Value-at-risk estimation with wavelet-based extreme value theory:
Evidence from emerging markets

Atilla Cifter
Faculty of Economics and Administrative Sciences, Banking and Finance Department, Okan University, Istanbul, Turkey

A R T I C L E   I N F O

Article history:
Received 12 November 2010
Received in revised form 24 January 2011
Available online 28 February 2011

Keywords:
Extreme value theory
Wavelet-based extreme value theory
Emerging markets

A B S T R A C T

This paper introduces wavelet-based extreme value theory (EVT) for univariate value-at-risk estimation. Wavelets and EVT are combined for volatility forecasting to estimate a hybrid model. In the first stage, wavelets are used as a threshold in generalized Pareto distribution, and in the second stage, EVT is applied with a wavelet-based threshold. This new model is applied to two major emerging stock markets: the Istanbul Stock Exchange (ISE) and the Budapest Stock Exchange (BUX). The relative performance of wavelet-based EVT is benchmarked against the Riskmetrics-EWMA, ARMA–GARCH, generalized Pareto distribution, and conditional generalized Pareto distribution models. The empirical results show that the wavelet-based extreme value theory increases predictive performance of financial forecasting according to number of violations and tail-loss tests. The superior forecasting performance of the wavelet-based EVT model is also consistent with Basel II requirements, and this new model can be used by financial institutions as well.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Value-at-risk (VaR) became a common tool for financial forecasting with the Riskmetrics-exponentially weighted moving average (EWMA), made famous by Morgan in the early 1990s. This model is actually a special case of Bollerslev’s [1] generalized autoregressive conditional heteroskedasticity (GARCH) model. More than 100 volatility models were developed during the last 28 years [2]. One of the most important features that made the conditional volatility models popular is their ability to capture many of the typical stylized facts of a financial time series, such as time-varying volatility, persistence and volatility clustering. However, conditional volatility models cannot capture extreme movements, as these models are based on past volatility rather than the extreme observations. Extreme value theory models can capture extreme movements, and the forecasting performance of these models is better than that of GARCH type models [3]. There are some important reasons to model the returns with extreme value theory. First, the distribution of returns is heavy-tailed or leptokurtic for most of the financial returns. Second, the right and left tails of returns are not symmetrical, and extreme value theory models can be applied to each tail with different parameters, as opposed to GARCH and other volatility models.

The methodology of forecast combination was introduced by Bates and Granger [4]. They urged that we should combine forecasts as a weighted average of the individual forecasts. Forecast combination can be estimated by a combination of individual forecasts or by a combination of models. Individual forecasts’ combination can be estimated with artificial intelligence techniques. Liu [5] and Ozun and Cifter [6] proposed neural networks for a combination of individual forecasts. Liu [5] combined neural networks with historical simulation and GARCH(1, 1) models and found that the combination of historical estimation with GARCH and neural networks significantly improved forecasting performance. Ozun and Cifter [6]
Hybrid models as a combination of EVT with conditional volatility models were proposed by McNeil and Frey [8]. This model uses a two-stage approach. In the first stage, a GARCH type model is applied to residuals. In the second stage, EVT is applied to standardized residuals. McNeil and Frey [8] found that the conditional EVT procedure gives a better one-day-ahead forecast than methods which ignore the heavy tails of the innovations or the stochastic nature of the volatility.

Wavelets can also be used to estimate hybrid financial forecasting models. The combination of wavelet transform and GARCH models was introduced by Chi and Kai-jian [9], Lai et al. [10,11], He et al. [12,13], and Tan et al. [14]. Chi and Kai-jian, Lai et al., and He et al. [12] proposed wavelet-decomposed value-at-risk, and He et al. [13] proposed wavelet denoising ARMA–GARCH models. This paper typically used Kupiec [15]’s test for backtesting, and although their model is superior to conventional ARMA–GARCH models, the number of violations is greater than that of ARMA–GARCH models. Tan et al. [14] combined wavelet transform with ARIMA and GARCH models and applied this model to one-day-ahead electricity price forecasting. They found that their model is far more accurate than other forecasting models.

Yamada and Honda [16] used wavelet analysis to predict business turning points of the Nikkei 225 index and found that wavelet analysis can capture business peaks and troughs (minimum points) as an alternative structural break analysis. Bowden and Zhu [17] combined wavelet analysis with structural breaks and applied this combined model to the agribusiness cycle. By using wavelets, they added the business cycle feature to structural break analysis.

In this paper, wavelet-based extreme value theory (EVT) is introduced for univariate value-at-risk estimation. A wavelet-based EVT model is proposed as a combination of wavelets and the EVT model following the approach of McNeil and Frey [8]. In the first stage, wavelets are used as a threshold in generalized Pareto distribution, and in the second stage, EVT is applied with wavelet-based thresholds. The relative performance of this new hybrid model is compared with conventional volatility models for one-day-ahead forecasts, and wavelet-based EVT is benchmarked against the Riskmetrics-EWMA, ARMA–GARCH, generalized Pareto distribution, and conditional generalized Pareto distribution models. This new model is applied to two major emerging stock markets: the Istanbul Stock Exchange (ISE) and the Budapest Stock Exchange (BUX). It is found that the wavelet-based EVT model increases predictive performance of financial forecasting according to the number of violations and tail-loss tests for emerging markets.

The remainder of the paper is organized as follows. Section 2 provides value-at-risk methodologies. Section 3 describes the data on daily index returns. Section 4 presents empirical results for the forecasting performance of the models. Section 5 concludes the study.

2. Value-at-risk models

Value-at-risk (VaR) became the standard benchmark for measuring risks in modern risk management. Financial institutions are required to report VaR according to Basel requirements and regulatory authorities, and most non-financial institutions also prefer to report VaR for trading and management purposes. VaR become popular in the 1990s, following well-known disasters such as Orange Country, Barings, Metallgesellschaft, Dawia, and many others, but are now used to measure credit, operational and liquidity risks. VaR measures the worst expected loss over a given horizon under normal market conditions at a given confidence level [18]. For a single asset, VaR is estimated as

\[
\text{VaR}_{t+1,q} = \mu_{t+1} + \sigma_{t+1} F^{-1}(p)\sigma_{t+1}
\]

where \(\mu_{t+1}\) is the forecast of the conditional mean, \(\sigma_{t+1}\) is the forecast of the conditional standard deviation, and \(F^{-1}(p)\) is the corresponding quantile of the assumed distribution [19].

In the following section, wavelet-based EVT and benchmarked VaR models are introduced. Riskmetrics-EWMA and ARMA–GARCH models are selected as conditional volatility models, where generalized Pareto distribution is selected as the EVT model, and conditional generalized Pareto distribution is selected as the hybrid GARCH-EVT model for benchmarking.

2.1. Riskmetrics-EWMA model

The simplest and well-known volatility forecasting model, called Riskmetrics-exponentially weighted moving average (EWMA), was developed by Morgan for computing market risks. This model may be defined as a special case of the GARCH model, where \(\alpha = 1 - \lambda\) and \(\beta = \lambda\) and \(\omega = 0\). Riskmetrics-EWMA variance model can be written as [20]:

\[
\sigma_t^2 = \omega \sigma_{t-1}^2 + (1 - \omega) r_{t-1}^2
\]

where \(\lambda\) is the decay factor that determines relative weights, \(r_{t-1}\) is the previous squared returns and \(\sigma_{t-1}^2\) is the previous variance of return. Optimal \(\lambda\) can be determined by the in-sample optimization algorithm. \(\lambda\) is usually set to 0.94 for daily data and 0.97 for monthly data [21]. The higher the decay factor, the longer the previous returns. If \(\lambda\) equals 1, volatility is explained by solely previous volatility. Therefore, the mean reverting can be transformed into unconditional variance.
2.2. ARMA(r, s)–GARCH(p, q) models

Engle [22] introduced the autoregressive conditional heteroskedastic (ARCH) model for time-varying volatility in a time series. ARCH model can be expressed as [22]:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 = \omega + (\epsilon_t)^2$$  \hspace{1cm} (3)

where $\sigma_t$ is conditional volatility of $\epsilon_t$ with the conditions of $\omega < 1$ and $\sum_{i=1}^{q} \alpha_i < 1$ and $\epsilon_t/\psi_{t-1} \sim N(0, \sigma_t)$ as $N(.)$ is a probability density function with mean 0 and conditional variance.

Bollerslev [1] extended the ARCH model by adding $p$ lags to the conditional variance in the linear ARCH model, and introduced the generalized autoregressive conditional heteroskedasticity model (GARCH) as

$$\sigma_t^2 = \sigma_0 + \alpha_1 \sum_{i=2}^{n} \epsilon_{t-i}^2 + \beta_1 \sum_{i=2}^{n} \sigma_{t-i}^2$$ \hspace{1cm} (4)

where $\alpha, \beta > 1$ and $\alpha_0 > 1$.

If there is an autoregressive moving average effect in a non-stationary time series, ARMA(r, s) should be added to the GARCH(p, q) model. The ARMA(r, s)–GARCH(p, q) model can be estimated by a two-step approach, where, in the first step, ARMA(r, s) is estimated as

$$r_t = c + \sum_{i=2}^{r} a_i r_{t-i} + \sum_{i=2}^{s} b_i \epsilon_{t-i} + \epsilon_t.$$ \hspace{1cm} (5)

In the second step ARMA(r, s)–GARCH(p, q) is estimated as

$$\sigma_t^2 = \sigma_0 + \sum_{i=2}^{n} \omega_i \epsilon_{t-i}^2 + \sum_{i=2}^{n} \nu_i \sigma_{t-i}^2$$ \hspace{1cm} (6)

where $\epsilon_t = \epsilon_t \eta_1, \alpha, i = 1, 2, 3, \ldots, p, b_i, i = 1, 2, 3, \ldots, q, w_i, i = 1, 2, 3, \ldots, r, v_i, i = 1, 2, 3, \ldots, s, C, \alpha_0$ and $\lambda$ are model parameters. The first equation is called the mean equation and the second equation is called the variance equation.

Bollerslev [1] suggested using Gaussian distribution, where the log-likelihood function can be estimated as

$$l_T = -\frac{1}{2} \sum_{t=1}^{T} \left[ \ln(2\pi) + \ln(h_t^2) + z_t^2 \right]$$ \hspace{1cm} (7)

where $\sigma_t^2$ is variance, $z_t = \epsilon_t/\sigma_t$, and $T$ is number of observations. Bollerslev [23] showed that a fat-tailed distribution such as Student-t performs better capturing higher observed kurtosis. The log-likelihood function of the student-t distribution is shown below:

$$l_T^{\text{student-t}}(\theta) = T \left\{ \ln \Gamma \left( \frac{v + 1}{2} \right) - \ln \Gamma \left( \frac{v}{2} \right) - \frac{1}{2} \ln \left[ \pi (v - 2) \right] \right\} - \frac{1}{2} \sum_{t=1}^{T} \ln(h_t) + (1 + v) \ln \left( 1 + \frac{\epsilon_t^2}{v - 2} \right).$$ \hspace{1cm} (8)

The main drawback of these two distributions is that although student-t may account for fat tails, it is still symmetric. Lambert and Laurent [24] applied skewed student-t distribution in value-at-risk estimation. Skewed student-t distribution is proposed by Fernandez and Stell [25]. The main advantage of this distribution is that it considers both asymmetry and fat-tailness. If $\Gamma(.)$ denotes the gamma function, the log-likelihood of a standardized skewed student-t distribution is:

$$l_T^{\text{skewed-st}} = T \left\{ \ln \Gamma \left( \frac{\eta + 1}{2} \right) - \ln \Gamma \left( \frac{\eta}{2} \right) - 0.5 \ln \left[ \pi (\eta - 2) \right] + \ln \left( \frac{2}{\xi + 1} \right) + \ln(s) \right\}$$

$$- \frac{1}{2} \sum_{t=1}^{T} \ln \sigma_t^2 + (1 + \eta) \ln \left[ 1 + \frac{(se + m)^2}{\eta - 2} \xi - 2t \right].$$ \hspace{1cm} (9)

where $\xi$ is skewness parameter, $\eta$ is degrees of freedom, $\sigma_t^2$ is variance, $\epsilon_t$ is error term and $T$ is number of observations.

2.3. Extreme value theory

Extreme value theory focuses on extremes rather than the average values. There are two widely used extreme value theory distributions in VaR estimation: generalized extreme value (GEV) and generalized Pareto distribution (GPD). GEV distribution focuses on the block (or per period) maxima values, where GPD focuses on exceeding a given threshold ($u$). Fig. 1 shows both methods [26]. In this paper, only GPD is applied for volatility forecasting.
GPD is estimated by the peak-over-threshold (POT) method to find the distribution of extreme values over a defined threshold value. Considering a distribution function $F$ of random variable $X$, we estimate the distribution function $F_u$ of values of $x$ above a certain threshold $u$. The distribution function $F$ and conditional distribution function $F_y$ are defined as [26]:

$$F_u(y) = P(X - u \leq y | X > u) = \frac{F(u + y) - F(u)}{1 - F(u)} = \frac{F(x) - F(u)}{1 - F(u)}$$

for $0 \leq y \leq y_f - u$ where $y_f$ is the right end point of $F$, $u$ is threshold and $y = x - u$ is the excess value. The realization of the random variable $X$ is between 0 and $u$. If the optimal threshold is determined, it means that a sufficient number of observations is available.

Pickands [27]'s and Balkama and Haan De [28]'s theory states that for a large class of distribution functions $F$, the conditional distribution function $F_u(y)$ is approximated by $F_u(y) \approx \text{GPD}_{\xi, \sigma}(y)$.

$\text{GPD}_{\xi, \sigma}(y)$ can be defined as

$$\text{GPD}_{\xi, \sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma} y\right)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - e^{-y/\sigma} & \text{if } \xi = 0 \end{cases}$$

where $y \in [0, (x_f - u)]$ if $\xi \geq 0$ and $y \in [0, (-\sigma/\xi)]$ if $\xi < 0$. If $x$ is defined as $x = u + y$, that GPD can also be shown as

$$\text{GPD}_{\xi, \sigma}(x) = 1 - \left(1 + \frac{\xi (x - u)}{\sigma}\right)^{-1/\xi}$$

from the equation of $F_u(y)$, $F(x)$ can be estimated as

$$F(x) = (1 - F(u))F_u(x) + F(u)$$

and the estimator of $F(x)$ with GPD parameters can be computed as

$$\hat{F}(x) = \frac{N_u}{n} \left(1 - \left(1 + \frac{\hat{\xi}}{\hat{\sigma}} (x - u)\right)^{-1/\hat{\xi}}\right) + \left(1 - \frac{N_u}{n}\right)$$

where $n$ is the total number of observations in the data set and $N_u$ is the number of violations (extremes) above $u$.

Eq. (14) can be simplified as

$$\hat{F}(x) = 1 - \frac{N_u}{n} \left(1 + \frac{\hat{\xi}}{\hat{\sigma}} (x - u)^{-1/\hat{\xi}}\right).$$

Finally, one-step-ahead VaR forecast of GPD for a determined probability $p$ can be estimated as

$$\text{VaR}_{t+1, \theta}^p = u + \frac{\hat{\sigma}}{\hat{\xi}} \left(\left(\frac{n}{N_u} p\right)^{-\hat{\xi}} - 1\right)$$

where $u$ is the threshold parameter, $\hat{\sigma}$ is the scale, and $\hat{\xi}$ is the shape parameter. Shape parameter determines fatness of the tail. If the shape parameter $\hat{\xi} > 0$, it becomes ordinary Pareto distribution, if $\hat{\xi} = 0$ it becomes exponential distribution, and if $\hat{\xi} < 0$ it becomes Pareto type II distribution. The case $\hat{\xi} > 0$ is the most relevant for financial data, since it corresponds to heavy tail [29].

Determining the threshold value is crucial in GPD. The choice of the threshold value is subject to trade-off between variance and bias [30]. If the threshold is determined to be over its optimal value, this makes the estimator more volatile.
but less biased. If the threshold is determined to be less than its optimal value, this makes the estimator less volatile but more biased. Optimal threshold can be determined with the graph of mean excess, the Hill estimator [7], the Q–Q graph suggested by Gencay and Selçuk [31] and Gencay and Selçuk [32], the graphical bootstrap method [33], and the sample percentile approach [34]. In this paper, threshold is estimated with the graph of mean excess that determined the quantile level, and it is found that 5th percentile of the sample should be chosen as a threshold value. The rolling regression approach is estimated for the conditional threshold in GPD.

2.4. Conditional extreme value theory

McNeil and Frey [8] have proposed the conditional extreme value theory as a combined method of GARCH and EVT process. This model makes minimal assumptions about the underlying innovation distribution and uses a two-stage approach:

Step 1: Fit a GARCH type model to the return data by using a quasi-maximum likelihood approach. Estimate $\mu_{t+1}$ and $\sigma_{t+1}$ from the fitted model and calculate the implied model residuals.

Step 2: Consider the standardized residuals as a result of a white noise process, use the extreme value theory to model the tail of $F_z(z)$ and use this EVT model to estimate $z_q$. The EVT model can be estimated with any of the extreme value distributions, but McNeil and Frey [8] suggest using the peak-over-threshold method.

One-step-ahead VaR forecast of conditional EVT is given by:

$$\text{VaR}_{t+1,q} = \mu_{t+1} + \sigma_{t+1} \text{VaR}_{t+1,q}(z)$$

where $\text{VaR}_{t+1,q}(z)$ is VaR forecasts with conditional EVT method determined with peak-over-threshold (POT).

2.5. Wavelet-based extreme value theory

Wavelet theory is based on Fourier analysis, in which any function can be represented as the sum of sine and cosine functions. A wavelet $\psi(t)$ is simply a function of time $t$ that obeys a basic rule, known as the wavelet admissibility condition [35]:

$$C_{\psi} = \int_0^\infty \frac{|\psi(f)|}{f} \, df < \infty$$

where $\psi(f)$ is the Fourier transform, a function of frequency $f$, of $\psi(t)$.

Based on the length of data, there are two types of wavelet transforms, namely, continuous wavelet transform (CWT) and discrete wavelet transform (DWT). Since most of the time series have a finite number of values, the discrete version of wavelet transform is used in finance and economics applications. Discrete wavelets are defined as [36]:

$$\phi_{j,k} = 2^{j/2} \phi(2^j t - k)$$

$$\psi_{j,k} = 2^{j/2} \psi(2^j t - k).$$

In practice, the DWT is implemented via the pyramid algorithm. For each iteration of the pyramid algorithm, three objects are required: the data vector $x$, the wavelet filter $h_l$ and the scaling filter $g_l$. At the first level, $j = 1$, DWT wavelet coefficients $\psi_{l,t}$ and scaling coefficients $v_{l,t}$ are obtained as follows:

$$\psi_{l,t} = \sum_{l=0}^{l-1} h_l x_{2^l t + 1 - l \mod N}$$

$$v_{l,t} = \sum_{l=0}^{l-1} g_l x_{2^l t + 1 - l \mod N}.$$
The wavelet and scaling filters, $\tilde{g}_j$, $\tilde{h}_j$, are scaled as $\tilde{g}_j = g_j / 2^{j/2}$, $\tilde{h}_j = h_j / 2^{j/2}$. Non-decimal wavelet coefficients represent differences between generalized averages of the data on a scale $\tau = 2^{-j}$ [36].

Wavelet-based GPD can be estimated by replacing standard threshold with wavelet threshold. This approach will add the business cycle feature to GPD, as wavelets include short–middle–long term features. Therefore, the robustness to filter lengths should be estimated prior to out-of-sample forecasting in wavelet-based GPD. In-sample forecasting data can be used to determine the filter lengths. In this paper, the first 250 observations are used for in-sample forecasting, as shown in the data section. Optimal filter lengths can be estimated with the number of violations and tail-loss tests. It should be noted that root mean squared error (RMSE) cannot be used to determine the optimal filter lengths in extreme value theory, as it does not consider tail losses.

The robustness to filter length test is applied to in-sample data from the ISE and BUX indexes. Lengths of filters 1–6 are taken as the threshold, and absolute values of scaled variables are considered, where filter 1 represents two days ($2^1$) and filter 6 represents 64 days ($2^6$). The previous filter is added ($\sum_{j=1}^{n} w_j$) to the current filter to decrease the extreme observations. Table 1 reports the filter length test for the ISE and BUX indexes. Filter 6 contains the minimum number of violations, which are one and zero for the ISE and BUX indexes respectively. Christoffersen's conditional coverage test shows that models with filter 3 and beyond are statistically significant. According to this tail-loss test, 6 is the best filter. RMSE statistics show that filter 1 is the best model, and filter 6 is the worst model. RMSE should not be used to determine filter lengths, as noted above. Since filter 6 reduces the number of violations and Christoffersen test's probability, filter 6 should be chosen to estimate threshold value in wavelet-based GPD.

### 2.6. Backtesting methodology

Selections of volatility models play a major role in value-at-risk estimation. The Basle Committee on Banking Supervision [37] offered to use number of violations or absolute failure rate to compare one year of profit and loss to the VaR forecast. The main drawback of this approach is that it does not consider failure rate or cumulative failure probability. In the literature, root mean squared error (RMSE) is used as the failure rate where the Kupiec [15] test and the Christoffersen [38] test are used as failure probability or tail-loss probability. There are also other backtesting models, but the Christoffersen [38] conditional coverage test is widely used and has superior forecasting performance for one-day-ahead VaR estimation.

The root mean squared error (RMSE) is a scaled dependent comparison algorithm for forecasts. The smaller its values, the more accurate its forecasts become. The test value is calculated as the deviation of the $h$-step ahead forecasts of a variable,
\[
E(y_{t+h}), \text{ from its observed time path, } y_{t+h}. \text{ The RMSE of } E(y_{t+h}) \text{ can be estimated as }
\]
\[
\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=T+1}^{T+h} (\hat{\sigma}_{f,x} - \sigma_{r,t})^2}
\]  
(24)

where \( T + 1 \) is the beginning of the testing sample, \( T + h \) is the end of the testing sample, \( \hat{\sigma}_{f,x} \) is forecasted volatility, and \( \sigma_{r,t} \) is realized volatility.

The likelihood ratio (LR) test of unconditional coverage, independence, and the joint test of coverage and independence named as conditional coverage, as proposed by Christoffersen [38], show whether VaR models have a correct coverage at each point in time. Conditional coverage covers the other tests. This test enables joint testing of randomness and correct coverage while retaining the individual hypothesis as a subcomponent. The LR test of conditional coverage can be estimated as [38]:
\[
\text{LR}_{cc} = -2 \ln \left[ \alpha^N (1 - \alpha)^{T-N} \right] + 2 \ln \left[ (1 - \pi_{01})^{n_{00}} \pi_{01} (1 - \pi_{11})^{n_{10}} \pi_{11} \right] \sim \chi^2_{(1)}
\]  
(25)

where \( T \) is the number of observations, \( N \) is the number of violations, \( \alpha \) is the confidence level, \( n_{ij} = 0, 1 \) if violations occur “1” if not “0”, and \( \pi_{ij} = \Pr(t_{i-1} = j/t_{i-1} = i) \) is the probability value. The first part of the test is the unconditional coverage test (LRUC), and the second part is independence (LRind) test.

This paper estimates a VaR value at 1% confidence level, which is the same as Basel II requirements. VaR forecasting performances are compared to the number of violations, RMSE and Christoffersen tests for left tail.

3. Data

Two major emerging markets from Europe are selected for VaR estimation. The first one is the ISE-100 index from the Istanbul Stock Exchange, Turkey, and the second index is the BUX from the Budapest Stock Exchange, Hungary. Global stock investors take positions in both markets. Both series are from Bloomberg, sampled at a daily frequency. The data set covers 1986 daily observations from May 10, 2002 to April 30, 2010. The first 250 observations, from May 10, 2002 to May 13, 2003, are used for starting parameters as an in-sample of forecasting, where the last 1736 observations, from May 14, 2003 to April 30, 2010, are used for out-of-sample forecasting. The daily returns and histograms of ISE and BUX are shown in Fig. 2.

In order to test volatility effects, a time series should not contain a unit root. There are two standard unit root tests, namely the augmented Dickey–Fuller [39] and Phillips–Perron [40] tests. Because of the size and power weaknesses of these two tests, other unit root tests were developed. Two of them are the Dickey–Fuller Test with GLS detrending-ADF-GLS [41] and the Ng–Perron [42] tests. Table 2 reports the unit root test results, and all the tests indicate that both the ISE and BUX indexes do not contain a unit root in log difference level. Therefore, univariate volatility models are set based on log-differenced series.
Table 2
Unit root test statistics of the time series.

<table>
<thead>
<tr>
<th></th>
<th>ADF test</th>
<th>P–P test</th>
<th>ADF-GLS test</th>
<th>Ng–Perron tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>MZ_α</td>
<td>MZ_γ</td>
</tr>
<tr>
<td>ISE</td>
<td>−0.563012</td>
<td>−0.642784</td>
<td>0.971895</td>
<td>1.18342</td>
</tr>
<tr>
<td>ISE(Log returns)</td>
<td>−43.1840</td>
<td>−43.1877</td>
<td>−2.29843</td>
<td>−7.07214</td>
</tr>
<tr>
<td>BUX</td>
<td>−1.074385</td>
<td>−1.084621</td>
<td>0.237744</td>
<td>0.30874</td>
</tr>
<tr>
<td>BUX(Log returns)</td>
<td>−44.1907</td>
<td>−44.1892</td>
<td>−7.44945</td>
<td>−37.5566</td>
</tr>
</tbody>
</table>

Notes: Tests contain a constant but not a time trend. The table reports results of the augmented Dickey–Fuller [39], Phillips–Perron [40], Dickey–Fuller Test with GLS detrending-ADF-GLS [41] and Ng–Perron [42] tests. The number of lags (nl) in the tests have been selected using the Schwarz information criterion with a maximum of twelve lags.

Table 3
Descriptive statistics.

<table>
<thead>
<tr>
<th></th>
<th>ISE(Log return)</th>
<th>BUX(Log return)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>−0.000999</td>
<td>−0.000639</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>0.019465</td>
<td>0.018162</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.049422</td>
<td>−0.92808</td>
</tr>
<tr>
<td>Excess Kurtosis</td>
<td>2.7067</td>
<td>21.442</td>
</tr>
<tr>
<td>Minimum</td>
<td>−0.12127</td>
<td>−0.24122</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.090137</td>
<td>0.12649</td>
</tr>
<tr>
<td>Jarque-bera test</td>
<td>530.64</td>
<td>33505.0</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>LM(2)</th>
<th>LM(5)</th>
<th>LM(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISE(Log return)</td>
<td>17.916</td>
<td>25.452</td>
<td>15.170</td>
</tr>
<tr>
<td>BUX(Log return)</td>
<td>36.401</td>
<td>36.401</td>
<td>27.142</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
</tbody>
</table>

Notes: Tests contain a constant but not a time trend. The table reports results of the augmented Dickey–Fuller [39], Phillips–Perron [40], Dickey–Fuller Test with GLS detrending-ADF-GLS [41] and Ng–Perron [42] tests. The number of lags (nl) in the tests have been selected using the Schwarz information criterion with a maximum of twelve lags.

The descriptive statistics for two market returns are provided in Table 3. It is observed that both markets are far from normal distribution according to diagnostic tests. Besides, skewness and excess kurtosis values confirm that both distributions are not normal, which shows that volatility models should be estimated with non-normal distributions like skewed student-t, extreme value theory, etc. LM statistics of Engle [22] indicates that ARCH effects exist for both market returns.

4. Empirical findings

In this section, the relative forecasting performances of wavelet-based EVT and benchmarked volatility models are illustrated. The univariate forecasting performance of volatility models is tested for the ISE and BUX indexes. Riskmetrics-EWMA, ARMA(r, s)–GARCH(p, q), generalized Pareto distribution, and conditional generalized Pareto distribution are selected as the benchmarked models for forecasting. ARMA(r, s)–GARCH(p, q) is estimated as conditional volatility model. Table 4 shows the results of ARMA(r, s)–GARCH(p, q) models for the ISE and BUX indexes. According to Akaike [43] selection criteria, the optimal GARCH type model is AR(1)–GARCH(1, 1) for normal, student-t and skewed student-t distributions. Besides, AR(1)–GARCH(1, 1) parameters are statistically significant. Therefore, the AR(1)–GARCH(1, 1) model is selected as the VaR model for backtesting. The other ARMA(r, s)–GARCH(p, q) models are not included in the empirical analysis.

Fig. 3 shows scale, shape, and threshold parameters in GPD. In this paper, GPD parameters are set on a rolling basis in out-of-sample forecasting and thus increase robustness of GPD in extreme value theory analysis. Fig. 3 shows that we should estimate all the parameters as time-varying, since none of the parameters is constant over time. There are certain breaks in parameters, which also supports using time-varying parameters. The other implication from Fig. 3 is that after 2008’s global financial crisis, the threshold values of the ISE and BUX indexes converged each other, which shows that reducing portfolio VaR with portfolio selection will become more difficult in emerging markets.

Wavelet-based GPD is estimated for both the ISE and BUX indexes, and 64 days (2^6) filter lengths is selected as the threshold in this hybrid model. Maximal overlap discrete wavelet transform (MODWT) is selected, and all parameters are set on a rolling basis in out-of-sample forecasting. The predictive performance of wavelet-based EVT (W-GPD) and benchmarked models are shown in Table 5, and a comparison of their out-of-sample forecasting graphs is shown in Fig. 4 for ISE and Fig. 5 for BUX. Among others, W-GPD has the minimum number of violations for both stock markets for left tail as an indicator of risk. Conditional GPD and AR(1)–GARCH(1, 1) with skewed student-t distribution models also have a lower number of violations, but their performance exceeds W-GPD. The best model is Riskmetrics-EWMA.
Table 4
ARMA(r, s)–GARCH(p, q) tests.

<table>
<thead>
<tr>
<th>Model</th>
<th>AR</th>
<th>MA</th>
<th>ω</th>
<th>α</th>
<th>β</th>
<th>ν</th>
<th>ξ</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH(1, 1)-n</td>
<td>−</td>
<td>−</td>
<td>−0.001†</td>
<td>0.08†</td>
<td>0.191†</td>
<td>−</td>
<td>−</td>
<td>−5.5956</td>
</tr>
<tr>
<td>GARCH(1, 1)-n</td>
<td>0.051†</td>
<td>−</td>
<td>−0.001</td>
<td>0.08</td>
<td>0.198</td>
<td>−</td>
<td>−</td>
<td>−5.5974</td>
</tr>
<tr>
<td>AR(1)–GARCH(1, 1)-n</td>
<td>(2.661)</td>
<td>−</td>
<td>−0.004†</td>
<td>0.08†</td>
<td>0.198</td>
<td>−</td>
<td>−</td>
<td>−5.5973</td>
</tr>
<tr>
<td>MA(1)–GARCH(1, 1)-n</td>
<td>−</td>
<td>(2.705)</td>
<td>−0.001</td>
<td>0.08†</td>
<td>0.198</td>
<td>−</td>
<td>−</td>
<td>−5.5973</td>
</tr>
<tr>
<td>ARMA(1, 1)–GARCH(1, 1)-n</td>
<td>0.247</td>
<td>−0.194†</td>
<td>−0.001</td>
<td>0.08†</td>
<td>0.198</td>
<td>−</td>
<td>−</td>
<td>−5.5973</td>
</tr>
<tr>
<td>GARCH(1, 1)-t</td>
<td>−</td>
<td>−</td>
<td>−0.001</td>
<td>0.08†</td>
<td>0.191†</td>
<td>6.834†</td>
<td>−</td>
<td>−6.4688</td>
</tr>
<tr>
<td>GARCH(1, 1)-t</td>
<td>0.0378†</td>
<td>−</td>
<td>−0.001</td>
<td>0.08†</td>
<td>0.191†</td>
<td>6.939†</td>
<td>−</td>
<td>−6.4689</td>
</tr>
<tr>
<td>MA(1)–GARCH(1, 1)-t</td>
<td>−</td>
<td>(2.074)</td>
<td>−0.001</td>
<td>0.08†</td>
<td>0.191†</td>
<td>6.935†</td>
<td>−</td>
<td>−6.4694</td>
</tr>
<tr>
<td>ARMA(1, 1)–GARCH(1, 1)-t</td>
<td>0.01†</td>
<td>0.018</td>
<td>−0.001</td>
<td>0.08†</td>
<td>0.191†</td>
<td>6.937†</td>
<td>−</td>
<td>−6.4697</td>
</tr>
<tr>
<td>GARCH(1, 1)-Skew t</td>
<td>−</td>
<td>−</td>
<td>−0.001</td>
<td>0.08†</td>
<td>0.191†</td>
<td>6.937†</td>
<td>0.02</td>
<td>−6.4682</td>
</tr>
<tr>
<td>AR(1)–GARCH(1, 1)-Skew t</td>
<td>0.037†</td>
<td>−</td>
<td>−0.001</td>
<td>0.08†</td>
<td>0.191†</td>
<td>6.933†</td>
<td>−</td>
<td>−6.4682</td>
</tr>
<tr>
<td>MA(1)–GARCH(1, 1)-Skew t</td>
<td>−</td>
<td>(2.054)</td>
<td>−0.001</td>
<td>0.08†</td>
<td>0.191†</td>
<td>6.930†</td>
<td>−</td>
<td>−6.4689</td>
</tr>
<tr>
<td>ARMA(1, 1)–GARCH(1, 1)-Skew t</td>
<td>0.021</td>
<td>0.016</td>
<td>−0.001</td>
<td>0.08†</td>
<td>0.191†</td>
<td>6.932†</td>
<td>−</td>
<td>−6.4689</td>
</tr>
<tr>
<td>BUX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH(1, 1)-n</td>
<td>−</td>
<td>−</td>
<td>−0.001</td>
<td>0.119†</td>
<td>0.900†</td>
<td>−</td>
<td>−</td>
<td>−5.3828</td>
</tr>
<tr>
<td>GARCH(1, 1)-n</td>
<td>0.048†</td>
<td>−</td>
<td>−0.001</td>
<td>0.119†</td>
<td>0.899†</td>
<td>−</td>
<td>−</td>
<td>−5.3842</td>
</tr>
<tr>
<td>AR(1)–GARCH(1, 1)-n</td>
<td>(2.342)</td>
<td>−</td>
<td>−0.001</td>
<td>0.119†</td>
<td>0.899†</td>
<td>−</td>
<td>−</td>
<td>−5.3841</td>
</tr>
<tr>
<td>MA(1)–GARCH(1, 1)-n</td>
<td>−</td>
<td>(2.382)</td>
<td>−0.001</td>
<td>0.119†</td>
<td>0.899†</td>
<td>−</td>
<td>−</td>
<td>−5.3841</td>
</tr>
<tr>
<td>ARMA(1, 1)–GARCH(1, 1)-n</td>
<td>−0.064</td>
<td>0.112</td>
<td>−0.001</td>
<td>0.09</td>
<td>0.900†</td>
<td>−</td>
<td>−</td>
<td>−5.3836</td>
</tr>
<tr>
<td>GARCH(1, 1)-t</td>
<td>−</td>
<td>−</td>
<td>−0.0008</td>
<td>0.099†</td>
<td>0.912†</td>
<td>5.497†</td>
<td>−</td>
<td>−5.4607</td>
</tr>
<tr>
<td>AR(1)–GARCH(1, 1)-t</td>
<td>(2.156)</td>
<td>−</td>
<td>−0.0008</td>
<td>0.099†</td>
<td>0.912†</td>
<td>5.497†</td>
<td>−</td>
<td>−5.4607</td>
</tr>
<tr>
<td>MA(1)–GARCH(1, 1)-t</td>
<td>−</td>
<td>(1.262)</td>
<td>−0.0008</td>
<td>0.099†</td>
<td>0.912†</td>
<td>5.497†</td>
<td>−</td>
<td>−5.4606</td>
</tr>
<tr>
<td>ARMA(1, 1)–GARCH(1, 1)-t</td>
<td>−0.292</td>
<td>0.318</td>
<td>−0.0008</td>
<td>0.099†</td>
<td>0.912†</td>
<td>5.497†</td>
<td>−</td>
<td>−5.4603</td>
</tr>
<tr>
<td>GARCH(1, 1)-Skew t</td>
<td>−</td>
<td>−</td>
<td>−0.0008</td>
<td>0.099†</td>
<td>0.912†</td>
<td>5.498†</td>
<td>−</td>
<td>−5.4601</td>
</tr>
<tr>
<td>AR(1)–GARCH(1, 1)-Skew t</td>
<td>0.042†</td>
<td>−</td>
<td>−0.0008</td>
<td>0.099†</td>
<td>0.912†</td>
<td>5.498†</td>
<td>−</td>
<td>−5.4604</td>
</tr>
<tr>
<td>MA(1)–GARCH(1, 1)-Skew t</td>
<td>−</td>
<td>(2.163)</td>
<td>−0.0008</td>
<td>0.099†</td>
<td>0.912†</td>
<td>5.498†</td>
<td>−</td>
<td>−5.4604</td>
</tr>
<tr>
<td>ARMA(1, 1)–GARCH(1, 1)-Skew t</td>
<td>−0.291</td>
<td>0.317</td>
<td>−0.0008</td>
<td>0.099†</td>
<td>0.912†</td>
<td>5.498†</td>
<td>−</td>
<td>−5.4600</td>
</tr>
</tbody>
</table>

r-statistics are shown in brackets. † 5% confidence level.

and the worst model is W-GPD, according to RMSE criteria. Since RMSE does not represent tail loss, we should use more appropriate backtesting procedures to test out-of-sample forecasting performance of the VaR models. Christoffersen’s [38] unconditional coverage, independence and conditional coverage tests are applied to estimate tail-loss performance of the models.

The W-GPD model is statistically significant based on all three tests, where the other models are statistically significant mostly two tests. According to conditional coverage test, only the W-GPD model is statistically significant for both stock markets. Besides, the W-GPD model’s Christoffersen [38] log-likelihood statistics are the smallest ones for all three tests. The most powerful backtesting procedure in the Christoffersen [38] test is conditional coverage one for tail-loss forecasting. This shows that the W-GPD model has superior forecasting performance compared to the conditional volatility, extreme value theory, and conditional extreme value theory models for volatility forecasting.
5. Conclusion

Value-at-risk (VaR) and conditional volatility models have become common tools for financial forecasting. However, conditional volatility models cannot capture extreme movements, as these models are based on past volatility rather than the extreme observations. On the other hand, extreme value theory models can capture extreme movements and forecasting performance of these models better than conventional volatility models. In this paper, wavelets and EVT are combined for volatility forecasting and wavelet-based threshold estimation adds the business cycle feature to extreme value theory. The filter length is determined as 64 days ($2^6$) in wavelet-based GPD. Since threshold estimation is one of the main problems
in generalized Pareto distribution, this approach can also be considered an improvement in non-parametric statistics. This combined feature increases the predictive out-of-sample forecasting performance.

The main contribution of this paper is to propose a combined EVT model and compare the predictive performance of this model with conventional conditional volatility models. This hybrid model is tested on two major emerging stock markets: the Istanbul Stock Exchange (ISE) and Budapest Stock Exchange (BUX). These two markets are selected since global stock investors frequently take positions in both markets. The relative performance of wavelet-based EVT is benchmarked against the conditional volatility and extreme value theory models. The comparison of the performance of wavelet-based EVT to that of the traditional volatility models shows that the wavelet-based EVT model is the most appropriate one among other value-at-risk models according to the number of violations and Christoffersen [38] tail-loss tests. The number of violations in the wavelet-based GPD model dramatically decreases, and the model is statistically significant for all Christoffersen [38] tail-loss tests for both stock markets. The superior forecasting performance of the wavelet-based EVT model is also consistent with Basel II requirements. Therefore, financial institutions can also estimate market value-at-risk by using wavelet-based EVT models for their single assets. For future research, this model can also be extended to multivariate cases for portfolio value-at-risk estimation.
References