

Ranking of VaR and ES Models: Performance in Developed and Emerging Markets

Saša Žiković

Randall K. Filer*

March 2011

Abstract

We introduce a new methodology for ranking and comparing the performance of VaR and ES models. The proposed ranking scheme is based on a nonparametric ANOVA test. The analysis of the relative performance of VaR and ES models is performed using daily returns for sixteen stock market indices (eight from developed and eight from emerging markets) prior to and during the global financial crisis. Simulation results show that for a large number of different VaR models there is no statistical difference in their performance measured by Lopez size adjusted score. Statistically significant top performers are conditional EVT GARCH, models based on volatility updating (HW and FHS) and nonparametric mirrored historical simulation. ES backtesting results are similar to VaR results with the models being even more closely matched. The same models that were the top performers in VaR comparison also perform significantly better in ES estimation compared to other models.

Key words: Ranking of models, Value at Risk, Expected shortfall, Extreme value theory, Financial crisis

JEL classification: G24, C14, C22, C52, C53

* Saša Žiković is an Assistant Professor at the Faculty of Economics, University of Rijeka and Visiting Professor at Humboldt University in Berlin, Faculty of Administration Ljubljana and Faculty of Economics Split. Randall K. Filer is Professor of Economics at Hunter College and the Graduate Center of the City University of New York and Visiting Professor of Economics at CERGE-EI in Prague and ISET in Tbilisi. He is a research fellow at IZA, Bonn; CESifo, Munich; and the William Davidson Institute of the University of Michigan. This paper was partially completed while he was on leave at the Economics Institute, Zagreb. The authors thank Paul Embrechts, Kevin Dowd, Oleh Havrylyshyn, Paul Wachtel and Evan Kraft for insightful comments and suggestions on closely related ideas, as well as the Croatian National Bank and conference participants at the 13th and 14th Dubrovnik Economic Conference in Dubrovnik, Croatia (June, 2007, 2008) for their helpful comments. We are grateful to Dejan Divjak from KD Bank for providing us with the data.

1. Introduction

The years leading up to the recent financial market turbulence have been characterized by exceptionally high growth of the world economy accompanied by moderate inflation. This strong performance resulted in unusually high returns in financial markets, especially in emerging and Anglo-Saxon countries. Risk premia and volatilities were exceptionally low across a very wide spectrum of assets including bonds, stocks, foreign exchange, and derivatives in general. Perception of a low risk environment and further growth prospects were further fuelled by historically low interest rates, booming real-estate prices and inflating monetary aggregates. The high level of asset prices kept leverage ratios low, while the combination of strong income flows and historically low interest rates did the same with debt service ratios.

However in 2005 Alan Greenspan (2005) noted: "...history has not dealt kindly with the aftermath of protracted periods of low risk premiums." Indeed, historically risk premia and Value at Risk (VaR) measures tend to be at their lowest immediately prior to the outbreak of a crisis or a period of exceptionally high market volatility. In 2007 Knight (2007) warned: "We might be witnessing the proliferation of... 'option-like' payoff patterns in the financial system," whereby investors assumed positions that yielded modest but steady income streams in times of prosperity but which could result in large, discontinuous losses in times of crisis. This "pattern" can be attributed to the introduction of new instruments and patterns of behaviour that raised the risk of extreme events while giving a false impression of a low-risk environment. In hindsight, it is clear that these warnings should have been heeded. The non-linear payoffs during worsening market conditions, combined with the assumptions of normality and IID behaviour widely used in VaR models, wrecked havoc on financial

institutions, led to a massive need for government intervention in financial markets and created wide-spread doubts about VaR models in the eyes of regulators and investors.

Since its introduction VaR as a risk measure has been criticized theoretically, especially for the fact that these models do not account for the extent of losses that could be suffered beyond the specified threshold. In the eyes of investors and regulators, these extreme losses are precisely what a risk measure should flag. VaR is, however, inherently incapable of distinguishing between situations where losses in the tail are only a slightly worse than the threshold, and those where they are overwhelming. It provides only a lower bound for losses in the tail and thus has a bias toward optimism instead of the conservatism that is generally thought to be beneficial in risk management.

An alternative measure of risk that has been adopted from the insurance industry and quantifies losses that might be encountered in the tail is the Expected Shortfall (ES). While VaR represents a minimum loss one expects at a determined confidence level, ES is the expected value of that loss, provided that the loss is equal to or greater than the VaR. Artzner, et al. (1997, 1999) have shown, using an axiomatic approach to define a satisfactory or “coherent” risk measure, that VaR fails a coherency test because it does not universally exhibit sub-additivity, whereby the risk of a combined portfolio cannot be greater than the sum of the risks associated with any possible division of that portfolio. VaR can only be made sub-additive if the implausible assumption that returns are elliptically distributed is imposed. In this case, however, VaR and ES are equivalent and give exactly the same information (see Embrechts, et al., 1997).

Even though VaR measures have substantial theoretical flaws, they have been imposed on financial institutions as a regulatory obligation under Basel I, II and III rules. ES, on the other hand, although a coherent measure of risk, has not been approved by regulators to calculate capital requirements. It is actually quite surprising given that VaR and ES are

inherently connected in the sense that ES figures can be easily calculated from the VaR surface in the tail. Perhaps because of this lack of approval, ES has not been as extensively studied as VaR in empirical research. Estimation techniques that have been developed for VaR measures in the past decades can easily be employed to yield superior ES forecasts. This means that advances in VaR estimation need not be lost with the adoption of coherent risk measures into regulatory framework. The inherent connection between VaR and ES is extremely helpful for financial institutions, since all the building blocks required for VaR estimation (databases, risk drivers, calculation routines, etc.) are also needed for estimation of ES. Thus, if an institution already has the capacity to calculate VaR, it needs only small adjustments to produce estimates of coherent risk measure, such as ES. Such a measure should be valuable for internal purposes even before it is required by regulators.

As opposed to purely VaR oriented literature the empirical literature that compares VaR and ES has been limited in both emerging and developed markets. Gencay, Selcuk, and Ulugulyagci (2003) and Gencay and Selcuk (2004) analyzed the performance of unconditional Extreme Value Theory (EVT) models against variance-covariance and historical simulation models in nine emerging countries. They found that an unconditional EVT model outperformed classical VaR models at extreme confidence levels. Maghyereh and Al-Zoubi (2006) investigated the relative performance of popular VaR models against an unconditional EVT methodology for seven Middle Eastern and North African countries. Again EVT models outperformed classical variance-covariance and historical simulation models in most cases. Similar results were reported by Mendes (2000) for Latin American countries. Cotter (2007 and 2004) tested a parametric EVT and Gaussian estimates of VaR and ES in six Asian markets during the Asian crisis and five equity indexes from European markets. He found that EVT estimates are superior under both VaR and ES risk measures looking at both the Kupiec and Christoffersen criterions, although it was hard to reach any

conclusion regarding the significance of these differences. Nyströmand and Skoglund (2002) tested the performance of VaR models on a wide range of assets in developed countries and found that for quantiles higher than the 98 percentile the use of unconditional EVT models made a substantial predictive contribution and that the generalized Pareto distribution more accurately modelled the empirically observed tails than the normal distribution. In contrast to these findings, however, Silva and Mendes (2003) found that the performance of an unconditional EVT model is not satisfactory in meeting Basel II criteria in Asian stock markets.

To remedy the problems of the unconditional estimation that is traditional in EVT, McNeil and Frey (2000) developed a conditional EVT approach to both VaR and ES estimation and showed empirically that the traditional parametric VaR models with normal density fail to accurately estimate losses during financial crises. They, along with many others (see Acerbi et al. 2001, Yamai and Yoshihara, 2002 and Inui and Kijima, 2005), advocated the use of ES as an alternative risk measure with good theoretical properties. Overall, the literature strongly suggests that although ES provides superior risk measures to VaR, these have not been as exhaustively studied as VaR measures.

Apart from these well known “technical” problems there is also a usually overlooked systemic problem with risk management. This problem is risk model comparison and ranking. We must recall that when evaluating and backtesting VaR/ES figures we are looking at only a single realization of the underlying data generation process. Consequently our judgement on the performance of a particular risk models is based only on the performance of the model with regards to only that single realization. VaR model comparison literature is truly vast but it rarely tangles the question of whether there is truly any significant statistical difference in the performance of different models. Is it only a game of chance that in a particular market

and in particular time period a certain model performs better than some other or does a certain model consistently and statistically significantly outperform some other model?

In this paper we make two contributions; first, develop a new methodology to VaR and ES model comparison which allows us to rank competing VaR/ES models. Second, we provide an empirical investigation and tail risk assessment of a wide array of VaR and ES models in both developed and emerging countries prior to and during the global financial crisis. Some of these issues have been tackled in previous research, but our analysis presents some further insights. The topics covered in this article are relevant to risk management of commercial banks, insurance companies, and investment banks around the world, which have to assess their portfolios risk periodically.

The following VaR models are analyzed in the paper:¹

- (a) Normal simple moving average (VCM) method,
- (b) RiskMetrics system,
- (c) Historical simulation,
- (d) Mirrored historical simulation²,
- (e) Kernel historical approach³,

¹ For a good overview of a wide range of VaR and ES models see, for example, Dowd (2005).

² Mirrored historical simulation (MHS) is a simple extension of historical simulation using “mirror” scenarios suggested by Holton (1998). This technique is rarely found in the academic literature although it enables the user to double the number of scenarios while reducing convergence error by a factor of $1/\sqrt{2}$. The implementation of the technique is straightforward and involves multiplying the historical return series by -1 and adding these mirror scenarios to the existing set of observations. Mirror scenarios can also be used to reduce error relating to non-stationary markets. For example, rather than using 1,000 historical daily returns, where a fourth will be more than three years old it might be more useful and reflective of the current market situation to use a set of 500 daily observations but increase the number of losses by using the mirror technique.

³ A kernel approach places mini-density functions around each data point, with the kernel itself being the sum of these “mini-densities” and has a total area underneath it of 1. The kernel estimator can be pictured as placing “bumps” around each of the recorded return observations. The shape of these bumps is determined by the kernel function $K(x)$ and the bandwidth h determines their width. As the sample size grows, the net sum of all the

- (f) BRW (time weighted) simulation with decay factors of 0.97 and 0.99,
- (g) GARCH model,
- (h) Filtered Historical simulation (FHS) method - Barone-Adesi et. al. (1999).
- (i) Unconditional EVT approach using Generalized Pareto distribution,
- (j) Conditional McNeil and Frey (2000) EVT approach and

The ES models analyzed in the paper are:

- (a) VCV with GPD
- (b) RiskMetrics with GPD
- (c) GARCH with GPD
- (d) Bootstrapped historical simulation,
- (e) Bootstrapped “mirrored” historical simulation,
- (f) Bootstrapped kernel historical approach,
- (g) Bootstrapped BRW simulation
- (h) FHS-ES approach
- (i) McNeil and Frey (2000) EVT approach,
- (j) Unconditional EVT approach and

2. Value at Risk and Expected Shortfall

VaR is usually defined as:

smoothed points approaches the true probability density function, whatever that may be, irrespective of the method of smoothing the data. This is because the influence of each point becomes arbitrarily small as the sample size grows, so the choice of kernel imposes no restrictions on the results asymptotically. In a small sample there may be differences, which can be examined by using different kernels. Butler and Schachter (1998) first introduced kernel estimation to VaR calculation. Similarly to Silverman (1986) they found that the best fits to the financial data were given by the Epanechnikov and adaptive Gaussian kernels. For this reason in our analysis we use the Epanechnikov kernel smoothing in our kernel historical approach.

“the maximum potential loss that a portfolio can suffer within a fixed confidence level (cl) during a holding period.”

Let $(X_t, t \in Z)$ be a strictly stationary time series representing daily observations of the negative log return for a financial asset. The dynamics of X are given by:

$$X_t = \mu_t + \sigma_t Z_t \quad (1)$$

where the innovations Z are IID with zero mean, unit variance and marginal distribution function $F_Z(z)$. It is typical to assume that μ_t and σ_t are measurable with respect to ψ_{t-1} (the information set up to time $t-1$) and that $F_X(x)$ denotes the marginal distribution of (X_t) . For a horizon hp , $F_{X_{t+1}+\dots+X_{t+hp}|\psi_t}(x)$ denotes the predictive distribution of the return over the next hp days, given the information set up to and including day t . From a tail event perspective, for a given confidence level cl ($0 < cl < 1$), the unconditional $VaR_{cl}(X)$ is a quantile of the marginal distribution denoted by:

$$VaR_{cl}(X) = \inf\{x \in R : F_X(x) \geq cl\} \quad (2)$$

while the conditional $VaR_{cl}(X)$ is a quantile of the predictive distribution for the return over the next hp days denoted by:

$$VaR_{cl, hp}^t(X) = \inf\{x \in R : F_{X_{t+1}+\dots+X_{t+hp}|\psi_t}(x) \geq cl\}. \quad (3)$$

This definition can sometimes be misleading because VaR does not actually represent maximum losses since, as we have seen, a portfolio can lose much more than suggested by VaR depending on the shape of the tail of the distribution. A more insightful definition of VaR, based on equation (2), is:

“VaR is the minimum potential loss that a portfolio can suffer in the $100(1-cl)\%$ worst cases during a holding period,”

or

“VaR is the maximum potential loss that a portfolio can suffer in the 100(cl)% best cases during a holding period.”

VaR can be thought of as “the best possible outcome among a set of the worst case scenarios” and, therefore, systematically underestimates the potential losses associated with any specific confidence level. Both VaR and ES contain implicit assumptions regarding agents’ risk aversion. If a user has a ‘well-behaved’ risk-aversion function, then the weights will rise smoothly, and the more risk-averse the user, the more rapidly the weights will rise. Given that VaR explicitly weights all losses greater than that at the confidence level as zero it actually assumes that agents are risk-loving (i.e., have negative risk-aversion) in the tail region. ES, in contrast, is characterized by all losses in the tail region (i.e., the 100(1-cl)% largest losses) having an identical weight. This implies that the investor is risk-neutral in the tail region. Both assumptions seem highly unlikely in real life.

Following equation (2), the unconditional ES is defined as:

$$ES_{cl}(X) = E[X | X > VaR_{cl}(X)] = -cl^{-1} \int_{-\infty}^{VaR} xf(x)dx \quad (4)$$

while the conditional ES can be expressed as:

$$ES'_{cl, hp}(X) = E \left[\sum_{j=1}^{hp} X_{t+j} \mid \sum_{j=1}^{hp} X_{t+j} > VaR'_{cl, hp}(X), \psi_t \right]. \quad (5)$$

ES is very appealing as a risk measure because it sums all values of x , weighted by $f(x)$, from minus infinity to VaR threshold, thus taking into account the magnitude of potential losses beyond VaR threshold. ES has been referred to in the literature under many names including Expected tail loss (ETL), Conditional VaR (CVaR), tail VaR, tail conditional expectation, and mean excess loss. ES has been used by insurance practitioners, especially casualty insurers for a long time as conditional average claim size. For continuous loss distributions, the ES at a given confidence level is the expected loss given that the loss is

greater or equal to the VaR at that level. For distributions with possible discontinuities it has a more subtle definition and can differ depending on whether the loss is strictly greater to the VaR ($CVaR^+$) or is greater than or equal to the VaR ($CVaR^-$). $CVaR^+$ is also known as “mean shortfall”, although the seemingly identical term “expected shortfall” has been interpreted by Acerbi, et al. (2001) as a synonym for $CVaR$ itself. $CVaR^-$ is also known as “tail VaR” (Artzner, et al. 1999).

Although, as discussed above, ES ($CVaR$) is a coherent measure of risk, it still has its own problems. Yamai and Yoshida (2002) find that even ES, although better at forecasting the true level of risk, it is not reliable during periods of market turmoil and can also give overly optimistic results. Kondor and Varga-Haszonits (2008) find that whenever there is an asset in a portfolio that dominates, with regards to risk and reward, over others in a given sample, the portfolio’s return cannot be maximized under any coherent measure on that sample, including ES. In periods of high volatility and/or extreme price spikes, classical, widely used VaR models, prove to be overly liberal and optimistic – a definite problem in risk management.

One possible avenue for improving risk model’s estimates lies in extreme value theory (EVT), which specifically models the extreme price changes (i.e., the tails of the return distribution). Focusing on extreme returns rather than the entire distribution seems natural since, by definition, risk management is concerned with measuring the economic impact of rare events.

EVT provides a framework for analyzing extreme (rare) events using historical data. By definition, extreme events are rare, meaning that their estimates are often required for levels of a process that are greater than those in the available data set. EVT is based on the Extreme Value Theorem, a relative of the widely used Central Limit Theorem. Suppose we

have a set of observed returns drawn from an unknown distribution. The EVT says that as the sample size increases, in the limit, the distribution of extreme returns converges to:

$$G_{\xi, \sigma, \mu}(x) = \begin{cases} 1 - \left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ 1 - e^{-(x - \mu) / \sigma} & \text{if } \xi = 0 \end{cases} \quad x \in \begin{cases} [\mu, \infty] & \text{if } \xi \geq 0 \\ [\mu, \mu - \sigma / \xi] & \text{if } \xi < 0 \end{cases} \quad (6)$$

where, μ is the distribution mean, σ is the dispersion of the distribution and ξ indicates the heaviness of the tails.

When $\mu = 0$ and $\sigma = 1$, the representation is known as the standard Generalized Pareto distribution (GPD). The GPD embeds a number of other distributions. For the analysis of financial time series the most relevant is the heavy-tailed Fréchet distribution in which case the tail index, $\xi > 0$.

It is important to be aware of the limitations implied by the EVT paradigm. EVT models are developed using asymptotic arguments, which can create difficulties when applied to finite samples. In order to estimate the tails of the loss distribution we use the result from asymptotic theory that for a sufficiently high threshold u , $F_u(y) \approx G_{\xi, \beta(u)}(y)$. An approximation of $F(x)$, for $X > u$, can be obtained as:

$$F(x) = [1 - F(u)]G_{\xi, \sigma, u}(x - u) + F(u) . \quad (7)$$

An estimate of $F(u)$ can be obtained also non-parametrically by means of the empirical cdf:

$$\hat{F}(u) = (n - k) / n \quad (8)$$

where k represents the number of observations exceeding the threshold u and n the number of observations. By substituting equation (7) into equation (8), the following estimate for $F(x)$ is obtained:

$$\hat{F}(x) = 1 - \frac{k}{n} \left(1 + \hat{\xi} \frac{x-u}{\hat{\sigma}} \right)^{-\frac{1}{\hat{\xi}}} \quad \text{given that } G_{\xi, \sigma, u}(x) = 1 - \left(1 + \xi \frac{x-u}{\sigma} \right)^{-\frac{1}{\xi}} \quad (9)$$

where $\hat{\xi}$ and $\hat{\sigma}$ are the maximum likelihood estimates of ξ and σ . This equation can be inverted to obtain a quantile of the underlying distribution, which is actually the VaR. For $cl \geq F(u)$ VaR is calculated as:

$$VaR_{cl} = q_{cl}(F) = u + \frac{\sigma}{\xi} \left(\left(\frac{1-cl}{\bar{F}(u)} \right)^{-\xi} - 1 \right) = u + \frac{\sigma}{\xi} \left(\left(\frac{1-cl}{k/n} \right)^{-\xi} - 1 \right) \quad (10)$$

Assuming that $\xi < 1$, ES is calculated as:

$$ES_{cl} = \frac{1}{1-cl} \int_{cl}^1 q_x(F) dx = \frac{VaR_{cl}}{1-\xi} + \frac{\sigma - \xi u}{1-\xi}. \quad (11)$$

The estimation of return distributions of financial time series using the EVT has been studied by McNeil, 1997; Embrechts, Resnick and Samorodnitsky, 1997; Danielsson and de Vries, 1997; and Danielsson, Hartmann and de Vries, 1998, among others. In all these papers, however, the focus has been on estimating an unconditional (stationary) distribution of asset returns. None of the unconditional EVT-based methods for quantile estimation yields estimates that are easily updated to reflect the recent volatility. Given the conditional heteroskedasticity of most financial data, McNeil and Frey (2000) developed a conditional EVT approach combining GARCH volatility forecasting with EVT tail estimation, which in empirical testing provides very good conditional and unconditional risk coverage.

EVT models are also plagued by problems in the estimation of tail index (see, for example, Diebold, Schuermann, and Stroughair (2000)). Although a number of methods have been proposed for estimation of tail indices, none provide robust results when analyzed over changing sample periods or with the inclusion or omission of extreme values (outliers). Parametric ES estimates, even those based on the GPD distribution, are highly sensitive to functional form misspecification. Simpler parametric models cannot adequately adapt to

sudden changes in volatility levels. Nonparametric ES models such as calculating the ES from historical data regarding tail losses are, by definition, unresponsive to shifts in market regimes and the occurrence of extreme events. An overview of models analyzed in the empirical part of the paper is given in Table 1.

3. Methodology for comparing and ranking VaR and ES models

In the literature there are a number of models that test the hypothesis whether a certain model is better than some other model, such as Diebold and Mariano (1995) equal predictive ability (EPA), White (2000) reality check (RC) and Hansen (2005) superior predictive ability (SPA). The question of interest in all of these models is whether any alternative forecast is better than the benchmark forecast, or equivalently, whether the best alternative forecasting model is better than the benchmark. This question can be addressed by testing the null hypothesis that the benchmark is not inferior to any alternative forecast. For a more complete discussion on this issue, see Sullivan, Timmermann, and White (2003) and references therein. Using such tests is useful for a forecaster who wants to explore whether a better forecasting model is available, compared to the model currently being used to make predictions. After a search over several alternative models, the relevant question is whether an observed excess performance by an alternative model is significant or not. Tests for equal predictive ability (EPA), in a general setting, were proposed by Diebold and Mariano (1995) and West (1996), where the framework of the latter can accommodate the situation where forecasts involve estimated parameters. A test for comparing multiple nested models was given by Harvey and Newbold (2000) and McCracken (2000) derived results for the case with estimated parameters and non-differentiable loss functions, such as the mean absolute deviation loss function. West and McCracken (1998) developed regression-based tests and other extensions

were made by Harvey, Leybourne, and Newbold (1998), West (2001), and Clark and McCracken (2001) who considered tests for forecast encompassing.

There is an inherent problem with comparing and ranking competing VaR and ES models since we are usually measuring only a single realization of the underlying data generation process. The question is whether there is any significant statistical difference in the performance of different models. Although at first it might seem that the difference between models is obvious to the plain sight we are often faced with situations where one model is preferred for one market or security and then after a few months some other. Is it only a game of chance that in a particular market and in particular time period a certain model performs better than some other? It all comes down to a question whether something that we subjectively perceive as different is actually statistically different. We propose a simple nonparametric approach to statistically sound comparison between competing risk models which allows us to rank different VaR and ES models depending on their performance under the metric of our choice.

The proposed ranking procedure for competing VaR/ES models is performed according to Lopez score for VaR models and modified Blanco-Ihle error statistic for ES models. Although we use these measures they can be easily replaced by any other measures depending on the users concerns and interests. The proposed ranking procedure consists of:

- 1) Fitting an ARMA-GARCH model to the time series in order to obtain IID observations.
- 2) Estimating the empirical CDF of each time series (applying it to the non-tail regions of distribution) with a Gaussian kernel. This smoothes the CDF estimates, eliminating the staircase pattern of unsmoothed sample CDFs. Finding the upper and the lower thresholds such that $x\%$ of the residuals is reserved for each tail. Fitting the amount by which those extreme residuals in each tail fall beyond the associated threshold to a parametric GPD.

- 3) Generating N simulated paths for the residuals from the obtained semi-parametric distribution (each path is T observations long)
- 4) Adding the ARMA-GARCH model to the residuals to obtain $N \times T$ simulated time series returns
- 5) Calculating VaR/ES for each of the $N \times T$ simulated returns for each VaR/ES model
- 6) Calculating N Lopez/modified Blanco-Ihle scores for each of the N VaR/ES - simulated return pairs, for each VaR/ES model
- 7) Comparing if the mean values of Lopez/modified Blanco-Ihle scores for different VaR/ES models are significantly different between each other. For this purpose one-way ANOVA approach is employed. The purpose of one-way ANOVA is to find out whether data from several groups have a common mean. The p-value returned by ANOVA depends on assumptions about the random disturbances in the model equation. For the p-value to be correct, these disturbances need to be independent, normally distributed, and have constant variance.
- 8) Checking for autoregression, heteroskedasticity and normality in Lopez/modified Blanco-Ihle scores. If the data is not normally distributed it needs to be transformed to uniform variates by empirical CDF and then to normal via inverse CDF after which one-way ANOVA can be calculated. Critical values used for the multiple comparisons are based on Tukey-Kramer honestly significant difference criterion since it is optimal for balanced one-way ANOVA. An alternative is to use a non-parametric Kruskal-Wallis test (a nonparametric version of one-way ANOVA) which makes only mild assumptions about the data, and is appropriate when the distribution of the data is non-normal. The assumption behind this test is that the measurements come from a continuous distribution, but not necessarily a normal distribution. The test is based on an analysis of variance using the ranks of the data values, not the data values themselves.

We have to point out an obvious limitation of this approach and that is the assumption that our description of the central mass and the tails of the process distribution are adequate i.e. that the underlying process is well described by the visible realization.

4. Data and backtesting methodology

We analyze the performance of various VaR and ES models using the log of daily returns of eight equity indices from developed markets (US - Dow Jones Industrial (DJIN), Nasdaq, S&P 500, Russell 2000 (RTY); Japan – Nikkei; Germany – DAX; France – CAC; and UK - FTSE) and eight emerging markets (Brazil – Bovespa; Russia - CRTX; India – Sensex; South Africa – Jalsh; Malaysia – KLCI; Mexico – Mexbol; Hong Kong - Heng Seng; and Taiwan - Taipei). Returns were collected from the Bloomberg website for the period January 1, 2000 through July 1, 2010. In order to differentiate between “normal” and stressed market conditions we choose two backtesting periods consisting of 750 observations each. The period between June/July 2004 and June/July 2007 forms the pre-crisis backtesting period, and the period between June/July 2007 and July 2010 forms the crisis backtesting period. VaR and ES figures were calculated for a one-day ahead horizon and 99 percent confidence level. Based using the proposed ranking procedure the VaR models are tested using: Kupiec test, Christoffersen Unconditional Coverage (UC), Conditional Coverage (CC) and Independence (IND) test, and Lopez and Blanco-Ihle tests as well as root mean squared error (RMSE) and mean average percentage error (MAPE) statistics. The Christoffersen UC test is problematic because it gives a distorted image of VaR models’ performance. Since it is chi-square distributed with one degree of freedom, deviations from the test’s expected value that occur on the conservative side (i.e. with number of exceedances lower than their expected value) are penalized more severely. This characteristic is not compatible with risk-averse or

risk-neutral assumptions. Thus, from the regulatory standpoint, the Kupiec binomial test is preferable to the Christoffersen UC test because it is more desirable to have positive than negative deviations. The same logic extends to Christoffersen conditional coverage (CC) test, which should also be treated sceptically since it automatically disadvantages VaR models that err on the conservative side.

Blanco and Ihle (1998) suggested evaluating forecasts according to a loss function equal to:

$$C_t = \begin{cases} \frac{L_t - VaR_t}{VaR_t} & \text{if } L_t > VaR_t \\ 0 & \text{if } L_t \leq VaR_t \end{cases} \quad (22)$$

This loss function allows for the sizes of tail losses to influence the rankings of VaR models. Models that generate higher tail losses would generate higher values under this size-adjusted loss function than models that generate lower tail losses, *ceteris paribus*. The problem with the Blanco-Ihle loss function is that it compares the calculated VaR with tail losses, which does not make sense since VaR forecasts only the least possible tail losses. Since VaR does not contain any information about the size of the expected tail loss, the Blanco-Ihle loss function only measures the discrepancy between the lowest possible tail loss and actual tail losses. Blanco-Ihle loss function can easily be modified to compare ES with the actual value of the tail loss, a more meaningful comparison. The modified function equals:

$$C_t = \begin{cases} \frac{|L_t - ES_t|}{ES_t} & \text{if } L_t > VaR_t \\ 0 & \text{if } L_t \leq VaR_t \end{cases} \quad (23)$$

In order to select superior ES models, each model will be graded by four symmetrical error statistics: the mean absolute error (MAE), two versions of the root mean squared error (RMSE), and the proposed ES modification of the Blanco-Ihle loss function. Among these error statistics, ES modification of the Blanco-Ihle loss function is probably the most

informative, since it compares the tail loss to ES while taking into account the relative size of the tail loss compared to the difference between the two. In our two-stage backtesting procedure, the best performing VaR/ES model must first satisfy both the Kupiec and Christoffersen independence (IND) tests and then provide superior tail loss forecasts, in the sense of minimizing the error statistics.

5. Backtesting results and findings

To secure the same out-of-the-sample backtesting period for all of the examined stock indices, the out-of-the-sample data sets are formed by removing the 1.500 most recent observations from each stock index and forming two sub-periods of 750 days. The first sub-period from June/July 2004 to June/July 2007 represents the pre-crisis period. The second sub-period from June/July 2007 to July 2010 represents the crisis period. The remaining observations are used to calculate the VaR and ES starting values and calibrate volatility. The length of the tail-loss data set used for backtesting depends on the number of errors generated by each VaR model. The quality of ES forecasts depends on both the ES estimation model and the quality of the VaR forecast. This dependence can be easily seen from the simple fact that a loss that might fall in the extreme range under one VaR model and, as such, be included in the ES forecast might not exceed another, more conservative, VaR measure.

Data from all the analyzed stock indices shows leptokurtosis, asymmetry and significant heteroskedasticity, with autoregression being especially pronounced in the emerging markets. Based on the Akaike and Bayesian information criterion asymmetric EGARCH representation of volatility with GED and Student's t distribution was used to capture the dynamics of data-generating processes. The asymmetry parameter in EGARCH

model was significantly different from zero for most of the indexes.⁴ The asymmetry parameter, which controls the asymmetric impact of positive and negative shocks on conditional variance, indicates significantly higher conditional volatility after negative shocks.

Estimation of the tail index parameter is crucial in applying EVT models, which are directly linked to threshold value u which defines the level above which returns are considered extreme. The threshold value for each index was determined by comparing the Hill estimator with the mean excess plot and the quantile-quantile (QQ) plot (Danielsson and de Vries, 1997). The same procedure of estimating the threshold value was also performed on IID innovations required for the implementation of the McNeil and Frey (2000) EVT-GARCH model. Maximum likelihood estimates (MLE) of the shape (tail index) and scale (sigma) parameters for the GPD, during the pre-crisis and crisis period, for the analyzed stock indices' threshold losses (losses surpassing the threshold value set by Hill estimator), and threshold innovations are presented in Table 2. The mean excess and QQ plots, Hill estimator and MLE all show that tail indexes for both developed and emerging countries are greater than zero, implying empirically fat tails and that the GPD belongs to the Fréchet and Gumbel domains of attraction. This suggests that the normal distribution is inappropriate for modelling tail returns. In the pre-crisis period the tail indexes vary between -0.195 (FTSE) and 0.129 (NIKKEI) for the developed markets and between -0.074 (MEXBOL) and 0.173 (HENG SENG) for the emerging markets. During the crisis most of the tail indexes changed significantly and ranged between -0.042 (NASDAQ) and 0.18 (S&P500) for the developed markets and between -0.099 (MEXBOL) and 0.251 (JALSH) for the emerging markets. The greatest changes in the size of the tail index between the two periods were recorded for FTSE

⁴ For the BOVESPA, CRTX, JALSH, KLCI and HENG SENG indices the asymmetric impact is not significantly different from zero. Results are available from the authors on request.

(0.253) and CRTX (0.214). The distribution of tail losses for the stock indices in South Africa, Hong Kong and Russia may not have a finite fourth moment, since the estimated tail index is around 0.25⁵. The tails of the innovations from the analysed time series are similar to the values of returns. The very high tail index values in majority of indexes are an indication of severe crashes experienced in these markets.

In the pre crisis period, in both developed and emerging markets, good performance is recorded for nonparametric models (HS, MHS and KHS) as well as FHS and extreme value based approaches. Very weak performance is recorded for BRW, VCV, RiskMetrics and Hull-White models. The performance of tested VaR models is significantly different during the tested crisis period. In developing markets good performance is recorded for Hull-White, GPD and EVT-GARCH model. In table 3b we make a distinction between the standard EVT GARCH model and EVT GARCH (L) model, where L stands for longer time series. We introduce two EVT GARCH models and this notation because we find a very interesting behaviour – if we use the standard rolling window for GARCH parameters (calculated just in the crisis period) and from that calculate standardized innovations EVT GARCH forecasts are very poor. This is due to the very thin tail indicated by fitting GPD. If instead we use a longer period of for example 6 years results are much better since the GPD tail is closer to what we might expect. At first sight this finding indicates a counter logical behaviour (if we are using newer information we should have better volatility forecasts ergo better risk measures). It comes down to a bit of a paradox when using EVT GARCH approach; if we are using a very accurate volatility forecast reflecting the current environment there is a lack of outliers and

⁵ For $\xi > 0$, $E[X^k]$ is infinite for $k > 1/\xi$. The number of finite moments is ascertained by the value of ξ : if $0.25 \leq \xi \leq 0.5$ the second and higher moments are infinite; if $\xi \leq 0.25$, the fourth and higher moments are infinite.

standardized innovations are bunched around zero with their tails exponentially decreasing. When calculating GPD parameters from such a innovation series it is quite logical to obtain even negative estimates of the tail parameter. The paradox lies in the fact that if the volatility estimate is fairly good but not perfect there will be outliers in the standardized innovations series which will lead to higher GPD tail index and thus actually increase the accuracy of EVT GARCH VaR forecasts. In emerging markets only the EVT GARCH (L) and MHS 250 models performed satisfactory, again with the clear distinction between standard and prolonged EVT GARCH models.

Overall we find good performance across both developed and emerging markets for extreme value based approaches. Mirrored historical simulation, a simple extension of the historical simulation, yielded surprisingly good risk coverage and satisfied the backtesting criteria for a great majority of stock indices tested. Backtest results also show that kernel historical approach VaR estimator, although inferior to mirrored historical simulation, delivers significant variance and mean square error reductions when compared to plain historical simulation. This difference is similar to that found by Chen and Tang (2005).

It is useful to analyze the averages of VaR forecasts for the models that satisfy the Basel II/III-required Kupiec test as well as the Christoffersen independence criterion. Rankings according to the minimum average VaR value (provided the Basel II/III criteria and Christoffersen independence test at a 5 percent significance level are satisfied) are presented in Table 4. For all of the indices in both developed and emerging markets, GPD and HW models provide the highest VaR estimates, with HW providing very high values during the crisis period. This characteristic makes them the most conservative but also the most expensive in terms of capital requirements for financial institutions. During the precrisis period for the developed markets, the GARCH model yielded the lowest average VaR four

out of eight times (Nikkei, DAX, CAC and FTSE index) followed by BRW simulation with a decay factor of 0.99, which was the best performer in two cases (DJIN and Nasdaq index). For emerging markets, the performance of GARCH model is even better, yielding the lowest average VaR for five out of eight indices (BOVE, Mexbol, KLCI, SENSEX and Heng Seng index). During the crisis period for the developed countries EVT GARCH (L) was the top performer (S&P 500, DJI, NASDAQ and FTSE index), followed by FHS model (RTY, Nikkei, DAX). For emerging markets, both GARCH and FHS models were the best performers for three indexes. In summary, among VaR models that satisfy the Basel criteria, the FHS and GARCH models provided the lowest average VaR in most cases, making them the models with the lowest opportunity cost of holding idle capital.

To backtest the various ES models, we ranked the models by their ability to yield minimal loss functions, i.e. the minimum departure from the reported tail loss values. Rankings of the ES models according to modified Blanco-Ihle error statistics at the 99 percent confidence level are presented in Table 6. According to the modified Blanco-Ihle statistic, in the pre-crisis period, both in the developed and emerging markets, bootstrapped MHS model was the best performing ES model, yielding the lowest statistic. In the developing markets following closely was the basic bootstrap historical simulation. In the emerging markets bootstrapped FHS and GPD model were ranked as second and third performers. The worst performers across all the markets were the VCV, RiskMetrics and GARCH models with GPD distribution. During the crisis period, both in the developed and emerging markets bootstrapped FHS, MHS and EVT GARCH were the best performing ES models. The worst performing models were again the VCV, RiskMetrics and GARCH models with GPD distribution.

In summary, according to the overall backtesting results, bootstrapped mirrored historical simulation is the superior ES measure. We find no benefit to using a kernel

approach instead of bootstrapped historical simulation. This finding is similar to that reported in Chen (2008) for plain historical simulation. The underlying reason that there is no benefit from kernel smoothing of ES estimates lies in the fact that the unconditional ES is a mean parameter, which can be estimated accurately by simple averaging and therefore does not call for additional data smoothing. It is also interesting to note that, although historical simulation models are clearly inferior to EVT models in VaR estimation, in ES estimation bootstrapping historical exceedances over VaR performs better than theoretically well-founded EVT models.

To take a step further from the standard VaR/ES performance reporting we apply the methodology presented in section 3 to test whether there is any statistically significant difference in the performance of tested VaR and ES models. The data is simulated based on the distribution of returns in the crisis period. For each index 2.000 simulations were performed with length of each simulated index being 1.000 data points. Analysing the VaR model performance on simulated data, presented in table 7a, discovers a number of interesting features. When looking at individual indexes, for a large number of different models there is no statistical difference in their performance measured by Lopez size adjusted score. When looking at overall performance on the developed markets the best performing VaR model that is statistically different from other tested models is conditional EVT GARCH model, followed by the unconditional GPD model. Even in the summary results across eight developed markets there is no statistical difference between FHS and BRW ($\lambda=0.99$) model. In the emerging markets overall the best model is MHS 250, followed by EVT GARCH model. Similarly to the developed markets there is again an overlapping between FHS and MHS 500 model. The statistically worst performing VaR models across both markets are the simplest models: Normal VCV, plain historical simulation and RiskMetrics model. Overall,

statistically significant top performers are conditional EVT GARCH, models based on volatility updating (HW and FHS) and nonparametric mirrored historical simulation. Since our metric of choice is the size adjusted Lopez score these models provide the closest fit to the actual level of risk encountered in the analysed markets.

ES backtesting results are similar to VaR results with the models being even more closely matched. A noticeable difference from VaR results is the mirrored historical simulation model being in the same rank or even superior to conditional EVT GARCH model. In the developed markets the best ES models were MHS 250 and FHS followed by conditional and unconditional extreme value based models. In the emerging markets EVT GARCH is the best model followed by both MHS models and FHS. Again the same models that were the top performers in VaR comparison perform significantly better than other tested models. We find no benefit to using a kernel approach instead of bootstrapped historical simulation. This finding is similar to that reported in Chen (2008) for plain historical simulation. The underlying reason that there is no benefit from kernel smoothing of ES estimates lies in the fact that the unconditional ES is a mean parameter, which can be estimated accurately by simple averaging and therefore does not call for additional data smoothing. It is also interesting to note that, although historical simulation models are clearly inferior to EVT models in VaR estimation, in ES estimation bootstrapping historical exceedances over VaR often perform better than theoretically better-founded EVT models. After performing our simulation based test we conclude that there is sufficient statistically significant evidence that models like conditional and unconditional extreme value approach, FHS and mirrored historical simulation perform superiorly to other analysed model both in VaR and ES estimation.

6. Conclusion

When evaluating and backtesting VaR/ES figures we are looking at only a single realization of the underlying data generation process. Consequently our judgement on the performance of a particular risk models is based only on the performance of the model with regards to only that single realization. Is it only a game of chance that in a particular market and in particular time period a certain model performs better than some other or does a certain model consistently and statistically significantly outperform some other model? Although it might seem that the difference between models is obvious to the plain sight we are often faced with situations where one model is preferred for one market or security and than after a few months some other. It all comes down to a question whether something that we subjectively perceive as different is actually statistically different. We develop a simple nonparametric approach to statistically sound comparison between competing risk models which allows us to rank different VaR and ES models depending on their performance under the metric of choice.

From our simulation results it is visible that for a large number of different VaR models there is no statistical difference in their performance measured by Lopez size adjusted score. Overall, statistically significant top performers are conditional EVT GARCH, models based on volatility updating (HW and FHS) and nonparametric mirrored historical simulation. Since our metric of choice is the size adjusted Lopez score these models provide the closest fit to the actual level of risk encountered in the analysed markets. ES backtesting results are similar to VaR results with the models being even more closely matched. A noticeable difference from VaR results is the mirrored historical simulation model being in the same rank or even superior to conditional EVT GARCH model. The results show that the same models that were the top performers in VaR comparison perform significantly better than

other tested models. Purely parametric models, such as the VCV, RiskMetrics and GARCH models, were the worst performers in both VaR and ES comparison. The wide use of such models may, indeed, have played a role in the recent under-appreciation of risks in financial markets. This points to a conclusion that the strengths and weaknesses of every model are preserved in both VaR and ES estimation. It is also interesting to note that although historical simulation based models are clearly inferior to EVT models in VaR estimation, in ES estimation bootstrapping historical exceedances over VaR is often superior to theoretically well-founded EVT models. This finding suggests that during the period we analyzed there were a large number of extreme events that were so far in the tail that they were left out by VaR measures but correctly picked up in ES estimation. After performing our simulation based test we conclude that there is sufficient statistically significant evidence that models like conditional and unconditional extreme value approach, FHS and mirrored historical simulation perform superiorly to other analysed model both in VaR and ES estimation.

References:

- 1) Acerbi C., Nardio C., Sirtori C. (2001), “Expected Shortfall as a Tool for Financial Risk Management”, Working Paper, <http://www.gloriamundi.org/var/wps.html>
- 2) Angelidis T., Degiannakis S. (2007), “Backtesting VaR Models: An Expected Shortfall Approach”, University of Crete, Department of Economics, Working Papers No. 0701, <http://ideas.repec.org/p/crt/wpaper/0701.html>
- 3) Artzner P., Delbaen F., Eber J.M., Heath D. (1997), “Thinking coherently”, *Risk*, Vol. 10, No. 11, p. 68–71.
- 4) Artzner P., Delbaen F., Eber J.M., Heath D. (1999), “Coherent measures of risk”, *Mathematical Finance*, Vol. 9, No. 3, p. 203–228.
- 5) Barone-Adesi G., Giannopoulos K., Vosper L. (1999), “VaR without Correlations for Portfolios of Derivative Securities”, *The Journal of Futures Markets*, Vol. 19, No. 5, p. 583-602.
- 6) Blanco C., Ihle G. (1998), “How Good is Your VaR Using Backtesting to Assess System Performance”, *Financial Engineering News*, August, p. 1-2.
- 7) Butler, J.S., Schachter, B. (1998), “Estimating Value-at-Risk with a Precision Measure by Combining Kernel Estimation with Historical Simulation”, *Review of Derivatives Research* 1, p. 371-390.
- 8) Chen, S. X., Tang, C. Y. (2005), “Nonparametric inference of Value at Risk for dependent financial returns.” *Journal of Financial Econometrics* 3, p. 227–255.
- 9) Chen, X.C. (2008), “Nonparametric Estimation of Expected Shortfall”, *Journal of Financial Econometrics*, p. 87–107.
- 10) Clark, T.E., McCracken, M.W. (2001), “Tests of Equal Forecast Accuracy and Encompassing for Nested Models,” *Journal of Econometrics*, 105, p. 85–110.

- 11) Cotter, J. (2007), "Extreme risk in Asian equity markets", MPRA Paper, <http://mpa.ub.uni-muenchen.de/3536/>
- 12) Cotter, J. (2004), "Downside Risk for European Equity Markets", *Applied Financial Economics*, Vol. 14, No. 10, p. 707-716.
- 13) Danielsson, J., de Vries, C. (1997), "Tail Index and Quantile Estimation with Very High Frequency Data", *Journal of Empirical Finance*, 4, p. 241-257.
- 14) Danielsson, J., Hartmann, P., de Vries, C. (1998), "The cost of conservatism", *Risk*, Vol. 1, No. 11, p. 101-103.
- 15) Diebold, F.X., Mariano, R. (1995), "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics*, 13, p. 253-265.
- 16) Diebold, F.X., Schuermann, T., Stroughair, J. (2000), "Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management", *Journal of Risk Finance*, 1, p. 30-36.
- 17) Dowd, K. (2005), "Measuring market risk", New York, John Wiley & Sons
- 18) Embrechts, P., Resnick, I. S., Samorodnitsky, G. (1997), "Extreme value theory as a risk management tool", 28th International ASTIN Colloquium, Cairns
- 19) Gencay, R., Selcuk, F., Ulugulyagci, A. (2003), "High volatility, thick tails and extreme value theory in Value-at-Risk estimations", *Insurance: Mathematics and Economics*, p.337-356.
- 20) Gencay, R., Selcuk, F. (2004), "Extreme value theory and Value-at-Risk: Relative performance in emerging markets", *International Journal of Forecasting*, 20, p. 287-303.
- 21) Greenspan, A. (2005), "Reflections on central banking", remarks at "The Greenspan Era: Lessons for the Future" – Symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming, 25-27 August.
- 22) Hansen, P.R. (2005), "A Test for Superior Predictive Ability. *Journal of Business and Economic Statistics*, Volume: 23, Issue: 4, p.365-380.

- 23) Harvey, D., Leybourne, S.J., Newbold, P. (1998), "Tests for Forecast Encompassing," *Journal of Business and Economic Statistics*, 16, p. 254–259.
- 24) Harvey, D., Newbold, P. (2000), "Tests for Multiple Forecast Encompassing," *Journal of Applied Econometrics*, 15, p. 471–482.
- 25) Holton, A. G. (1998), "Simulating Value-at-Risk", *Risk*, Vol. 11, No. 5, p. 60-63.
- 26) Hull, J., White, A. (1998), "Incorporating volatility updating into the Historical Simulation method for Value at Risk", *Journal of Risk* (1), p.1-19.
- 27) Inui, K., Kijima, M. (2005), "On the Significance of Expected Shortfall as a Coherent Risk Measure", *Journal of Banking and Finance*, 29, p. 853–864.
- 28) Knight, M. D. (2007), "Now you see it, now you don't: risk in the small and in the large", speech delivered at the Eighth Annual Risk Management Convention of the Global Association of Risk Professionals, 27–28 February
- 29) Kondor, I., Varga-Haszonits, I. (2008), "Feasibility of Portfolio Optimization under Coherent Risk Measures", <http://xxx.tau.ac.il/abs/0803.2283>
- 30) Maghyereh, I. A., Al-Zoubi, A. H. (2006), "Value-at-risk under extreme values: the relative performance in MENA emerging stock markets", *International Journal of Managerial Finance*, Vol. 2, No. 2, p. 154-172.
- 31) McCracken, M.W. (2000), "Robust Out-of-Sample Inference," *Journal of Econometrics*, 99, p. 195–223.
- 32) McNeil, A. (1997), "Estimating the tails of loss severity distributions using extreme value theory", *ASTIN Bulletin* 27, p. 117–137.
- 33) McNeil, A. J., Frey, R. (2000), "Estimation of Tail-related Risk Measures for Heteroscedastic Financial Time Series: An extreme value approach", *Journal of Empirical Finance*, 7, p. 271-300.

- 34) Mendes, B. (2000), "Computing robust risk measures in emerging equity markets using extreme value theory", *Emerging Markets Quarterly*, Vol. 4, p. 25-41.
- 35) Nyströmand, K., Skoglund, J. (2002), "Univariate Extreme Value Theory, GARCH and Measures of Risk", Swedbank, Working Paper, Group Financial Risk Control, Sweden
- 36) Silva, A., Mendes, B. (2003), "Value-at-risk and extreme returns in Asian stock markets", *International Journal of Business*, Vol. 8, p. 17-40.
- 37) Silverman, B.W. (1986), "Density Estimation for Statistics and Data Analysis", London, Chapman and Hall
- 38) Sullivan, R., Timmermann, A., White, H. (2003), "Forecast Evaluation with Shared Data Sets," *International Journal of Forecasting*, 19, p. 217–227.
- 39) Yamai, Y., Yoshihara, T. (2002), "On the Validity of Value-at-Risk: Comparative Analyses with Expected-Shortfall", Bank of Japan, Institute of Monetary and Economic studies, p. 57-86.
- 40) West, K. D. (1996), "Asymptotic Inference About Predictive Ability," *Econometrica*, 64, p. 1067–1084.
- 41) West, K.D., McCracken, M.W. (1998), "Regression Based Tests of Predictive Ability," *International Economic Review*, 39, p. 817–840.
- 42) West, K.D. (2001), "Tests for Forecast Encompassing When Forecasts Depend on Estimated Regression Parameters," *Journal of Business and Economic Statistics*, 19, p. 29–33.
- 43) White, H. (2000), "A Reality Check for Data Snooping," *Econometrica*, 68, p. 1097–1126.
- 44) Žiković, S. (2007), "Measuring Market Risk in EU New Member States", 13th Dubrovnik Economic Conference, Croatian National Bank, Dubrovnik, Croatia, June 26-27.

Table 1: Overview of VaR and ES models

Model	VaR	ES	Description
Historical simulation	$VaR^{cl} = F^{-1}(cl) = X_{(i)}$	$ES^{cl} = \left(\sum_{i=[ncl]}^n X_{n(i)} \right) / (n - [ncl])$	$F_n(t) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq t)$
Mirrored HS	$Y_i = X_i $ $VaR^{cl} = F^{-1}(cl) = Y_{(i)}$	$ES^{cl} = \left(\sum_{i=[ncl]}^n Y_{n(i)} \right) / (n - [ncl])$	
Kernel HS	$G(x) = \sum_k^N \frac{N!}{k!(N-k)!} g(x)$ $VaR^{cl} = (r \in \{r_{t-1}, \dots, r_{t-N}\} G(r; t, N) \geq cl)$	$ES^{cl} = \left(\sum_{i=[ncl]}^n X_{n(i)} \right) / (n - [ncl])$	$\hat{f}(x) = (1/nh) \sum_{i=1}^n K((x-X_i)/h)$ $g(x) = F(x)^k (1-F(x))^{N-k}$
BRW simulation	$G(x; t, N) = \sum_{i=1}^N 1_{\{r_{t-i} \leq x\}} w_{t-i}$ $VaR_t^{cl} = (r \in \{r_{t-1}, \dots, r_{t-N}\} G(r; t, N) \geq cl)$	$ES_t^{cl} = \left(\sum_{i=[ncl]}^n X_{n(i)} \right) / (n - [ncl])$	$\{w\} = \frac{1-\lambda}{1-\lambda^N}, \dots, \left(\frac{1-\lambda}{1-\lambda^N} \right) \lambda^{N-1}$
VCV	$VaR_t^{cl} = \mu_t + \sigma_t \alpha_{cl}$	$ES_t^{cl} = \mu_t + \sigma_t E[Z Z < z_{cl}]$	$\sigma_t = \sqrt{\frac{1}{T} \sum_{i=1}^T (r_i - \bar{r})^2}$
RiskMetrics	$VaR_t^{cl} = \mu_t + \sigma_t \alpha_{cl}$	$ES_t^{cl} = \mu_t + \sigma_t E[Z Z < z_{cl}]$	$\sigma_t = \sqrt{0.94\sigma_{t-1}^2 + 0.06\varepsilon_t^2}$
GARCH	$VaR_t^{cl} = \mu_t + \sigma_t \alpha_{cl}$	$ES_t^{cl} = \mu_t + \sigma_t E[Z Z < z_{cl}]$	$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2$
FHS	$VaR_t^{cl} = (r \in \{r_{t-1}, \dots, r_{t-N}\} G(r; t, N) \geq cl)$	$ES_t^{cl} = \left(\sum_{i=[ncl]}^n \hat{Z}_{n(i)} \right) / (n - [ncl])$	$z_t = \varepsilon_t / \sigma_t$ $\hat{z}_{t+1} = z_t \times \hat{\sigma}_{t+1}$ $\hat{r}_{t+1} = \alpha_0 + \sum_{i=1}^p \alpha_i r_{t-i+1} + \sum_{i=1}^q \theta_i \hat{z}_{t-i+1} + \hat{z}_{t+1}$
Unconditional GPD	$VaR^{cl} = q^{cl}(F) = u + \frac{\sigma}{\xi} \left(\left(\frac{1-cl}{\bar{F}(u)} \right)^{-\xi} - 1 \right)$	$ES_t^{cl} = \frac{1}{1-cl} \int_{cl}^1 q_x(F) dx = \frac{VaR_{cl}}{1-\xi} + \frac{\sigma - \xi u}{1-\xi}$	
Conditional GPD (McNeil, Frey)	$VaR_t^{cl} = \mu_t + \sigma_t VaR(Z)_{cl}$	$ES_t^{cl} = \mu_t + \sigma_t ES(Z)^{cl}$ $ES(Z)^{cl} = \frac{VaR_{cl}}{1-\xi} + \frac{\sigma - \xi u}{1-\xi}$	$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2$ $Z = \left(\frac{x_{t-n+1} - \mu_{t-n+1}}{\sigma_{t-n+1}}, \dots, \frac{x_t - \mu_t}{\sigma_t} \right)$ $VaR(Z)^{cl} = u_z + \frac{\sigma_z}{\xi_z} \left(\left(\frac{1-cl}{\bar{F}(u_z)} \right)^{-\xi_z} - 1 \right)$

Table 2a: Maximum likelihood estimates of shape and scale parameter of the GPD for negative returns and innovations, 750 observations, period June/July 2004 - June/July 2007

	Returns			Innovations			Returns			Innovations		
	estimate	s.e.	threshold	estimate	s.e.	threshold	estimate	s.e.	threshold	estimate	s.e.	threshold
DAX (09.08.2004 - 16.07.2007)						BOVESPA (03.06.2004 - 14.06.2007)						
Tail index	-0,037	0,094	2,518	0,017	0,137	1,610	0,061	0,147	3,569	0,042	0,144	1,508
Sigma	1,189	0,161		0,531	0,102		1,010	0,204		0,535	0,107	
CAC (20.08.2004 - 24.07.2007)						SENSEX (17.06.2004 - 14.06.2007)						
Tail index	0,014	0,096	2,175	0,199	0,166	2,017	0,055	0,146	3,273	-0,092	0,126	1,476
Sigma	0,993	0,134		0,431	0,093		1,266	0,255		0,828	0,155	
FTSE (23.07.2004 - 12.07.2007)						CRTX (01.06.2004 - 18.06.2007)						
Tail index	-0,195	0,112	2,246	-0,029	0,120	1,344	-0,012	0,137	4,934	-0,100	0,125	1,510
Sigma	1,231	0,217		0,645	0,111		1,987	0,387		0,775	0,144	
NIKKEI (20.05.2004 - 06.06.2007)						JALSH (01.07.2004 - 29.06.2007)						
Tail index	0,129	0,157	2,757	0,016	0,141	1,593	0,146	0,159	2,209	0,014	0,141	1,509
Sigma	0,741	0,154		0,660	0,130		0,800	0,168		0,636	0,126	
DJIN (16.07.2004 - 10.07.2007)						KLCI (08.06.2004 - 19.06.2007)						
Tail index	0,102	0,112	1,625	0,167	0,145	1,396	0,049	0,162	1,781	0,263	0,175	1,448
Sigma	0,682	0,101		0,475	0,090		0,911	0,250		0,438	0,097	
SP500 (16.07.2004 - 10.07.2007)						MEXBOL (20.07.2004 - 05.07.2007)						
Tail index	0,062	0,147	2,252	0,124	0,156	1,502	-0,074	0,128	2,659	0,078	0,150	1,632
Sigma	0,656	0,133		0,516	0,107		1,230	0,232		0,499	0,102	
NASDAQ (16.07.2004 - 10.07.2007)						HENG SENG (02.06.2004 - 13.06.2007)						
Tail index	-0,013	0,116	3,335	0,045	0,130	1,425	0,173	0,119	2,050	-0,086	0,127	1,558
Sigma	1,031	0,224		0,522	0,093		0,762	0,118		0,627	0,118	
RTY (16.07.2004 - 10.07.2007)						TAIPEI (15.06.2004 - 27.06.2007)						
Tail index	0,082	0,107	2,127	0,028	0,142	1,568	0,002	0,139	3,212	0,027	0,142	1,419
Sigma	0,630	0,092		0,475	0,094		0,944	0,196		0,772	0,153	

Table 2b: Maximum likelihood estimates of shape and scale parameter of the GPD for negative returns and innovations, 750 observations, period June/July 2007 – July 2010

	Returns			Innovations			Returns			Innovations		
	estimate	s.e.	threshold	estimate	s.e.	threshold	estimate	s.e.	threshold	estimate	s.e.	threshold
DAX (17.07.2007 - 01.07.2010)						BOVESPA (15.06.2007 - 01.07.2010)						
Tail index	0,111	0,078	2,697	0,007	0,116	1,250	0,132	0,157	4,263	-0,146	0,118	1,622
Sigma	1,472	0,172		0,622	0,102		1,393	0,291		0,566	0,103	
CAC (25.07.2007 - 01.07.2010)						SENSEX (15.06.2007 - 01.07.2010)						
Tail index	0,031	0,143	3,851	0,250	0,173	2,172	0,067	0,148	4,094	0,015	0,141	1,448
Sigma	1,130	0,225		0,420	0,092		1,367	0,277		0,620	0,122	
FTSE (13.07.2007 - 01.07.2010)						CRTX (19.06.2007 - 01.07.2010)						
Tail index	0,058	0,125	2,722	-0,118	0,162	1,524	0,202	0,167	6,226	-0,071	0,129	1,394
Sigma	1,185	0,203		0,647	0,145		2,407	0,518		0,674	0,127	
NIKKEI (07.06.2007 - 01.07.2010)						JALSH (02.07.2007 - 01.07.2010)						
Tail index	0,076	0,149	3,421	0,048	0,206	1,879	0,251	0,173	3,077	-0,122	0,107	1,431
Sigma	1,633	0,332		0,425	0,109		0,781	0,171		0,510	0,083	
DJIN (11.07.2007 - 01.07.2010)						KLCI (20.06.2007 - 01.07.2010)						
Tail index	0,000	0,139	2,911	-0,234	0,069	1,019	0,093	0,152	1,983	0,077	0,149	1,381
Sigma	1,365	0,268		0,831	0,097		1,151	0,236		0,668	0,136	
SP500 (11.07.2007 - 01.07.2010)						MEXBOL (06.07.2007 - 01.07.2010)						
Tail index	0,180	0,164	3,128	-0,011	0,170	2,020	-0,099	0,115	3,368	-0,089	0,098	1,204
Sigma	1,189	0,253		0,560	0,085		1,201	0,206		0,686	0,100	
NASDAQ (11.07.2007 - 01.07.2010)						HENG SENG (14.06.2007 - 01.07.2010)						
Tail index	-0,042	0,133	4,264	-0,065	0,139	1,688	0,218	0,169	3,776	-0,073	0,109	1,376
Sigma	1,536	0,295		0,614	0,105		1,235	0,267		0,504	0,081	
RTY (11.07.2007 - 01.07.2010)						TAIPEI (28.06.2007 - 01.07.2010)						
Tail index	0,061	0,147	3,608	0,152	0,204	1,966	-0,047	0,094	3,129	-0,169	0,093	1,309
Sigma	1,562	0,316		0,350	0,094		0,989	0,135		0,732	0,106	

Table 3a: Number of VaR model successes according to Kupiec and Christoffersen independence tests at 5 and 10% significance level, 99% confidence level, 750 observations (June/July 2004 - June/July 2007)

Developed markets (8)								
	HS 250	HS 500	MHS 250	MHS 500	KHS 250	KHS 500	BRW $\lambda=0,97$	BRW $\lambda=0,99$
Kupiec test*	5	8	8	8	8	8	2	7
Kupiec test**	2	8	8	8	8	8	1	7
Independence***	8	7	7	7	7	7	7	5
	VCV	Risk Metrics	GARCH	FHS	HW	EVT GARCH	GPD	
Kupiec test*	2	1	6	8	2	8	8	
Kupiec test**	2	1	6	8	1	8	8	
Independence***	7	3	8	8	6	8	8	
Emerging markets (8)								
	HS 250	HS 500	MHS 250	MHS 500	KHS 250	KHS 500	BRW $\lambda=0,97$	BRW $\lambda=0,99$
Kupiec test*	7	6	8	8	8	8	3	8
Kupiec test**	5	6	8	8	8	7	2	6
Independence***	6	6	6	6	7	5	8	7
	VCV	Risk Metrics	GARCH	FHS	HW	EVT GARCH	GPD	
Kupiec test*	2	0	5	8	3	8	8	
Kupiec test**	0	0	2	8	2	8	8	
Independence***	5	5	8	8	7	8	8	

* 5% significance level

** 10% significance level

*** Christoffersen (1998) independence tests at 5% significance level

HS n – historical simulation model with n day moving window; MHS n - “mirrored” historical simulation model with n day moving window; KHS n – kernel historical approach with n day moving window; BRW - Boudoukh, Richardson, Whitelaw (time weighted) simulation model, λ - decay factor; VCV – normally distributed variance-covariance model; GARCH – parametric EGARCH(p, q) model with GED or T distributed innovations; FHS – Filtered historical simulation Barone-Adesi et. al. (1999); HW – Hull-White (1998) model; EVT-GARCH – McNeil, Frey (2002) conditional EVT model, GPD – unconditional EVT model using Generalized Pareto distribution;

Table 3b: Number of VaR model successes according to Kupiec and Christoffersen independence tests at 5 and 10% significance level, 99% confidence level, 750 observations (June/July 2007 – July 2010)

Developed markets (8)								
	HS 250	HS 500	MHS 250	MHS 500	KHS 250	KHS 500	BRW $\lambda=0,97$	BRW $\lambda=0,99$
Kupiec test*	0	0	1	1	1	1	0	3
Kupiec test**	0	0	1	0	1	0	0	2
Independence***	7	6	8	8	8	7	8	8
	VCV	Risk Metrics	GARCH	FHS	HW	EVT GARCH	EVT GARCH (L)	GPD
Kupiec test*	0	0	1	3	7	2	8	8
Kupiec test**	0	0	1	0	7	2	8	7
Independence***	6	8	8	8	5	8	7	8
Emerging markets (8)								
	HS 250	HS 500	MHS 250	MHS 500	KHS 250	KHS 500	BRW $\lambda=0,97$	BRW $\lambda=0,99$
Kupiec test*	1	1	8	6	5	1	0	5
Kupiec test**	1	1	7	6	2	1	0	5
Independence***	6	5	7	6	7	6	7	6
	VCV	Risk Metrics	GARCH	FHS	HW	EVT GARCH	EVT GARCH (L)	GPD
Kupiec test*	0	0	4	6	5	3	8	6
Kupiec test**	0	0	4	5	4	3	6	6
Independence***	6	6	8	8	5	8	8	7

* 5% significance level

** 10% significance level

*** Christoffersen (1998) independence tests at 5% significance level

Table 4a: VaR ranking according to minimal average VaR values, 99% confidence level, 750 observations (June/July 2004 - June/July2007)

	S&P 500	DJIN	NASDAQ	RTY	NIKKEI	FTSE	DAX	CAC
HS 250	1,55	1,57	2,07	2,40	2,85	1,81	2,35	2,25
HS 500	1,70	1,73	2,35	2,50	3,12	1,98	2,85	2,51
MHS 250	1,72	1,76	2,26	2,66	3,05	1,94	2,48	2,35
MHS 500	2,00	2,04	2,64	2,69	3,22	2,30	3,08	2,81
KHS 250	1,63	1,66	2,23	2,59	3,06	1,92	2,53	2,37
KHS 500	1,78	1,78	2,44	2,59	3,26	2,08	2,93	2,60
BRW $\lambda=0,97$	1,50	1,48	2,01	2,30*	2,67	1,62	2,21	2,02
BRW $\lambda=0,99$	1,60**	1,62**	2,13	2,46	2,97	1,85	2,44	2,33
Normal VCV	1,51	1,49	2,11	2,47	2,56	1,51	2,04	1,86
Risk Metrics	1,46*	1,43*	1,95**	2,32	2,33*	1,47*	1,93*	1,79
GARCH RM	1,52	1,49	1,99	2,38	2,48**	1,51**	2,00**	1,86*
HW	1,76	1,70	2,10	2,59	3,38	2,24	2,60	2,54
FHS	1,81	1,75	2,30	2,67	2,68	1,83	3,07	2,06
EVT GARCH	2,07	2,06	2,24	2,37	3,34	1,78	2,16	2,73
GPD	3,44	3,29	3,90	3,45	4,22	2,84	3,44	3,28
	JALSH	BOVE	MEXBOL	KLCI	CRTX	SENSEX	H SENG	TAIPEI
HS 250	2,65	3,94	2,92	1,53	6,04	3,68	2,37	3,39
HS 500	2,51	4,07	2,74	1,58	5,82	3,58	2,38	3,42
MHS 250	3,26	4,35	3,40	1,82	6,92	4,37	2,54	3,63
MHS 500	3,15	4,61	3,28	1,96	6,50	4,01	2,71	4,07
KHS 250	2,85	4,23	3,12	1,64	6,56	4,03	2,51	3,58
KHS 500	2,75	4,23	2,83	1,66	6,06	3,74	2,47	3,57
BRW $\lambda=0,97$	2,41	3,75	2,64	1,45	5,52	3,35	2,27	2,85
BRW $\lambda=0,99$	2,69	4,08	2,94	1,59	6,28	3,83	2,47	3,41
Normal VCV	2,32*	3,64	2,53	1,35	4,85	3,18	2,05	2,54**
Risk Metrics	2,24	3,37*	2,47*	1,31*	4,45*	2,79*	1,95*	2,18*
GARCH RM	2,33	3,49**	2,59**	1,36**	4,62	2,84**	2,08**	2,23
HW	3,80	4,20	4,33	2,29	7,06	4,93	2,71	3,08
FHS	2,61**	4,08	3,00	1,75	5,37**	3,51	2,31	2,74
EVT GARCH	3,00	4,03	3,05	2,24	5,65	3,77	2,11	3,58
GPD	5,18	5,29	4,74	4,66	8,80	6,56	4,00	4,31

Grey areas mark VaR models which satisfied the Kupiec (1995) and the Christoffersen (1998) independence test at 5% significance level, * lowest average VaR value, ** lowest average VaR value for a model which satisfies the Kupiec and the Christoffersen independence test

Table 4b: VaR ranking according to minimal average VaR values, 99% confidence level, 750 observations (June/July 2007 - July 2010)

	S&P 500	DJIN	NASDAQ	RTY	NIKKEI	FTSE	DAX	CAC
HS 250	4.98	4.40	4.61	5.40	5.87	4.43	4.67	4.74
HS 500	4.70	4.29	4.47	5.43	5.46	4.39	4.61	4.99
MHS 250	5.53	4.88	5.52	5.94	6.10	5.29	5.54	5.75
MHS 500	5.69	5.22	5.15	6.14	6.87	5.80	5.47	6.18
KHS 250	5.15	4.57	4.99	5.76	6.05	4.85	4.87	5.03
KHS 500	4.98	4.43	4.64	5.62	5.77	4.59	4.84	5.07
BRW $\lambda=0,97$	4.06	3.60	3.94	4.56	4.47	3.80	3.92	4.09
BRW $\lambda=0,99$	4.92	4.37	4.72	5.40	5.65	4.55	4.65	4.75
Normal VCV	4.07	3.70	4.19	4.97	4.55	3.84	4.00	4.20
Risk Metrics	3.90	3.54	4.05	4.86	4.32	3.72	3.86	4.15
GARCH RM	4.00	3.65	4.16	4.90	4.32	3.75	3.94	4.19
HW	13.70	12.38	11.18	14.08	12.75	10.76	10.20	10.69
FHS	3.85	3.68	3.88	4,73**	4,20**	3.67	3,84**	3.93
EVT GARCH	4.32	3.62	4.35	4.91	4.06	3.70	4.81	6.44
EVT GARCH (L)	4,89**	5,00**	4,91**	5.27	5.03	3,78**	4.85	5.08
GPD	8.43	6.23	6.37	8.22	8.92	6.19	5.23	5,53**
	JALSH	BOVE	MEXBOL	KLCI	CRTX	SENSEX	H SENG	TAIPEI
HS 250	4.25	5.83	4.28	2,96**	9.37	5.27	5.62	4.41
HS 500	4.29	6.10	4.67	3.13	9.52	5.44	5.67	4.37
MHS 250	4.80	7.24	5.18	3.23	10.53	5.96	7.06	4.84
MHS 500	5.21	7.22	5.44	3.35	12.16	6.22	7.09	4.91
KHS 250	4.55	6.34	4.61	3.09	9.90	5.62	5.98	4.64
KHS 500	4.58	6.22	4.78	3.22	10.31	5.71	5.92	4.54
BRW $\lambda=0,97$	3.60	5.06	3.93	2.41	7.76	4.73	4.71	4.11
BRW $\lambda=0,99$	4.22	6.06	4.57	2.97	9.30	5.39	5.55	4.52
Normal VCV	3.87	5.29	4.06	2.46	7.56	4.98	5.28	3.86
Risk Metrics	3.70	4.96	3.75	2.20	7.27	4.66	5.02	3.81
GARCH RM	3.79	5,14**	3.96	2.26	7.31	4,79**	5,11**	3.98
HW	7.81	9.98	8.40	5.28	21.66	11.39	13.04	8.21
FHS	3,87**	5.36	4.25	2.59	7,71**	5.26	4.78	4,34**
EVT GARCH	2.90	3.99	4.34	3.54	8.24	6.02	4.28	3.97
EVT GARCH (L)	4.07	5.52	5,09**	3.54	8.71	7.49	5.29	4.65
GPD	6.69	8.74	4.39	6.56	18.11	7.30	9.68	4.03

Grey areas mark VaR models which satisfied the Kupiec (1995) and the Christoffersen (1998) independence test at 5% significance level, * lowest average VaR value, ** lowest average VaR value for a model which satisfies the Kupiec and the Christoffersen independence test

Table 5a: VaR ranking according to minimal Lopez size adjusted score, 99% confidence level, 750 observations (June/July 2004 - June/July2007)

	S&P 500	DJIN	NASDAQ	RTY	NIKKEI	FTSE	DAX	CAC
HS 250	4,05	3,04	1,04	1,05	4,06	4,05	3,05	3,05
HS 500	1,04	0,03**	-1,97	0,04	1,04	0,06	0,04	0,05
MHS 250	-0,98	-0,98	-2,97	0,03**	-0,96	0,04**	-1,97	-1,97
MHS 500	-2,98	-1,98	-3,98	-0,97	-1,97	0,04	-1,97	0,03**
KHS 250	2,04	1,03	-1,97	-1,97	1,05	1,04	1,03	-0,96
KHS 500	-0,97**	-0,97	-2,98	0,04	-0,96	0,05	0,03**	0,04
BRW $\lambda=0,97$	7,06	5,06	4,06	2,06	6,08	6,05	3,06	8,06
BRW $\lambda=0,99$	2,03	-0,96	-0,97	-0,96	0,05	5,04	0,04	0,03
Normal VCV	4,05	5,05	0,04**	1,05	9,10	11,08	9,08	9,08
Risk Metrics	9,06	7,07	2,06	5,06	7,09	8,07	7,09	7,07
GARCH RM	4,04	4,05	1,04	0,04	2,06	2,04	2,04	2,05
HW	9,06	8,07	15,08	6,08	7,06	3,02	5,06	1,04
FHS	-4,98	-1,97	-3,97	-1,97	-0,96**	-2,98	-6,99	-1,97
EVT GARCH	-6,98	-5,98	-2,97	0,05	-5,98	-1,98	-1,97	-6,99
GPD	-7,00	-7,00	-7,00	-7,00	-7,00	-7,00	-7,00	-7,00
	JALSH	BOVE	MEXBOL	KLCI	CRTX	SENSEX	H SENG	TAIPEI
HS 250	6,11	2,07	1,09	3,10	0,16	2,11	3,05	-0,94**
HS 500	5,13	-2,94	2,11	4,11	-0,86	2,12	1,05	-2,96
MHS 250	2,07	-2,95	-1,95	0,07	-1,88	-1,92	1,04	-0,95
MHS 500	0,07**	-4,96	-0,94	-1,93	-2,90	0,09	-0,97	-2,97
KHS 250	2,09	-1,95	-0,93	2,08**	-1,87	-0,92	2,04	-1,95
KHS 500	3,12	-2,95	1,10	0,10	-0,87	1,11	1,04	-2,97
BRW $\lambda=0,97$	9,11	4,09	6,09	5,08	2,18	2,11	4,07	3,11
BRW $\lambda=0,99$	3,09	0,06**	3,07	1,07	0,11	0,08	0,04**	-1,96
Normal VCV	9,16	3,08	7,12	6,12	9,29	11,18	10,09	3,11
Risk Metrics	7,10	7,11	10,11	7,08	9,24	10,16	6,07	5,12
GARCH RM	5,09	2,07	-0,94	3,06	6,16	3,11	3,05	5,11
HW	4,04	4,11	0,05**	6,04	1,10	3,05	5,04	6,06
FHS	2,05	-3,96	-3,96	-2,97	0,08**	0,05	-1,97	1,05
EVT GARCH	-2,97	-3,96	-3,96	-6,00	-3,94	0,03**	1,05	-5,00
GPD	-6,98	-6,98	-6,99	-7,00	-5,97	-7,00	-7,00	-7,00

Reported figures represent Lopez (1998) test scores. Grey areas mark VaR models satisfying the Kupiec (1995) and the Christoffersen (1998) independence test at 5% significance level,, ** lowest Lopez score i.e. smallest deviation from expected values.

Table 5b: VaR ranking according to minimal Lopez size adjusted score, 99% confidence level, 750 observations (June/July 2007 - July 2010)

	S&P 500	DJIN	NASDAQ	RTY	NIKKEI	FTSE	DAX	CAC
HS 250	15,25	14,23	16,26	14,30	8,29	10,22	6,17	10,23
HS 500	21,36	18,30	22,31	15,37	11,37	9,27	8,26	9,28
MHS 250	10,18	10,15	8,15	6,20	4,23	4,15	1,12	6,16
MHS 500	14,26	11,22	10,21	6,25	8,32	5,21	3,18	5,21
KHS 250	11,21	10,19	12,20	8,23	4,24	5,18	2,14	7,19
KHS 500	20,33	17,27	15,27	14,33	8,34	8,24	3,24	7,26
BRW $\lambda=0,97$	16,24	11,19	17,22	15,27	9,28	9,22	11,18	12,22
BRW $\lambda=0,99$	11,21	7,17	12,20	9,24	3,24	2,19	2,15	5,18
Normal VCV	29,41	27,34	22,31	22,39	20,41	20,32	13,29	18,35
Risk Metrics	16,14	11,12	14,13	10,09	7,18	11,14	9,12	5,13
GARCH RM	9,09	8,08	10,09	5,06	2,10	7,12	4,10	6,10
HW	-1,97	4,04	1,05**	-1,96	2,05	0,03**	1,05**	1,06
FHS	15,11	8,07	14,13	3,67	3,11	10,14	3,47	8,13
EVT GARCH	4,06	8,08	4,07	5,06	4,12	7,13	-5,95	-6,98
EVT GARCH (L)	-2,97**	-6,00	-3,96	0,03**	-1,94**	3,09	-5,95	-4,95
GPD	-4,97	-3,94**	-3,91	-4,93	-3,94	-4,94	1,11	1,12**
	JALSH	BOVE	MEXBOL	KLCI	CRTX	SENSEX	H SENG	TAIPEI
HS 250	6,15	7,24	10,14	1,09	7,55	4,16	13,25	6,11
HS 500	5,16	6,29	5,13	1,12	7,74	7,20	18,37	6,13
MHS 250	2,11	0,15**	2,07	0,08**	3,41	1,12	2,13	2,08
MHS 500	0,10	0,20	-1,94	1,10	2,57	2,13	11,22	4,09
KHS 250	3,12	4,19	4,09	0,09	3,46	3,12	7,17	1,09
KHS 500	4,14	5,27	5,11	2,11	6,67	7,18	17,33	5,11
BRW $\lambda=0,97$	8,17	11,25	8,13	7,14	12,48	8,19	13,27	10,14
BRW $\lambda=0,99$	2,14	4,19	1,08	1,09	5,41	2,13	9,21	0,09**
Normal VCV	13,22	13,32	16,21	8,18	16,76	15,26	18,33	17,25
Risk Metrics	12,11	10,16	11,16	11,17	10,37	12,19	5,17	14,18
GARCH RM	4,06	2,09	8,13	4,11	2,26	2,12	2,09	9,11
HW	4,04	6,12	0,06**	3,03	0,07	4,08	1,05**	2,08
FHS	0,03**	1,07	7,09	-2,92	1,22	-0,91**	7,13	3,07
EVT GARCH	27,22	26,32	7,08	-5,95	0,17	-1,95	11,21	9,12
EVT GARCH (L)	3,05	-0,94	-1,98	-5,95	0,16**	-5,98	3,10	1,06
GPD	-5,99	-4,94	8,14	-6,97	-5,95	-3,95	-6,96	9,14

Reported figures represent Lopez (1998) test scores. Grey areas mark VaR models satisfying the Kupiec (1995) and the Christoffersen (1998) independence test at 5% significance level,, ** lowest Lopez score i.e. smallest deviation from expected values.

Table 6a: Ranking of ES model according to modified Blanco-Ihle error statistic, 99% confidence level, 750 observations (June/July 2004 - June/July2007)

	S&P500	DJIN	NASDAQ	RTY	NIKKEI	FTSE	DAX	CAC	Total
VCV GPD	11	12	11	12	12	13	13	12	12
RM GPD	13	13	13	13	13	12	11	13	13
GARCH GPD	12	11	12	11	11	11	12	11	11
Bootstr FHS	4	4	5	1	1	5	1	4	2
Bootstr HS250	1	3	7	9	4	6	3	5	4
Bootstr HS500	8	7	2	7	5	8	5	7	6
Bootstr KHS250	5	6	9	5	9	10	8	9	9
Bootstr KHS500	6	5	4	8	8	9	10	8	8
Bootstr MHS250	2	1	1	2	6	4	9	2	3
Bootstr MHS500	3	2	3	3	3	3	2	3	1
Bootstr BRW	7	8	6	6	7	7	6	6	7
EVT GARCH	10	10	8	4	2	2	7	1	5
GPD	9	9	10	10	10	1	4	10	10
	JALSH	BOVE	MEXBOL	KLCI	CRTX	SENSEX	H SENG	TAIPEI	Total
VCV GPD	11	12	11	9	12	13	12	8	12
RM GPD	13	13	13	11	13	12	13	5	13
GARCH GPD	12	11	12	6	11	5	11	7	11
Bootstr FHS	1	1	2	1	3	6	1	3	1
Bootstr HS250	6	2	7	5	9	9	7	12	7
Bootstr HS500	7	7	5	10	5	8	5	9	6
Bootstr KHS250	9	8	9	7	10	11	8	11	10
Bootstr KHS500	8	10	8	12	6	10	6	10	9
Bootstr MHS250	4	5	1	2	2	3	4	2	2
Bootstr MHS500	5	6	6	3	8	2	2	1	3
Bootstr BRW	10	3	3	8	7	7	9	13	8
EVT GARCH	2	9	10	13	1	1	3	6	5
GPD	3	4	4	4	4	4	10	4	4

Lowest value marks the most successful ES model

Table 6b: Ranking of ES model according to modified Blanco-Ihle error statistic, 99% confidence level, 750 observations (June/July 2007 - July 2010)

	S&P500	DJIN	NASDAQ	RTY	NIKKEI	FTSE	DAX	CAC	Total
VCV GPD	12	12	12	12	13	12	11	12	12
RM GPD	13	13	13	13	10	13	14	13	13
GARCH GPD	14	14	14	14	12	14	13	14	14
Bootstr FHS	2	4	1	1	4	1	2	2	1
Bootstr HS250	6	6	5	6	8	6	5	7	6
Bootstr HS500	5	7	6	5	7	8	8	8	7
Bootstr KHS250	9	8	8	9	11	9	9	10	9
Bootstr KHS500	8	9	9	11	9	10	10	11	10
Bootstr MHS250	1	3	3	2	2	7	3	6	2
Bootstr MHS500	4	5	4	4	6	5	4	4	5
Bootstr BRW	10	11	10	7	14	11	12	9	11
EVT GARCH	3	2	2	8	3	2	6	1	3
EVT GARCH (L)	7	10	11	10	5	3	7	5	8
GPD	11	1	7	3	1	4	1	3	4
	JALSH	BOVE	MEXBOL	KLCI	CRTX	SENSEX	H SENG	TAIPEI	Total
VCV GPD	11	13	14	10	12	13	13	13	12
RM GPD	14	14	13	9	11	14	12	12	13
GARCH GPD	13	12	12	14	9	12	14	14	14
Bootstr FHS	1	5	3	2	3	4	5	2	2
Bootstr HS250	9	7	7	5	7	7	10	7	7
Bootstr HS500	6	8	6	13	8	3	9	8	8
Bootstr KHS250	10	10	8	6	10	9	11	9	9
Bootstr KHS500	8	11	10	12	13	6	8	10	11
Bootstr MHS250	2	3	5	4	5	10	2	5	4
Bootstr MHS500	4	6	2	3	4	5	6	6	5
Bootstr BRW	7	9	11	11	14	8	7	11	10
EVT GARCH	5	4	1	7	2	2	3	1	1
EVT GARCH (L)	3	1	9	8	1	1	4	4	3
GPD	12	2	4	1	6	11	1	3	6

Lowest value marks the most successful ES model

Table 7a: Rankings of simulated VaR models performance (N = 2.000) according to Lopez

size adjusted score

	S&P 500	DJIN	NASDAQ	RTY	NIKKEI	FTSE	DAX	CAC	Average
HS 250	5	5	6	6	5	6	4	6	12
HS 500	6	6	8	6	7	6	4	5	13
MHS 250	4	3	4	4	4	3	1	4	6
MHS 500	4	3	4	4	5	4	2	4	8
KHS 250	4	3	5	4	4	4	1	4	7
KHS 500	7	6	6	6	5	6	2	4	10
BRW $\lambda=0,97$	6	4	7	6	6	6	5	6	11
BRW $\lambda=0,99$	3	2	6	5	2	2	1	3	4
Normal VCV	7	6	8	7	7	8	5	7	14
Risk Metrics	5	4	6	5	5	7	5	3	9
GARCH RM	2	3	5	3	1	5	3	3	5
HW	2	2	3	2	3	5	2	1	3
FHS	4	2	1	2	2	6	2	5	4
EVT GARCH	1	1	2	1	1	1	3	2	1
GPD	3	1	3	2	2	1	1	1	2
	JALSH	BOVE	MEXBOL	KLCI	CRTX	SENSEX	H SENG	TAIPEI	Average
HS 250	5	5	6	1	6	4	5	4	11
HS 500	5	4	3	1	6	5	6	5	10
MHS 250	2	1	2	1	4	1	1	2	1
MHS 500	1	1	2	2	4	1	3	3	3
KHS 250	3	3	3	1	4	2	3	2	5
KHS 500	4	4	3	2	5	5	5	4	9
BRW $\lambda=0,97$	6	6	4	5	7	5	5	6	12
BRW $\lambda=0,99$	2	3	1	2	5	2	4	1	4
Normal VCV	7	6	6	5	7	6	6	7	14
Risk Metrics	7	6	6	5	6	6	3	6	13
GARCH RM	4	2	5	4	3	2	1	5	7
HW	4	5	1	4	2	3	3	3	6
FHS	1	2	4	3	2	1	3	2	3
EVT GARCH	2	2	1	3	1	4	1	1	2
GPD	5	3	5	3	5	3	2	4	8

Lowest value marks the most successful VaR model

Table 7b: Rankings of simulated ES models performance (N = 2.000) according to modified Blanco Ihle score

	S&P500	DJIN	NASDAQ	RTY	NIKKEI	FTSE	DAX	CAC	Average
VCV GPD	8	7	7	8	7	7	5	6	10
RM GPD	8	7	8	8	7	7	6	6	11
GARCH GPD	8	7	8	8	7	7	6	6	11
Bootstr FHS	1	3	1	1	2	4	2	2	1
Bootstr HS250	4	5	3	4	4	3	3	4	5
Bootstr HS500	3	5	3	4	4	5	4	4	6
Bootstr KHS250	5	6	5	7	6	5	4	5	7
Bootstr KHS500	5	6	5	7	5	6	4	5	8
Bootstr MHS250	1	2	2	2	1	2	2	4	1
Bootstr MHS500	3	4	2	3	3	2	2	3	4
Bootstr BRW	6	7	6	5	8	6	6	5	9
EVT GARCH	2	2	1	6	2	1	3	1	2
GPD	7	1	4	2	1	1	1	2	3
	JALSH	BOVE	MEXBOL	KLCI	CRTX	SENSEX	H SENG	TAIPEI	Average
VCV GPD	6	6	6	4	6	6	5	7	9
RM GPD	7	6	6	4	6	7	5	7	10
GARCH GPD	7	6	6	6	5	6	5	7	10
Bootstr FHS	1	3	2	1	1	5	2	4	3
Bootstr HS250	4	4	3	2	4	4	4	3	5
Bootstr HS500	3	4	3	5	4	2	4	3	5
Bootstr KHS250	5	5	4	2	6	4	4	5	6
Bootstr KHS500	4	5	4	5	7	3	3	5	7
Bootstr MHS250	1	2	2	2	2	3	2	3	2
Bootstr MHS500	2	4	1	1	2	3	3	1	2
Bootstr BRW	4	5	5	4	8	4	3	6	8
EVT GARCH	1	1	1	3	1	1	1	2	1
GPD	6	1	2	1	3	5	1	2	4

Lowest value marks the most successful ES model