low degree of robustness and univocality in identifying a global minimum.

Notable in Figure 2 is the variety in co-movement of CVaR and joint loss functions depending on the index, which calls attention to CES forecast quality. For instance, Nikkei scores diverge to an appreciably larger extent than do NASDAQ scores. Since CES forecasts are based solely on CVaR forecasts and the corresponding Weissman (1978) estimator, this hints at distributional concerns that affect the tail index and lead to deviations from the asymptotic tail shape from which the CES estimator is derived, rather than inaccuracies in CVaR estimates. To this end, we may consider the standardised residuals in Figure 3 corresponding to the 5% most extreme losses, which by construction, regardless of k^*, are located in the tail and ought to follow a Pareto shape. Whereas NASDAQ extreme residuals do seem to exhibit the polynomial decay characteristic of the Pareto distribution, the Nikkei contains a single exceptionally large outlier that is tremendously unlikely given the remaining observations and theoretical Pareto shape. This aberrant observation corresponds to the unexpected burst of the Japanese asset price bubble that preluded Japan's 'Lost Decade' — the period from 1991 to 2000 in which stagnation in economic growth followed the evaporation of half of the market capitalisation of the Nikkei index (Hayashi & Prescott, 2002). The discrepancy between Quantile and AL log scores arises from the fact that quantile forecasts vary less with changes in the Weissman (1978) estimator than the asymptotic tail

volume beyond that threshold. These findings emphasise the importance of residual analysis and the benefits of robustification of the extreme value index estimator used. Interestingly, implementation of robust Hill estimator $\hat{\gamma} = \frac{1}{k^* - \lfloor \alpha T \rfloor} \sum_{t=\lfloor \alpha T \rfloor}^{k^*} \ln U_{T-t:T} / U_{T-k^*:T}$, where the robust k^* follows as $k^* := \underset{k \in \{k_{\min}, \ldots, k_{\max}\}}{\arg\min} \left(\sup_{t \in \{\lfloor \alpha T \rfloor, \ldots, k_{\max}\}} \left| U_{T-t:T} - U_{T-k:T} \left(\frac{t}{k}\right)^{-\hat{\gamma}} \right| \right)$, does improve all Nikkei joint losses for the CVaR and CES forecast evaluation in Table 3, but further investigation is beyond the scope of this paper.



(a) NASDAQ (b) Nikkei

Figure 3: Negated standardised residuals of the left tail (5%) from 1990 to 2000 for two indices.

6 CVaR and CES forecast evaluation

After having treated the results based on the methods in Hoga (2019), model selection and extrapolation location choices, we turn to forecast evaluation to explore the performance of our generalised estimators and contrast with recent literature.

6.1 Simulation study

Before going into all differences in performance for various choices of l, which is of primary interest, we observe from Table 2 that the various methods are more similar than different in that the choice of the extreme index estimator and k^* are more influential on the resulting performance than the choice of l within any one such method. However, notable dissimilarities are present and relevant for the choice of l. The analysis is carried out for both CVaR and CES forecasts, which are generally in accordance, so deviations are mentioned explicitly. The same holds for the three extreme quantiles under consideration. Firstly, the bias can be seen to increase as l grows, hinting at larger deviations from the Pareto shape as extrapolation is done from less extreme quantiles. Secondly, RMSE tends to be stable regardless of l. Since Mean-Squared Error is comprised of the

variance and the squared bias, the increasing bias implies that variance diminishes for larger l. This contradicts the asymptotic result in Hoga (2019) that the uncertainty in the extrapolation factor dominates, but we may hypothesise that αT and $2\alpha T$ are excessively small for this finite sample of a thousand observations, such that the model assumptions are not satisfied adequately. Thirdly, coverage levels are highly similar across the various choices of l and display minor to moderate undercoverage (74.5%-94.5% for nominal 95%) with differences arising depending on the model, quantile, and risk measure. The most substantial impact is from the model choice, where (M1) yields coverage levels considerably poorer than (M2) and (M3). Moreover, coverage improves as α shrinks, which is in line with expectations because the CIs are derived from asymptotic theory where $\alpha \to 0$ as T becomes large. Regarding risk measures, we find that coverage levels for CES are superior to those for CVaR, which is an advantageous feature as regulatory bodies shift their focus more towards CES (see Section 1). Finally, confidence intervals become more narrow for relatively large l, embodying the lower forecast uncertainty as deduced from the RMSE, and for less extreme quantiles.

Because the performance measures for CVaR and CES are predominantly similar, findings remain largely in line with Hoga (2019). However, for the purposes of evaluating the most appropriate choice of l, several desiderata are of importance. If one aims to obtain unbiased point forecasts, opting for relatively small l is most appropriate, whereas minimising forecast uncertainty would require implementing l closer to k^*, which by extension also holds for applications where narrow confidence intervals are sought after. Based on performance in the simulation study, we therefore advise to set out the purpose of the risk measure estimates before selecting the most appropriate extrapolation location accordingly.

6.2 Empirical application

With our goal of real-life application in mind, an empirical investigation for six major stock indices from around the world yields the results reported in Table 3. The two main topics treated in this section are the improvements brought about by allowing for asymmetries in the effect of shocks based on their sign and the comparison of the various potential extrapolation location choices.

As evident from Table 3, the Quantile scores and AL log scores are markedly lower based on a TGARCH-type specification vis-à-vis the GARCH-type specification used in Table 7 for all indices except the NASDAQ. In conjunction with the residual adequacy tests that help detect misspecification issues in Table 8, these findings provide persuasive evidence in favour of an AR(1)-TGARCH(1,1) specification for the data under consideration. Moreover, they call attention to

Table 2: CVaR and CES bias, RMSE, coverage probabilities (in %, nominal coverage is 95%) and interval length at three extreme quantiles using the Hill estimator with average k^* and various choices of l.

Model	k^*	α	l	CVaR				CES			
				Bias	RMSE	Cov.	Len.	Bias	RMSE	Cov.	Len.
(M1)	67	2.5%	k^*	-0.04	0.58	74.5	0.54	-0.09	0.71	83.2	1.01
			αT	-0.03	0.97	76.6	0.65	-0.08	1.11	84.7	1.11
			$2\alpha T$	-0.03	0.96	76.9	0.65	-0.07	1.11	84.2	1.10
		1%	k^*	-0.09	0.64	79.8	0.88	-0.10	0.80	86.7	1.59
			αT	-0.04	0.58	83.5	1.15	-0.04	0.70	87.8	1.77
			$2\alpha T$	-0.09	0.58	83.6	1.16	-0.09	0.71	88.9	1.86
		0.5%	k^*	-0.16	0.78	84.3	1.30	-0.09	0.97	89.2	2.19
			αT	-0.03	0.78	86.2	1.94	0.06	0.98	87.7	2.71
			$2\alpha T$	-0.09	0.76	87.6	1.96	-0.01	0.95	89.6	2.31
(M2)	59	2.5%	k^*	0.03	0.26	86.3	0.66	-0.14	0.51	93.5	1.80
			αT	-0.01	0.23	88.5	0.86	-0.20	0.52	93.4	2.24
			$2\alpha T$	0.05	0.22	88.5	0.83	-0.12	0.46	94.5	2.01
		1%	k^*	0.00	0.35	90.7	1.25	-0.39	0.83	94.2	3.41
			αT	0.01	0.35	89.9	1.63	-0.37	0.81	93.3	3.76
			$2\alpha T$	-0.06	0.36	90.2	1.66	-0.47	0.88	94.0	3.87
		0.5%	k^*	-0.10	0.50	93.1	2.06	-0.71	1.27	93.4	5.27
			αT	0.05	0.51	89.5	2.77	-0.47	1.11	92.9	5.77
			$2\alpha T$	-0.11	0.51	90.5	2.88	-0.71	1.26	92.7	5.67
(M3)	58	2.5%	k^*	0.02	0.10	82.7	0.28	-0.04	0.17	91.7	0.63
			αT	0.01	0.11	86.0	0.36	-0.06	0.20	92.5	0.76
			$2\alpha T$	0.03	0.11	85.5	0.35	-0.03	0.17	93.5	0.69
		1%	k^*	0.00	0.16	89.1	0.49	-0.13	0.30	93.8	1.13
			αT	0.01	0.16	87.8	0.64	-0.13	0.30	92.5	1.25
			$2\alpha T$	-0.02	0.16	89.0	065	-0.16	0.33	92.9	1.28
		0.5%	k^*	-0.04	0.21	92.8	0.78	-0.25	0.45	92.8	1.70
			αT	0.03	0.21	89.1	1.02	-0.15	0.38	93.4	1.83
			$2\alpha T$	-0.03	0.21	90.0	1.06	-0.24	0.44	92.4	1.78

Table 3: p-values of tests for (un)conditional CVaR coverage, Quantile scores (CVaR), and AL log scores (CVaR and CES jointly) all at $\alpha = 0.005$ for six indices and four choices of l from 2001 to 2020 with an AR(1)-TGARCH(1,1) specification.

Index	l	UC	CC	Quantile score	AL log score
S&P 500	k*	0.2497	0.9998	0.7466	-9592.5
	αT	0.9724	0.1255	**0.7287**	**-9644.3**
	$2\alpha T$	0.4811	0.9990	0.7451	-9603.4
	l^*	0.2497	0.9998	0.7454†	-9605.1†
NASDAQ	k*	0.4817	0.9986	0.9488	-8366.3
	αT	0.1261	0.1525	0.9290	-8379.7
	$2\alpha T$	0.7973	0.9953	**0.9422**	**-8415.9**
	l^*	0.4817	0.9986	0.9484	-8363.4
Nikkei	k*	0.9132	0.0038**	1.0721	-7529.2
	αT	0.2099	0.0079**	**1.0611**	**-7565.3**
	$2\alpha T$	0.9059	0.0044**	1.0717	-7559.7
	l^*	0.9132	0.0038**	1.0725	-7526.6
HSI	k*	0.4132	0.0999	**0.8990**	**-8293.6**
	αT	0.4435	0.1309	0.8999	-8284.5
	$2\alpha T$	0.7147	0.1101	0.9008	-8281.8
	l^*	0.5539	0.1051	0.8995	-8286.5
FTSE 100	k*	0.5485	0.9916	0.7910	-9740.2
	αT	0.9517	0.9916	**0.7748**	**-9851.3**
	$2\alpha T$	0.2074	0.9998	0.7910	-9750.5
	l^*	0.5485	0.9979	0.7894†	-9746.0
DAX 30	k*	0.0322*	0.0159*	**0.9686**	-8573.3
	αT	0.0097**	0.1580	0.9714	-8522.6
	$2\alpha T$	0.0097**	0.0186*	0.9715	-8565.0
	l^*	0.0181*	0.0172*	0.9694	**-8574.4**

Note. *p<0.05, **p<0.01, ***p<0.001; The p-values are calculated using tests for unconditional coverage (UC) and conditional coverage (CC) as described in Christoffersen (1998). A dagger (†) indicates a statistically significant two-sided Diebold-Mariano (DM) test (at 5%), for superior or inferior predictive performance with respect to $l = k^*$, as described in Diebold & Mariano (2002). A Bonferroni correction for multiple comparisons (m=36) is applied for the significance of the DM tests. For each data set and scoring function, the lowest score is in bold.

the importance of misspecification in extreme quantile estimation, in contrast to previous findings (see Section 2.1). For instance, FHS with a TGARCH filter for the ARMA residuals convincingly outperforms the more elaborate EVT approach without allowing for asymmetries. For most indices, this holds to such an extent that the change in the loss function brought about by the alternative heteroskedasticity specification exceeds the performance differential when comparing the Hill and MR estimators as in Hoga (2019).

Concerning the extrapolation location, correct unconditional coverage is only rejected for the DAX 30, for all four choices of l. Similarly, conditional coverage is rejected at 5% for the Nikkei for each method, and at 10% for the DAX 30 except when $l = \alpha T$. No multiple comparisons correction is applied, so these may be due to type I errors (false positive), although the clustering of the rejections adds to their credibility. The loss functions that quantify the point forecast quality indicate no unambiguously superior choice of l, with $l = \alpha T$ attaining the lowest score most frequently. This goes to show that asymptotic EVT methods are only worth so much, since FHS is more appropriate for large sample sizes, while asymptotic EVT results may not be achieved for smaller sample sizes. However, when we take k^* as a benchmark, statistically significant outperformance at 5%, accounting for multiple comparisons by means of a Bonferroni correction, does occur; in three instances performance is significantly different, exclusively in favour of $l = l^*$. This result is counterintuitive because the average loss differential between k^* and l^* tends to be low compared to that between k^* and αT or $2\alpha T$ and as such requires additional elaboration. Upon further inspection, the dissimilarity between l^* and k^*-based forecasts on the one hand, and those using αT and $2\alpha T$ on the other, proves to be the underlying cause. For example, the S&P 500 Quantile score DM statistic is based on standard deviations in the loss differentials, taking values of 2.5×10^{-6}, 5.7×10^{-7}, and 4.5×10^{-8} when comparing the loss for k^* to $l = \alpha T$, $l = 2\alpha T$, and $l = l^*$ respectively. Thus, for the difference in performance to be statistically significant, the loss differential must be more than a factor ten larger for $l = \alpha T$ and $l = 2\alpha T$ as compared to l^*, which is not the case. l^* regularly taking on values substantially closer to k^* is likely at the root of the increased degree of similarity in the losses attained for each observation.

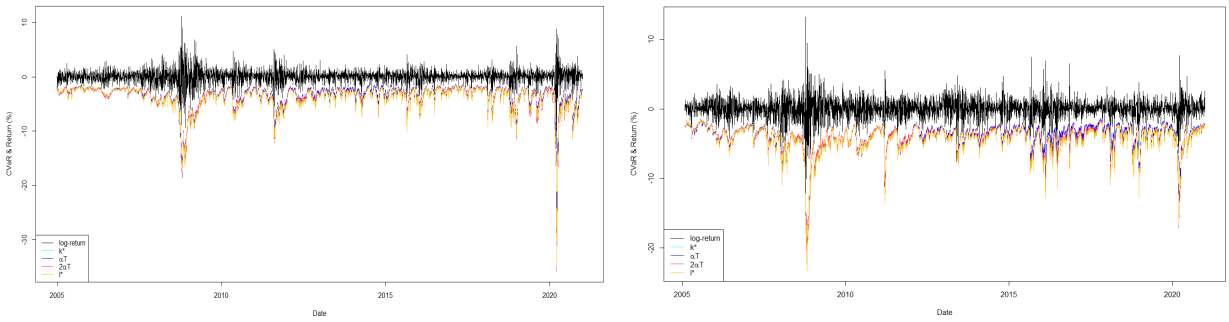
From an economic perspective, variation in the correctness of (un)conditional coverage across indices suggests over- or underidentification of the proportion of CVaR violations and excessive clustering of violations (and non-violations) for certain indices. Specifically, both conditional and unconditional coverage are inadequate for the DAX, whereas only CC is rejected for the Nikkei. To compare against a benchmark with correct coverage, consider for instance the NASDAQ, where

correct coverage is not rejected. Recall that UC concerns the fraction of CVaR violations, whereas CC is regarding the probability of such violations conditional on the past. The latter contains an element of independence of violations, which is violated in case correct UC is not rejected, whilst correct CC is. Note that this does not apply vice versa; correct CC may be rejected absent temporal dependence between CVaR (non-)violations as long as unconditional coverage is inadequate.

Applying these principles to the NASDAQ, Nikkei and DAX, we observe satisfactory coverage for the NASDAQ, while the Nikkei exhibits violation clustering. This can be inferred from correct UC but inadequate CC and is illustrated in Figure 12. The proportion of violations — 19, 25, 20, and 19 out of 3896 for l equal to k^*, αT, $2\alpha T$, and l^* respectively — is exceptionally close to the desired 0.5%, although less so for $l = \alpha T$ at 0.64%, but clustering in the CVaR violations can be observed when including five lags, which is a typical working week. In particular, in the aftermath of the Tōhoku earthquake and tsunami that effected the Fukushima Daiichi nuclear plant meltdowns, the Nikkei experienced two consecutive days of severe losses. In contrast to recessions such as the Great Recession and COVID-19 Recession, where volatility estimates adjust sufficiently rapid to avoid violation clustering, the nuclear disaster was a black swan event pur sang and proved a challenge for the ARMA-TGARCH filter. This is visible in Figure 4 (enlarged in Figures 14 and 15), which displays Nikkei returns and CVaR forecasts, where recessions lead to major spikes in CVaR forecasts with long memory that are much less outspoken around the Fukushima meltdowns. Similarly, the Nikkei plunged with two extreme negative returns that violated CVaR forecasts in late May of 2013 as concerns over the sustainability and effectiveness of Abenomics prompted sell-offs in anticipation of reduced government spending on securities. Both events that contributed to the volatility clustering observed in Figure 12 are specific to Japan and are not observed for the NASDAQ in Figure 14, which could explain why only the Nikkei has excellent UC, but inadequate CC. This is confirmed by inspecting the NASDAQ CVaR violations in Figure 11, which have similar violation fractions, but these are more evenly spaced over time. Such unexpected events with recurring effect on stock markets occur for many indices, but are scarce. For example, the September 11 attacks caused record losses in New York Stock Exchange history but their impact is not included in the forecast evaluation for American indices as it predates the time period evaluated.

An analogous comparison for the DAX (see Figure 13), however, does not yield conclusive results because violations are not as evenly spaced as for the NASDAQ, particularly during the COVID-19 pandemic, but are not as close in succession as those for the Nikkei. On the other hand, insufficient UC could be the primary culprit behind the rejection of correct CC since UC is a constituent part

of CC, with violation fractions being between 0.74% and 0.79%, which is well in excess of the 0.5% target. For these occur disproportionately often during the COVID-19 pandemic, behaviour in this period warrants further investigation. Comparing the six indices illuminates a likely explanation as to why violations for the DAX occur comparatively frequently around the start of 2020; the response to worsening prospects owing to the spread of COVID-19 was most abrupt and severe for the DAX (-29.3% in ten trading days vis-à-vis -17.0% to 24.2% for the other indices), resulting in violation clustering before conditional volatility estimates could be adjusted appropriately.



(a) NASDAQ (b) Nikkei

Figure 4: CVaR forecasts and log-returns for two indices.

In summary, the empirical application suggests that confidence interval coverage is satisfactory for most indices, that opting for a TGARCH-type heteroscedasticity specification considerably improves CVaR and CES forecasts, and that performance for the various extrapolation locations considered is comparable, but significant outperformance with respect to $l = k^*$ as in Hoga (2019), if any, occurs in favour of l^*.

7 Conclusion

In this paper, we set out to validate the findings in Hoga (2019) and improve the Extreme Value Theory (EVT) approach to extreme quantile estimation by introducing and implementing an extrapolation location generalisation to the Weissman (1978) estimator with new asymptotic results and highlighting the importance of misspecification issues in the modelling of temporal dependencies in the series under consideration. We find replication results with a high degree of agreement with those in Hoga (2019) and all statistically significant results reported remain as such. Moreover, residual adequacy tests that focus on asymmetry (Engle & Ng, 1993), which is crucial since our empirical application is restricted to the left tail only, indicate that AR(1)-GARCH(1,1) residuals as used in Hoga (2019) fall short, whereas AR(1)-TGARCH(1,1) proves appropriate. Finally, our sim-

ulation study reveals that the preferable extrapolation location is determined by the study-specific desiderata and our empirical application reveals that for most of the major global stock indices, conditional and unconditional coverage are satisfactory and point forecasts for various choices of the tuning threshold are comparable, but outperformance with respect to $l = k^*$, if any, occurs exclusively for $l = l^*$. To wit, the optimal l is chosen every one hundred forecasts to minimise the realised CVaR loss in the second half of the total estimation window based on parameters fitted for the first half of the estimation window. Given the resulting l^*, parameters are retrieved from the full estimation window and extrapolation occurs from l^* to the desired extreme quantile.

In conclusion, we have validate the findings in Hoga (2019), provide a significant improvement by implementing an alternative data generating process specification, present extensive economic backgrounds to interpret surprising results, and explore and conjecture new asymptotic theory regarding an extrapolation location generalisation that allows for extrapolation from anywhere within the tail as defined for the estimation of the extreme value index. Although loss functions for the original Weissman (1978) estimator ($l = k^*$), Filtered Historical Simulation (FHS, with $l = \alpha T$), extrapolation from a data-driven choice of fixed l ($l = 2\alpha T$ for $T = 1000$), and extrapolation from some optimal location ($l = l^*$) are similar, Diebold-Mariano tests with a Bonferroni correction identify three instances of statistically significant outperformance as compared to $l = k^*$, all of which arise in case $l = l^*$. As such, evidence in favour of our extrapolation location generalisation is not overwhelming but does provide a promising prospect for further research with the potential of consistent outperformance. However, the fact that FHS does not perform significantly different from, and generally scores lower than, $l = k^*$ as in Hoga (2019) is alarming since the author's intricate EVT-based approach is only warranted if it provides a significant improvement over more straightforward methods.

An important limitation of our research is that no indisputable framework exists for several methodological decisions, thereby introducing a degree of arbitrariness that may affect the findings derived from extensions that build on these fundamental decisions. Examples are the choice of the extreme value index estimator relied upon, the loss functions used to evaluate point forecasts, and the heuristic to select k^* and the range from which to choose it. Such ancillary decisions to some extent obscure subsequent results. Additionally, we execute empirical performance evaluation for data ranging from 2001 to 2020. Although this range is fairly large, containing about five thousand observations for each index, time- and market-varying features imply that conclusions regarding model specification cannot be extrapolated directly to other time periods or indices. However, under

the appropriate filter specification, findings that rely on independent and identically distributed residuals remain relevant.

Besides addressing these limitations, we have several suggestions for future research. Fistly, implementation of a robustification of the Hill (1975) estimator and the selection of k^* yields ambiguous results. Further investigation is outside the scope of this paper, but may well prove beneficial in alternative applications. Both methods rely on omitting the tail segment beyond the extreme quantile of interest. This segment is where most variation in observations lies and is not extrapolated to, so may be omitted if the gains in robustness outweigh the loss of information from excluding these observations. Specifically, the robust Hill estimator as implemented is given by $\hat{\gamma} = \frac{1}{k^* - \lfloor \alpha T \rfloor} \sum_{t=\lfloor \alpha T \rfloor}^{k^*} \ln U_{T-t:T} / U_{T-k^*:T}$, where the robust k^* follows as $k^* := \underset{k \in \{k_{\min}, ..., k_{\max}\}}{\arg\min} \left(\sup_{t \in \{\lfloor \alpha T \rfloor, ..., k_{\max}\}} \left| U_{T-t:T} - U_{T-k:T} \left(\frac{t}{k}\right)^{-\hat{\gamma}} \right| \right)$. Under appropriate conditions on k, l, and α as $T \to \infty$, we can expect this estimator to converge and attain the Cramer-Rao lower bound, as does the standard Hill (1975) estimator, asymptotically. In a finite sample, however, investigation of the robustness-variance tradeoff can uncover potential uses of a more robust approach. Secondly, the moving window size in this paper is restricted to a thousand. Investigating the role of sample size on the optimal fixed choices of the extrapolation location would enable practitioners to utilise our generalised estimator in a much wider range of applications. Lastly, we find that improved model specification provides more substantial improvements than the extrapolation location decision in terms of point forecast loss scores and that FHS performs well. As such, future research may want to redirect its attention towards improved model specification before exploring more advanced EVT techniques. Importantly, this caveat is subject to the desired extreme quantile. Hoga (2019) provides an economic basis for using α equal to 0.025, 0.01, and 0.005. However, more extreme quantiles for applications such as dyke resistance against floodings uncover the potential of the extrapolation location generalisation. Evaluating l^* for $\alpha = 0.001$ yields lower loss scores for all six indices and two functions and the difference is statistically significant for five indices. FHS, on the other hand, performs much less consistently. The loss score discrepancies are even more outspoken for $\alpha = 0.0001$, with l^* being significantly superior in all twelve comparisons. Therefore, a promising objective for future research is exploring the generalised Weissman estimator in various disciplines with less available data or more extreme quantiles of interest, necessitating extrapolation and facilitating more opportunity for improvement for an extrapolation location within the tail.

8 Bibliography

Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1997). *A characterization of measures of risk* (Tech. Rep.). Cornell University Operations Research and Industrial Engineering.

Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, *9*(3), 203–228.

Azzalini, A., & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(2), 367–389.

Black, F. (1976). Studies of stock market volatility changes. *1976 Proceedings of the American Statistical Association Business and Economic Statistics Section*.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Chan, N. H., Deng, S.-J., Peng, L., & Xia, Z. (2007). Interval estimation of value-at-risk based on GARCH models with heavy-tailed innovations. *Journal of Econometrics*, *137*(2), 556–576.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 841–862.

Cont, R., et al. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, *1*(2), 223–236.

Danielsson, J., de Haan, L., Peng, L., & de Vries, C. G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis*, *76*(2), 226–248.

Danielsson, J., & De Vries, C. G. (1997). Tail index and quantile estimation with very high frequency data. *Journal of Empirical Finance*, *4*(2-3), 241–257.

Danielsson, J., Ergun, L. M., Haan, L. d., Vries, C. G. d., et al. (2016). Tail index estimation: quantile driven threshold selection. Working paper, London School of Economics and Erasmus University Rotterdam.

Daníelsson, J., Jansen, D. W., & De vries, C. G. (1996). The method of moments ratio estimator for the tail shape parameter. *Communications in Statistics-Theory and Methods*, *25*(4), 711–720.

Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *20*(1), 134–144.

Ding, Z., Granger, C. W., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, *1*(1), 83-106.

Drees, H., & Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and their Applications*, *75*(2), 149–172.

Drees, H., Resnick, S., & de Haan, L. (2000). How to make a Hill plot. *The Annals of Statistics*, *28*(1), 254–274.

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events: for insurance and finance* (Vol. 33). Springer.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987–1007.

Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, *48*(5), 1749–1778.

Ferreira, M. (2018). Risk seeker or risk averse? Cross-country differences in risk attitudes towards financial investment. *The Behavioral Economics Guide 2018*, 86–95.

Finkenstadt, B., & Rootzén, H. (2003). *Extreme values in finance, telecommunications, and the environment*. CRC Press.

Francq, C., & Zakoian, J.-M. (2019). *GARCH Models: structure, statistical inference and financial applications*. John Wiley & Sons.

Gadea, M. D., Gómez-Loscos, A., & Pérez-Quirós, G. (2018). Great Moderation and Great Recession: From plain sailing to stormy seas? *International Economic Review*, *59*(4), 2297–2321.

Gandelman, N., Hernandez-Murillo, R., et al. (2015). Risk aversion at the country level. *Federal Reserve Bank of St. Louis Review*, *97*(1), 53–66.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, *48*(5), 1779–1801.

Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis*, *32*(2), 177–203.

Harvey, A., & Chakravarty, T. (2008). Beta-*t*-(E) GARCH. Working papers in Economics, Faculty of Economics, University of Cambridge.

Hasanhodzic, J., & Lo, A. W. (2011). Black's leverage effect is not due to leverage. Working paper, Boston University and MIT.

Hayashi, F., & Prescott, E. C. (2002). The 1990s in Japan: A lost decade. *Review of Economic Dynamics*, *5*(1), 206–235.

Hens, T., & Steude, S. C. (2009). The leverage effect without leverage. *Finance Research Letters*, *6*(2), 83–94.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, *3*(5), 1163–1174.

Hoga, Y. (2019). Confidence intervals for conditional tail risk measures in ARMA–GARCH models. *Journal of Business & Economic Statistics*, *37*(4), 613–624.

Jalal, A., & Rockinger, M. (2008). Predicting tail-related risk measures: The consequences of using GARCH filters for non-GARCH data. *Journal of Empirical Finance*, *15*(5), 868-877.

Kawa, L. (2020). *Stock market volatility tops financial crisis with VIX at record.* Bloomberg. Retrieved 02-05-2021, from `https://www.bloomberg.com/news/articles/2020-03-16/stock-market-volatility-tops-financial-crisis-with-vix-at-record`

Kim, S.-W., & Lee, B.-S. (2008). Stock returns, asymmetric volatility, risk aversion, and business cycle: some new evidence. *Economic Inquiry*, *46*(2), 131–148.

Laas, D., & Siegel, C. F. (2017). Basel III versus Solvency II: An analysis of regulatory consistency under the new capital standards. *The Journal of Risk and Insurance*, *84*(4), 1231–1267.

Liu, F., & Wang, R. (2020). A theory for measures of tail risk. *Mathematics of Operations Research*, *forthcoming*. Retrieved from `https://papers.ssrn.com/abstract_id=2841909`

McNeil, A. J. (1999). Extreme value theory for risk managers. *Departement Mathematik ETH Zentrum*, *12*(5), 121–237.

McNeil, A. J., & Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, *7*(3-4), 271–300.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, *160*(1), 246–256.

Perron, P., & Yamamoto, Y. (2021). The Great Moderation: Updated evidence with joint tests for multiple structural changes in variance and persistence. *Empirical Economics*, 1–26.

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Reiss, R.-D., Thomas, M., & Reiss, R. (1997). *Statistical analysis of extreme values* (2nd ed.). Springer.

Shao, X. (2010). A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*(3), 343–366.

Sun, P., & Zhou, C. (2014). Diagnosing the distribution of GARCH innovations. *Journal of Empirical Finance*, *29*, 287–303.

Taylor, J. W. (2019). Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. *Journal of Business & Economic Statistics*, *37*(1), 121-133.

Tsourti, Z., & Panaretos, J. (2001). *Extreme value index estimators and smoothing alternatives: review and simulation comparison* (Tech. Rep. No. 149). Athens University of Economics and Business.

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, *73*(364), 812–815.

Weisstein, E. W. (2004). *Bonferroni correction.* Wolfram Research, Inc. Retrieved 15-06-2021, from `https://mathworld.wolfram.com/BonferroniCorrection.html`

Wilhelmsson, A. (2006). GARCH forecasting performance under different distribution assumptions. *Journal of Forecasting*, *25*(8), 561–578.

World Bank. (2012). *Global financial development report 2013: Rethinking the role of the state in finance*. The World Bank.

Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, *18*(5), 931–955.

A Replication results

A.1 Simulation study

Table 4: τ-quantiles of V_{t_0}.

t_0 \ τ	0.50	0.60	0.70	0.80	0.90	0.95	0.975	0.99	0.995
0.1	3.710	5.994	9.531	15.82	29.58	48.21	73.32	115.0	148.8
0.2	4.004	6.662	10.71	17.71	34.22	56.64	85.19	127.1	169.2
0.3	4.727	7.566	12.36	20.80	40.58	67.57	100.9	161.3	203.9

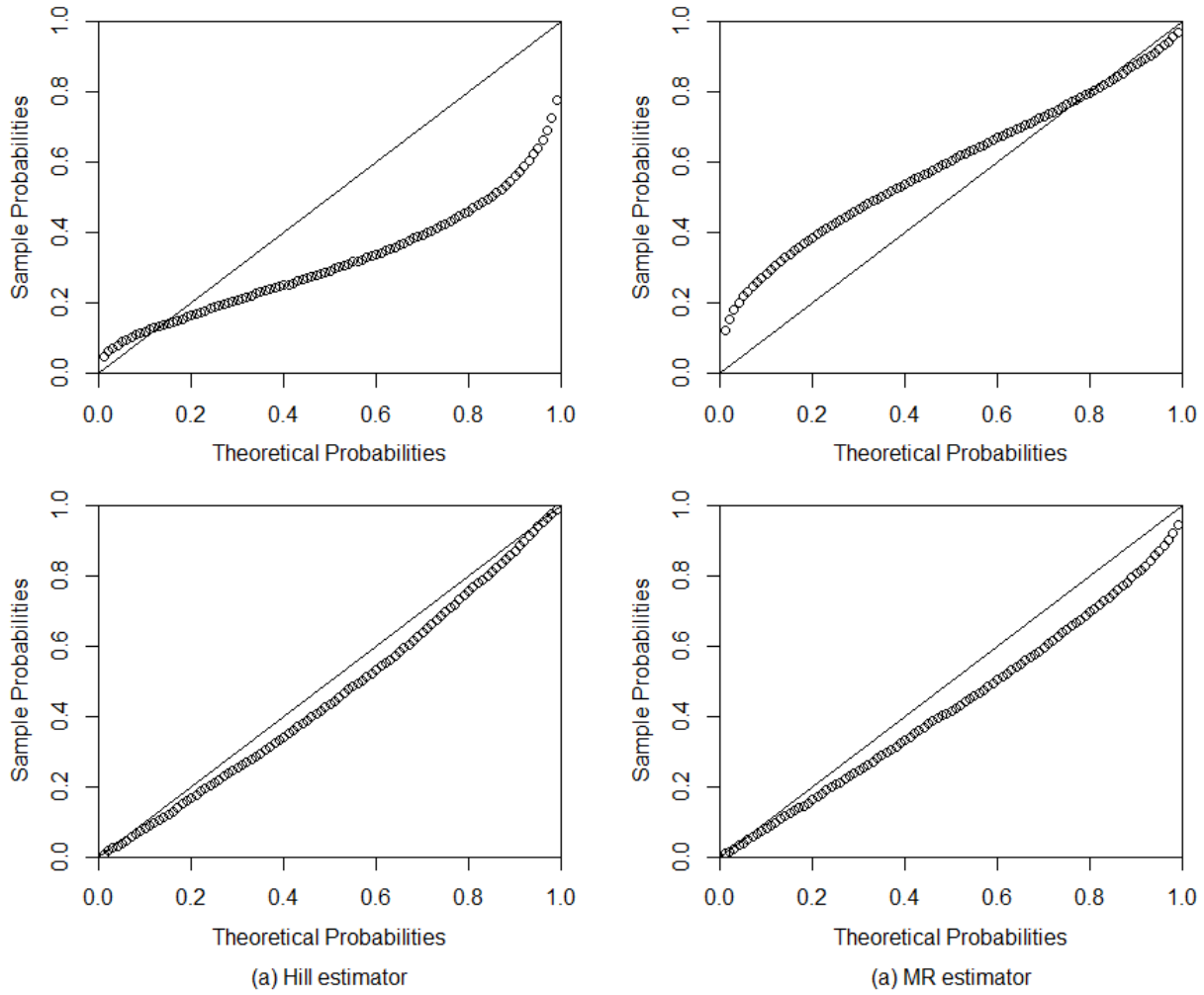


Figure 5: PP plots for the first (top) and second (bottom) asymptotic distributions in Hoga (2019) (Eq. (24) & (25) respectively) for (M2) based on a hundred thousand points with a hundred thousand realisations each for the Hill (a) and MR (b) estimators.

Table 5: Bias, RMSE, coverage probabilities (in %, nominal coverage is 95%) and interval length for CVaR and CES at three extreme quantiles using the Hill and MR estimators with average k^* for models (M1), (M2) and (M3). Bias, RMSE, and interval length are $\times 100$ for (M2) and (M3).

Model	Measure	Estimator	k^*	α	Bias	RMSE	Coverage (%)		Int. length	
							$I_{\text{na}}^{0.95}$	$I_{\text{sn}}^{0.95}$	$I_{\text{na}}^{0.95}$	$I_{\text{sn}}^{0.95}$
(M1)	CVaR	Hill	67	2.5%	-0.04	0.58	49.6	74.5	0.26	0.54
				1%	-0.09	0.64	69.2	79.8	0.59	0.88
				0.5%	-0.16	0.78	76.7	84.3	0.90	1.30
		MR	98	2.5%	0.06	0.56	60.8	75.7	0.41	0.70
				1%	0.01	0.63	80.5	81.7	0.82	1.08
				0.5%	-0.05	0.77	86.5	83.6	1.21	1.51
	CES	Hill	67	2.5%	-0.09	0.71	39.0	83.2	0.30	1.01
				1%	-0.10	0.80	59.0	86.7	0.68	1.59
				0.5%	-0.09	0.97	67.6	89.2	1.04	2.19
		MR	98	2.5%	0.02	0.69	52.1	81.6	0.49	1.22
				1%	0.00	0.81	68.9	83.3	0.96	1.82
				0.5%	0.01	1.01	74.9	83.0	1.42	2.44
(M2)	CVaR	Hill	59	2.5%	0.03	0.26	59.7	86.3	0.30	0.66
				1%	0.00	0.35	85.2	90.7	0.85	1.25
				0.5%	-0.10	0.50	91.9	93.1	1.51	2.06
		MR	64	2.5%	0.14	0.28	55.8	79.4	0.37	0.71
				1%	0.20	0.39	82.6	80.0	1.03	1.20
				0.5%	0.21	0.50	91.0	82.8	1.79	1.82
	CES	Hill	59	2.5%	-0.14	0.51	44.6	93.5	0.45	1.80
				1%	-0.39	0.83	64.1	94.2	1.30	3.41
				0.5%	-0.71	1.27	69.4	93.4	2.30	5.27
		MR	64	2.5%	0.13	0.47	51.8	84.4	0.55	1.56
				1%	0.05	0.68	78.1	87.4	1.52	2.74
				0.5%	-0.07	0.96	87.0	88.7	2.65	4.10
(M3)	CVaR	Hill	58	2.5%	0.02	0.10	56.3	82.7	0.13	0.28
				1%	0.00	0.16	84.5	89.1	0.35	0.49
				0.5%	-0.04	0.21	91.5	92.8	0.59	0.78
		MR	66	2.5%	0.08	0.13	50.3	73.3	0.17	0.30
				1%	0.01	0.18	79.6	75.7	0.42	0.47
				0.5%	0.09	0.22	90.9	81.6	0.70	0.69
	CES	Hill	58	2.5%	-0.04	0.17	46.9	91.7	0.17	0.63
				1%	-0.13	0.30	67.3	93.8	0.47	1.13
				0.5%	-0.25	0.45	70.3	92.8	0.80	1.70
		MR	66	2.5%	0.07	0.17	51.7	81.5	0.22	0.56
				1%	0.04	0.25	80.7	86.6	0.57	0.92
				0.5%	-0.02	0.34	89.1	89.0	0.94	1.35

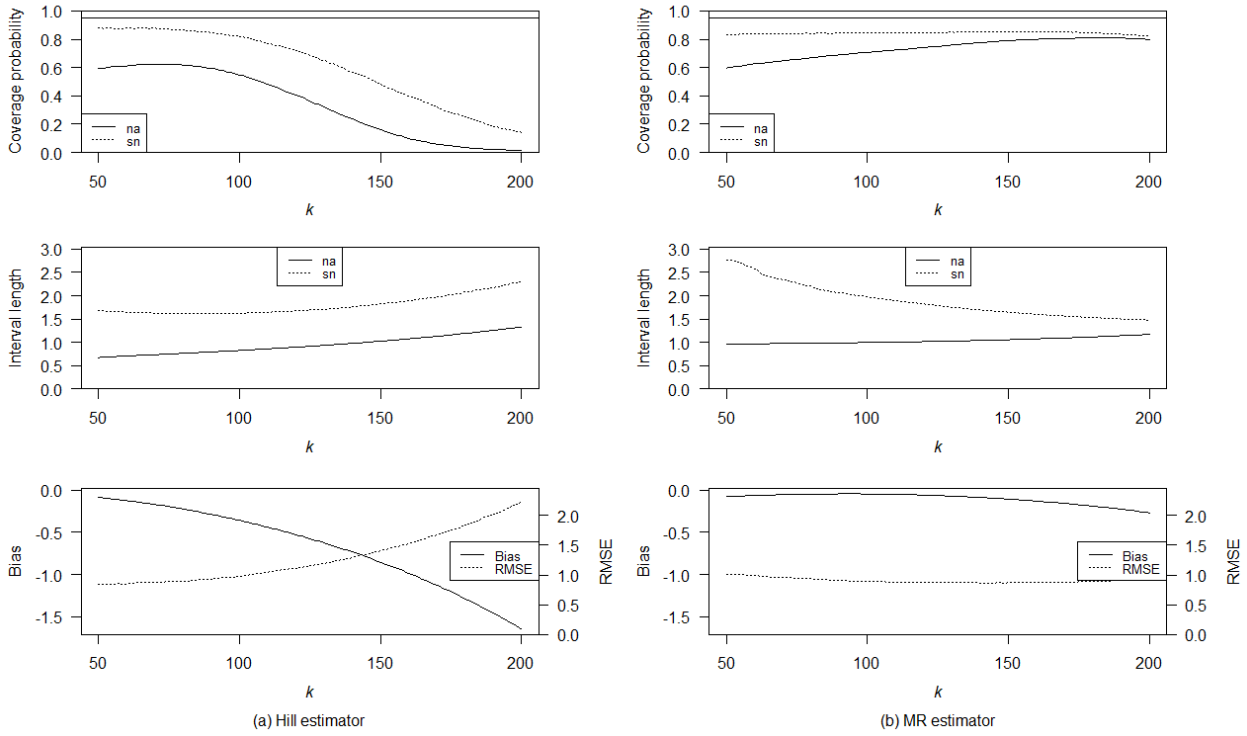


Figure 6: Coverage probability and interval length for $I_{\text{na}}^{0.95}$ and $I_{\text{sn}}^{0.95}$ using the Hill (a) and MR (b) estimators with corresponding bias and RMSE as functions of $k \in \{50, \ldots, 200\}$ for model (M1). The horizontal line (top) indicates the desired 95% coverage.

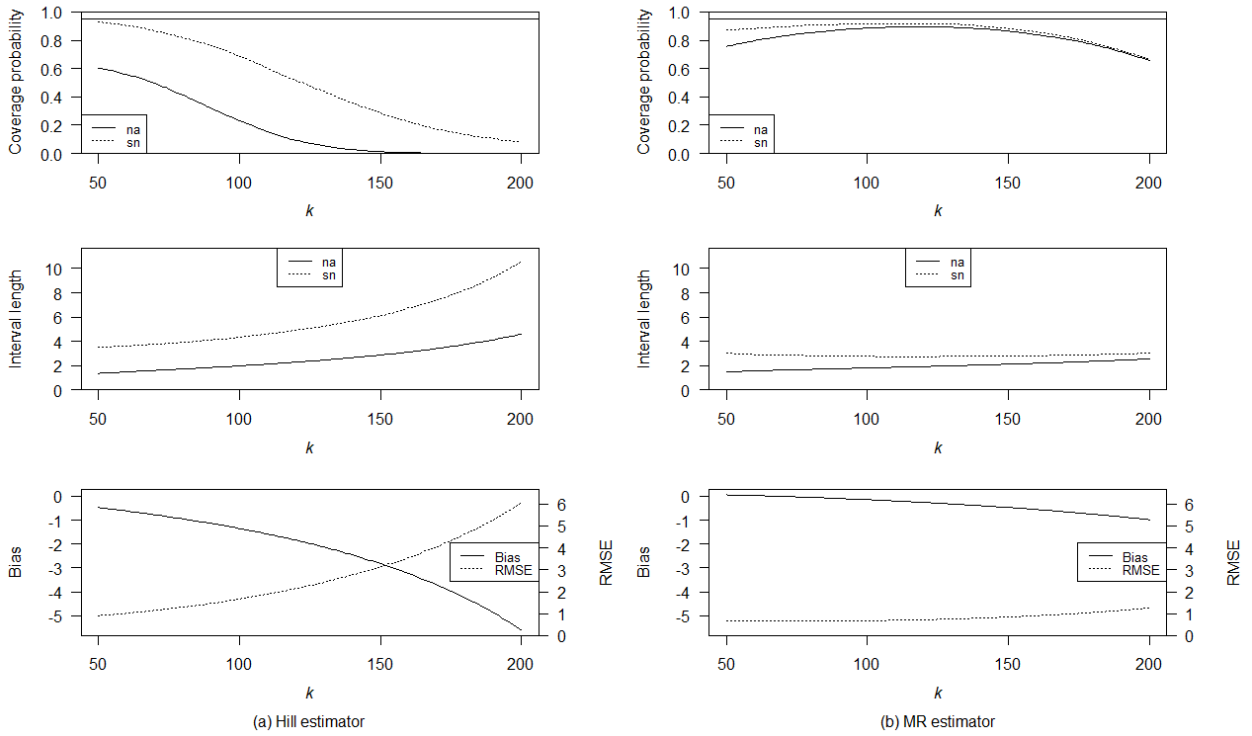


Figure 7: Coverage probability and interval length for $I_{\text{na}}^{0.95}$ and $I_{\text{sn}}^{0.95}$ using the Hill (a) and MR (b) estimators with corresponding bias and RMSE as functions of $k \in \{50, \ldots, 200\}$ for model (M2). Interval length, bias, and RMSE are multiplied by 100. The horizontal line (top) indicates the desired 95% coverage.

A.2 Empirical application

Table 6: p-values of tests for (un)conditional CVaR coverage, Quantile scores (CVaR), and AL log scores (CVaR and CES jointly) for six stock indices and three methods from 1997 to 2016 and $\alpha = 0.005$.

Index	Method	UC	CC	Quantile score	AL log score
DJIA	G-k-H	0.3525	0.1081	0.7881	-9141.4
	AG-k^*-H	0.9706	0.1288	0.7702†	**-9269.8**†
	AG-k^*-MR	0.0042***	0.0020***	**0.7675**	-9259.1
NASDAQ	G-k-H	0.1097	0.0909*	0.9527	-8489.0
	AG-k^*-H	0.4798	0.1136	**0.9366**†	**-8605.1**†
	AG-k^*-MR	0.1927	0.0006***	0.9427	-8515.7
Nikkei	G-k-H	0.5675	0.1038	1.1342	-7365.2
	AG-k^*-H	0.7312	0.1088	**1.1283**	**-7402.5**
	AG-k^*-MR	0.2137	0.0000***	1.1376	-7313.8
HSI	G-k-H	0.1321	0.0844*	0.9800	-8049.9
	AG-k^*-H	0.4079	0.1006	**0.9663**	**-8167.5**†
	AG-k^*-MR	0.1517	0.0000***	0.9676	-8140.4
CAC 40	G-k-H	0.2255	0.9998	1.0245	-8324.6
	AG-k^*-H	0.9171	0.1336	**1.0162**	**-8376.8**
	AG-k^*-MR	0.3155	0.0102**	1.0172	-8304.8
DAX 30	G-k-H	0.1587	0.0993*	1.0104	-8253.6
	AG-k^*-H	0.6030	0.1220	0.9844†	**-8383.7**†
	AG-k^*-MR	0.0178**	0.0171**	**0.9816**	-8327.7

Note. *p<0.1, **p<0.05, ***p<0.01; The p-values are calculated using tests for unconditional coverage (UC) and conditional coverage (CC) as described in Christoffersen (1998). A dagger (†) indicates a statistically significant two-sided Diebold-Mariano test (at 5%), for superior predictive performance of AG-k^*-H with respect to G-k-H, as described in Diebold & Mariano (2002). For each data set and scoring function, the lowest score is in bold.

Table 7: p-values of tests for (un)conditional CVaR coverage, Quantile scores (CVaR), and AL log scores (CVaR and CES jointly) for six stock indices and three methods from 2001 to 2020 and $\alpha = 0.005$.

Index	Method	UC	CC	Quantile score	AL log score
S&P 500	G-k-H	0.6304	0.9980	0.7799	-9324.1
	AG-k^*-H	0.7965	0.9964	**0.7617**†	**-9480.2**†
	AG-k^*-MR	0.0004***	0.1458	0.7660	-9412.8
NASDAQ	G-k-H	0.3540	0.1078	0.9394	-8491.9
	AG-k^*-H	0.6795	0.1378	**0.9221**	**-8629.8**†
	AG-k^*-MR	0.0004***	0.0031***	0.9553	-8412.6
Nikkei	G-k-H	0.4293	0.0981*	1.1397	-7302.1
	AG-k^*-H	0.4293	0.0981*	**1.1282**	**-7368.9**
	AG-k^*-MR	0.0876*	0.0000***	1.1298	-7333.3
HSI	G-k-H	0.1343	0.0838*	0.9429	-8112.8
	AG-k^*-H	0.7147	0.1101	**0.9249**	**-8264.5**†
	AG-k^*-MR	0.0336**	0.0000***	0.9270	-8256.0
FTSE 100	G-k-H	0.0889*	0.9782	0.8091	-9544.6
	AG-k^*-H	0.6164	0.9782	**0.7947**	**-9705.5**†
	AG-k^*-MR	0.0000***	0.4445	0.8217	-9569.7
DAX 30	G-k-H	0.9363	0.0057***	1.0150	-8377.3
	AG-k^*-H	0.4187	0.0090***	**1.0071**	**-8438.7**
	AG-k^*-MR	0.0005***	0.0000***	1.0189	-8340.8

Note. *p<0.1, **p<0.05, ***p<0.01; The p-values are calculated using tests for unconditional coverage (UC) and conditional coverage (CC) as described in Christoffersen (1998). A dagger (†) indicates a statistically significant two-sided Diebold-Mariano test (at 5%), for superior predictive performance of AG-k^*-H with respect to G-k-H, as described in Diebold & Mariano (2002). For each data set and scoring function, the lowest score is in bold.

B Extension results

B.1 Simulation study

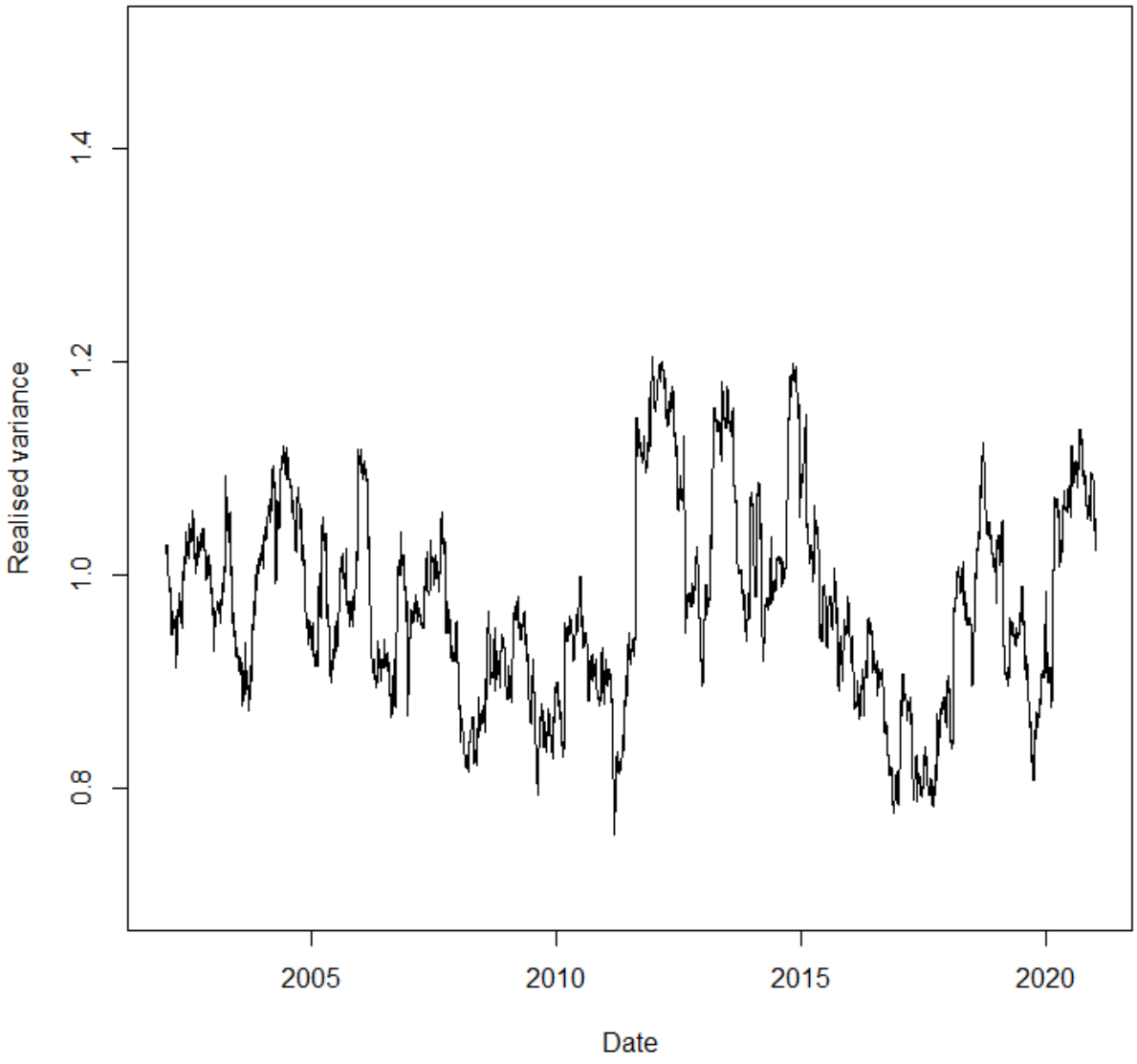


Figure 8: Moving average of squared pseudo-random innovations from the distribution associated with NASDAQ shocks (Hoga, 2019) with a window size of 253 trading days (∼1 year).

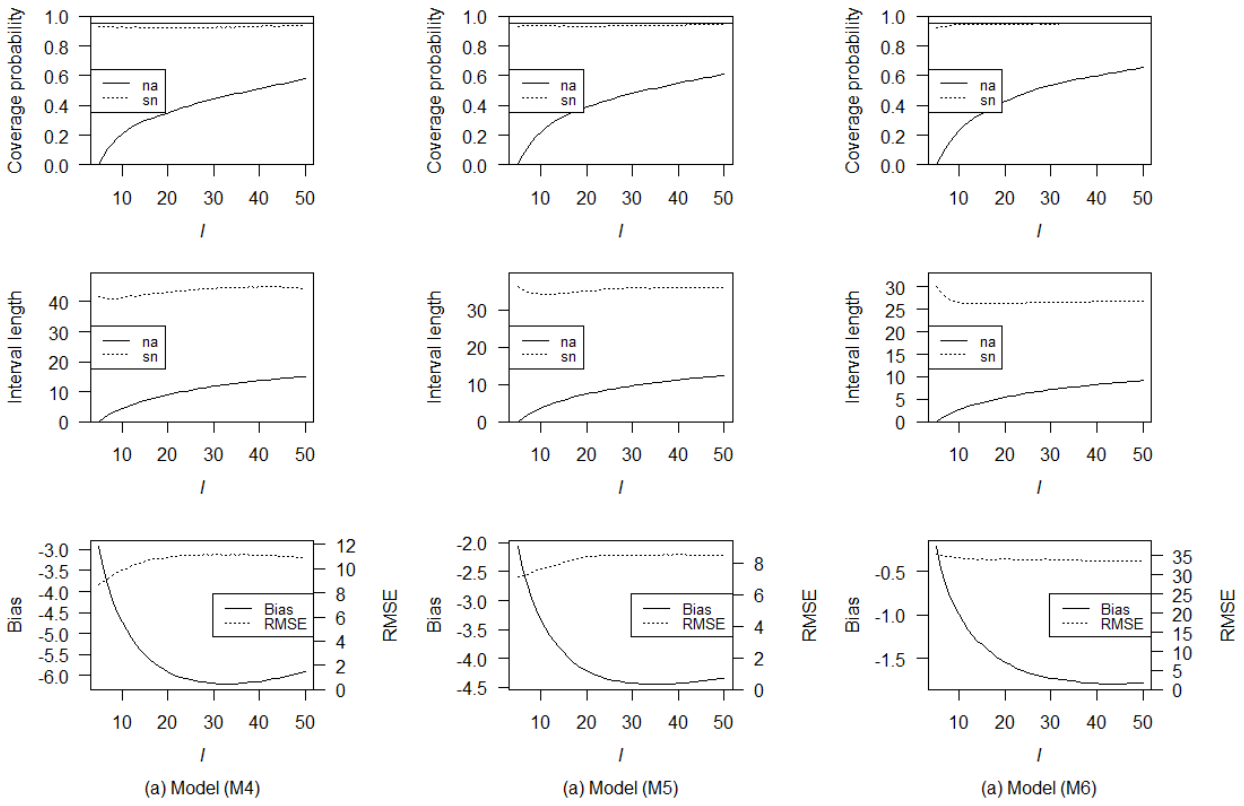


Figure 9: CVaR forecast performance measures for the alternative models in Appendix B of Hoga (2019).

B.2 Empirical application

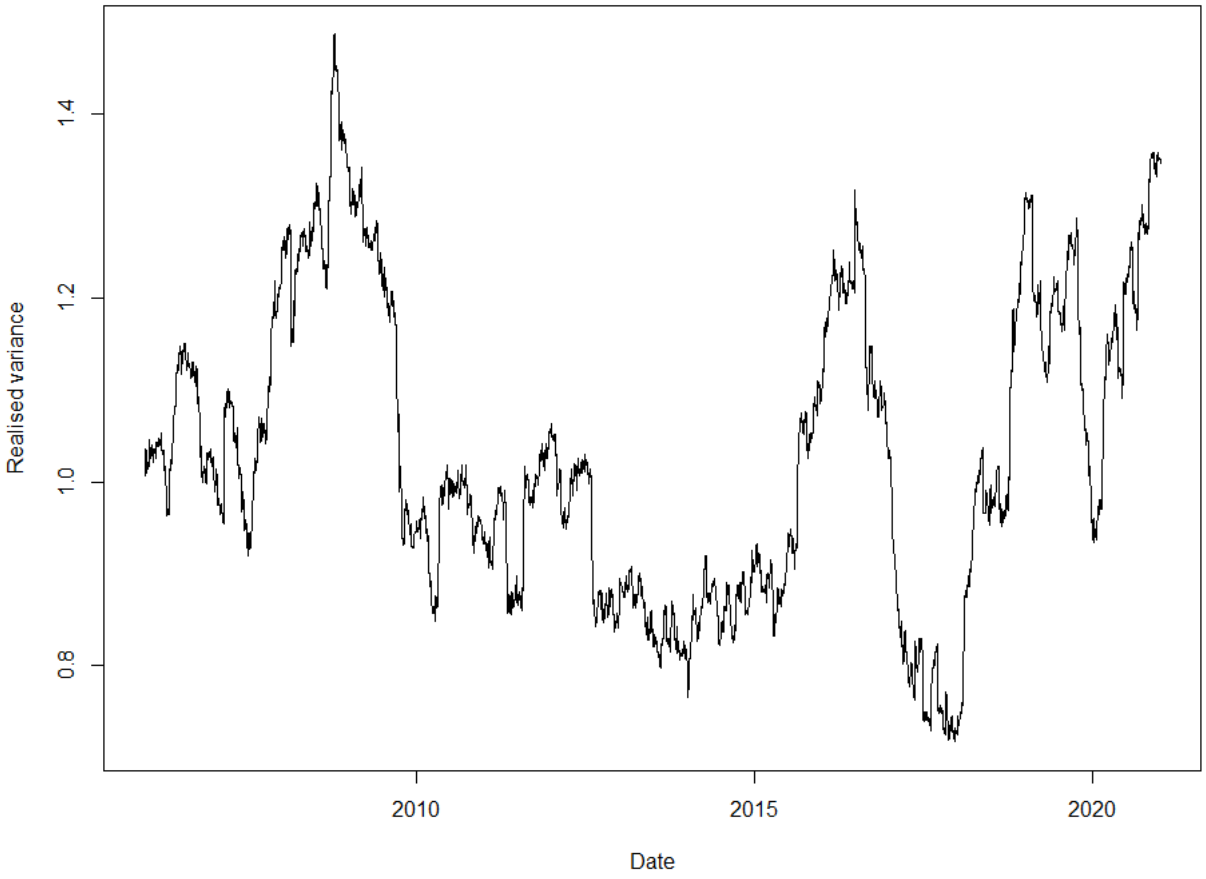


Figure 10: Moving average of NASDAQ AR(1)-GARCH(1,1) squared standardised residuals with a window size of 253 trading days (∼1 year).

Table 8: p-values for ARCH-type adequacy tests (Engle & Ng, 1993) for six indices from 1990 to 2000.

	AR(1)-GARCH(1,1)			AR(1)-TGARCH(1,1)		
Index	Sign	Pos. size	Neg. size	Sign	Pos. size	Neg. size
S&P 500	3.2e-4	9.5e-6	< 2.2e-16	0.34	0.04	0.65
NASDAQ	8.3e-6	5.6e-2	< 2.2e-16	0.25	0.89	0.84
Nikkei	4.6e-4	3.7e-3	7.8e-5	0.40	0.91	0.49
HSI	7.8e-2	2.4e-1	< 2.2e-16	0.66	0.21	0.37
FTSE 100	9.4e-1	6.2e-2	6.5e-2	0.74	0.53	0.97
DAX 30	4.3e-3	4.8e-1	2.9e-14	0.21	0.01	0.91

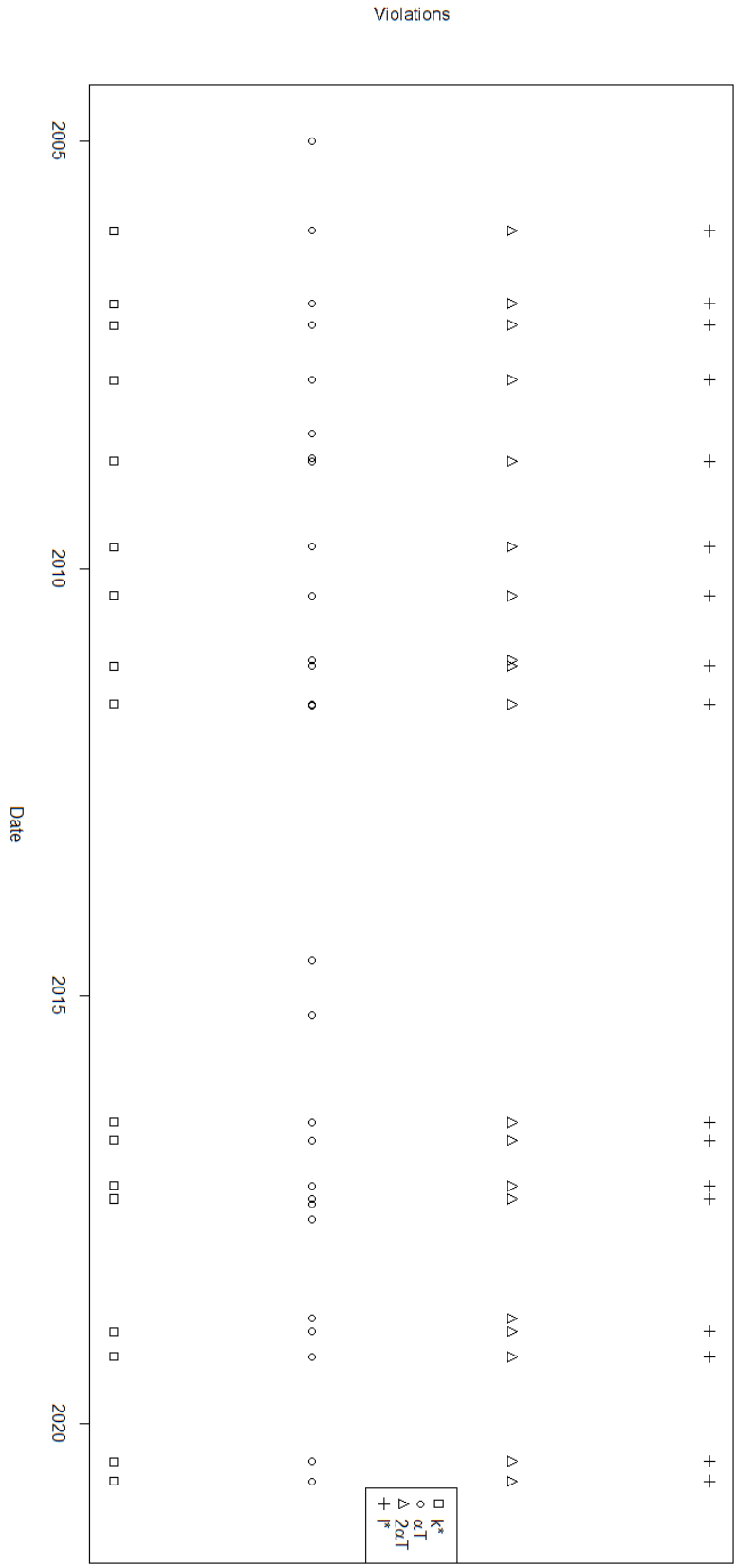


Figure 11: CVaR violations for NASDAQ and four choices of l.

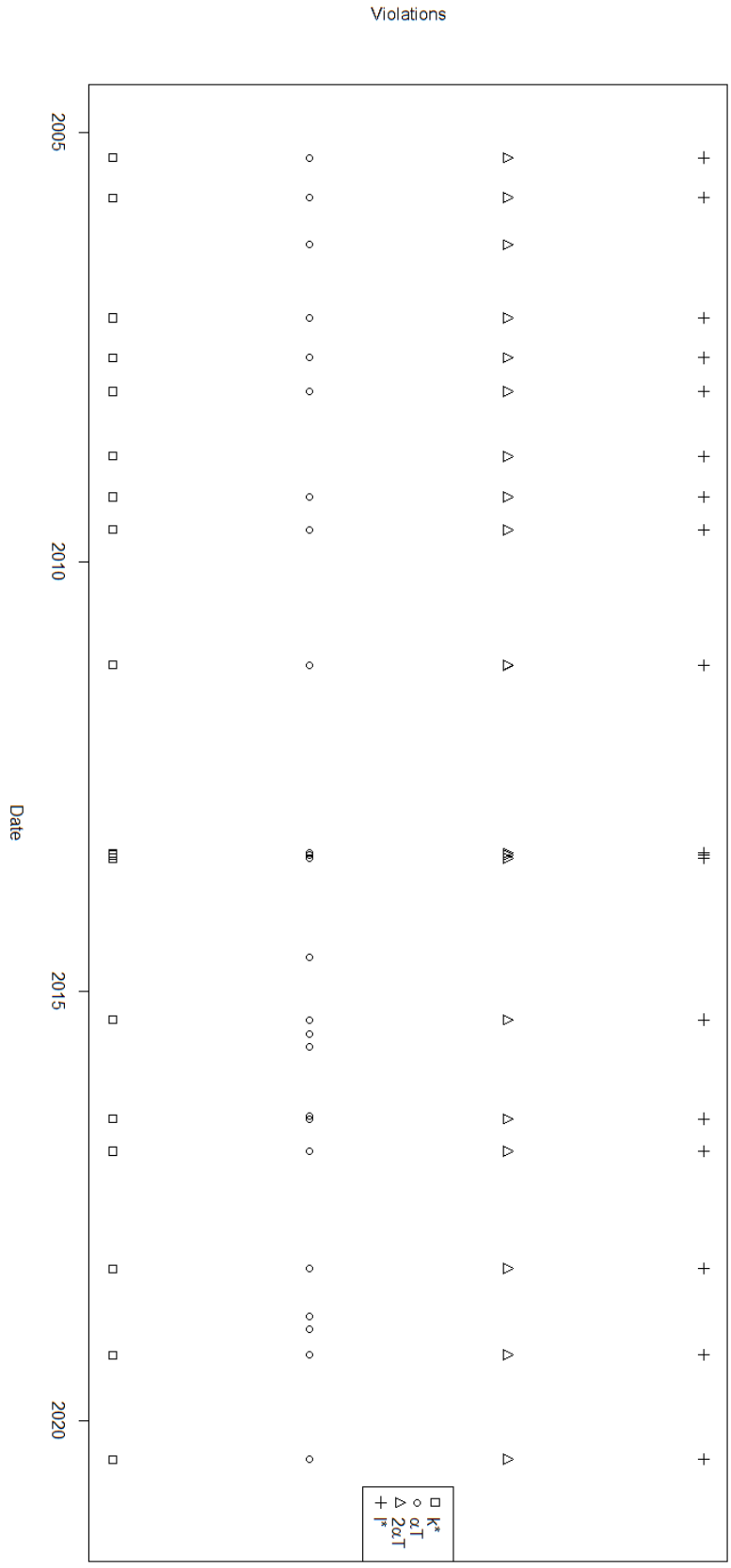


Figure 12: CVaR violations for Nikkei and four choices of l.

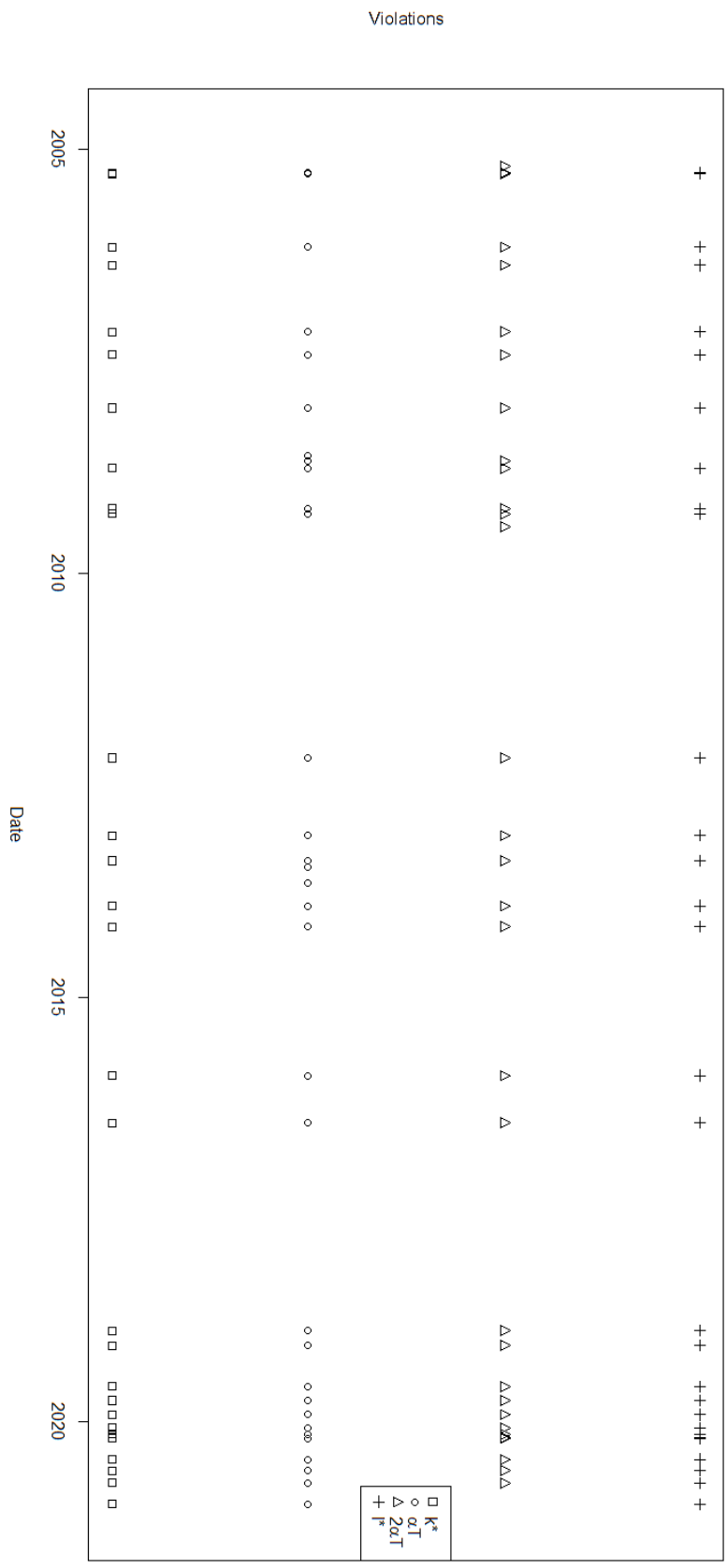


Figure 13: CVaR violations for DAX 30 and four choices of l.

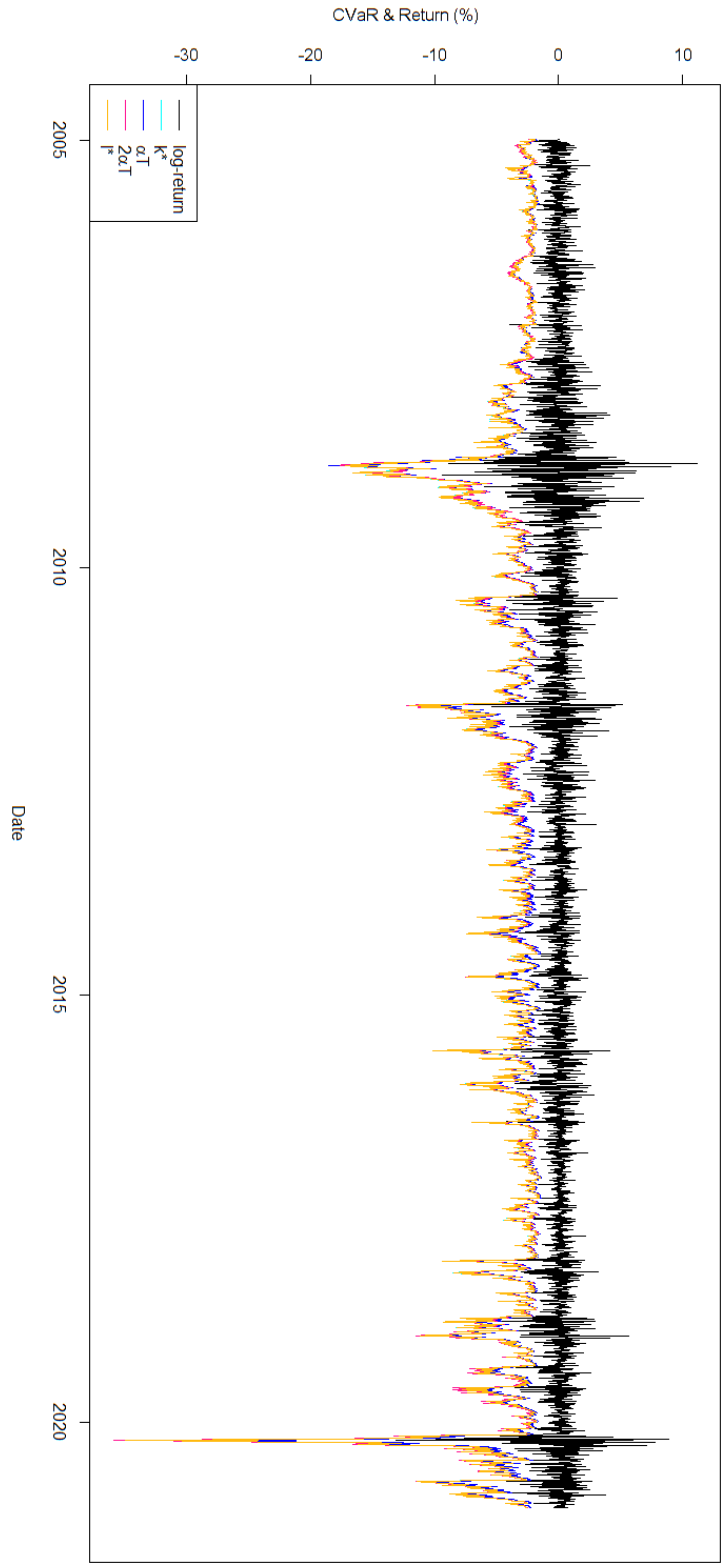


Figure 14: Log-returns and CVaR forecasts for four choices of l for the NASDAQ.

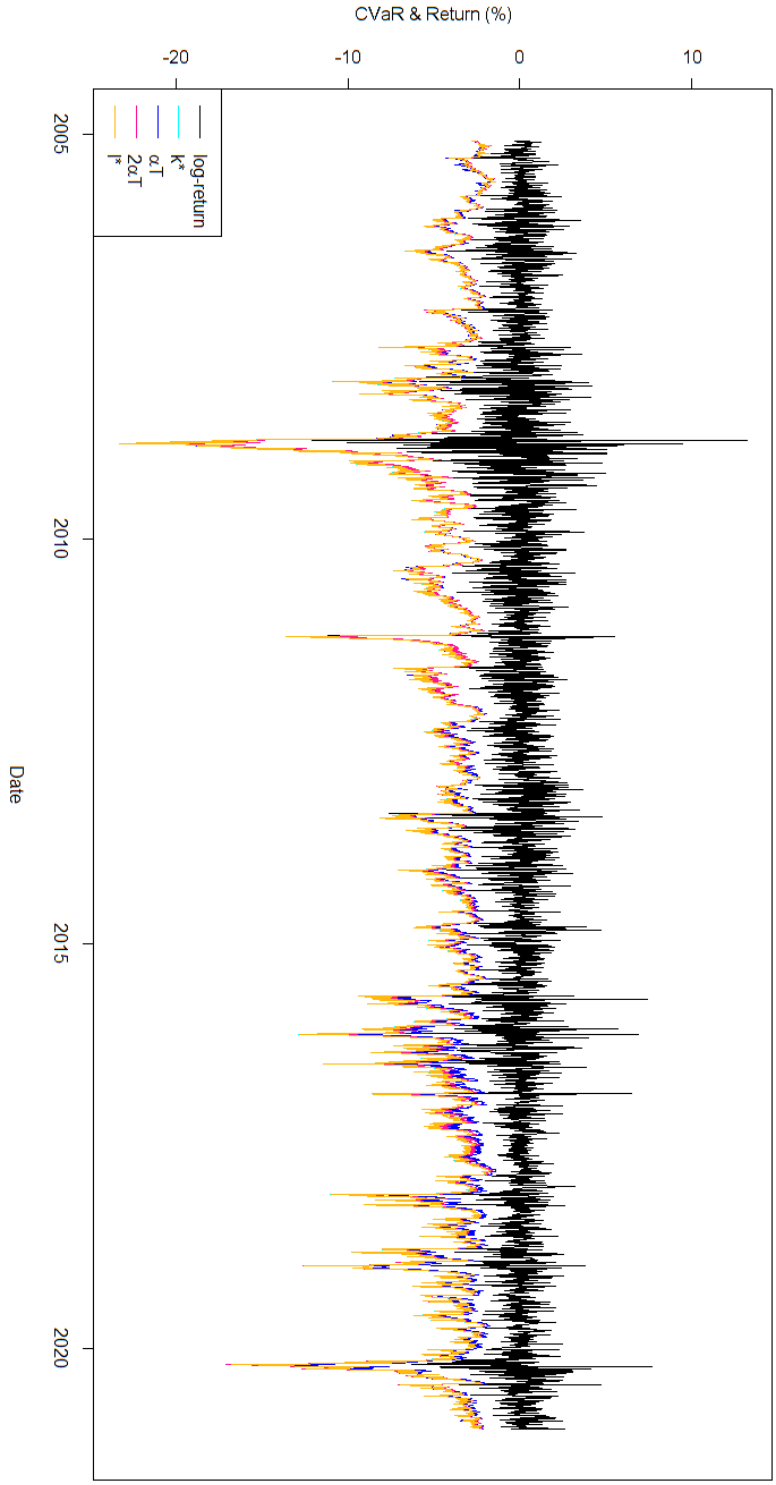


Figure 15: Log-returns and CVaR forecasts for four choices of l for the Nikkei.

C Weissman extrapolation factor variance for extreme quantiles

This section is devoted to a proof that uncertainty in the Weissman (1978) estimator grows superlinearly as α shrinks to zero. Essentially, the proof of Theorem 1 in Hoga (2019) revolves around demonstrating that, asymptotically, the uncertainty in the extrapolation factor $(\alpha T/k)^{-\hat{\gamma}}$ dominates that of the order statistic $U_{T-k:T}$, so it suffices to restrict the analysis to extrapolation factor. To this end, we consider $\lim_{\alpha \to 0^+} \frac{\partial}{\partial \alpha} \mathrm{V} \left(\alpha T/k\right)^{-\hat{\gamma}}$. For finite $\hat{\gamma}$, non-zero α, and because k is an integer sequence, such that $(\alpha T/k)^{-\hat{\gamma}}$ and $(\alpha T/k)^{-2\hat{\gamma}}$ are bounded for any $T = 1, 2, \ldots$, it follows that

$$
\begin{aligned}
\frac{\partial}{\partial \alpha} \mathrm{V} \left((\alpha T/k)^{-\hat{\gamma}} \right) &= \frac{\partial}{\partial \alpha} \mathbb{E} \left((\alpha T/k)^{-2\hat{\gamma}} \right) - \frac{\partial}{\partial \alpha} \mathbb{E} \left((\alpha T/k)^{-\hat{\gamma}} \right)^2 \\
&= \mathbb{E} \left(\frac{\partial}{\partial \alpha} (\alpha T/k)^{-2\hat{\gamma}} \right) - 2 \mathbb{E} \left((\alpha T/k)^{-\hat{\gamma}} \right) \mathbb{E} \left(\frac{\partial}{\partial \alpha} (\alpha T/k)^{-\hat{\gamma}} \right) \\
&= -\frac{2T}{k} \mathbb{E} \left(\hat{\gamma} (\alpha T/k)^{2\hat{\gamma}-1} \right) + \frac{2T}{k} \mathbb{E} \left((\alpha T/k)^{-\hat{\gamma}} \right) \mathbb{E} \left(\hat{\gamma} (\alpha T/k)^{-\hat{\gamma}-1} \right) \\
&= -\frac{2T}{k} \mathrm{COV} \left((\alpha T/k)^{-\hat{\gamma}}, \hat{\gamma} (\alpha T/k)^{-\hat{\gamma}-1} \right),
\end{aligned}
$$

where the final expression contains the covariance of two functions that are increasing in $\hat{\gamma}$, such that their covariance is positive. It has thus been shown that the variance of the extrapolation factor decreases (increases) as α grows (shrinks). Moreover, we may consider the slope of the variance relative to a linear trend using some arbitrary shrinkage factor $1/a$, where $a > 1$, to find that

$$
-\frac{2T}{k} \mathrm{COV} \left(((\alpha/a) T/k)^{-\hat{\gamma}}, \hat{\gamma} ((\alpha/a) T/k)^{-\hat{\gamma}-1} \right) = a^{2\hat{\gamma}+1} \left[-\frac{2T}{k} \mathrm{COV} \left((\alpha T/k)^{-\hat{\gamma}}, \hat{\gamma} (\alpha T/k)^{-\hat{\gamma}-1} \right) \right].
$$

Ergo, if α changes by a factor $1/a$, the slope of the variance changes by $a^{2\hat{\gamma}+1} > a$ using that almost surely $\hat{\gamma} = \gamma > 0$ for sufficiently large T. Recall that linear growth is characterised by $f(cx) = cf(x)$ for any constant c. This concludes the proof of the superlinear growth of the Weissman (1978) estimator as α shrinks to zero from above.