

**A COMPARISON OF THE EXTREME VALUE THEORY AND GARCH
MODELS IN TERMS OF RISK MEASURES
UNA COMPARACIÓN DE LA TEORIA DE VALORES
EXTREMOS Y DE LOS MODELOS GARCH EN TÉRMINOS DE
MEDIDAS DE RIESGO**

Ezgi Nevruz^{1*} y Şule Şahin^{1,2}

Abstract

In this paper, we apply extreme value theory (EVT) and time series models to eight developed and emerging stock markets published in the Morgan Stanley Capital International (MSCI) Index. Based on the Human Development Index (HDI) rankings, which are consistent with the MSCI index, we analyse Singapore, Spain, UK and US for developed stock markets and Chile, Russia, Malaysia and Turkey for emerging stock markets. We use the daily prices (in USD) of eight countries for the period from January 2014 to December 2017 and examine the performances of the models based on in-sample testing. Calculating the value-at-risk (VaR) as a risk measure for both right and left tails of the log-returns of the selected models, we compare these countries in terms of their financial risks. The obtained risk measures enable us to discuss the grouping and the ranking of the stock markets and their relative positions.

Keywords: Extreme value theory, GARCH models, Human development index, Risk measures, Value-at-risk.

Resumen

En este trabajo aplicamos Teoría de Valores Extremos (EVT) y modelos de series temporales a ocho mercados bursátiles desarrollados y emergentes, publicados en el índice Morgan Stanley Capital International (MSCI). Basándonos en los rankings del Human Development Index (HDI), que son consistentes con el índice MSCI, analizamos Singapur, España, Reino Unido y Estados Unidos para los mercados bursátiles desarrollados, y Chile, Rusia, Malasia y Turquía para los mercados bursátiles emergentes. Utilizamos los precios diarios (en dólares USD) de estos ocho países para el período desde enero de 2014 a diciembre de 2017, y examinamos el funcionamiento de los modelos basados en el testeo en muestras. Calculando el valor en riesgo (VaR) como medida de riesgo para las colas derecha e izquierda del retorno en logaritmos de los modelos seleccionados, comparamos estos países en términos de sus riesgos financieros. Las medidas de riesgo obtenidas nos permiten debatir la agrupación y el ranking de los mercados bursátiles y sus posiciones relativas.

Palabras clave: Teoría de valores extremos, Modelos GARCH, Índice de desarrollo humano (HDI), Medidas de riesgo, Valor en riesgo

¹*Department of Actuarial Sciences, Hacettepe University, Ankara, Turkey*

²*Institute for Financial and Actuarial Mathematics, University of Liverpool, Liverpool, UK*

*Corresponding author: ezginevruz@hacettepe.edu.tr

1. Introduction

In this paper, we compare financial risks of different countries according to a commonly used risk measure Value-at-Risk (VaR). Market risk management enables us to evaluate the potential risks and their possible effects in the future. Risk measures have been developed and have taken an increasingly important role in financial risk management and regulatory purposes after the financial disasters in 1990s. In this context, forecasting the risk of the stock markets requires a comprehensive information and the models which reflect the characteristics of the portfolios best should be determined. Time series models and extreme value theory (EVT) are two popular methods used for financial modelling.

The aim of examining financial time series is to investigate the behaviour of the prices and to manage the future risks according to the price movements. Since the future price is uncertain, it must be considered as a random variable with a probability distribution. Therefore, we use models describing how consecutive prices are determined statistically in order to investigate the prices. By analyzing stock returns, time series models are used for forecasting, pricing and financial risk management (Aas and Dimakos, 2004). Since generalized autoregressive conditional heteroskedasticity (GARCH) models capture price volatilities effectively, we prefer using GARCH models to represent stock returns.

On the other hand, the EVT becomes one of the most popular methods to forecast extreme financial risks since it is a well-developed approach under the probability theory. In this approach, the asymptotical distribution of extreme events, which are rare but severe, are taken into consideration. Statistical methods obtained from the EVT are very useful when they are applied to finance, particularly within the frame of risk measurement (Rocco, 2014).

Gençay *et al.* (2003) divide risk models into two groups according to the volatility of quantile forecasts. The first group consists of GARCH(1,1) and GARCH(1,1)-t whereas the second group includes historical simulation, variance-covariance approach, adaptive generalized Pareto distribution (GPD) and non-adaptive GPD models. When models are compared in terms of the VaR, the GPD model appears as a robust quantile forecasting tool for Turkish Stock Exchange index. Ekşi *et al.* (2006) compare the EVT, which is used to generate VaR estimates and provide the tail forecasts of daily returns, with other risk computation methods for Turkish Stock Market. The relative performance of the expected shortfall is measured with respect to other risk measures in their study. Gençay and Selçuk (2004) investigate the relative performance of the VaR models with the daily stock market returns of nine different emerging markets including Turkey.

Lawler (2003) provides a time series analysis of the Shanghai and New York Stock Exchange composite price indices to compare their weekly rates of return and volatility. Co-movement of these two markets in 1992-2002 are also analysed in this study.

Gilli and K ellezi (2000) present a practical application where the observations of 31 years of daily returns on an index representing the Swiss market are analysed. Point and interval estimates of the tail risk measures are computed by modelling the loss tail.

We analyse the stock indices of different developed (Singapore, Spain, UK and US) and emerging (Chile, Russia, Malaysia and Turkey) markets using both the EVT and GARCH models. These models are compared by in-sample testing and the best model is chosen for each country. Finally, we compare these countries in terms of their financial risks by means of VaR and discuss the groupings and the relative positions of the countries based on the Human Development Index (HDI).

The paper is organised as follows. Section 2 introduces the data. Section 3 presents a brief background of the models fitted to the financial data. We discuss the results of the application in Section 4. The risk measures are obtained and analysed by comparing the countries in Section 5. Finally, Section 6 concludes.

2. Data

Emerging and developed financial markets have different dynamics. Emerging markets have experienced more severe financial disasters than developed economies had. The structure of prices must hence be investigated carefully especially when emerging and developed countries are compared. A relative comparison is also beneficial for the investors in their decisions (Gen ay and Sel uk, 2004).

We use the daily prices published in Morgan Stanley Capital International (MSCI) index (MSCI, 2018). According to this index, countries are classified into two main groups as “developed” and “emerging”. Within each group, countries are also categorized according to their regional characteristics. We choose developed economies and emerging markets which can be comparable with each other according to the HDI rankings. United Nations Develop-

ment Programme (UNDP) publishes Human Development Reports annually (UNDP, 2014). Since the starting date of the country indices is 2014, we consider 2014 HDI rankings in order to be consistent with the estimation process. The HDI reflects not only financial but also socio-economic situation of a country. As a composite indicator of human development, the HDI consists of index of life expectancy at birth, education index and gross national income per capita. The chosen countries with respect to HDI rankings are Singapore, Spain, UK and US for developed stock markets and Chile, Russia, Malaysia and Turkey for emerging stock markets. The daily prices for the chosen countries are obtained from January 2014 to December 2017 in terms of USD.

According to the common tendency in both the time series and the EVT literature, the rate of return is measured by the change in the natural logarithm of the price index in a given period (Lawler, 2003). The log-return is obtained as

$$r_t = \log \left(\frac{P_t}{P_{t-1}} \right) = p_t - p_{t-1}$$

where P_t is the price level and $p_t = \log(P_t)$ is the natural logarithm of the price level (Camilleri, 2006). Using the logarithmic transformations of prices has important advantages: (i) a non-linear relationship can be transformed into a linear one, (ii) the use of linear regressions with logarithmic series provides an immediate interpretation of the estimated coefficients such as elasticities, and (iii) the series are usually compressed and a constant variance is obtained for the transformed series.

We apply the Augmented Dickey-Fuller (ADF) unit root test to check the stationarity of the series. The results in Table 1 and Table 2 show that prices should be log-differenced to obtain the stationary data for all countries by testing the null hypothesis “ H_0 : The data needs to be differenced to make it stationary”.

Table 1. The ADF test results for the prices

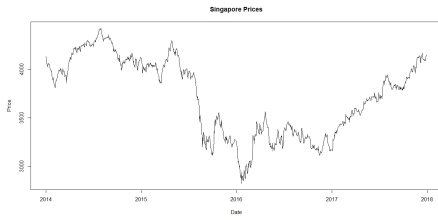
Developed Countries	Singapore	Spain	UK	US
ADF statistic	-1.1361	-1.2374	-1.2475	-1.4752
p-value (ADF)	0.9168	0.9004	0.8968	0.8005
Emerging Countries	Chile	Russia	Malaysia	Turkey
ADF statistic	-1.0706	-2.1625	-0.9231	-2.2046
p-value (ADF)	0.9273	0.5095	0.9506	0.4917

Table 2. The ADF test results for the log-returns

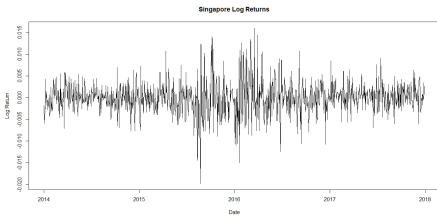
Developed Countries	Singapore	Spain	UK	US
ADF statistic	-8.3864	-10.5130	-10.9550	-10.6780
p-value (ADF)	0.01	0.01	0.01	0.01
Emerging Countries	Chile	Russia	Malaysia	Turkey
ADF statistic	-9.6925	-9.7817	-9.7318	-9.9976
p-value (ADF)	0.01	0.01	0.01	0.01

In addition to the statistical tests, graphical representations show that the log-returns are more appropriate for the time series analysis due to the stationarity property. EVT is also applicable for log-returns as the volatility can be seen in the graphs. The graphs also indicate that emerging countries experience more jumps in the returns comparing with the developed countries which might cause higher risk measures. Figure 1 and 2 represent the daily prices and log-returns of the selected countries. Since the daily prices are not stationary, we will use the log-returns for future analysis.

Figure 1. The prices and log-returns of the developed countries



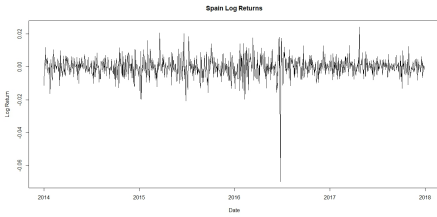
(a) Prices-Singapore



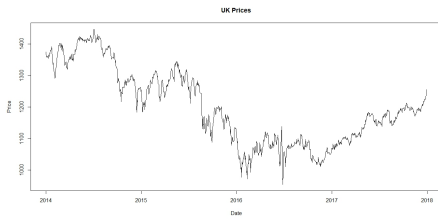
(b) Log-returns-Singapore



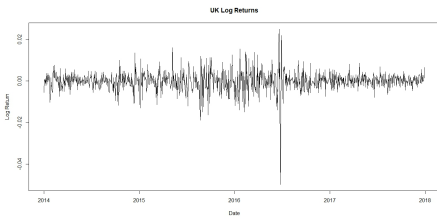
(c) Prices-Spain



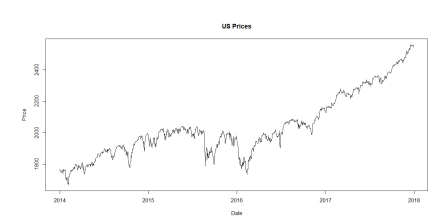
(d) Log-returns-Spain



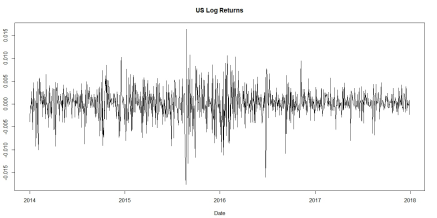
(e) Prices-UK



(f) Log-returns-UK

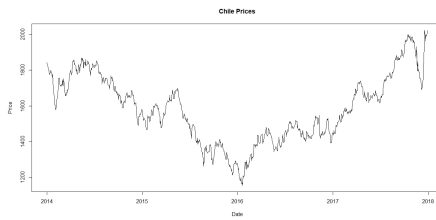


(g) Prices-US

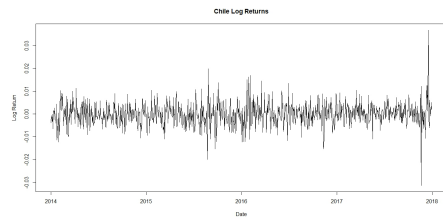


(h) Log-returns-US

Figure 2. The prices and log-returns of the emerging countries



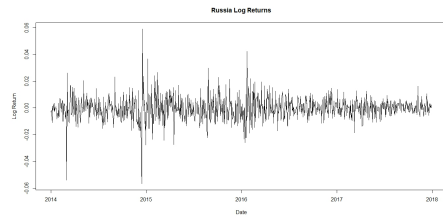
(a) Prices-Chile



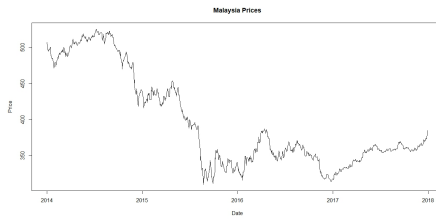
(b) Log-returns-Chile



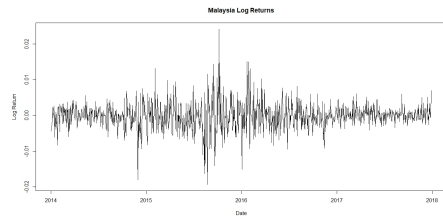
(c) Prices-Russia



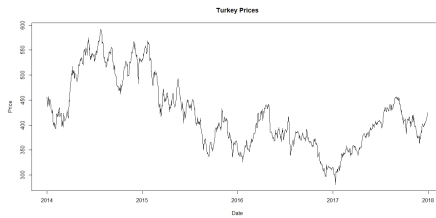
(d) Log-returns-Russia



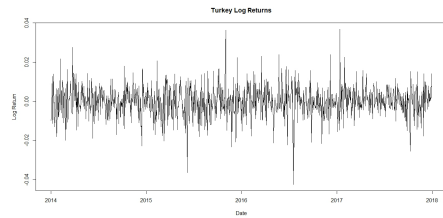
(e) Prices-Malaysia



(f) Log-returns-Malaysia



(g) Prices-Turkey



(h) Log-returns-Turkey

3. Models for Index Returns

The risk measurement in this paper is based on the tail estimation and the variability of the index returns. We thus choose two modelling approaches, EVT and GARCH models, which are commonly used in financial studies. EVT is used to model high severity and low frequency events whereas GARCH considers the heteroskedasticity which appears when both high and low volatility occur in different periods.

3.1. Extreme Value Theory

EVT is very effective for modelling financial data in order to examine the limit behaviours of fat-tailed distributions (Gençay and Selçuk, 2004; McNeil, 1997). Since one of the ways that we assess the risk is the analysis of the tail risk, we only handle the randomness in the tail using limit laws rather than fitting a single distribution to the whole sample.

There are two principal types of methods to determine extreme events (Gençay and Selçuk, 2004; Gilli and Këllezi, 2000):

- (i) **Block maxima (BM) method:** In order to define the asymptotic distribution of extreme events, *Fisher-Tippet Theorem* suggests that the limit distribution of the maximum values of a random variable in specified periods belongs to an extreme value distribution family.
- (ii) **Peaks over threshold (POT) method:** According to *Pickands-Balkema-de Haan Theorem*, the conditional distribution of the values above a given high threshold is approximately GPD.

BM is a traditional method which is used to analyse data with seasonality. On the other hand, POT is preferred by many financial applications due to the fact that this method uses data more efficiently (Gilli and Këllezi, 2000). In this paper, we analyse the behaviour of large observations which exceed a high threshold instead of maximum values of observations. We thus use POT approach in order to model log-returns.

In the Pickands-Balkema-de Haan theorem, the conditional excess distribution

function $F_u(y)$ converges to the GPD for a high value of threshold u as follows:

$$F_u(y) \approx G_{\xi, \sigma}(y) = \begin{cases} 1 - (1 + \frac{\xi y}{\sigma})^{-\frac{1}{\xi}} & , \xi \neq 0 \\ 1 - \exp(-\frac{y}{\sigma}) & , \xi = 0 \end{cases} \quad (3.1)$$

where σ is the scale parameter and ξ is the shape parameter of the distribution. Here, $y \geq 0$ when $\sigma > 0$ and $\xi \geq 0$, and $0 \leq y \leq -\sigma/\xi$ when $\xi < 0$ (Gilli and K llezi, 2000).

3.2. Time Series Models

In the statistical analysis of time series, autoregressive-moving-average (ARMA) models provide a parsimonious description of a (weakly) stationary stochastic process in terms of two polynomials, one for the autoregression and the second for the moving average. The ARMA model is described as a tool for understanding and predicting future values in time series. The model consists of two parts, an autoregressive (AR) part and a moving average (MA) part. The AR part involves regressing the variable on its own lagged values. The MA part involves modelling the error term as a linear combination of error terms occurring contemporaneously and at various times in the past.

ARMA(p, q) model (where p is the order of the AR terms ϕ and q is the order of the MA terms ψ) is given as

$$\phi(B)Z_t = \psi(B)\epsilon_t \quad (3.2)$$

where unknown parameters are $\theta = (\sigma_\epsilon^2, \phi_1, \dots, \phi_p, \psi_1, \dots, \psi_q)'$ (Hamilton, 1994).

The autoregressive conditional heteroskedasticity (ARCH) model is a statistical model for time series data that describes the variance of the current error term or innovation as a function of the actual sizes of the previous time periods' error terms. The ARCH model is appropriate when the error variance in a time series follows an AR model; if an ARMA model is assumed for the error variance, the model is a GARCH model. For forecasting, combining autoregressive integrated moving average (ARIMA) and ARCH models could also be considered.

The GARCH(p, q) model (where p is the order of the ARCH terms w^2 and q

is the order of the GARCH terms σ^2) is given by

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \alpha_1 w_{t-1}^2 + \cdots + \alpha_p w_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2 \\ &= \alpha_0 + \sum_{i=1}^p \alpha_i w_{t-i}^2 + \sum_{i=1}^q \beta_i \sigma_{t-i}^2\end{aligned}\quad (3.3)$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$ and $\beta_i \geq 0$ for all i , and $\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) < 1$. Here w_t is a white noise (WN) with time-varying conditional variance (conditional heteroskedasticity) (Hamilton, 1994). Specifically, we write

$$w_t = \sigma_t \epsilon_t \quad (3.4)$$

where ϵ_t is a strong WN with zero mean and unit variance, i.e., $\epsilon_t \sim \text{iid } N(0, 1)$. And the conditional variance of the time series X_t is calculated as

$$\sigma_t^2 = \mathbb{V}(X_t | X_{t-1}, X_{t-2}, \dots) \quad (3.5)$$

After selecting the model, residuals of the fitted model should be tested in several ways. According to the *independence of residuals* assumption, once the model is developed and residuals are computed, there should be no remaining autocorrelations or partial autocorrelations at the lags in the autocorrelation functions (ACFs) and partial autocorrelation functions (PACFs). In addition to the graphical interpretation, the Ljung-Box test is a quantitative way to test for autocorrelation at multiple lags simultaneously by testing the hypotheses below:

$$\begin{aligned}H_0 : \rho_1 = \rho_2 = \dots = \rho_m = 0 & \quad \text{i.e. the residuals are independent} \\ H_1 : \rho_j \neq 0, \exists j \in \{1, 2, \dots, m\} & \quad \text{i.e. the residuals are not independent}\end{aligned}$$

4. Application

We analyse the data by separating the positive and negative returns to apply the EVT for both upper and lower tails. Similarly, time series analyses are applied for both positive and negative returns too. Although it might be a concern that separating as positive and negative returns might distort the dependency structure of the data, which is important for the regression and time series analysis, it is a common application in the financial and implied volatility literature (Davidson *et al*, 2001; Simon, 2003; Ahoniemi, 2008; Ma,

2013; Bugge *et al.*, 2016; Rizvi *et al.*, 2017). Negative and positive returns affect the volatility differently and using absolute or squared returns does not address this fact. To capture this effect, volatility studies treat the positive and negative returns as two data series.

Dividing the stock return data into two parts as a gain or loss might be useful to consider an investor's position as a short or long seller. In finance, being at a short position in a stock means that an increase in the stock price causes a loss. Hence, negative log-returns appear as a risk for a short position. On the other hand, positive log-returns yield riskiness for the long position in a stock since a price decline affects the long position negatively.

4.1. Selection of the Optimal EVT Model

The choice of the optimal threshold u is a crucial step for parameter estimation of the GPD. We first obtain the mean excess plots (MEP) to determine the threshold intervals for each country. In order to choose the best threshold carefully, we select four candidate intervals from the MEP. After that, we choose two potential optimal thresholds for each interval yielding stable scale and shape parameter graphs simultaneously. The potential thresholds are determined graphically according to threshold stability plots and finally the efficiency of the relevant GPD models are evaluated by the goodness-of-fit tests.

4.1.1 Estimation of the Upper Tail

In this subsection, we obtain the optimal thresholds for the positive log-returns for each country. After choosing the best threshold for the right tail, we estimate the parameters of the GPD distribution that fits the log-returns exceeding this threshold.

Table 3. The threshold intervals obtained by MEP of upper tail of log-returns for developed countries

Country	Singapore		Spain		UK		US	
	LB	UB	LB	UB	LB	UB	LB	UB
Interval 1	0.0021	0.0025	0.0034	0.0042	0.0014	0.0024	0.0008	0.0012
Interval 2	0.0039	0.0044	0.0054	0.0063	0.0062	0.0072	0.0036	0.0040
Interval 3	0.0075	0.0079	0.0131	0.0138	0.0087	0.0094	0.0053	0.0057
Interval 4	0.0120	0.0130	0.0178	0.0188	0.0108	0.0116	0.0066	0.0070

Table 4. The threshold intervals obtained by MEP of upper tail of log-returns for emerging countries

Country	Chile		Russia		Malaysia		Turkey	
	LB	UB	LB	UB	LB	UB	LB	UB
Interval 1	0.0040	0.0050	0.0045	0.0055	0.0025	0.0033	0.0017	0.0034
Interval 2	0.0074	0.0086	0.0099	0.0117	0.0078	0.0086	0.005	0.0059
Interval 3	0.0096	0.0101	0.0240	0.0260	0.0098	0.0104	0.0085	0.0095
Interval 4	0.0115	0.0122	0.0263	0.0300	0.0106	0.0119	0.0135	0.0145

Table 3 and Table 4 present the candidate intervals which consist of the optimal threshold for each country. These intervals are determined by MEP based on the regions where the graph is stable.

Having obtained four threshold intervals for each country, threshold stability plots help us to find two potential u for each interval, which are candidates for the optimal threshold, so that 8 optimal threshold candidates are found for each country. The test results of significance for the optimal thresholds are displayed in the following table for developed and emerging countries.

The bootstrap goodness-of-fit test is used to test the null hypothesis “ H_0 : A random sample has a GPD with unknown shape parameter ξ , which is a real number” (Villasenor-Alva and Gonzalez-Estrada, 2009). After optimal thresholds are determined visually, a goodness-of-fit test is needed for an efficient decision. Since the chosen threshold gives the estimations of the GPD parameters, we simulate values according to the number of exceedances with respect to the potential threshold and compare them with the real values. The optimal threshold for each country is chosen by the root mean square error (RMSE), the mean absolute percentage error (MAPE), the log-likelihood and the Bayesian Information Criteria (BIC) given in Table 5. Although the RMSE and MAPE are important performance criteria, BIC has been considered as the primary criterion to decide the optimal threshold since the EVT models will

be compared with the time series models.

Table 5. Test of significance results for the best threshold of upper tail of log-returns for developed and emerging countries

Country	Threshold	RMSE	MAPE	ln (L)	BIC
Singapore	0.00238	0.0039982	5.509185	-1004.8813	-1999.126
Spain	0.00348	0.0047664	4.137666	-1092.1563	-2173.393
UK	0.00144	0.0044701	9.031777	-1723.5261	-3435.331
USA	0.00084	0.0030425	8.975681	-1886.8295	-3761.821
Chile	0.00420	0.0048671	4.327342	-778.3624	-1546.537
Russia	0.00488	0.0090434	4.861209	-958.3332	-1905.808
Malaysia	0.00279	0.0041654	9.283499	-793.2761	-1576.402
Turkey	0.00255	0.0068810	17.622579	-1555.1105	-3098.432

4.1.2 Estimation of the Lower Tail

The analyses in Section 4.1.1 are repeated for negative log-returns in this subsection. Table 6 and Table 7 present the candidate intervals which consist of the optimal threshold for each country.

Table 6. The threshold intervals obtained by MEP of lower tail of log-returns for developed countries

Country	Singapore		Spain		UK		US	
	LB	UB	LB	UB	LB	UB	LB	UB
Interval 1	-0.0030	-0.0020	-0.0040	-0.0010	-0.0080	-0.0070	-0.0034	-0.0026
Interval 2	-0.0058	-0.0049	-0.0080	-0.0070	-0.0125	-0.0113	-0.0054	-0.0046
Interval 3	-0.0075	-0.0068	-0.0110	-0.0100	-0.0155	-0.0146	-0.0096	-0.0091
Interval 4	-0.0088	-0.0083	-0.0190	-0.0160	-0.0260	-0.0195	-0.0115	-0.0105

Table 7. The threshold intervals obtained by MEP of lower tail of log-returns for emerging countries

Country	Chile		Russia		Malaysia		Turkey	
	LB	UB	LB	UB	LB	UB	LB	UB
Interval 1	-0.0035	-0.0025	-0.0036	-0.0021	-0.0018	-0.0008	-0.0050	-0.0038
Interval 2	-0.0070	-0.0050	-0.0078	-0.0057	-0.0040	-0.0035	-0.0100	-0.0094
Interval 3	-0.0120	-0.0110	-0.0190	-0.0175	-0.0054	-0.0051	-0.0140	-0.0132
Interval 4	-0.0154	-0.0140	-0.0260	-0.0240	-0.0069	-0.0063	-0.0178	-0.0170

Similar as upper tail results, the potential optimal thresholds and significance test results are displayed in Table 8 for developed and emerging countries.

Table 8. Test of significance results for the best threshold of lower tail of log-returns for developed and emerging countries

Singapore					
Country	Threshold	RMSE	MAPE	ln (L)	BIC
Singapore	-0.00234	0.003572139	7.737261	-1032.6614	-2054.6478
Spain	-0.00112	0.006324746	30.169172	-1851.1370	-3690.2465
UK	-0.00744	0.008804776	4.249693	-199.3943	-391.1753
USA	-0.00263	0.003751820	8.758290	-727.8205	-1445.6466
Chile	-0.00288	0.004398696	4.587791	-1144.7742	-2278.6038
Russia	-0.00234	0.008235762	3.949035	-1599.8502	-3187.8306
Malaysia	-0.00098	0.003671441	19.969454	-1826.7269	-3641.6377
Turkey	-0.00399	0.007412375	5.186079	-1111.9504	-2212.7414

Since the chosen threshold gives the estimations of the GPD parameters, we simulate values according to the number of exceedances with respect to the potential threshold and compare them with the real values. The optimal threshold for each country is examined by the RMSE, the MAPE, the log-likelihood and the BIC. As in the upper tail results, the BIC is considered as the key criterion to compare the EVT models with the time series models in the next section.

In order to forecast the future returns, we choose the optimal parameters given in Table 9 for EVT method based on the best fitted models.

Table 9. The optimal GPD parameters for upper and lower tail of log-returns for developed and emerging countries

Upper Tail			Lower Tail		
Country	σ	ξ	Country	σ	ξ
Singapore	0.002506	0.063103	Singapore	-0.002606	-0.014909
Spain	0.003686	-0.044275	Spain	-0.003772	0.054015
UK	0.002380	0.130314	UK	-0.003586	0.199612
US	0.002478	-0.071686	US	-0.002733	-0.015439
Upper Tail			Lower Tail		
Country	σ	ξ	Country	σ	ξ
Chile	0.002905	0.066022	Chile	-0.002915	0.027725
Russia	0.004879	0.119886	Russia	-0.005047	0.056531
Malaysia	0.002319	0.108751	Malaysia	-0.002584	-0.005169
Turkey	0.005264	-0.037195	Turkey	-0.005821	-0.049697

4.2. Selection of the Optimal Time Series Model

We use both graphical tools and statistical tests in order to choose the best time series models for positive and negative log-returns of each country.

4.2.1 Model Estimation of the Positive and Negative Returns

Primarily we analyse the ACF and PACF graphs of the log-returns to see if they can be modelled by ARMA(p,q). We also check the goodness of fit of the models using some statistics such as BIC, adjusted R^2 and Ljung-Box test results. After trying several ARMA, ARIMA, GARCH and ARMA-GARCH models for the positive returns for each country, we see that the GARCH or ARMA-GARCH models fit the data best for all countries.

Table 10 presents the best models and the BIC values for the positive and negative log-returns for developed and emerging countries. GARCH and ARMA-GARCH models seem as the best models for the returns but the orders of the models change from country to country. Although the models presented in the table have been chosen as the best, all adjusted- R^2 values are very low and close to zero which might indicate that the increasing number of parameters provide very few information. Results also show that higher order GARCH models are required for negative log-returns for most of the countries comparing with the models proposed for the positive log-returns.

Table 10. The time series model selection results for the positive and negative log-returns for developed and emerging countries

Country	Positive log-returns		Negative log-returns	
	Model	BIC	Model	BIC
Singapore	GARCH(1,1)	-4269.58	GARCH(1,1)	-4642.52
Spain	GARCH(1,1)	-3972.78	GARCH(2,2)	-3882.57
UK	GARCH(1,1)	-4803.99	GARCH(2,2)	-4163.34
USA	ARMA(1,1)-GARCH(1,1)	-4573.73	GARCH(1,1)	-4316.71
Chile	GARCH(0,1)	-3935.23	GARCH(1,1)	-4265.17
Russia	GARCH(1,1)	-3403.32	GARCH(2,1)	-3824.06
Malaysia	ARMA(1,1)-GARCH(1,1)	-4176.37	GARCH(2,1)	-4953.98
Turkey	GARCH(1,0)	-3634.15	GARCH(0,1)	-3529.32

The parameters for the optimal time series models for positive and negative returns of all countries are presented in Table 11 and 12, respectively. These parameters are used to simulate the log-return data to compare with the observed values in order to analyse the goodness-of-fit of the proposed models.

Table 11. Parameters for the optimal time series models for positive returns for developed and emerging countries

Country	Best Model	Parameters					
		α_0	α_1	β_1			
Singapore	GARCH(1,1)	9.714e-07	0.1308	0.7963			
Spain	GARCH(1,1)	1.658e-05	0.1771	0.2667			
UK	GARCH(1,1)	1.707e-07	0.06408	0.9281			
		μ	ϕ_1	ψ_1	ω	α_1	β_1
US	ARMA(1,1)-GARCH(1,1)	3.418e-05	9.911e-01	-9.568e-01	4.696e-07	7.812e-02	9.051e-01

Country	Best Model	Parameters					
		α_0	α_1	β_1			
Chile	GARCH(0,1)	1.869e-05	0.2633				
Russia	GARCH(1,1)	8.373e-07	0.0585	0.932			
		μ	ϕ_1	ψ_1	ω	α_1	β_1
Malaysia	ARMA(1,1)-GARCH(1,1)	-8.856e-05	9.794e-01	-8.796e-01	2.987e-07	2.235e-01	8.08e-01
		α_0	α_1	β_1			
Turkey	GARCH(1,0)	5.572e-05		0.02166			

Table 12. Parameters for the optimal time series models for negative returns for developed and emerging countries

Country	Best Model	Parameters				
		α_0	α_1	α_2	β_1	β_2
Singapore	GARCH(1,1)	1.513e-7	0.06683		0.9214	
Spain	GARCH(2,2)	5.093e-6	0.00462	0.2698	0.629	1.781e-7
UK	GARCH(2,2)	2.564e-6	0.1951	0.4363	1.519e-13	0.3666
USA	GARCH(1,1)	5.038e-6	0.3139		0.2442	
Country	Best Model	α_0	α_1	α_2	β_1	β_2
Chile	GARCH(1,1)	4.647e-6	0.06678		0.7153	
Russia	GARCH(2,1)	1.595e-6	0.258		0.01046	0.7393
Malaysia	GARCH(2,1)	7.014e-8	0.1354		0.3963	0.4694
Turkey	GARCH(0,1)	6.004e-5	0.04515			

When we compare the BIC values for the best models chosen in time series analysis with the EVT models, we see that the time series models display better fits due to producing lower BIC values. Thus, by looking at these results we might conclude that GARCH and ARMA-GARCH models with specified parameters represent the log-return data better than the EVT models.

5. Risk Measures

VaR is one of the most commonly used risk measures which reflects the maximum loss within a given confidence level (Jorion, 2001). Despite its simplicity, VaR is not coherent because of not fulfilling subadditivity property except under the normal distribution assumption.

VaR can be expressed as

$$\text{VaR}_q(X) = F_X^{-1}(q) = \sup \{x \mid \mathbb{P}(X > x) > q\}$$

where F^{-1} is the *quantile function* of random variable X defined as the inverse of the distribution function F . Although it would be better to present both VaR and expected shortfall to compare the models using different risk measures, we only calculate VaR to provide a benchmark comparison to keep the simplicity. However, we will consider the other risk measures for future studies.

The methods used for computing VaR can be grouped as the parametric and non-parametric approaches. In this paper, we estimate VaR values for EVT and GARCH models both of which are parametric approaches. In addition to the above definition that is used to calculate VaR for the selected EVT models, we use another calculation of VaR suggested by Gençay *et al.* (2003). This definition, which is proposed as a variance-covariance approach, is useful for our study due to its applicability to GARCH models.

Let r_t , $t = 1, 2, \dots, n - 1$ be the log-return of the prices which follows a martingale process as $r_t = \mu_t + \epsilon_t$ where ϵ_t has a distribution function F with zero mean and variance σ_t^2 . The VaR in this case can be calculated as

$$\text{VaR}_t(\alpha) = \hat{\mu}_t + F^{-1}(\alpha)\hat{\sigma}_t$$

where $F^{-1}(\alpha)$ is the q -th quantile ($q = 1 - \alpha$) value of the unknown distribution function F . An estimate of μ_t and σ_t^2 can be obtained from the sample mean and the sample variance. Instead of the sample variance, the standard deviation in this equation can also be estimated by a statistical model (Gençay *et al.*, 2003).

According to the optimal EVT and GARCH models, we estimate the VaR of developed and emerging countries for positive and negative log-returns, respectively. Figures 3 and 4 display the calculated VaR values of the chosen methods for each country for 95% and 99.5% confidence levels.

Figure 3. Risk measure results for positive log-returns for developed and emerging countries

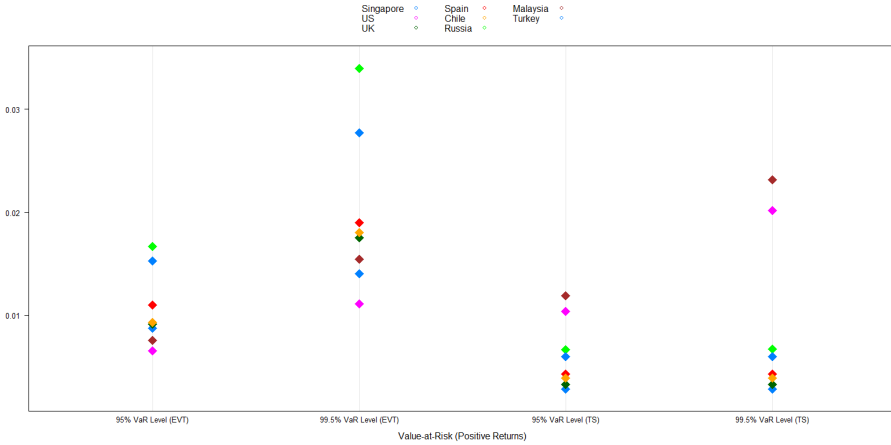


Figure 4. Risk measure results for negative log-returns for developed and emerging countries

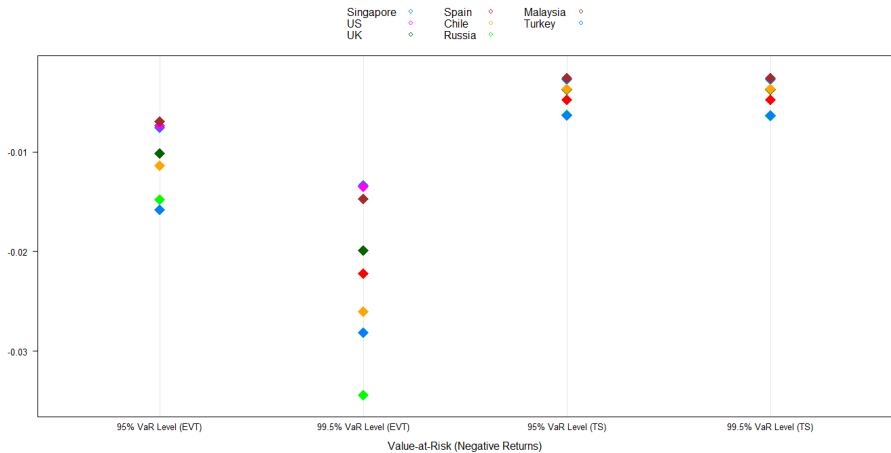


Figure 3 shows that the risk measures obtained from the EVT models are generally higher than the risk measures obtained from the time series models for positive returns which indicates that the EVT models provide more conservative results. The ranking of the countries according to the VaR are consistent within the same model (EVT or time series) for different confidence levels (95% and 99.5%) except for a few discrepancies.

Figure 3 displays the VaR analysis for positive returns of the EVT models, which shows that Russia and Turkey have the highest risk measures and the US has the smallest. However, the VaRs obtained from the time series models indicate that Malaysia and the US have the highest risk measures while Singapore has the smallest. Different models might lead different country rankings in terms of risk measures. Moreover, the graphs do not suggest a significant boundary between the developed and emerging markets which makes it difficult to analyse the relative positions of the countries.

By analysing Figure 4 we see that the absolute value of the VaR results are higher compared to the positive returns for both EVT and time series models. The rankings of the countries based on the calculated risk measures for two sets of models are different. Time series models are more consistent based on the country ranking for two different confidence levels while the EVT models indicate some discrepancies. Based on the negative returns data, EVT models show that Russia and Turkey have the highest risk while Malaysia and Singapore have the lowest. Time series models for two confidence levels present almost identical numbers for the risk measures which change on the fifth or sixth decimal places and also indicates Russia and Turkey are the most risky countries while the US and Malaysia are the least. There is no indication of a separation of the developed and emerging countries in terms of the risk measures obtained for both sets of models.

The results obtained for different confidence levels might indicate that the structure of the tail distributions of these countries are similar due to the increments on the risk measures between different confidence levels.

When we compare the rankings of these countries based on HDI and the riskiness of their stock markets, we see that there is a correlation between those two. Higher HDI indicates lower riskiness for emerging markets and the countries have been ordered from the highest to lowest HDI as Chile, Russia, Malaysia and Turkey. However, the ranking is not that strict for the developed countries since Spain has the second lowest VaR just after Singapore while it is the fourth in terms of the HDI index. Table 13 shows the rankings of the countries for both the HDI and the VaR obtained from the analysis of the returns.

Table 13: Rankings of the countries in terms of HDI and VaR

Country	HDI Ranking	VaR Ranking
Singapore	1	1
USA	2	4
UK	3	3
Spain	4	2
Chile	5	5
Russia	6	6
Malaysia	7	7
Turkey	8	8

The rankings presented in Table 13 are consistent with the graphs of the prices and the log-returns displayed in Figure 1 and 2. Since the only difference in ranking exists for the US and Spain, the log-return graph of the US confirms the relative riskiness of its stock markets due to the higher volatility.

6. Conclusions

We have compared developed and emerging countries based on the HDI index and the VaR results obtained from EVT and GARCH and ARMA-GARCH models. The results show that the time series models fit the log-return data better for all the countries chosen (Singapore, Spain, UK, US, Chile, Russia, Malaysia and Turkey) based on the BIC values. The risk measures obtained from EVT models are more conservative with respect to the risk measures obtained from the time series models. They are also sensitive to the confidence levels due to the changing rankings of the countries based on the calculated risk measures. However, the time series models provide more consistent results for both negative and positive log-returns. Due to the increments of the risk measures for 95% and 99.5% confidence levels, the tail structures are similar for all countries. The risk measures obtained from different models for different confidence intervals might indicate different rankings which makes us to consider more sophisticated ranking/ordering approaches. Although the HDI rankings and the VaR rankings seem consistent for most of the countries, the main components of the HDI which are life expectancy index, education index and gross national income per capita might have different affects on the rankings with different weights. We intend to consider the stochastic ordering approaches to compare the countries based on different risk measures and

HDI for further studies.

Acknowledgment

The authors would like to thank to the anonymous referee(s) who have contributed to the development of the contents with helpful suggestions.

References

- Aas, K. and X. K. Dimakos (2004). Statistical modelling of financial time series: An introduction. *Technical report, Norwegian Computing Center, Applied Research and Development. Project no: 220194.*
- Ahoniemi, K. (2008). Modeling and forecasting the VIX index. *SSRN*, 1-22.
- Bugge, S. A., Guttormsen, H. J., Molnár, P., and Ringdal, M. (2016). Implied volatility index for the Norwegian equity market. *International Review of Financial Analysis* 47; 133-141.
- Camilleri, S. J. (2006). An analysis of stock index distributions of selected emerging markets. *Bank of Valletta Review* 33, 33-49.
- Davidson, W. N., Kim, J. K., Ors, E., and A. Szakmary (2001). Using implied volatility on options to measure the relation between asset returns and variability. *Journal of Banking & Finance*, 25(7); 1245-1269.
- Ekşi, Z., Yıldırım, İ., and K. Yıldırak (2006). Alternative risk measures and extreme value theory in finance: Implementation on ISE 100 index. *International Conference on Business, Economics and Management*, İzmir, Turkey, Yaşar University.
- Gençay, R. and F. Selçuk (2004). Extreme value theory and value-at-risk: Relative performance in emerging markets. *International Journal of Forecasting* 20(2): 287-303.
- Gençay, R., Selçuk, A., and A. Ulugülyağcı (2003). High volatility, thick tails and extreme value theory. *Insurance: Mathematics and Economics* 33(1): 337-356.

- Gilli, M. and E. K llezi (2000). Extreme value theory for tail-related risk measures. *International Center for Financial Asset Management and Engineering FAME Research Paper Series* rp18.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
- Jorion, P. (2001). *Value at Risk*. McGraw-Hill.
- Lawler, C. C. (2003). A time-series analysis of the Shanghai and New York stock price indices. *Annals of Economics and Finance*, 4(1):17-35.
- Ma, L. (2013). Mutual fund flows and performance: A survey of empirical findings. Technical report, Working Paper.
- McNeil, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *Astin Bulletin* 27(1): 117-137.
- MSCI (2018). MSCI Indexes: End of day data (Country). <https://www.msci.com/end-of-day-data-country>. Accessed: 2018-07-02.
- Rizvi, S. A. A., Roberts, S. J., Osborne, M. A., and F. Nyikosa (2017). A novel approach to forecasting financial volatility with gaussian process envelopes. *Technical report, Machine Learning Research Group, Oxford-Man Institute of Quantitative Finance, University of Oxford*.
- Rocco, M. (2014). Extreme value theory for finance: A survey. *Journal of Economic Surveys*, 28(1). 82-108
- Simon, D. P. (2003). The Nasdaq volatility index during and after the bubble. *The Journal of Derivatives*, 11(2):9-24.
- UNDP (2014). United Nations Development Programme: Human development reports, Human Development Data 2014. <http://hdr.undp.org/en/data>. Accessed: 2018-07-02.
- Villasenor-Alva, J. A. and E. Gonzalez-Estrada (2009). A bootstrap goodness of fit test for the generalized Pareto distribution. *Computational Statistics and Data Analysis* 53(11): 3835-3841.