Assessing The Performance Of The Extremal Random Forest In Estimating The Value-At-Risk In Emerging Markets

Eline Telgenkamp (493935et)
Supervisor: Chen Zhou
Second assessor: Aishameriane Venes Schmidt

July 3, 2022

Abstract

This paper studies the extremal random forest (ERF) of Gnecco et al. (2022) and investigates whether the ERF improves the estimation of the extreme conditional Value-at-Risk (VaR) in emerging stock markets. We assess the performance of the ERF relative to the generalized random forest (GRF) by Athey et al. (2019), the gradient boosting for extremes algorithm (GBEX) by Velthoen et al. (2021), and the unconditional generalized Pareto distribution (GPD). Here, we apply the different methods both on the MSCI Emerging Market Index and on the individual countries it includes. We find that the behavior of the stock returns over time has a significant effect on the performance of the ERF, where the performance decreases when the year the model is trained on is less volatile compared to the year it is subsequently tested on. The ERF is also less accurate for smaller sample sizes. However, we find that the ERF outperforms the other methods for negative stock returns with fat tails. This suggests that the ERF is the better method among those assessed to predict extreme conditional VaR for emerging markets where negative returns occur more frequently. Lastly, the expected worldwide market volatility and sovereign credit risk have the biggest impact on the extreme conditional VaR for the emerging equity market. With a probability of 0.0005, we estimate a maximum potential daily loss of around 2.5% for low market volatility and 12.5% for high volatility.

Bachelor Thesis for BSc^2 in Econometrics and Economics

ERASMUS UNIVERSITEIT ROTTERDAM

Contents

1 Introduction

Recently, the world experienced several rare events, such as a shut down of the New York Stock Exchange due to a significant decrease in stock prices (Frazier, 2021), and negative US crude oil prices as oil demand was cut down substantially worldwide (Brower et al., 2020). Extreme events such as the latter create more awareness among investors about the risks they face when investing (Demos & Mackenzie, 2012). By using risk measures, such as the Value-at-Risk, banks and other financial institutions try to quantify these risks. Value-at-Risk, VaR(α), gives the expected maximum loss of an investment or portfolio during a set time horizon with a given confidence level $100 \times (1 - \alpha)\%$ (Dimitrakopoulos et al., 2010).

The wide adoption of the VaR combined with the occurrence of extreme events, potentially harmful to banks and other financial institutions, requires the tail distribution of stock returns to be studied well. More specifically, extreme Value-at-Risk should be modeled appropriately to incorporate these rare events (McNeil, 1999). Financial institutions have significant amounts of money invested in emerging equity markets, and, as emerging markets are relatively volatile compared to developed markets, risk management is crucial (Gencay & Selçuk, 2004). Considering the impact financial institutions can have on our everyday lives, it is socially relevant to research the extreme Value-at-Risk in emerging markets. For this reason, we study the estimation of the extreme Value-at-Risk for stock returns in emerging equity markets conditional on external covariates.

McNeil (1999) points out that the tail risk is underestimated if a normal distribution for the returns is assumed, as financial time series experience fat tails. However, the estimation of these extreme quantiles poses a challenge, because there exists only a few to no data points for them. Extreme value theory proposes the use of tail approximation (e.g., De Haan and Ferreira, 2006). Here, an intermediate quantile is first estimated with classical regression methods, and, thereafter, the estimate is extrapolated to the extreme quantile level. This extrapolation can be done with a generalized Pareto distribution (GPD) (Gencay and Selçuk, 2004; Gnecco et al., 2022; Embrechts et al., 1998).

Additionally, there are several studies using machine learning models for the formerly mentioned extrapolation (Gnecco et al., 2022; Velthoen et al., 2021; Shrivastava et al., 2014). Velthoen et al. (2021) propose a gradient boosting model to estimate the generalized Pareto distribution. They conclude that their gradient boosting model outperforms classical methods from quantile regression and extreme value theory. However, Gnecco et al. (2022) claim that forest-based approaches are more favorable than gradient boosting methods and neural networks, as forest-based approaches require less tuning and have more understandable statistical properties. The researchers consider an extremal random forest, and, consequently, create a bridge between the literature on extreme value theory and the literature on random forests.

In this paper, we study the extremal random forest of Gnecco et al. (2022) for the estimation of the extreme Value-at-Risk in emerging stock markets conditional on external covariates. Therefore, the research question is as follows:

Does the extremal random forest algorithm help to improve estimating the extreme conditional Value-at-Risk in emerging stock markets?

To model the extreme conditional VaR, we require appropriate covariates that describe the variation in the emerging equity market. Koepke (2019) explains five main drivers for the inflow and outflow of capital in emerging markets. These main drivers are global risk aversion, the global interest rate, global and domestic economic output growth, and country risk. Accordingly, Hooker (2004) agrees that changes in the gross domestic product (GDP) growth and country risk (more specifically, sovereign credit risk) affect the emerging equity market, and adds the impact of changes in the local exchange rate. Furthermore, Ahmed and Zlate (2014) include domestic interest rates as they find that interest rate differentials between emerging and developed markets are significant drivers of capital flow into emerging markets.

To answer the research question, we estimate the extreme Value-at-Risk for emerging stock markets conditional on the aforementioned covariates. For the estimation, we use the extremal random forest (ERF) algorithm of Gnecco et al. (2022). We assess the performance of the ERF relative to other machine learning models, such as the generalized random forest of Athey et al. (2019) and the gradient boosting for extremes algorithm of Velthoen et al. (2021). We perform an explanatory data analysis and an out-of-sample analysis both on an index of the emerging market as a whole and on the emerging countries, separately. Here, we conclude that the extreme conditional VaR positively correlates with the Volatility Index, suggesting that the potential loss increases when investors expect more market volatility. The negative returns of Malaysia, the United Arab Emirates and Saudi Arabia have fat tails, and, for these countries, the ERF outperforms the other methods. Therefore, we conclude that the ERF is the best method to predict the extreme conditional VaR for emerging markets where negative returns occur more frequently.

This paper is structured in the following way. First, we give an overview of the relevant literature, see Section 2. In Section 3, we discuss the data and its key characteristics. Then, we discuss the analysis in more detail in Section 4, and, thereafter, the results in Section 5. Finally, in Section 6, we draw conclusions and try answer the research question.

2 Literature Review

Following Gnecco et al. (2022), our paper creates a bridge between the literature on extreme value theory and the literature on random forests. This makes the relevance of the research twofold, and, therefore, the literature review is split into two sections. We start by discussing the relevant literature on extreme value theory in Section 2.1. Thereafter, Section 2.2 discusses the machine learning models researched thus far to predict extreme conditional Value-at-Risk (VaR).

2.1 Extreme Value Theory

By estimating the extreme conditional quantiles, we study the tail behavior of the conditional distribution. This poses challenges as there are only a few to no observations for these quantiles potentially leading to large biases in the empirical estimators that are based on quantile losses (Gnecco et al., 2022). For this reason, extreme value theory researches the asymptotic results of the extrapolation to the extreme quantiles (Velthoen et al., 2021). In the upcoming paragraphs, we elaborate on extreme value theory and its role in predicting extreme conditional quantiles.

In extreme value theory, two main approaches exist, namely modeling of (1) the distribution of maximum realizations, and (2) the exceedances of a particular threshold (Gencay & Selçuk, 2004). The first approach covers the *block maxima models* which use the Fisher-Tippett theorem to deal with the convergence of maxima. McNeil (1998) uses a block maxima model as it has an easy interpretation in terms of time horizons. The researcher applies the model to daily stock returns to estimate the tail index and to calculate quantiles. Moreover, the block maxima models are often used to estimate the probability of extreme weather events (McNeil, 1998; Gilleland and Katz, 2006). Gumbel (1958) introduced the block maxima approach, and, as it dates back to 1958, it is the older approach.

However, for estimating extreme conditional quantiles, we are interested in more than solely the behavior of the maxima. We are interested in the distribution of observations exceeding a certain threshold, and, therefore, we discuss the second approach in more depth. The second approach covers the *peaks-over-threshold (POT) models*. The POT models study the distribution of the observations exceeding a predefined threshold u, hence, the exceedances $Y_t > u$ (Gencay & Selçuk, 2004). There are two different types of POT models, namely semi-parametric models and fully parametric models. The semi-parametric models rely on the Hill estimator that makes an inference about the tail behavior (Hill, 1975). The fully parametric models are built on the generalized Pareto distribution (GPD) from Pickands (1975).

As mentioned earlier, there are only a few or no observations suitable for estimating extreme conditional quantiles. Therefore, extreme value theory proposes the use of tail approximation (e.g., De Haan and Ferreira, 2006), and, in this paper, we follow Gnecco et al. (2022) by applying the GPD for tail approximation. Here, an intermediate quantile level is first estimated with classical regression methods, and, thereafter, the estimate is extrapolated to the extreme quantile level with the GPD (Gencay and Selçuk, 2004; Gnecco et al., 2022; Embrechts et al., 1998). Section 4 elaborates more on the use of the GPD and extreme value theory in this paper.

2.2 Extreme Conditional Value-at-Risk Estimation

In this section, we discuss the use of models to predict extreme conditional Value-at-Risk (VaR). First, we elaborate on methods that do not necessarily rely on the generalized Pareto distribution (GPD) parameters. Thereafter, we discuss different models that do use these GPD parameters for extrapolation, as discussed in Section 2.1. There have been several studies using machine learning models for this extrapolation such as neural networks (Shrivastava et al., 2014; Taylor, 2000; Diagne, 2002), a quantile regression forest (Meinshausen & Ridgeway, 2006), a generalized random forest (Athey et al., 2019), a gradient boosting method (Velthoen et al., 2021), and an extremal random forest (Gnecco et al., 2022). In the upcoming paragraphs, we elaborate on these different machine learning models.

Artificial neural networks combined with extreme value theory have been used by several researchers, such as Taylor (2000) and Diagne (2002). Taylor (2000) explains the importance of the estimation of the tail distribution for risk management, more specifically for Value-at-Risk models. Therefore, the researcher uses a neural network to estimate non-linear quantile models for multiperiod returns. Also, Diagne (2002) emphasizes the importance of accurate estimation of the tail distribution for risk management. The researcher claims that neural networks combined with extreme value theory can provide fundamental insights for VaR estimation.

As an alternative to neural networks, there are several types of random forests proposed to estimate extreme conditional quantiles. The random forest was initially introduced by Breiman (2001), and Meinshausen and Ridgeway (2006) generalize this random forest by proposing a quantile regression forest (QRF). As the QRF looks beyond the conditional mean, it is able to capture more of the aspects of the conditional distribution of the response variables. Meinshausen and Ridgeway (2006) conclude that the QRF outperforms linear and tree-based methods, and, therefore, is a competitive algorithm for estimating conditional distributions. Note that the QRF does not rely on the GPD parameters for extrapolation to the extreme conditional quantiles (Gnecco et al., 2022).

Also, Athey et al. (2019) base their approach on the random forest of Breiman (2001) and propose a generalized random forest (GRF). They consider an estimator for the extreme conditional quantile with forest-based weights $w_n(\cdot, \cdot)$. These forest-based weights are derived from the trees in which an observation appears. The regression forest proposed by Breiman (2001) takes the average prediction across the different trees to predict an observation. In contrast, the GRF uses adaptive nearest neighbor estimation to allow for statistical extensions. Velthoen et al. (2021) and Gnecco et al. (2022) extend the work of Athey et al. (2019) by applying the GRF for extreme conditional quantiles, and by incorporating it into their 2-step approach. Note that, similar to the QRF, the GRF does not depend on the GPD parameters for the extrapolation.

Additionally, Velthoen et al. (2021) propose a gradient boosting model for their approach, named gradient boosting for extremes algorithm (GBEX). Here, they first use the generalized random forest of Athey et al. (2019) to estimate the intermediate conditional quantiles at τ_0. Thereafter, they estimate the GPD parameters with a gradient boosting algorithm where the optimal GPD parameters minimize a negative log-likelihood. In their simulations, they conclude that the gradient boosting model outperforms classical methods from quantile regression and extreme value theory.

However, Gnecco et al. (2022) claim that forest-based approaches are more favorable than gradient boosting methods and neural networks as forest-based approaches require less tuning and have more understandable statistical properties. Therefore, they propose an extremal random forest (ERF) for predicting the extreme conditional quantiles. First, they estimate the set of weights $w_n(\cdot, \cdot)$ with the GRF. Next, they predict the extreme conditional quantiles by estimating the GPD parameters, and, similar to Velthoen et al. (2021), the optimal GPD parameters minimize a weighted negative log-likelihood. Gnecco et al. (2022) show that the ERF and the GBEX outperform the GRF for both the simulation study and a US wage data set. As the ERF either outperforms or performs equally well to the GBEX, we study the performance of the ERF on a new data set. More specifically, we apply the ERF to emerging stock market data where the Value-at-Risk is predicted.

3 Data

We consider two different data sets. In Section 3.1, we discuss the data set for the replication of Gnecco et al. (2022). Thereafter, we discuss the data set used for answering our research question, see Section 3.2.

3.1 US Wage Data

The first data set consists of 65,023 observations and Table 1 shows the descriptive statistics. The data set contains information on US-born men, namely their weekly wage, the years of education they have received, their age and their race. The variable race is 1 if the person is black and 0 if white. Furthermore, the variable age ranges between 40 and 49 years with a mean of approximately 44 years and 4 months. The variable education ranges from 5 to 20 years with a mean of 12.888 years. Gnecco et al. (2022) use the weekly wage as the response variable Y, and age, education and race as the covariates for the extremal random forest. The weekly wage has a mean of around 719.494 US Dollar and a standard deviation of approximately 623.074. Moreover, the weekly wage is positively skewed with a skewness of 43.383 and has fat tails with a kurtosis of 4549.361.

Table 1. *Descriptive Statistics Of the Weekly Wages, Age, Years Of Education and Race For 65,023 US-Born Men.*

	Descriptive Statistics					
Variable	Mean	Minimum	Maximum	Std. Dev.	Skewness	Kurtosis
Weekly wage	719.494	0.157	80155	623.074	43.383	4549.361
Age	44.351	40	49	2.901	0.076	1.767
Education	12.888	5	20	3.098	0.260	3.028
Race	0.076	0	1	0.266	3.188	11.161

Note. The numbers are rounded to three decimal places. Race is 0 for white and 1 for black individuals. Std. Dev. stands for standard deviation.

3.2 Emerging Markets and Covariates

We use the second data set to study the extreme conditional Value-at-Risk (VaR) for emerging equity markets. Here, we consider a response variable and several relevant covariates. For the response variable, we use the Morgan Stanley Capital International (MSCI) Emerging Market Index. To model the extreme conditional VaR, we require appropriate covariates that describe the variation in the emerging equity markets. The relevant covariates are the Volatility Index, the US 3-month Treasury-Bill rate, the world GDP growth, the emerging market GDP growth, the JP Morgan Emerging Market Bond Index, and the Emerging Market Economies US Dollar Index. To deepen the research, we consider the countries in which the MSCI Emerging Market

Index is invested, separately (BlackRock, 2022b). Here, we take the corresponding MSCI Index for China, Taiwan, India, South Korea, Brazil, Saudi Arabia, South Africa, Mexico, Thailand, Indonesia, Malaysia, the United Arab Emirates and Qatar as response variables, for more details on these variables see Table A1 Appendix A. In the upcoming paragraphs, we first justify the use of these variables. Thereafter, we discuss the time period and frequency considered, and, lastly, we explain how we cleaned the data to arrive at the total number of observations.

To start, we elaborate on the response variable used, namely the Morgan Stanley Capital International (MSCI) Emerging Market Index. This index is a proxy for the emerging equity markets. More specifically, we take the negative returns of the MSCI Emerging Market Index as the response variable, so that the Value-at-Risk is positive. We collect data from an Exchange Traded Fund (ETF) offered by BlackRock tracking the MSCI Emerging Market Index. This ETF is named the iShares MSCI Emerging Markets ETF. We use the ETF to ensure we have enough data points as the data on the index itself is released monthly while the data on the ETF is released daily. The same approach applies to the MSCI Indices of the separate countries, for which we also collect data from the corresponding iShares ETF offered by BlackRock.

Next, we discuss six relevant covariates. Koepke (2019) and Hooker (2004) identify a number of drivers for the emerging market capital flow, and, here, we follow their choice of relevant covariates and the appropriate proxies thereof. The first covariate is the Volatility Index (VIX) retrieved from the Chicago Board Options Exchange (CBOE) website. The VIX is used as a proxy for global risk aversion. Second, we use the US 3-month Treasury-Bill (TB) rates from the US Federal Reserve website. The US TB rates are used as a proxy for the global interest rates. The third covariate is the Emerging Market US Dollar Index (EMUSDI) retrieved from the Board of Governors of the Federal Reserve System website. The EMUSDI is a proxy for the emerging markets' currency relative to the US Dollar. The US Dollar Index describes the value of the US Dollar against currencies used in international trade, in this case, currencies from emerging markets (Logue & Rasure, 2022). Next, we consider the JP Morgan Emerging Market Bond Index (EMBI) which measures the performance of government and corporate bonds in emerging markets (Hayes & Scott, 2020). Therefore, we use the EMBI as a proxy for both sovereign credit risk and emerging market interest rates. Here, we approach the data gathering in a similar way as for the response variable, namely we collect data from a BlackRock ETF tracking the EMBI. The ETF is named the iShares J.P. Morgan USD Emerging Markets Bond ETF. Lastly, the fifth and sixth covariates are the emerging market GDP growth and the world GDP growth which are both retrieved from the International Monetary Fund (IMF) website.

For all the time series, we consider the period from December 13, 2007, until April 29, 2022. Furthermore, we use daily data so that there are enough data points for the algorithm. The emerging market GDP growth rate and the world GDP growth rate are released annually, and, therefore, we transform them. This transformation results in a data set with daily world GDP growth rates and another data set with daily emerging market GDP growth rates. Here, we assume that the percentage growth is equally distributed over the trading days in a year. We transform the yearly GDP growth rates as follows: $daily\ growth\ rate = \sqrt[n]{yearly\ growth\ rate}$ where $n = 253$ is the average number of trading days in a year (NYSE, 2022).

Lastly, we clean the data by correcting for missing observations; if an observation is missing, then the average of the four neighbour observations is taken to smooth out the data. As the data is from financial markets, observations on weekend days and banking holidays are missing, however, we disregard this and assume the data to be continuous after cleaning it. All prices are in US Dollar and the data bases are accessed on 06/05/2022. Table 2 shows the descriptive statistics of the MSCI Emerging Markets ETF and the relevant covariates. As we use daily data, we have 3620 observations for each time series.

The MSCI Emerging Markets ETF has a mean of 41.498 and a standard deviation of 6.237. Furthermore, the Emerging Market ETF is negatively skewed with a skew of -0.551. To estimate the extreme conditional VaR, we consider the negative returns of the MSCI Emerging Markets ETF. The negative returns range between approximately -13.7% and 10.8% with a mean of -0.004%, suggesting a positive daily return on average. Furthermore, the negative Emerging Markets ETF returns are negatively skewed at -0.114, and the returns have fat tails with a kurtosis of 15.592. The VIX and the emerging market GDP growth are also fat tailed with kurtoses of 10.737 and 12.379, respectively. Moreover, the TB rate has a negative minimum of -0.05% meaning that the return on a Treasury-Bill was negative at least one day in the period of the time series. Also, the GDP growth rates have negative minimum values of aprroximately -98.6% and -98.5% for the world and the emerging markets, respectively. These values arose during the COVID19 crisis as the lockdowns had massive effects on the GDP of countries around the world. Lastly, the Emerging Market US Dollar Index has a high standard deviation of 12.658 and a mean of 108.765 which is approximately equal to the mean of the Emerging Market Bond Index at 108.017.

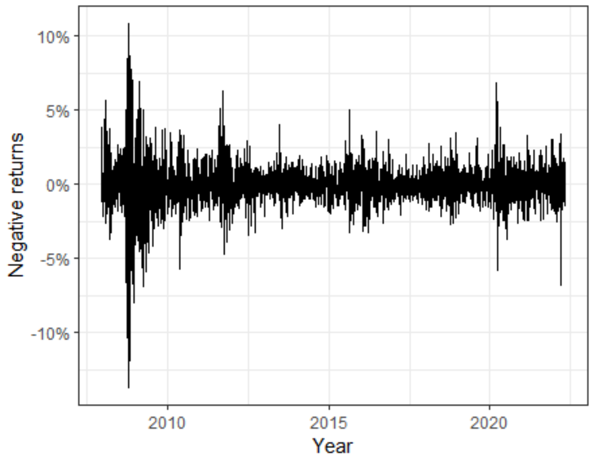
Table 2. *Descriptive Statistics Of MSCI Emerging Markets (EM) ETF, Its Negative Returns and the Relevant Covariates For the Period December, 2007, Until April, 2022.*

	Descriptive Statistics					
Variable	Mean	Minimum	Maximum	St.Dev.	Skewness	Kurtosis
MSCI EM ETF	41.498	19.118	58.141	6.237	-0.551	4.213
Negative returns	-0.004	-13.682	10.793	1.360	-0.114	15.592
VIX	20.246	9.14	82.69	9.610	2.363	10.737
EM Bond Index	108.017	68.233	122.414	8.059	-1.517	6.221
TB rate	0.546	-0.050	3.290	0.763	1.409	3.604
EM USD Index	108.765	89.152	137.310	12.658	0.162	1.527
World GDP growth	71.267	-98.636	98.901	68.163	-2.081	5.329
EM GDP growth	85.023	-98.466	99.026	50.312	-3.373	12.379

Note. The numbers are rounded to three decimal places. Std. Dev. stands for standard deviation. Negative returns, TB rate, world and EM GDP growth are in percentages. The data is from BlackRock (2022a), Federal Reserve Bank of St. Louis (2022), Organization for Economic Cooperation and Development (2022), International Monetary Fund (2022).

Furthermore, Figure 1 plots the negative returns of the MSCI Emerging Markets ETF against time, and Figure A15 in Appendix A plots the covariates against time. In Figure 1, we observe time-varying volatility as some periods show more up-and-down movement in the negative returns compared to other periods. For example, around the Great Financial Crisis in 2008, the negative returns range approximately between 10% and -15%. Additionally, we observe more volatile negative returns around 2020 with the start of the COVID19 pandemic.

Figure 1. Negative Returns Of the MSCI Emerging Market ETF For the Period December, 2007, Until April, 2022.



4 Methodology

In this paper, we follow the extremal random forest algorithm and the corresponding notation as proposed by Gnecco et al. (2022). To start, we will define and explain the notation used in this paper, see Section 4.1. In Section 4.2, we explain the functioning of the extremal random forest (ERF) algorithm. Thereafter, we conduct a simulation study to assess the performance of the ERF and other algorithms, see Section 4.3. Lastly, in Section 4.4, we assess the performance of the ERF and other algorithms on US wage data and emerging market data.

4.1 Extreme Conditional Quantiles

Following the notation used by Gnecco et al. (2022), we consider response variable $Y \in \mathbb{R}$ and predictor X from a potential range of predictors $\chi \subset \mathbb{R}^p$, where p is large to ensure large dimensions. For the first data set, the response variable is the weekly wage; the set of predictors contains the variables education, age and race. For the second data set, the response variable is the MSCI Emerging Market Index, and the set of predictors contains all the relevant covariates discussed in Section 3.2. Additionally, let $(X_1, Y_1), ..., (X_n, Y_n)$ be n independent copies of the random vector (X, Y). We define $Q_x(\tau)$ as the quantile at level $\tau \in (0, 1)$ of the conditional distribution of $Y|X = x$, in short: $Q_x(\tau) = F^{-1}_{Y|X=x}(\tau)$. We are interested in the conditional quantile $Q_x(\tau)$ for $\tau \approx 1$ as this is the extreme quantile.

Furthermore, Gnecco et al. (2022) find two challenges for estimating the conditional quantile $Q_x(\tau)$. First of all, there are only a few to no observations for these extreme conditional quantiles, resulting in a large bias in the empirical estimators that are based on quantile losses. To address this first challenge, they propose the use of tail approximation as is motivated by extreme value theory (e.g., De Haan and Ferreira, 2006). Here, they first estimate the quantile $Q_x(\tau_0)$ for an intermediate quantile level $\tau_0 < \tau$, and, thereafter, they extrapolate it to the extreme quantile $Q_x(\tau)$, going beyond the range of the data, more details in Section 4.2. We define the observations in the data that exceed the quantile $Q_x(\tau_0)$, in other words, the exceedances, as $Z_i = \max(0, Y_i - Q_x(\tau_0))$. To extrapolate, Gnecco et al. (2022) use the approximation by the generalized Pareto distribution (GPD) of the exceedances. Following Pickands (1975), the generalized Pareto distribution is:

$$G(z;\theta) = 1 - \max\left(0, \left(1 + \frac{\xi}{\sigma}z\right)^{-\frac{1}{\xi}}\right), \quad z > 0, \tag{1}$$

where $\theta = (\sigma, \xi) \in (0, \infty) \times \mathbb{R}$ is the parameter vector consisting of scale parameter $\sigma : \chi \to (0, \infty)$ and shape parameter $\xi : \chi \to \mathbb{R}$. Consequently, we extrapolate as follows:

$$Q_x(\tau) \approx Q_x(\tau_0) + \frac{\sigma(x)}{\xi(x)}\left[\left(\frac{1-\tau}{1-\tau_0}\right)^{-\xi(x)} - 1\right], \tag{2}$$

which approximates the conditional quantile of $Y|X = x$ at $\tau \approx 1$. The extrapolation to the extreme conditional quantile holds under mild assumptions (Balkema and De Haan, 1974; Pickands, 1975). Here, the response variable Y is heavy-tailed when $\xi > 0$ (e.g., Student's t), light-tailed when $\xi = 0$ (e.g., Gaussian), and Y has a finite upper endpoint when $\xi < 0$ (e.g., uniform) (Velthoen et al., 2021).

The second challenge is regarding the dimensional of the predictor space $\mathbb{R}^p$ which might get relatively large. For this reason, several solutions and alternative models that can cope with these high-dimensional spaces have been proposed in the literature, see Taylor (2000) and Friedman (2002). Gnecco et al. (2022) focus on forest-based approaches which are methods based on the original random forest developed by Breiman (2001). These methods require little tuning and their statistical properties are understandable compared to, for example, gradient boosting and neural networks (Athey et al., 2019). Therefore, our paper models and estimates the extreme conditional quantile $Q_x(\tau)$ of $Y|X = x$ with a random forest, as is done by Gnecco et al. (2022).

4.2 The Extremal Random Forest Algorithm

In this section, we discuss the extremal random forest algorithm of Gnecco et al. (2022) and elaborate on the necessary steps. The independent copies $(X_1, Y_1), ..., (X_n, Y_n)$ are the training data. First, we specify both an intermediate quantile level τ_0 and an extreme quantile level τ. The intermediate quantile level $\tau_0 \in (0, 1)$ is chosen so that the classical quantile regression techniques can be used to obtain estimator $\hat{Q}_x(\tau_0)$. We follow Gnecco et al. (2022) by using a generalized random forest (GRF) with quantile loss to estimate $\hat{Q}_x(\tau_0)$ for $\tau_0 = 0.8$.

After estimating the intermediate conditional quantile, we will extrapolate to the extreme conditional quantile. Gnecco et al. (2022) assume that $Y - Q_x(\tau_0) \mid Y > Q_x(\tau_0)$ approximately follows a generalized Pareto distribution (GPD) so that $G(z; \theta(x))$ from Equation (1) is the cumulative distribution function. Therefore, we extrapolate by fitting a GPD. To do so, we estimate the parameter vector $\theta(x) = (\sigma(x), \xi(x))$ with maximum-likelihood. The log-likelihood function is as follows:

$$\ell_\theta(Z_i) = \begin{cases} \log \sigma + \left(1 + \frac{\xi}{\sigma} Z_i\right) & \text{if } Z_i > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Note that the log-likelihood only gives positive values for the exceedances, $Y_i > \hat{Q}_x(\tau_0)$. Next, Gnecco et al. (2022) estimate localizing weight functions $w_n(x, X_i)$ with a GRF (which may be different from the GRF used to estimate $\hat{Q}_x(\tau_0)$). They formalize the weighted log-likelihood function, that is:

$$L_n(\theta; x) = \sum_{i=1}^{n} w_n(x, X_i)\ \ell_\theta(Z_i)\ \mathbb{1}\{Z_i > 0\}, \tag{4}$$

where $w_n(x, X_i)$ are the weight functions. To get the parameter vector estimator $\hat{\theta}(x)$, we minimize the weighted log-likelihood function:

$$\hat{\theta}(x) = \underset{\theta \in \Theta}{\arg\min}\, L_n(\theta; x). \tag{5}$$

Next, we use the estimated intermediate quantile $\hat{Q}_x(\tau_0)$ and parameter vector $\hat{\theta}(x) = (\hat{\sigma}(x), \hat{\xi}(x))$ to calculate the estimated extreme conditional quantile $\hat{Q}_x(\tau)$ with Equation (2). The estimators are consistent under certain assumptions, see Gnecco et al. (2022).

Furthermore, the GRF has several parameters that need to be tuned, thus, we discuss a cross-validation scheme following Gnecco et al. (2022). The researchers find that the minimum node size $\kappa \in \mathbb{N}$ is the most critical parameter for ERF. Therefore, for κ, we consider a sequence $\alpha_1, .., \alpha_J$ of possible values. To tune this parameter, we perform a 5-fold cross-validation three times and grow forests of 50 trees at each fold. Let $\mathcal{N}_1, ..., \mathcal{N}_M$ be a random partitioning of $1, ..., n$ into M equally sized folds of the training data, in this case, $M = 5$. We fit the ERF on the training set $(X_i, Y_i), i \notin \mathcal{N}_m$ for each fold m and each $\alpha_j,\ j \in 1, ..., J$. Next, we estimate the GPD parameter vector $\hat{\theta}(X_i; \alpha_j)$ on $(X_i, Y_i), i \in \mathcal{N}_m$. The cross-validation error is as follows:

$$CV(\alpha_j) = \sum_{m=1}^{M} \sum_{i \in \mathcal{N}_m} \ell_{\hat{\theta}(X_i; \alpha_j)}(Z_i) \mathbb{1}\{Z_i > 0\}, \tag{6}$$

and is minimized for tuning parameter $\alpha*$. We repeat this procedure 50 times. Over these 50 simulations, we measure the performance as the square root of the mean integrated squared error ($\sqrt{\text{MISE}}$), see Section 4.3 for more details on the MISE.

Additionally, the maximization of the likelihood could experience convergence problems in small samples (Coles & Dixon, 1999). Therefore, Gnecco et al. (2022) propose the penalized log-likelihood:

$$\hat{\theta}(x) = \underset{(\sigma,\xi)=\theta \in \Theta}{\arg\min} \frac{1}{1 - \tau_0} L_n(\theta; x) + \lambda(\xi - \xi_0)^2, \tag{7}$$

where $\lambda \geq 0$ is a tuning parameter and ξ_0 is a constant shape parameter. For λ, we take the same steps to cross-validate this parameter and to assess the MISE as described in the previous paragraph for κ. Lastly, we follow Gnecco et al. (2022) and set the constant shape parameter ξ_0 equal to the unconditional fit $\hat{\xi}$ which is obtained by minimizing the GPD with constant weights $w_n(x, y) = 1$ for all $x, y \in \chi$.

4.3 Simulation Study

To assess the performance of the extremal random forest, we conduct a Monte Carlo simulation and compare different approaches, following Gnecco et al. (2022). First, we simulate the training data by generating $(X_1, Y_1), ..., (X_n, Y_n)$ independent copies of a random vector (X, Y). The response variable $Y|X = x$ follows a distribution with fat tails such as a Student's t distribution. We perform two types of simulations as done by Gnecco et al. (2022), for more details, we refer to Experiments 1 and 2 in Sections 4.3 and 4.4 of Gnecco et al. (2022). We assess the performance of the algorithm by computing the integrated squared error (ISE) as follows: $ISE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{Q}_{x_i}(\tau) - Q_{x_i}(\tau)\right)^2$. We simulate $m = 50$ times, and, thereof, we obtain a mean integrated squared error (MISE) which is the average of the ISEs.

We compare the extremal random forest (ERF) of Gnecco et al. (2022) with the generalized random forest (GRF) by Athey et al. (2019), the gradient boosting for extremes algorithm (GBEX) by Velthoen et al. (2021), and the unconditional GPD. For the ERF, we follow Gnecco et al. (2022) and repeat three times 5-fold cross-validation where $\kappa \in \{10, 40, 100\}$ is the minimum node size and $\lambda \in \{0, 0.01, 0.001\}$ is the penalty for the shape parameter. The other tuning parameters are set at their default values. For GRF, we follow Gnecco et al. (2022) by setting the tuning parameters of the GRF to the default values and fitting $\hat{Q}_x^{GRF}(\tau)$ to the training data. For GBEX, we follow Velthoen et al. (2021) and use 5-fold cross-validation to determine the optimal number of trees with a maximum per fold of 500. Furthermore, we set the depth of each gradient tree to $D = 2$ and the learning rate for the scale parameter to $\lambda^\sigma = 0.1$. We again set the other tuning parameters to their default values.

4.4 US Wage Quantiles and Emerging Markets Value-at-Risk

We compare the performance of the ERF, GRF, GBEX and the unconditional GPD on the US wage data and on the emerging markets data, see Section 3 for information on the data sets. For the GRF and GBEX, we use the *grf package* offered by Tibshirani et al. (2022) and the *gbex package* offered by Velthoen (2022) in Rstudio, respectively. Furthermore, we use the functions of the *erf package* offered by Gnecco (2022) for the ERF. As we would like to assess the methods for higher dimensional data sets, we add ten and seven random predictors drawn independently from a uniform distribution on $[-1, 1]$ to the US wage data set and the emerging markets data set, respectively, resulting in $p = 13$. First, we split the data sets into two subsets of equal size. For the US wage data, we split the data set in the middle. For the emerging market data, we make a random split to avoid biases, as stock prices experience time-varying volatility. Following Gnecco et al. (2022), we thereafter use the first subset to perform an exploratory data analysis and the second subset to evaluate the different methods. Throughout the whole analysis, we take $\tau_0 = 0.8$, $\lambda = 0.01$ and perform a three time 5-fold cross-validation with minimum node size $\kappa \in \{5, 40, 100\}$ for the ERF. Furthermore, we consider the same tuning parameters as discussed in Section 4.3 for the other methods. In the upcoming paragraphs, we elaborate on this approach in more detail.

We start with the exploratory data analysis on the first subset, called data set A from here onward. From data set A, we start by taking a random subset of 10% for the US wage data (3,251 observations) and 20% for the emerging market data (362 observations). The emerging market data set contains fewer observations, and, therefore, to ensure enough data points for the algorithm, we take a bigger subsample. We predict the GPD parameters $\hat{\theta}(x) = (\hat{\sigma}(x), \hat{\xi}(x))$ on the remaining data from data set A, hence, we do an out-of-sample prediction. Here, we assess the shape of the GPD parameters. Thereafter, we assess the shape of the extreme conditional quantiles for the different methods by predicting the extreme conditional quantiles with the ERF, GRF, GBEX and the unconditional GPD.

Afterwards, we assess the performance of the ERF compared to the GRF, GBEX and the unconditional GPD on the second subset, called data set B from here onward. As the actual extreme conditional quantiles are unknown, we need a different performance measure than for the simulation study in Section 4.3. We use the performance measure proposed by Wang and Li (2013):

$$\mathcal{R}_n(\hat{Q}_{\cdot}(\tau)) = \frac{\sum_{i=1}^{n} \mathbb{1}\{Y_i < \hat{Q}_{X_i}(\tau)\} - n\tau}{\sqrt{n\tau(1-\tau)}}, \tag{8}$$

where n is the number of test observations and $\hat{Q}_{\cdot}(\tau)$ is the τ-th conditional quantiles estimated on the training data set. According to the central limit theorem, the performance measure $\mathcal{R}_n(\hat{Q}_{\cdot}(\tau))$ is asymptotically standard normal. Following Gnecco et al. (2022), we partition data set B into ten random folds. We fit the different methods on each fold, the training data, and predict the extreme quantiles for the left-out observations, the test data. We assess the performance with Equation (8). To deepen the insides in the emerging equity market, we repeat this analysis two more times in different settings after performing the analysis on data set B. First, we partition the MSCI Emerging Market Index over time so that the methods are trained on year t and are estimating the extreme conditional VaR for year $t + 1$. Second, we assess the performance of the countries in which the MSCI Emerging Market Index is invested, separately. Note that we do not add the random predictors for the different countries so that we can better investigate the influence of the economic variables on the extreme conditional VaR. Finally, we give an economic interpretation of these extreme conditional quantiles and the variables driving them.

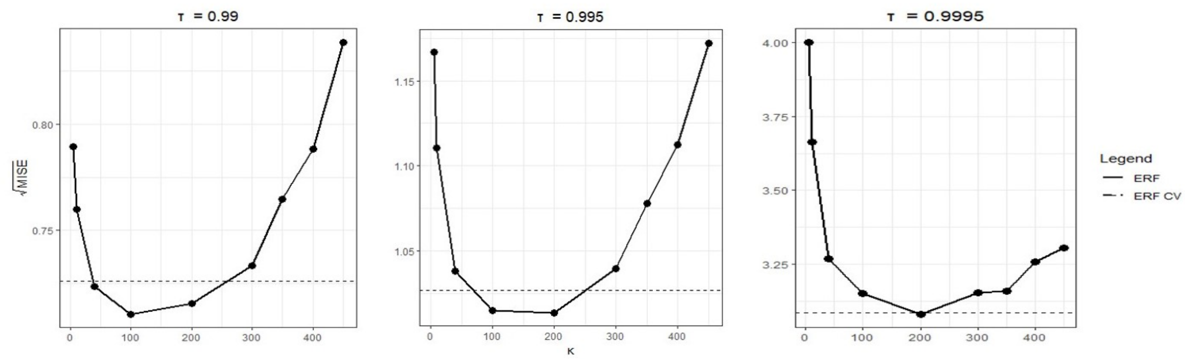
5 Results

In this section, we start by discussing the results of the parameter tuning, see Section 5.1. Thereafter, we compare the performance of the extremal random forest (ERF) to other methods. Here, we apply the ERF and the other methods to simulated data in Section 5.2, US wage data in Section 5.3, and emerging stock markets data in Section 5.4.

5.1 Parameter Tuning

The generalized random forest (GRF) has several parameters that need to be tuned which we assess over 50 simulations. Here, we discuss the results of tuning the minimum node size κ and the penalty parameter λ. Figure 2 shows the results for the minimum node size for quantile levels $\tau = 0.99$, 0.995, 0.9995. We observe that the $\sqrt{\text{MISE}}$ is minimized for minimum node sizes between 100 and 200, where the minimum node size increases to 200 for a more extreme quantile level $\tau = 0.9995$. Furthermore, the cross-validated ERF is in all cases relatively close to the minimum $\sqrt{\text{MISE}}$, and, for $\tau = 0.9995$, the cross-validated ERF gives the exact minimum $\sqrt{\text{MISE}}$. Therefore, we conclude that the cross-validated ERF performs relatively well, which is in line with the conclusions of Gnecco et al. (2022).

Figure 2. The Square Root Of the MISE For ERF With Different Minimum Node Sizes κ and For the Cross-Validated ERF Over 50 Simulations.



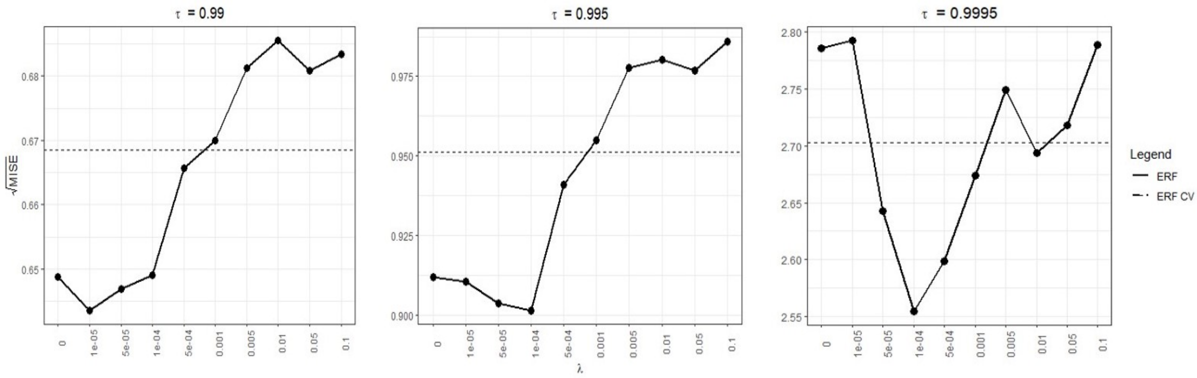
Note. The solid and dashed lines show $\sqrt{\text{MISE}}$ for the ERF with different minimum node sizes κ and with cross-validation, respectively. The data is generated following Gnecco et al. (2022).

The results for the penalty parameter λ are shown in Figure 3. Here, we set the minimum size nodes κ equal to the minimum of the previous analysis, namely 200, and perform the cross-validation purely on the penalty parameter λ. For $\tau = 0.99$ and $\tau = 0.995$, the $\sqrt{\text{MISE}}$ increases with the penalty parameter λ, suggesting that a heavier penalty results in worse predictions of the extreme conditional quantiles. For $\tau = 0.9995$, the plot has a different shape, where the $\sqrt{\text{MISE}}$ first decreases as λ increases, after which it increases again. Furthermore, the performance measure is minimized for λ between 0.00001 and 0.0001. Lastly, we observe that the cross-validated ERF is relatively often above the minimum $\sqrt{\text{MISE}}$.

5.2 Simulation Study

Next, we assess the performance of the ERF compared to the generalized random forest (GRF), the gradient boosting for extremes algorithm (GBEX) and the unconditional generalized Pareto distribution (GPD). We simulate the data 50 times following Experiments 1 and 2 of Gnecco et al. (2022).

Figure 3. The Square Root Of the MISE For ERF With Different Penalties λ and For the Cross-Validated ERF Over 50 Simulations.

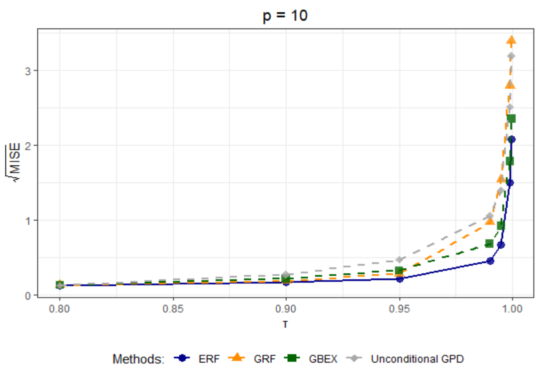


Note. The solid and dashed lines show $\sqrt{\text{MISE}}$ for the ERF with different penalties λ and with cross-validation, respectively. The data is generated following Gnecco et al. (2022).

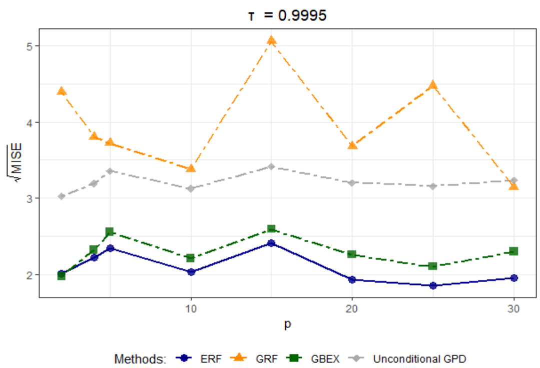
For Experiment 1 of Gnecco et al. (2022), we investigate the performance of the methods for (a) varying quantile levels with a fixed dimension, and (b) varying dimensions with a fixed quantile level. Figure 4a shows the $\sqrt{\text{MISE}}$ for the different methods with $p = 10$ and varying quantile levels τ. We observe that the performance measure increases almost exponentially for extremer quantile levels, $\tau \to 1$, suggesting that all methods have difficulty appropriately predicting extreme conditional quantiles. Over the range of τ, ERF outperforms the other methods by having the lowest $\sqrt{\text{MISE}}$. Furthermore, Figure 4b shows the $\sqrt{\text{MISE}}$ for the different methods with $\tau = 0.9995$ and varying dimensions p. Again, the ERF outperforms all methods in predicting the extreme conditional quantiles. The GRF shows a relatively volatile pattern for the $\sqrt{\text{MISE}}$ compared to the other methods. Lastly, we observe that the $\sqrt{\text{MISE}}$ of both the ERF and the GBEX slightly decreases for higher dimensions (higher p).

Figure 4. The Square Root Of the MISE For ERF, GRF, GBEX and the Unconditional GPD For Varying Quantile Levels τ and For Varying Model Dimensions p Over 50 Simulations.

(a) Varying Quantile Levels τ



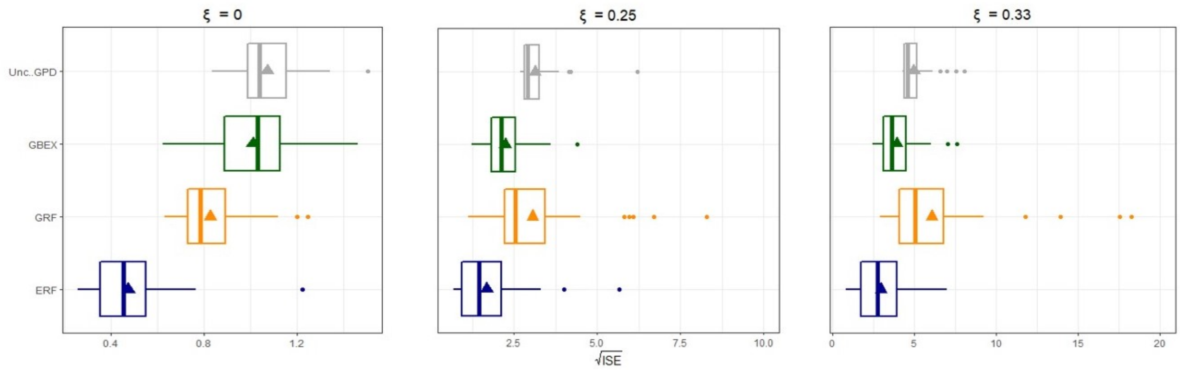
(b) Varying Model Dimensions p



Note. The data is generated following Experiment 1 of Gnecco et al. (2022).

In Experiment 2 of Gnecco et al. (2022), we assess the same performance for varying shape parameters ξ. For the different methods and different shape parameters, Figure 5 shows the square roots of the ISEs from each of the 50 simulations. The ERF gives the lowest average $\sqrt{\text{ISE}}$, especially for the Gaussian distribution ($\xi = 0$). Furthermore, the GBEX is the second best method for the scaled Student t-distribution where $\xi = 0.25$ or 0.33. For the Gaussian distribution, the GRF performs better than both the GBEX and the unconditional GPD, when looking at the average $\sqrt{\text{ISE}}$. However, the GRF has relatively more spread results as can be seen by the outliers for all values of ξ.

Figure 5. The Square Root Of the ISEs For ERF, GRF, GBEX and the Unconditional GPD Over 50 Simulations For Quantile Level $\tau = 0.9995$ and Dimension $p = 40$.



Note. The triangles are the average values. The data is generated following Experiment 2 of Gnecco et al. (2022).

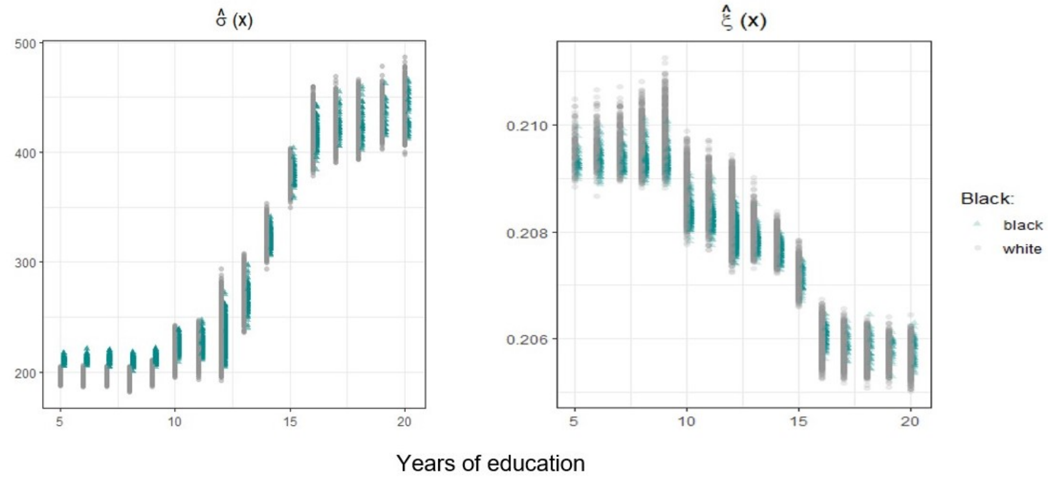
5.3 US Wage Extreme Conditional Quantiles

Furthermore, we compare the performance of the ERF, GBEX, GRF and the unconditional GPD on the US wage data. Following Gnecco et al. (2022), the data set is split into two equal parts. The first part of the data set is used for the exploratory data analysis and the second part is used to fit the different methods and assess their performance.

First, we discuss the exploratory data analysis. We start by fitting the ERF and estimating the GPD parameters. Figure 6 shows the estimated GPD parameters $\hat{\sigma}(x)$ and $\hat{\xi}(x)$ plotted against the years of education and the race. We observe that for both GPD parameters the spread of black and white men is relatively equal over the years of education. Furthermore, the scale parameter $\hat{\sigma}(x)$ increases with the years of education and shows a jump around 15 years of education. The shape parameter $\hat{\xi}(x)$ depends negatively on the years of education and shows a dip around 15 years of education.

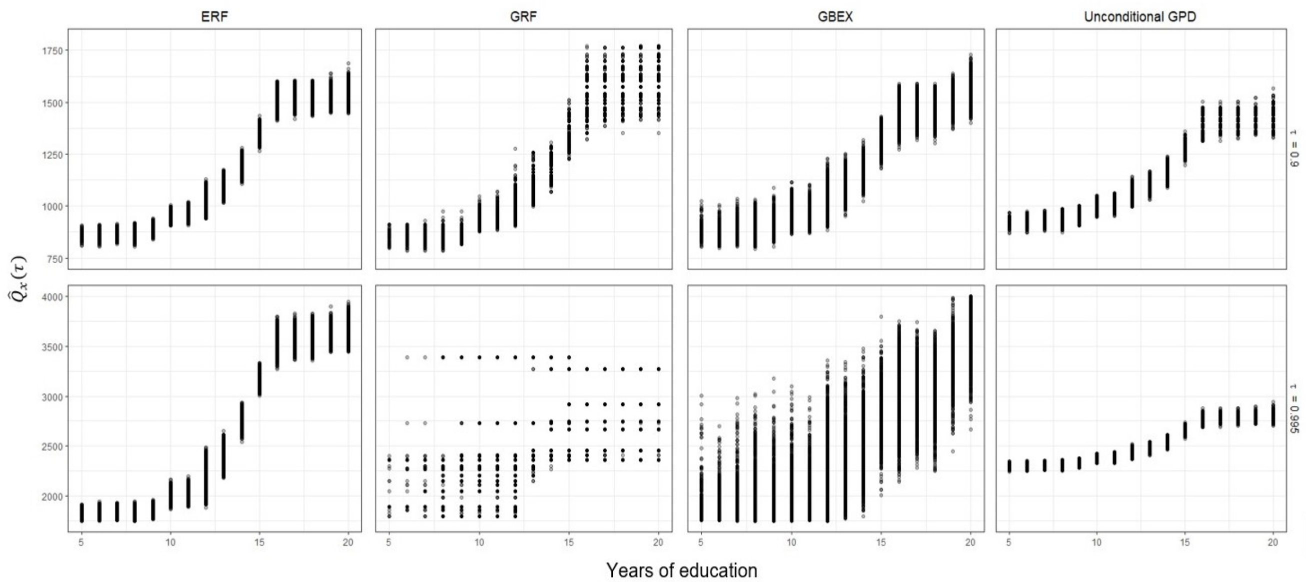
After fitting the ERF and estimating the GPD parameters, we estimate the extreme conditional quantiles using the different methods. Figure 7 shows the extreme conditional quantiles estimated by the ERF, GRF, GBEX and the unconditional GPD for $\tau = 0.9$, 0.995. For $\tau = 0.9$, all methods predict that the extreme conditional quantiles $\hat{Q}_x(\tau)$ increase with the years of education. This implies that higher educated American males earn more in the extremes than

Figure 6. GPD Parameters $\hat{\sigma}(x)$ and $\hat{\xi}(x)$ Plotted Against the Years Of Educations and Race.



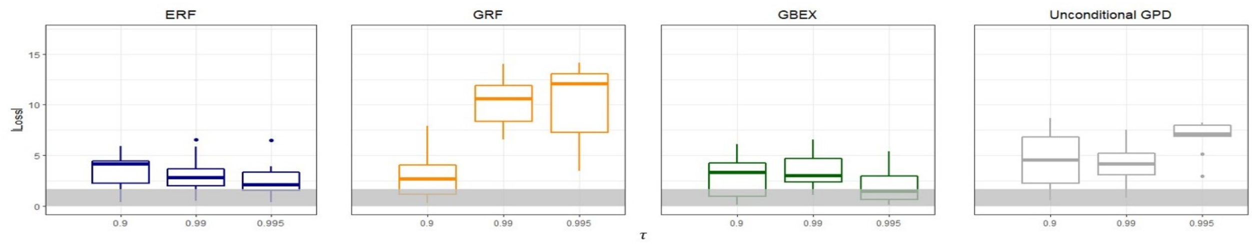
lower educated American males. As the quantile level increases to $\tau = 0.995$, both the GRF and GBEX show a very scattered prediction of the extreme conditional quantiles suggesting that these methods are unable to capture a clear pattern due to reduced flexibility. Furthermore, the unconditional GPD flattens, and, hereby, also loses it flexibility to capture the previous mentioned pattern. Here, the ERF is able to model the variability of the extreme conditional quantiles. Note that we need a formal performance measure to conclude on the best performing model, thus, the behavior of the methods in Figure 7 does not lead to conclusions on their performances.

Figure 7. Extreme Conditinal Quantiles Predicted By ERF, GRF, GBEX and the Unconditional GPD At Quantile Levels $\tau = 0.9$, 0.995 For US Wage Data.



For the second part of the data set, we fit the different methods and assess their performance based on the loss function in Equation (8). Figure 8 shows the absolute values of this loss function for the ERF, GRF, GBEX and the unconditional GPD. The grey area is used as a reference as it represents the 95% interval of the absolute value of a standard normal distribution. Again, the GRF and unconditional GPD perform poorly for increasing extreme quantile levels, where the absolute loss of the GRF makes a jump between $\tau = 0.9$ and $\tau = 0.99$. Gnecco et al. (2022) claim that the unconditional GPD is less flexible compare to the ERF and GBEX as the unconditional GPD cannot produce different scale parameters. Accordingly, we observe that the ERF and GBEX perform very well, and even seem to perform better for more extreme quantile levels.

Figure 8. Absolute Value Of the Loss Function For ERF, GRF, GBEX and the Unconditional GPD On the US Wage Data For τ = 0.9, 0.99, 0.9995.



Note. The grey area is the 95% interval of the absolute value of a standard normal distribution.

5.4 Emerging Stock Markets Extreme Conditional Value-At-Risk

In this section, we evaluate the performance of the ERF, GRF, GBEX and unconditional GPD in predicting the extreme conditional Value-at-Risk (VaR) for emerging equity markets. For Sections 5.4.1 and 5.4.2, the data set is split in two, where the first part of the data set is used for an exploratory data analysis, and the second part is used to fit the different methods and assess their performance. Thereafter, we study the performance of the out-of-sample predictions over time on the full data set in Section 5.4.3. Lastly, in Section 5.4.4, we assess the performance of the different methods for the countries in which the MSCI Emerging Market Index is invested.

5.4.1 Explanatory Data Analysis

We start by fitting the ERF on the first part of the data set, and, thereafter, we estimate the GPD parameters. Figure 9 shows the estimated GPD parameters, $\hat{\sigma}(x)$ and $\hat{\xi}(x)$, as a function of the Volatility Index (VIX), the Emerging Market Bond Index ETF (EMBI ETF) and the Emerging Market US Dollar (EM USD) Index. The different Treasury-Bill (TB) rate ranges are also shown in the plot, where the TB-rate mostly moves in the '$1 - 2\%$' range followed by the '$2 - 3\%$' range and '$> 3\%$' range. Furthermore, Figure 9, with the EM USD Index on the horizontal axis, appears to be scattered, meaning that the GPD parameters as a function of the EM USD Index do not show a clear pattern.

In contrast, we observe that the scale parameter, $\hat{\sigma}(x)$, increases with the VIX and decreases with the EMBI ETF price. Here, the shape parameter, $\hat{\xi}(x)$, has almost the same pattern as it slightly increases with the VIX. Thus, as the VIX represents the expected volatility over the next 30 days (Chicago Board Options Exchange, 2022), both the scale and shape parameters increase in more volatile markets, however, the increase is different in behavior. Moreover, we observe that both the scale and shape parameters decrease with the EMBI ETF, where the pattern of the scale parameter is more visible. The EMBI ETF measures the performance of government and corporate bonds in the emerging markets (Hayes & Scott, 2020), and, thus, the GPD parameters decrease when bonds in the emerging market perform better. Lastly, the scale parameter ranges from 0.371 to 1.866, with an average of 0.748. The shape parameter ranges between 0.202 and 0.271, with an average of 0.233, implying heavy-tails throughout the predictor space.

Figure 9. Estimated GPD Parameters $\hat{\sigma}(x)$ and $\hat{\xi}(x)$ Plotted Against the Volatility Index (VIX), the Emerging Market Bond Index ETF (EMBI ETF), the Emerging Market US Dollar Index (EM USD Index), and the Treasury-Bill (TB) Rate.

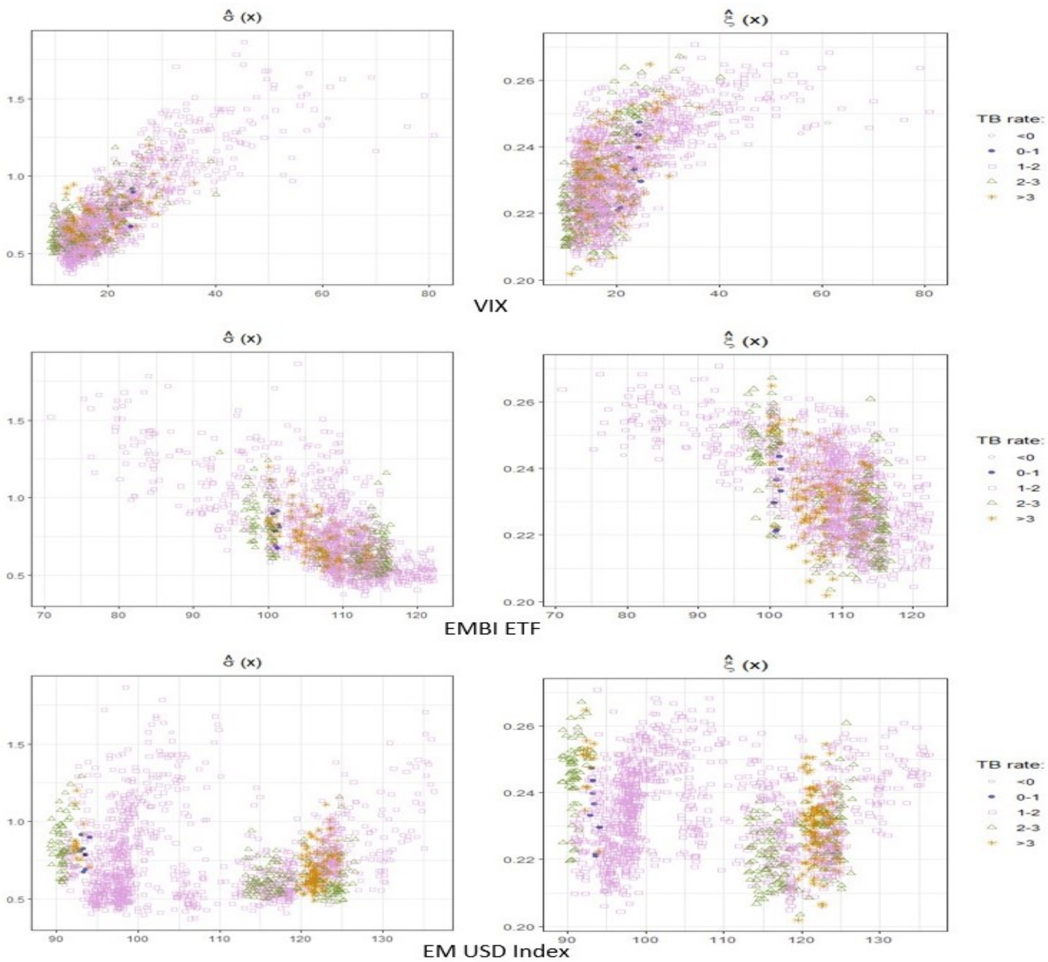


Table 3 shows the importance of each variable in estimating the GPD parameters, with the GBEX making a distinction between the scale and shape parameter. The ERF and GRF have equal values, as these methods are based on the same estimated GRF. For the ERF and GRF, the VIX and the Emerging Market Bond Index have the highest variable importance score, while the Treasury-Bill rate has the lowest. This suggests that expected worldwide market volatility and sovereign credit risk have the biggest impact on the extreme conditional VaR. However, the variables containing information on developed countries, with the US as a proxy, have a less significant effect on the extrapolation with the GPD parameters. Furthermore, the GBEX gives the VIX relatively high importance scores at 0.493 and 0.965 for the scale and shape parameters, respectively, suggesting that the VIX has the largest effect on the GPD parameters. Therefore, we plot the extreme conditional VaR against the VIX in Figure 10.

Table 3. *Variable Importance For the GPD Parameters $\sigma(x)$ and $\xi(x)$ Per Method.*

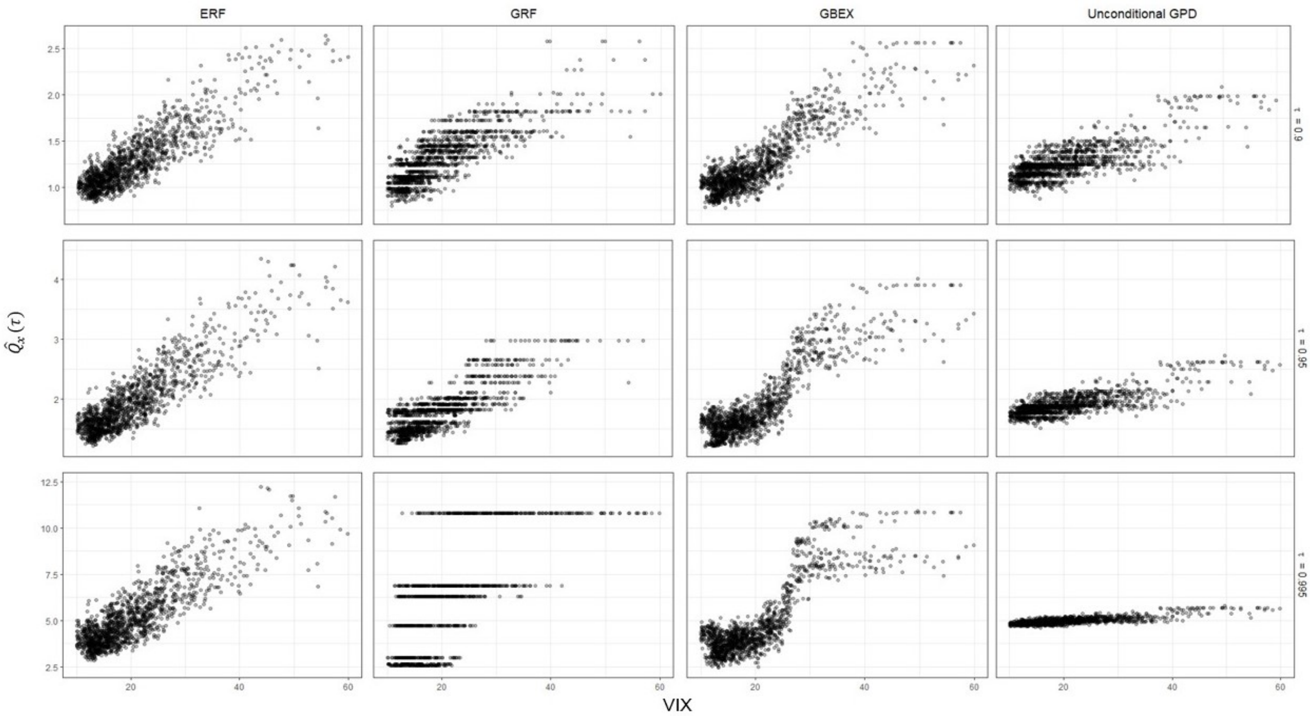
	Variable Importance Per Method			
Variable	ERF	GRF	GBEX, $\sigma(x)$	GBEX, $\xi(x)$
VIX	0.195	0.195	0.493	0.965
EM Bond Index	0.122	0.122	0.000	0.000
TB rate	0.027	0.027	0.061	0.011
EM USD Index	0.054	0.054	0.060	0.000
World GDP growth	0.092	0.092	0.058	0.000
EM GDP growth	0.080	0.080	0.025	0.000

Note. The numbers are rounded to three decimal places. The values of the ERF and GRF represent the combined variable importance of the scale and shape parameters. The values of the GBEX are scaled to sum up to one. The random predictors are excluded from the table.

Next, we predict the extreme conditional VaR. In Figure 10, we observe that the different methods maintain the shape of the scale parameter for quantile level $\tau = 0.9$, in other words, the extreme conditional VaR increases with the VIX. This increase slows down for higher values of the VIX where the extreme conditional VaR also seems more scattered. This could imply that the VaR only maintains a correlation with the VIX in markets that are not extremely volatile. Furthermore, the VaR with $\tau = 0.9$ ranges between 0.7 and 2.6, meaning that the maximum potential loss is between 0.7% and 2.6% with a probability of 0.1. In the middle quantile level $\tau = 0.95$, we observe slight a transition of the plot shape, for example, the GRF does not show an equally spread data plot anymore. However, for the ERF, the plot shape for $\tau = 0.95$ barely changes compared to $\tau = 0.9$. The VaR with $\tau = 0.95$ ranges between 1.2 and 4.5.

For the extremer quantile level $\tau = 0.995$ in Figure 10, we observe more discrepancies between the plots of the different methods. The GRF is unable to capture the shape of the extreme conditional VaR as it only shows horizontal lines instead of an evenly spread plot. Furthermore, the unconditional GPD plot becomes flatter as $\tau \to 1$ suggesting that the unconditional GPD converges to a value, and, thus, loses its flexibility. Here, both the ERF and GBEX still capture the upward trend of the extreme conditional VaR plotted against the VIX. Around a VIX of 25, the GBEX jumps in the estimated extreme conditional VaRs from approximately 5.0 to 7.5 suggesting that this is the tipping point of volatility for which more extreme losses will be incurred. Lastly, the extreme conditional VaRs estimated by the ERF range between 2.5 and 12.5, suggesting that the maximum potential loss is between 2.5% and 12.5% with a probability of 0.005. This potential loss increases when the investor expects a more volatile market in the upcoming 30 days.

Figure 10. Extreme Conditional Quantiles Predicted By ERF, GRF, GBEX and the Unconditional GPD At Quantile Levels $\tau = 0.9$, 0.95, 0.995 For Emerging Markets Stock Data.

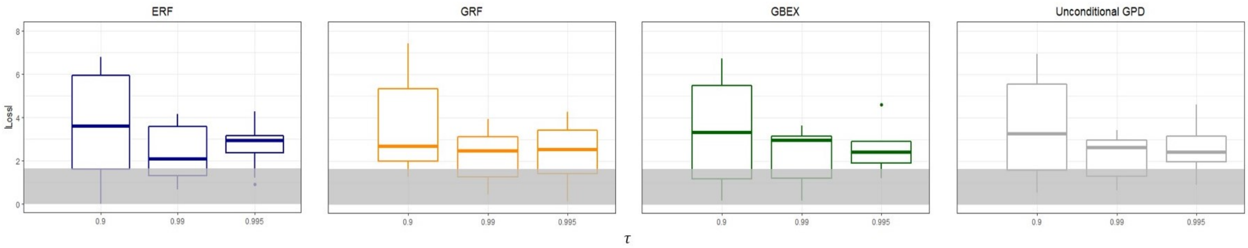


5.4.2 Performance For Out-Of-sample Predictions

In this section, we investigate the performance for out-of-sample predictions of the ERF, GRF, GBEX and the unconditional GPD on the second part of the data set. Here, we use the absolute value of the loss function in Equation (8). Figure 11 shows the absolute losses for the different methods, where the grey area is the 95% interval of the absolute value of a standard normal distribution. Based on the median absolute loss, shown as a fatter line, the ERF outperforms the other methods for $\tau = 0.99$ and the GBEX outperforms the other methods for $\tau = 0.995$. Furthermore, the different methods have a relatively similar spread in absolute loss, and the spread in the absolute loss decreases with the quantile level τ for all methods except the GRF. The GRF does not rely on the extrapolation from extreme value theory, and, thus, is less flexible for extremer quantile levels resulting in a larger spread of absolute losses.

However, from Figure 11, we are unable to draw a clear conclusion on the relative performance of the different methods. Therefore, we continue to investigate their performance while altering the circumstances. For example, the estimated shape parameter $\hat{\xi}(x)$ in Figure 9 does not show a clear pattern, especially when plotted against the EM USD Index. This scattered behavior of the shape parameter could lead to less accurate estimates resulting in more spread in the absolute losses of Figure 11. Thus, we set the shape parameter ξ to a fixed value. More specifically, we fix it to the mean value of the estimated shape parameters in Figure 9, resulting in $\hat{\xi}_{fixed} = 0.233$. By taking the shape parameters of the first part of the data set, we ensure that the methods do not have any initial information on the second part of the data set for the out-of-sample predictions.

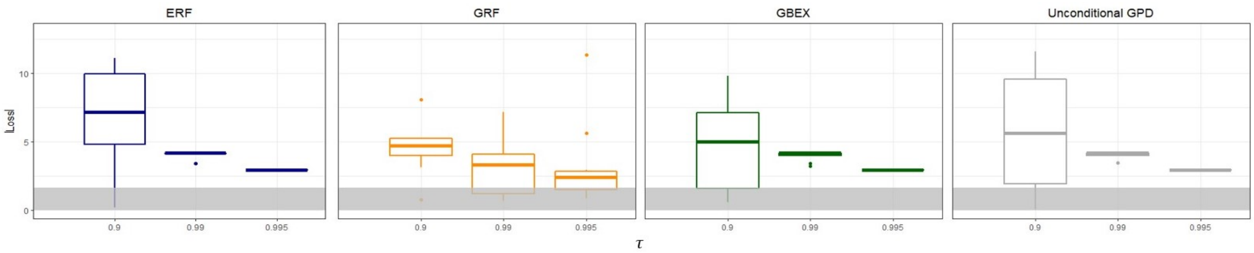
Figure 11. Absolute Value Of the Loss Function For ERF, GRF, GBEX and the Unconditional GPD On the Emerging Markets Stock Data For $\tau = 0.9, 0.99, 0.9995$.



Note. The grey area is the 95% interval of the absolute value of a standard normal distribution.

Figure 12 shows the absolute losses of the different methods with the fixed shape parameter. Here, the GRF outperforms the other methods as the GRF has the lowest median absolute loss and the absolute loss values fall partly in the 95% interval. The GRF does not rely on the extrapolation, thus, its performance is not affected by a change in the GPD parameters. In contrast, the performance of the ERF, GBEX and the unconditional GPD worsens with $\hat{\xi}_{fixed}$ as these methods rely on the GPD parameters for the extrapolation. For quantile levels $\tau = 0.99, 0.995$, the absolute losses of the ERF, GBEX and the unconditional GPD show few to no spread suggesting that the spread in the performance measure for these methods correlates with the movement of the shape parameter. This movement could have a stronger effect for more extreme quantile levels. Furthermore, the ERF, GBEX and the unconditional GPD with $\hat{\xi}_{fixed}$ perform worse compared to methods with a flexible shape parameter because the median absolute losses increase and no absolute loss falls in the 95% interval anymore. We conclude that the flexible shape parameter is necessary for the ERF, GBEX and unconditional GPD to perform well, and, thus, in the upcoming analyses, we use the flexible shape parameter as in Figure 11.

Figure 12. Absolute Loss Of the ERF, GRF, GBEX and the Unconditional GPD On the Emerging Markets Stock Data With Fixed Shape Parameter For $\tau = 0.9, 0.99, 0.9995$.



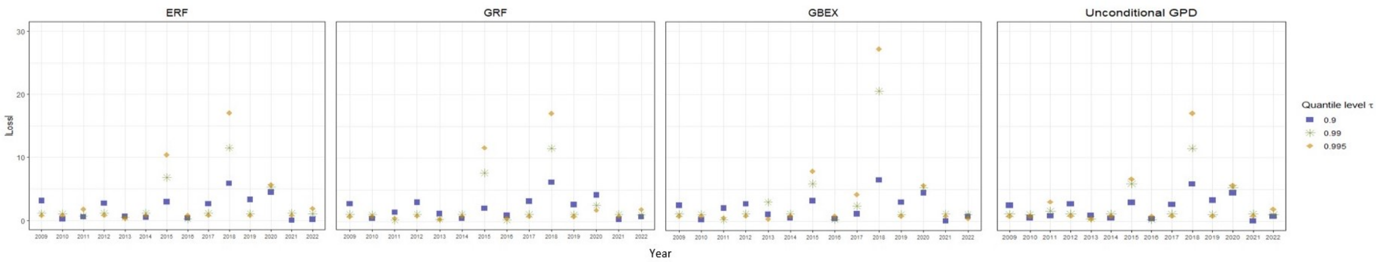
Note. The grey area is the 95% interval of the absolute value of a standard normal distribution.

5.4.3 Out-Of-Sample Predictions Over Time

To study the performance over time, we train the methods on year t and predict the extreme conditional Value-at-Risk (VaR) for year $t+1$. This results in training and testing subsets with between 250 and 253 observations. Furthermore, the year 2007 is excluded from the analysis as the data set only contains 11 observations in this year.

Figure 13 shows the absolute loss values plotted against the years, here, the years on the x-axis correspond with the year for which the prediction is made (year $t+1$). For all methods, we observe two to three spikes in the absolute losses where the spike in 2018 is larger than in 2015 and 2020. In Figure 1, the negative returns show more volatile behavior around the years 2015, 2018 and 2020 with standard deviations of 1.025, 1.003 and 1.544, respectively, while the negative returns show less volatile behavior around the years 2014, 2017 and 2019 with standard deviations of 0.712, 0.598 and 0.748, respectively. In Figure A15 in Appendix A, the VIX also increases in the predicting years compared to the training years suggesting an increase in overall market volatility. Thus, we conclude that the performance worsens when the models are trained on a less volatile year and the prediction of the extreme conditional VaR is done on a more volatile year. This decline of performance is more severe for extremer quantile levels τ, and is the most severe for the GBEX where the absolute loss in 2017 is 20.542 and 27.221 for $\tau = 0.99,\ 0.995$, respectively.

Figure 13. Absolute Value Of the Loss Function For ERF, GRF, GBEX and the Unconditional GPD On the Emerging Markets Stock Data For $\tau = 0.9,\ 0.99,\ 0.9995$ Plotted Against the Years.



Furthermore, in Figure 13, the GRF seems to perform better for extreme quantile levels than the ERF, GBEX and the unconditional GPD. Table 4 shows the mean absolute loss of each method, here, the GRF also has the lowest mean absolute loss for quantile levels $\tau = 0.99,\ 0.995$. Here, the same reasoning as for the fixed shape parameter could apply, in other words, the better performance of the GRF could be due to it not being affected by the movement of the GPD parameters. Moreover, Table 4 shows that the ERF performs best for the more moderate quantile level $\tau = 0.9$. Figure B16 in Appendix B shows the absolute losses per quantile level τ. In contrast with Figure 11, the majority of median absolute losses fall in the 95% interval and the spread in absolute losses is reduced for all methods in Figure B16 in Appendix B. This leads us to believe that the behavior of the negative emerging market stock returns over time has a significant effect on the performance of all methods.

Table 4. *Mean Of the Absolute Losses Trained and Tested Each Year.*

	Quantile level		
Method	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.995$
ERF	1.952	2.418	3.100
GRF	2.111	2.241	2.802
GBEX	2.020	3.233	3.695
Unconditional GPD	2.012	2.412	2.925

Note. The numbers are rounded to three decimal places.

5.4.4 Extreme Conditional Value-At-Risk Per Country

To further analyse the performance, we take the countries in which the MSCI Emerging Markets Index is invested, namely China, Taiwan, India, South Korea, Brazil, Saudi Arabia, South Africa, Mexico, Thailand, Indonesia, Malaysia, the United Arab Emirates and Qatar (BlackRock, 2022b). Some country-level data sets contain fewer observations than the MSCI Emerging Markets ETF data set, and, therefore, we partition the data sets with less than 1,500 observations into five random folds instead of the ten random folds, as done in Section 5.4.2.

First, we fit the different methods and estimate the GPD parameters. Table 5 shows the descriptive statistics of the different countries along with their ranking. The fatness of the tails is ranked in descending order, based on the shape parameter $\hat{\xi}(x)$. Malaysia has the fattest tails, followed by the United Arab Emirates and Saudi Arabia. Additionally, in Table A1 in Appendix A, Malaysia has the highest maximum negative return at 27.706%, and a relatively high positive skewness of 4.681 suggesting that extreme negative returns occur more frequently for Malaysia. A similar conclusion can be drawn for Saudi Arabia. In contrast, China and South Africa have a negative mean shape parameter. Furthermore, the negative returns of China are negatively skewed at -0.632, suggesting that positive returns occur more frequently.

Table 5. *Descriptive Statistics Of the Shape Parameter* $\hat{\xi}(x)$ *Per Country.*

	Descriptive Statistics				
Country	Mean	Minimum	Maximum	Standard Deviation	Rank
China	-0.094	-0.180	0.031	0.051	13
Taiwan	0.072	0.072	0.072	0	9
India	0.113	0.031	0.217	0.038	8
South Korea	0.144	0.138	0.150	0.003	5
Brazil	0.049	0.036	0.067	0.007	11
Saudi Arabia	0.212	0.163	0.269	0.021	3
South Africa	-0.044	-0.048	-0.036	0.002	12
Mexico	0.118	0.092	0.150	0.010	7
Thailand	0.134	0.118	0.150	0.007	6
Indonesia	0.175	0.145	0.218	0.012	4
Malaysia	0.292	0.284	0.305	0.003	1
United Arab Emirates	0.243	0.230	0.261	0.006	2
Qatar	0.049	0.028	0.064	0.008	10

Note. The numbers are rounded to three decimal places. Rank orders the fatness of the tails.

Next, we assess the importance of each variable for the fitted ERF to find the variables driving the extreme conditional VaR per country. Table 6 shows the combined variable importance score for the estimated GPD parameters. We observe that the world GDP growth and the GDP growth of each country have a low variable importance score. Moreover, the variable importance score of Taiwan is zero for each variable, and, in Table 5, the shape parameter of Taiwan has a standard deviation of 0. The Taiwan data set contains the fewest observations (212 observations) suggesting that the ERF does not perform well for samples of this size.

Furthermore, in contrast to Table 3, the Treasury-Bill (TB) rate and the Emerging Market US Dollar Index (EMUSDI) are relevant drivers of the GPD parameters. That way, the VIX, EMBI, TB rate and EMUSDI combined explain between 86.1% and 92.1% of the GPD parameters for all countries. Here, the Middle Eastern countries show two things: on one hand, that the VIX, EMBI and EMUSDI are the most important variables, and, on the other hand, that the TB rate has a slightly lower importance score. Therefore, we conclude that the Middle Eastern countries do not only depend on the expected market volatility and the sovereign credit risk, but also on the US 3-month TB rate and on the value of the emerging market currencies relative to the US Dollar.

Comparatively, in the other geographical areas, there is no clear pattern for the TB rate and EMUSDI. For example, South Korea, Malaysia and Mexico give a relatively low importance score to these variables. Here, the VIX alone explains 60.3% and 57.1% of the GPD parameters for Mexico and South Korea, respectively. Therefore, we conclude that the VIX and the EMBI remain the most important variables overall.

Table 6. *Variable Importance For the GPD Parameters Estimated By the ERF Per Country.*

	ERF Variable Importance Per Country												
Variable	China	Taiwan	India	South Korea	Brazil	Saudi Arabia	South Africa	Mexico	Thailand	Indonesia	Malaysia	United Arab Emirates	Qatar
VIX	0.250	0.000	0.227	0.571	0.439	0.201	0.346	0.603	0.269	0.338	0.397	0.211	0.204
EMBI	0.268	0.000	0.283	0.195	0.228	0.262	0.192	0.132	0.233	0.248	0.241	0.266	0.281
TB rate	0.155	0.000	0.150	0.051	0.114	0.179	0.168	0.056	0.214	0.109	0.075	0.200	0.162
EMUSI	0.214	0.000	0.261	0.081	0.144	0.220	0.210	0.125	0.145	0.187	0.161	0.217	0.216
World GDP	0.066	0.000	0.034	0.055	0.037	0.085	0.044	0.046	0.038	0.048	0.066	0.049	0.057
Country GDP	0.046	0.000	0.045	0.046	0.038	0.053	0.040	0.037	0.100	0.070	0.060	0.057	0.080

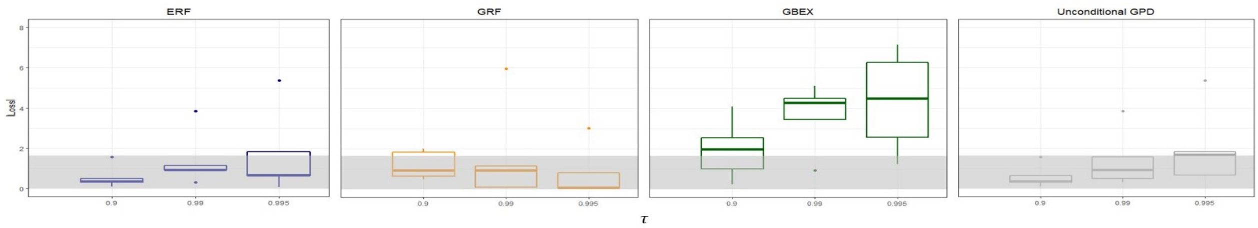
Note. The numbers are rounded to three decimal places. The values represent the variable importance for the scale and shape parameters combined. World GDP and Country GDP represent the corresponding GDP growth.

Lastly, we refit the different methods and use Equation (8) to assess the prediction performance. Here, we discuss the performance of the methods for the three lowest and the three highest ranked countries, see Figure 14. Figure B17 in Appendix B shows the absolute losses for all 13 countries. Figure 14a shows, for China, that the ERF outperforms the other methods for $\tau = 0.9$ and 0.99 based on the median and the spread of the absolute losses. Moreover, both the ERF and GRF perform well for $\tau = 0.995$. The South Africa plot in Figure 14d shows a slightly different pattern where the ERF and GBEX perform well for more extreme quantile levels $\tau = 0.99$ and 0.995. In contrast, for Brazil in Figure 14b, the GRF is the best performing method for all the different quantile levels. Furthermore, Table A1 in Appendix A shows that the negative returns of the MSCI Brazil Index have the highest standard deviation (2.364), suggesting that the GRF performs well for more volatile time series.

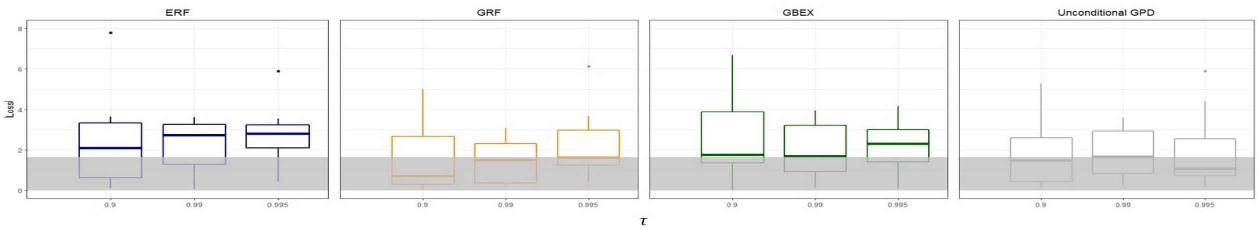
Malaysia, the United Arab Emirates and Saudi Arabia are the highest ranked countries based on their high estimated shape parameters $\hat{\xi}$, which are above 0.2. For both Malaysia and the United Arab Emirates in Figures 14e and 14f, the GRF outperforms the other methods for $\tau = 0.9$, and the ERF outperforms for the more extreme quantile levels $\tau = 0.99$ and 0.995. The Saudi Arabia plot in Figure 14c shows a slightly different pattern, where the GRF outperforms the other methods for $\tau = 0.9$ and 0.99 and the ERF and GBEX perform well for $\tau = 0.995$. Considering, on one hand, that these countries have the largest average shape parameter, and, on the other hand, that the ERF outperforms the other methods for the extreme quantile levels, we conclude that the ERF performs best for high average shape parameters. In other words, the ERF is the best method to predict the extreme conditional Value-at-Risks for emerging markets with fat-tailed negative returns.

Figure 14. Absolute Value Of the Loss Function For ERF, GRF, GBEX and the Unconditional GPD With $\tau = 0.9, 0.99, 0.9995$ For the Lowest and Highest Ranked Countries.

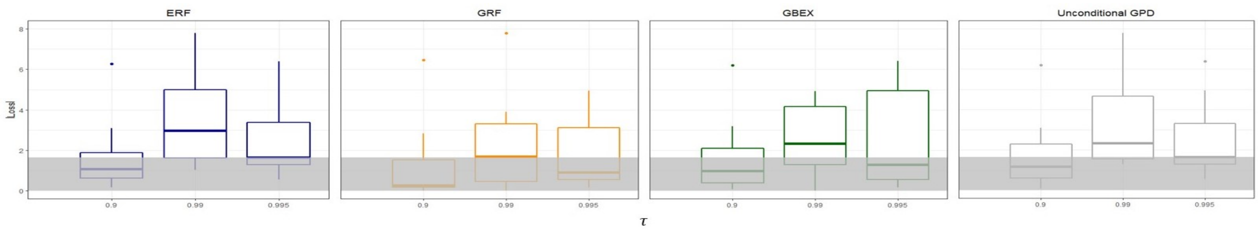
(a) China



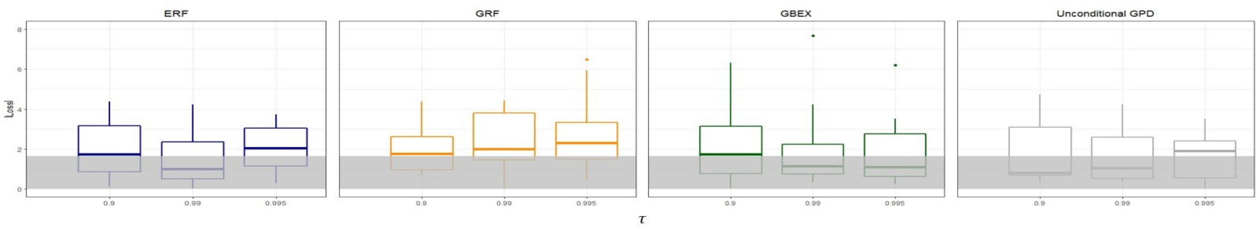
(b) Brazil



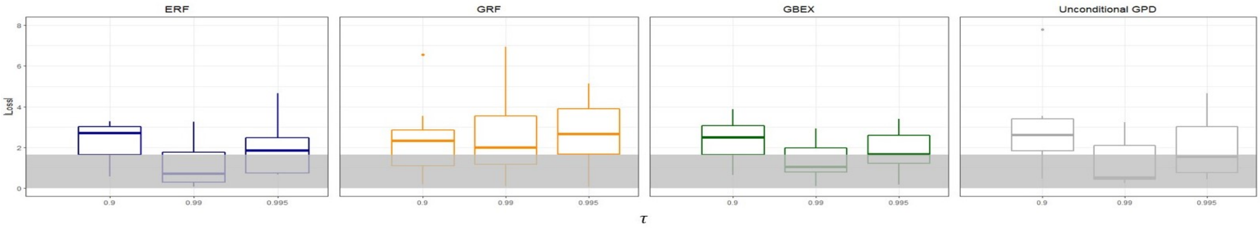
(c) Saudi Arabia



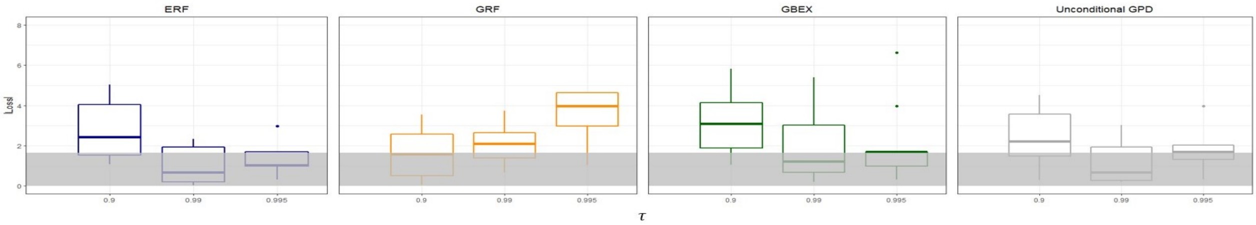
(d) South Africa



(e) Malaysia



(f) United Arab Emirates



Note. The grey area is the 95% interval of the absolute value of a standard normal distribution.

6 Conclusion and Economic Discussion

In this paper, we studied the extremal random forest (ERF) algorithm of Gnecco et al. (2022), and investigated whether the ERF improves the estimation of the extreme conditional Value-at-Risk (VaR) in emerging stock markets. To answer the research question, we assessed the performance of the ERF in comparison with (1) the generalized random forest (GRF) by Athey et al. (2019), (2) the gradient boosting for extremes algorithm (GBEX) by Velthoen et al. (2021), and (3) the unconditional generalized Pareto distribution (GPD). We trained and tested all methods on both the MSCI Emerging Market Index, and the countries the index is composed of. In the following paragraphs, we start by answering the research question, and follow with a discussion of the economic inferences that can be drawn from our results.

We assessed the performance of the methods in different settings and concluded that, overall, the ERF outperforms the other methods for stock data with a high average shape parameter. In other words, the ERF offers a better method to predict the extreme conditional VaR for emerging markets with fat-tailed negative returns. Furthermore, we found that the behavior of the negative stock returns over time has a significant effect on the ERF. Here, training and testing the ERF on consecutive years increased the performance significantly in terms of absolute loss, compared to taking random observations from the data set. However, the performance worsened when the ERF was trained on a comparatively less volatile year while the prediction of the extreme conditional VaR was done on a comparatively more volatile year. Additionally, for the ERF to perform well and keep its flexibility, each observation requires an individual value of the shape parameter. Lastly, the ERF did not perform as well for smaller sample sizes, as it disregarded all the given covariates and assigned them an importance score of zero.

Furthermore, our results offer several economic insights for both the emerging stock market as a whole and for the stock markets of the individual countries. First, we found that the expected worldwide market volatility and sovereign credit risk have the biggest impact on the extreme conditional VaR for the emerging equity market. Second, we found an upward pattern between the Volatility Index (VIX) and the conditional VaR, which deteriorates with increasing VIX-values. For the emerging market as a whole, we estimated a maximum potential daily loss of approximately 2.5% for low values of the VIX, and 12.5% for high values of the VIX, both with a probability of 0.005. Hence, extreme losses seem more likely to occur when investor sentiment worsens, possibly resulting in a bear market. Another factor for such potential losses can be found in unstable financial institutions and/or governments, ultimately affecting the credit risk of their countries.

Based on these findings, it seems particularly relevant for financial institutions to look at the worldwide market volatility, measured with the Volatility Index (VIX), along with the attractiveness of government and corporate bonds, measured through the Emerging Market Bond Index (EMBI). For instance, if the VIX increases and the EMBI decreases, it seems reasonable to assume that the likelihood of an extreme price movement increases. Under these conditions, the potential magnitude of an extreme conditional VaR increases. This would in turn decrease the attractiveness for market players to invest in emerging markets, as the majority of them can be expected to be rather risk-averse. Following this logic, the emerging countries would receive less capital inflow and their equity market would presumably become less liquid, further reducing their attractiveness from an investor perspective. Ultimately, the scenario described above would create a vicious circle. A lower market liquidity also could further increase the likelihood of an extreme price movement. An example of such price behavior in recent history can be found in the oil market, where a one-sided market led to negative prices, and the supply drastically exceeded the demand (Brower et al., 2020). Therefore, the observation of this type of pattern in the VIX and EMBI could be used as a warning signal for financial institutions that are invested in the emerging market, but also for the governments of these very markets.

Finally, we looked at the emerging countries on an individual basis, and ranked them based on the fatness of the tails of their negative returns. We concluded that Saudi Arabia, the United Arab Emirates and Qatar (i.e., the Middle Eastern countries) did not only depend on the expected market volatility and the sovereign credit risk, but were also significantly affected by both the US 3-month Treasury-Bill rate, and the value of the emerging market currencies relative to the US Dollar. To measure the latter variable, we used the Emerging Market US Dollar Index (EMUSDI) as a proxy. However, it should be noted that the currencies of all these Middle Eastern countries being pegged to the US Dollar (Zucchi et al., 2021), the EMUSDI could also indirectly indicate the tendency of the US to trade with these countries. This in turn affects the negative returns of all Middle Eastern countries in our sample. If the US trades less with these emerging countries, then their capital inflow would likely decrease, resulting in larger credit constraints. Consequently, these countries would have less capital available to invest in, for example, infrastructure, education and healthcare. Accordingly, the equity in these emerging markets would become less attractive, as companies that are active there would have to grow and perform in a less favorable environment. As a conclusion, the EMUSDI seems to be an important indicator for financial institutions invested in the Middle Eastern countries, as it can be used as a signal or proxy for the future economic growth in these markets.

References

Ahmed, S., & Zlate, A. (2014). Capital flows to emerging market economies: A brave new world? *Journal of International Money and Finance*, *48*, 221–248.

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *Annals of Statistics*, *47*(2), 1148–1178.

Balkema, A. A., & De Haan, L. (1974). Residual life time at great age. *Annals of Probability*, *2*(5), 792–804.

BlackRock. (2022a). iShares [Database accessed 6 May 2022]. https://www.ishares.com/nl

BlackRock. (2022b). iShares MSCI Emerging Markets ETF [Database accessed 6 May 2022]. https://www.ishares.com/us/products/239637/

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32.

Brower, D., Sheppard, D., Raval, A., & Meyer, G. (2020). *What negative US oil prices mean for the industry*. https://www.ft.com/content/88997d67-bf69-409e-8155-911fc1f2fd6f

Chicago Board Options Exchange. (2022). *VIX Volatility Suite*. https://www.cboe.com/tradable_products/vix/

Coles, S. G., & Dixon, M. J. (1999). Likelihood-based inference for extreme value models. *Extremes*, *2*(1), 5–23.

De Haan, L., & Ferreira, A. (2006). *Extreme value theory: an introduction* (Vol. 21). Springer.

Demos, T., & Mackenzie, M. (2012). *Cost of 'tail risk' protection to soar*. https://www.ft.com/content/79459ea0-4cfa-11e1-bdd1-00144feabdc0

Diagne, M. (2002). *Financial Risk Management and Portfolio Optimization Using Artificial Neural Networks and Extreme Value Theory* (Doctoral dissertation). Technische Universität Kaiserslautern.

Dimitrakopoulos, D. N., Kavussanos, M. G., & Spyrou, S. I. (2010). Value at risk models for volatile emerging markets equity portfolios. *Quarterly Review of Economics and Finance*, *50*(4), 515–526.

Embrechts, P., Resnick, S., & Samorodnitsky, G. (1998). Living on the edge. *Risk*, *11*(1), 96–100.

Federal Reserve Bank of St. Louis. (2022). Nominal Emerging Market Economies U.S. Dollar Index [Database accessed 6 May 2022]. https://fred.stlouisfed.org/tags/series

Frazier, L. (2021). *The Coronavirus Crash of 2020, and the Investing Lesson It Taught Us*. https://www.forbes.com/sites/lizfrazierpeck/2021/02/11/the-coronavirus-crash-of-2020-and-the-investing-lesson-it-taught-us/?sh=419bb60346cf

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378.

Gencay, R., & Selçuk, F. (2004). Extreme value theory and Value-at-Risk: Relative performance in emerging markets. *International Journal of Forecasting*, *20*(2), 287–303.

Gilleland, E., & Katz, R. W. (2006). Analyzing seasonal to interannual extreme weather and climate variability with the extremes toolkit. *18th Conference on Climate Variability and Change, 86th American Meteorological Society (AMS) Annual Meeting*, *29*.

Gnecco, N., Terefe, E. M., & Engelke, S. (2022). Extremal Random Forests. *arXiv preprint arXiv:2201.12865*.

Gnecco, N. (2022). *erf: Extremal Random Forest* [R package]. https://github.com/nicolagnecco/erf

Gumbel, E. J. (1958). *Statistics of Extremes*. Columbia University Press. https://doi.org/doi:10.7312/gumb92958

Hayes, A., & Scott, G. (2020). *Emerging Markets Bond Index (EMBI)*. https://www.investopedia.com/terms/e/emerging-markets-bond-index.asp#:~:text=The%5C%20emerging%5C%20markets%5C%20bond%5C%20index%5C%20(EMBI)%5C%20is%5C%20a%5C%20benchmark%5C%20index,specific%5C%20liquidity%5C%20and%5C%20structural%5C%20requirements.

Hill, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *Annals of Statistics*, 1163–1174.

Hooker, M. A. (2004). Macroeconomic factors and emerging market equity returns: a Bayesian model selection approach. *Emerging Markets Review*, *5*(4), 379–387.

International Monetary Fund. (2022). IMF Country Focus [Database accessed 6 May 2022]. https://www.imf.org/en/Countries/

Koepke, R. (2019). What drives capital flows to emerging markets? A survey of the empirical literature. *Journal of Economic Surveys*, *33*(2), 516–540.

Logue, A., & Rasure, E. (2022). *What Is the Trade-Weighted US Dollar Index?* https://www.thebalance.com/what-is-the-trade-weighted-us-dollar-index-5216766

McNeil, A. J. (1998). *Calculating quantile risk measures for financial return series using extreme value theory* (tech. rep.). ETH Zurich.

McNeil, A. J. (1999). Extreme value theory for risk managers. *Departement Mathematik ETH Zentrum*, *12*(5), 217–37.

Meinshausen, N., & Ridgeway, G. (2006). Quantile regression forests. *Journal of Machine Learning Research*, *7*(6).

NYSE. (2022). *Holidays Trading Hours*. https://www.nyse.com/markets/hours-calendars

Organization for Economic Cooperation and Development. (2022). Nominal GDP forecast [Database accessed 6 May 2022]. https://data.oecd.org/gdp/nominal-gdp-forecast.htm#indicator-chart

Pickands, J. I. (1975). Statistical Inference Using Extreme Order Statistics. *Annals of Statistics*, 119–131.

Shrivastava, U., Prakash, G., & Dhar, J. (2014). Extreme value theory and neural network for catastrophic fall prediction: a study of year 2008–2009. *International Journal of Information and Decision Sciences*, *6*(2), 193–210.

Statista. (2022). GDP growth rate in Taiwan 2000-2022 [Database accessed 6 May 2022]. https://www.statista.com/statistics/328535/gross-domestic-product-gdp-annual-growth-rate-in-taiwan/

Taylor, J. W. (2000). A quantile regression neural network approach to estimating the conditional density of multiperiod returns. *Journal of Forecasting*, *19*(4), 299–311.

Tibshirani, J., Athey, S., Sverdrup, E., & Wager, S. (2022). *grf: Generalized Random Forests* [R package version 2.1.0]. https://CRAN.R-project.org/package=grf

Velthoen, J. (2022). *gbex: Estimate a non-stationary GPD using gradient boosting* [R package version 0.0.1.000]. https://github.com/JVelthoen/gbex/

Velthoen, J., Dombry, C., Cai, J.-J., & Engelke, S. (2021). Gradient boosting for extreme quantile regression. *arXiv preprint arXiv:2103.00808*.

Wang, H. J., & Li, D. (2013). Estimation of extreme conditional quantiles through power transformation. *Journal of the American Statistical Association*, *108*(503), 1062–1074.

Zucchi, K., Scott, G., & Kvilhaug, S. (2021). *Top Exchange Rates Pegged to the U.S. Dollar*. https://www.investopedia.com/articles/forex/061015/top-exchange-rates-pegged-us-dollar.asp

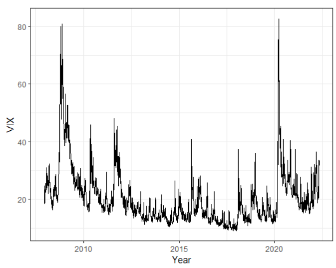
Appendix

A Additional Information on the Data Sets

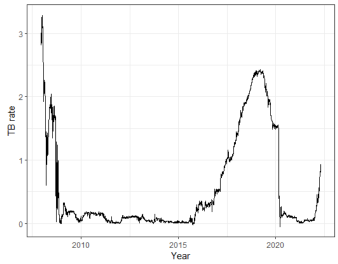
This section provides additional information on the covariates and the response variables per country. Figure A15 plots the Volatility Index (VIX), 3-month US Treasury-Bill (TB) rate, the Emerging Market Bond Index (EMBI) and the Emerging Market US Dollar Index (EMUSI) from December, 2007 until April, 2022. The VIX shows the biggest spikes in volatility around 2008 and 2020, corresponding to the Great Financial Crisis and the COVID19 crisis. We observe a similar pattern in the TB rate and the EMBI, where TB rate drops to zero and the EMBI ETF drops relatively significant in price, in 2008 and 2020. In contrast, the EMUSI ETF experiences a relatively significant price increase in those years.

Figure A15. The Volatility Index (VIX), Treasury-Bill (TB) Rate, Emerging Market Bond Index (EMBI) ETF and Emerging Market US Dollar (EM USD) Index From December, 2007 Until April, 2022.

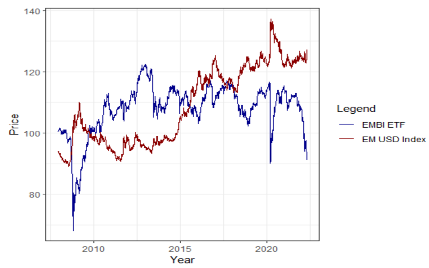
(a) VIX



(b) TB rate



(c) EMBI and EM USD Index



Furthermore, Table A1 shows the MSCI Index and the corresponding GPD growth for the 13 countries in which the MSCI Emerging Market Index is invested (BlackRock, 2022b). The negative returns of Malaysia have the highest kurtosis and skewness at 117.074 and 4.681, respectively. The negative returns of Taiwan have the lowest kurtosis at 4.232, and, additionally, the data set for Taiwan is the smallest data set with 212 observations. Lastly, the negative returns of Brazil are the most volatile with at standard deviation of 2.364.

Table A1. *Descriptive Statistics Of the MSCI ETF Of Each the Country and Their Corresponding Gross Domestic Product (GDP) Growth.*

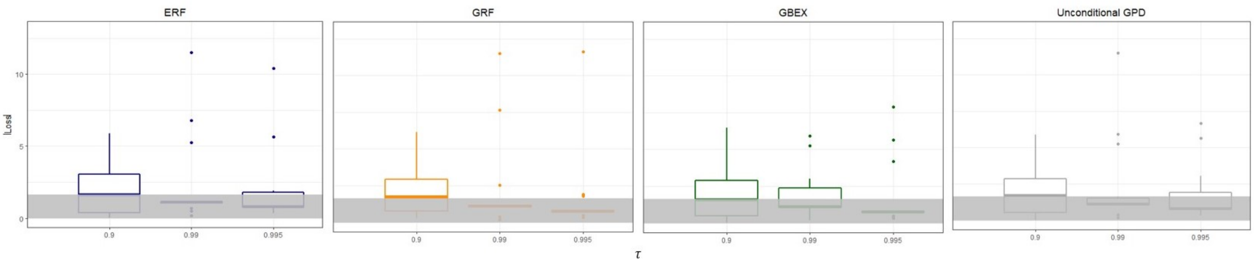
	Descriptive Statistics							
Variable	Mean	Median	Min.	Max.	St.Dev.	Skewness	Kurtosis	Numb. Obs.
China								
Negative returns	0.000	-0.037	-14.536	7.660	1.673	-0.632	12.150	721
GDP growth	98.919	98.970	98.622	99.175	0.234	-0.254	1.421	721
Taiwan								
Negative returns	0.040	0.0156	-3.958	3.783	1.108	0.049	4.232	212
GDP growth	98.369	98.922	98.354	98.379	0.012	-0.465	1.216	212
India								
Negative returns	-0.049	-0.091	-9.587	14.350	1.431	1.323	19.130	988
GDP growth	48.33	98.994	-98.565	99.069	86.228	-1.118	2.249	988
South Korea								
Negative returns	-0.016	-0.005	-28.237	18.519	1.806	-0.441	27.505	3619
GDP growth	98.73	98.822	97.887	99.083	0.283	-1.867	6.066	3619
Brazil								
Negative returns	-0.003	-0.033	-23.370	17.690	2.364	0.029	13.205	3619
GDP growth	98.967	98.999	98.220	99.292	0.272	-1.187	4.424	3619
Saudi Arabia								
Negative returns	-0.048	0.0001	-6.818	15.632	1.105	2.067	32.260	1667
GDP growth	38.902	98.396	-98.746	98.987	90.424	-0.861	1.741	1667
South Africa								
Negative returns	-0.027	-0.089	-8.706	11.565	1.788	0.312	6.308	3089
GDP growth	82.815	98.951	-98.353	99.179	54.120	-3.049	10.299	3089
Mexico								
Negative returns	-0.010	-0.037	-16.980	17.995	1.690	0.339	14.735	3619
GDP growth	71.403	98.977	-98.780	99.173	68.460	-2.080	5.329	3619
Thailand								
Negative returns	-0.022	-0.005	-9.538	13.037	1.429	0.637	12.534	2016
GDP growth	70.553	98.581	-98.901	98.982	68.847	-2.048	5.196	2016
Indonesia								
Negative returns	-0.014	0.0001	-15.755	10.466	1.520	0.025	12.104	3018
GDP growth	82.616	99.133	-98.317	99.364	54.740	-3.003	10.020	3018
Malaysia								
Negative returns	0.014	0.001	-7.390	27.706	1.104	4.681	117.074	3619
GDP growth	71.275	98.810	-98.870	98.977	68.425	-2.081	5.329	3619
United Arab Emirates								
Negative returns	0.006	-0.009	-10.259	13.768	1.261	0.832	21.095	2016
GDP growth	73.815	98.622	-98.903	98.828	65.446	-2.261	6.112	2016
Qatar								
Negative returns	-0.005	0.001	-10.239	12.920	1.076	0.805	22.016	2016
GDP growth	49.233	98.354	-98.690	98.848	85.328	-1.155	2.333	2016

Note. The numbers are rounded to three decimal places. Min., Max., Std. Dev. and Numb. Obs. stand for minimum, maximum, standard deviation and number of observations, respectively. Negative returns and GDP growth are in percentages. The data is from BlackRock (2022a), International Monetary Fund (2022), Organization for Economic Cooperation and Development (2022), and Statista (2022).

B Additional Absolute Loss Figures In Different Settings

In this section, we provide additional figures regarding the performance of the ERF relative to the GRF, GBEX and unconditional GPD. Figure B16 shows the absolute loss of the different methods when trained and tested on consecutive years, more specifically, trained on year t and estimated on year $t+1$. For $\tau = 0.9$, the performance of the different methods is relatively similar. When the quantile levels become more extreme, we observe that the GRF performs best in terms of median absolute loss and spread of the absolute losses.

Figure B16. Absolute Value Of the Loss Function For ERF, GRF, GBEX and the Unconditional GPD On the Emerging Markets Stock Data For $\tau = 0.9, 0.99, 0.9995$ Over the Years.

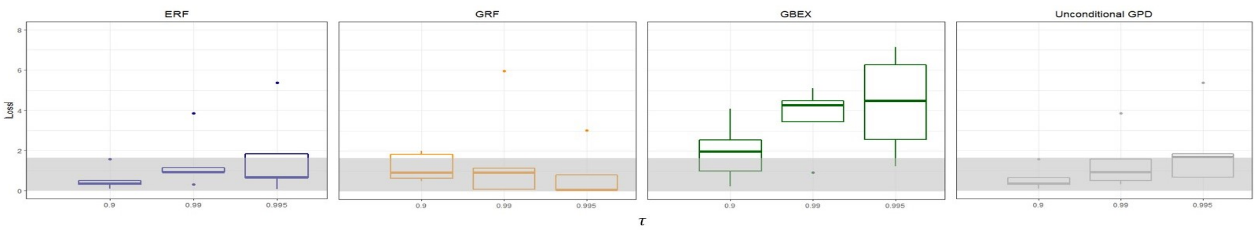


Note. The extreme absolute loss values of the GBEX are left out for visualization purposes.

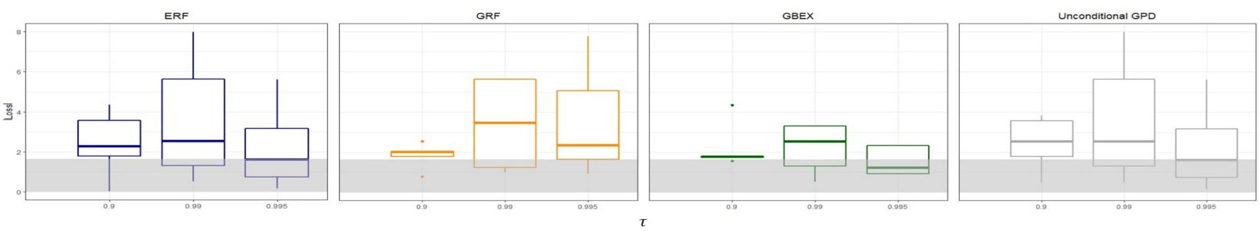
Figure B17 shows the absolute loss for the 13 countries in which the MSCI Emerging Market Index is invested (BlackRock, 2022b). The relative performance of the methods differs per country. For example, the GBEX performs bad for the China data set, while it outperforms the other methods for the India data set. Lastly, we observe that the ERF performs best for the Malaysia and United Arab Emirates data sets, as discussed in Section 5.

Figure B17. Absolute Value Of the Loss Function For ERF, GRF, GBEX and the Unconditional GPD With $\tau = 0.9, 0.99, 0.9995$ For the Countries Of the MSCI Emerging Market Index.

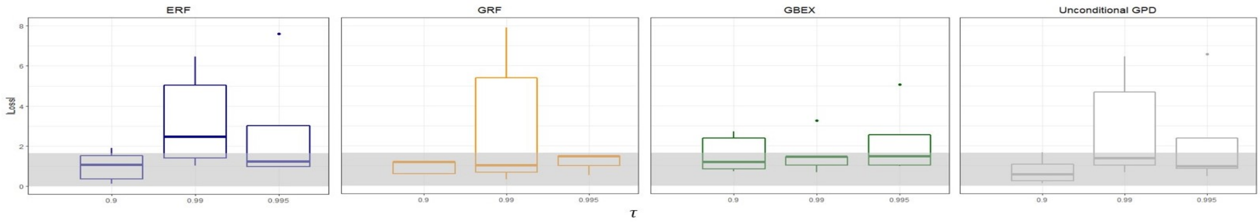
(a) China



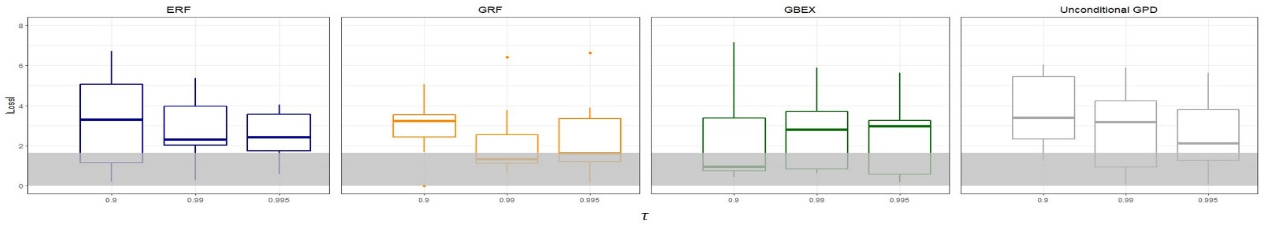
(b) Taiwan



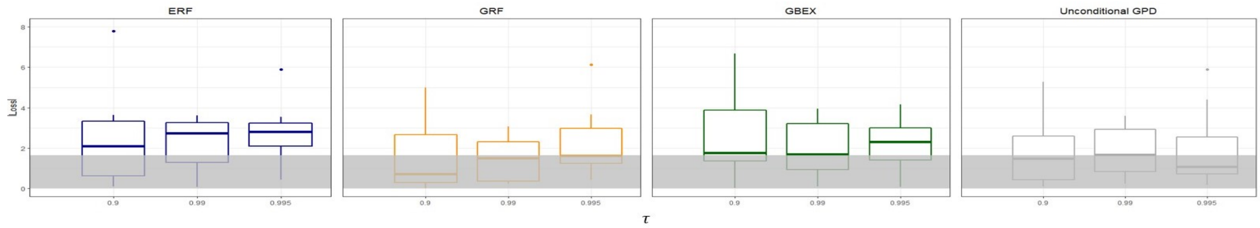
(c) India



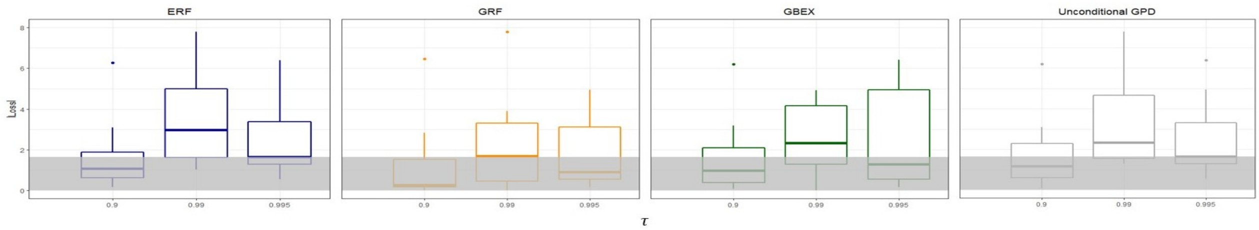
(d) South Korea



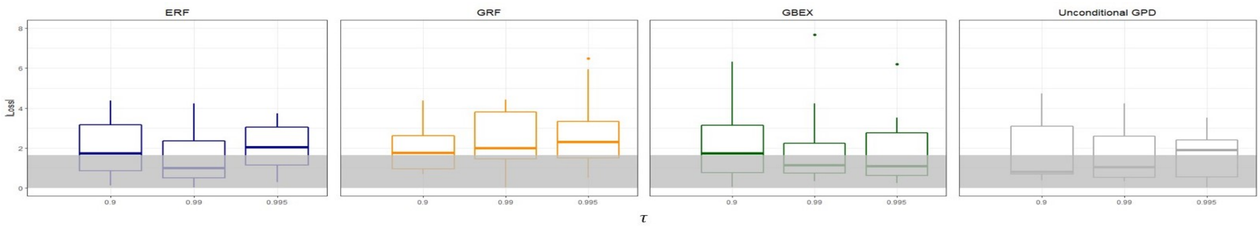
(e) Brazil



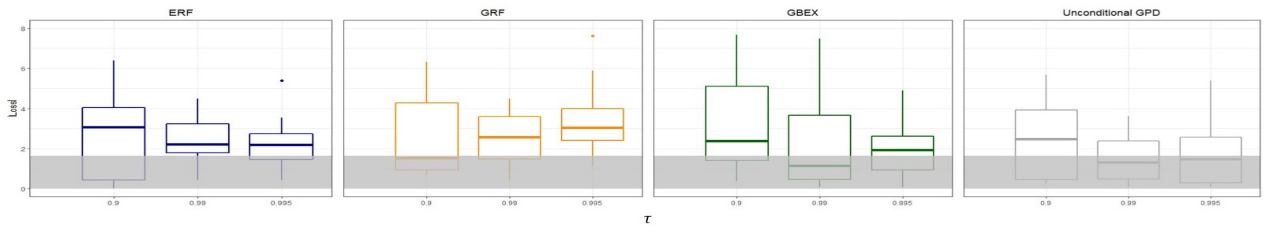
(f) Saudi Arabia



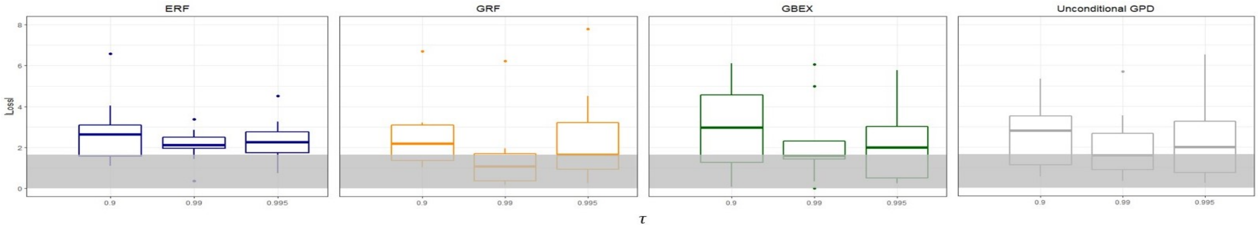
(g) South Africa



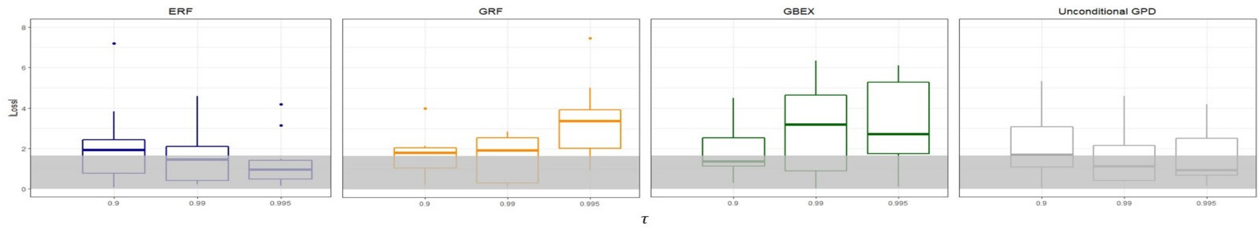
(h) Mexico



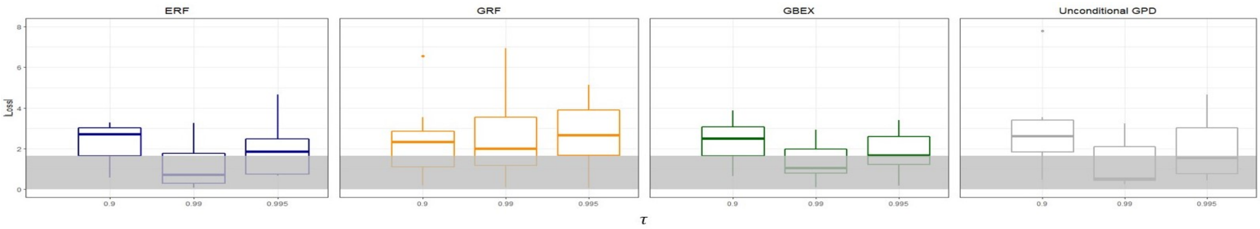
(i) Thailand



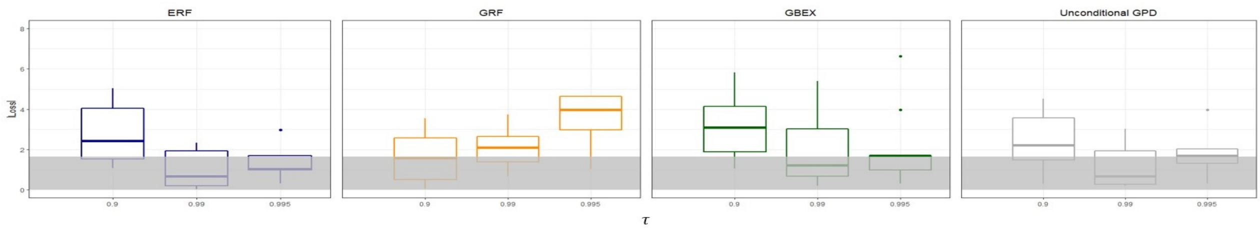
(j) Indonesia



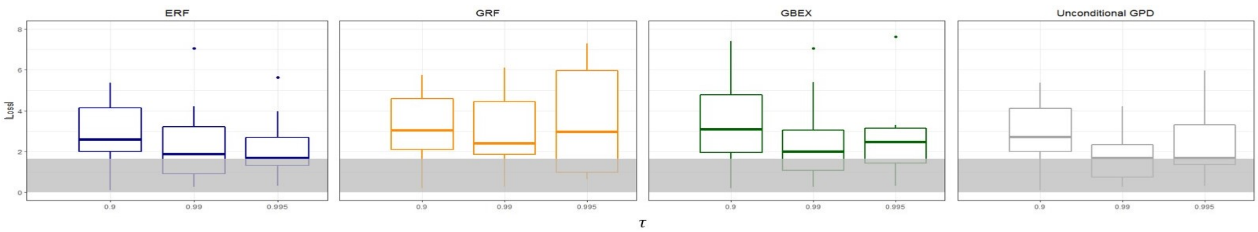
(k) Malaysia



(l) United Arab Emirates



(m) Qatar



Note. The grey area is the 95% interval of the absolute value of a standard normal distribution.