

Hybrid Historical Simulation VaR and ES: Performance in Developed and Emerging Markets

Saša Žiković
Randall K. Filer

CESIFO WORKING PAPER NO. 2820
CATEGORY 12: EMPIRICAL AND THEORETICAL METHODS
OCTOBER 2009

An electronic version of the paper may be downloaded

- *from the SSRN website:* www.SSRN.com
- *from the RePEc website:* www.RePEc.org
- *from the CESifo website:* www.CESifo-group.org/wp

Hybrid Historical Simulation VaR and ES: Performance in Developed and Emerging Markets

Abstract

We introduce a new hybrid approach to joint estimation of Value at Risk (VaR) and Expected Shortfall (ES) for high quantiles of return distributions. We investigate the relative performance of VaR and ES models using daily returns for sixteen stock market indices (eight from developed and eight from emerging markets) prior to and during the 2008 financial crisis. In addition to widely used VaR and ES models, we also study the behavior of conditional and unconditional extreme value (EV) models to generate 99 percent confidence level estimates as well as developing a new loss function that relates tail losses to ES forecasts. Backtesting results show that only our proposed new hybrid and Extreme Value (EV)-based VaR models provide adequate protection in both developed and emerging markets, but that the hybrid approach does this at a significantly lower cost in capital reserves. In ES estimation the hybrid model yields the smallest error statistics surpassing even the EV models, especially in the developed markets.

JEL Code: G24, C14, C22, C52, C53.

Keywords: value at risk, expected shortfall, hybrid historical simulation, extreme value theory, bootstrapping.

Saša Žiković

Faculty of Economics, University of Rijeka
szikovic@efri.hr

Randall K. Filer

City University of New York, Hunter College
rfiler@hunter.cuny.edu

September 2009

Saša Žiković is an Assistant Professor at the Faculty of Economics, University of Rijeka and Visiting Professor of Economics at Faculty of Economics Split and Faculty of Administration Ljubljana. 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. Perception of a low risk environment and further growth prospects were further fueled 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.

As Alan Greenspan (2005) noted, however: "...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 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 behavior 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 behavior 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 and II rules. ES, on the other hand, although a coherent measure of risk, has not been approved by regulators to calculate capital requirements. Perhaps because of this lack of approval, ES has not been as extensively studied as VaR in empirical research. Given that VaR and ES are inherently connected in the sense that ES figures can be easily calculated from the VaR surface in the

distribution tail. Estimation techniques that have been developed for VaR measures in the past decade can easily be employed to yield superior ES forecasts. This means that recent 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.

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, 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 modeled 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.

The current paper extends the advances that have been made analyzing VaR to ES estimation. To the best of our knowledge this is the first extensive study comparing VaR and ES model performance in both developed and emerging countries under the increased market stress of the recent financial crisis.¹ We provide an empirical investigation and tail risk assessment of a wide array of VaR and ES models for leading developed and emerging stock indexes, as well as a new hybrid approach to estimating VaR and ES, and a new loss function that relates realized to forecasted tail losses and is, therefore, more suited to the evaluation of ES forecasts.

The following VaR models are analyzed in this paper:²

¹ Angelidis and Degiannakis (2007) compared the performance of various parametric VaR and ES model using the S&P500 index, Gold Bullion price per Troy Ounce, and US dollar/British pound exchange rate although, since they tested the impact of different volatility forecasting models within a strictly parametric framework, their results are not comparable with the current paper.

² For a good overview of a wide range of VaR and ES models see, for example, Dowd (2005).

- (a) Normal simple moving average (VCV) VaR,
- (b) RiskMetrics system,
- (c) Historical simulation,
- (d) Mirrored historical simulation³,
- (e) Kernel historical approach⁴,
- (f) BRW (time weighted) simulation with decay factors of 0.97 and 0.99,
- (g) RiskMetrics system augmented with GARCH type volatility forecasting,
- (h) Unconditional EVT approach using Generalized Pareto distribution,
- (i) Conditional McNeil and Frey (2000) EVT approach and
- (j) Žiković's (2007) Hybrid Historical simulation (HHS) method.

The ES models analyzed in the paper are:

- (a) Bootstrapped historical simulation,
- (b) Bootstrapped mirrored historical simulation,
- (c) kernel historical approach,

³Mirrored 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.

⁴ 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 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.

- (d) McNeil and Frey (2000) EVT approach,
- (e) Unconditional EVT approach and
- (f) HHS ES approach newly developed in this paper.

2. Value at Risk and Expected Shortfall

VaR is usually defined as:

“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(1-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:

$$ES_{cl, hp}^t(X) = E \left[\sum_{j=1}^{hp} X_{t+j} \mid \sum_{j=1}^{hp} X_{t+j} > VaR_{cl, hp}^t(X), \psi_t \right]. \quad (5)$$

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 Yoshihara (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:

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

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 heteroscedasticity 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.

3. Hybrid Historical Simulation (HHS)

In view of the problems outlined above, we suggest a new technique that we will call “Hybrid Historical Simulation” (HHS), based on a combination of nonparametric bootstrapping and parametric GARCH volatility forecasting. This model is designed to combine the best features of nonparametric and parametric approaches in a simple and straightforward way. The leptokurtosis and asymmetry typically seen in financial data are accounted for by the nonparametric part of the model, while the parametric (GARCH) part of the model removes heteroskedasticity from the data. Although it is capable of dealing with leptokurtosis, asymmetry, autocorrelation and heteroskedasticity, the HHS model is not computationally intensive when compared with nonparametric techniques. The number of parameters to be estimated in the model is small, and determined by the GARCH specification structure, which can be kept simple to keep the model robust to misspecification. HHS differs from VaR models that account for time varying volatility and excess kurtosis relative to normal distribution but involve a relatively large number of parameters that typically cannot be solved in a closed, analytical form and can result in negative scale parameters, both of which exacerbate the numeric computation of the maximum likelihood estimates. HHS estimates is also less likely to yield unstable parameters that can generate misspecification and model risk.

The simplest nonparametric approach, historical simulation, provides a flexible and intuitive framework for risk analysis, but its basic version uses only the realized path of returns and therefore produces risk indicators with high variance. When the goal is to model returns for a horizon longer than the data frequency, simulation approaches, such as, bootstrapping are a sensible choice. Bootstrapping guarantees that the multivariate properties of the original data are preserved yet is flexible enough to incorporate updating of both mean and volatility.

The HHS model is based on modification of the recursive bootstrap procedure developed by Freedman and Peters (1984) and Hull and White (1998) volatility updating procedure. The model uses the observed distribution of the return series but does not impose any theoretical distribution on the data. In order to correctly implement bootstrapping, returns should not exhibit either heteroskedasticity or autocorrelation, meaning that they should be IID. In the modeling of residuals the following general specification is used:

$$r_t = \varphi(x) + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma_t)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2 \quad (16)$$

$$z_t = \varepsilon_t / \sigma_t$$

where φ is some functional form (usually ARMA), x is a vector of explanatory variables (observed at time t or lagged), ε_t is the disturbance term with zero mean and standard deviation (σ_t) that follows a GARCH process. Based on this general specification, the HHS model can be implemented in the following manner:

- 1) Autocorrelation is removed by fitting an ARMA(p,q) model to historical returns:

$$r_t = \alpha_0 + \sum_{i=1}^p \alpha_i r_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t$$

$$\varepsilon_t = \eta_t \sqrt{\sigma_t^2} \quad \eta_t \sim IID N(0,1) \quad (17)$$

2) A GARCH(p,q) model is fitted to the obtained residuals:

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2 \quad (18)$$

3) To obtain standardized residuals $\{z_t\}$, residuals $\{\varepsilon_t\}$ are then divided by the conditional GARCH(p,q) volatility forecasts:

$$z_t = \varepsilon_t / \sigma_t \quad (19)$$

Under the GARCH hypothesis this set of standardized residuals are IID and therefore suitable for bootstrapping.

4) The standardized residual returns $\{z_t\}$ are bootstrapped to obtain a standardized historical time series Θ . Since bootstrapping is applied to IID residuals results are unbiased:

$$z = \{z_1, z_2, \dots, z_t\} \quad z_i \in \Theta \quad (20)$$

5) After obtaining the bootstrapped standardized residuals, the calculation of the VaR is straightforward. A modification of Hull and White's (1998) framework for volatility updating the standardized residuals $\{z_t\}$ is used to scales them by the latest GARCH volatility forecast ($\hat{\sigma}_{t+1}$) to obtained a series of historical residuals that have been updated by forecasted volatility to reflect the current market conditions $\{\hat{z}_{t+1}\}$.

$$\hat{z}_{t+1} = z_t \times \hat{\sigma}_{t+1} \quad (21)$$

6) Simulated returns \hat{r}_{t+1} are then obtained by using the updated bootstrapped residuals

$\{\hat{z}_{t+1}\}$:

$$\hat{r}_{t+1} = \alpha_0 + \sum_{i=1}^p \alpha_i r_{t-i+1} + \sum_{i=1}^q \theta_i \hat{z}_{t-i+1} + \hat{z}_{t+1} \quad (22)$$

7) The VaR is approximated from $G(., t;N)$, the empirical cumulative distribution function of $\{\hat{r}_t\}$ based on return observations $\hat{r}_{t-1}, \dots, \hat{r}_{t-N}$. VaR can also be calculated by applying a smooth density estimator such as the kernel density. By modeling VaR to reflect current market conditions through nonparametric bootstrapping, we can choose between letting the observation period freely grow with the passing of time, resulting in slightly more conservative VaR estimates that are resilient to extreme events, or setting the length of the observation period arbitrarily, thereby allowing the VaR estimates to be less conservative but also less attuned to extreme events. The length of the observation period is purely arbitrary but should in no case be shorter than three years of daily data in order to capture an adequate number of extremes.

The Hybrid Historical Simulation (HHS) VaR model can be used as a basis for a semi-parametric approach to ES estimation. Such a model standardizes the tail losses in excess of HHS VaR by the latest GARCH volatility update for that point in time to form a series of standardized tail losses:

$$z_t = \frac{\text{Tail loss}_t}{\sigma_t} \quad (23)$$

Since these standardized tail losses are now IID they are suitable for bootstrapping. The new discrete pdf's of tail losses derived through bootstrapping can be updated by the latest GARCH volatility forecasts:

$$\hat{F}_{(n)}(t) = F_{(n)}(t) \times \sigma_{t+1} \quad (24)$$

By taking averages over a number of volatility-updated tail pdf's $(\hat{F}_{(n)}(t))$, ES forecasts can react to the latest market developments through the use of GARCH volatility updating. This HHS ES approach provides an elegant way of calculating confidence intervals for ES estimates, based on bootstrapping that is free of distributional assumptions. The only

requirement is that the underlying data-generating process can be described by a GARCH process. Unlike EVT ES models, the HHS ES model does not impose any distributional assumptions about the behavior of the tail losses and allows the empirical distribution of the tails to evolve over time.

The Hybrid Historical Simulation (HHS) ES can be expressed as:

$$ES_{cl} = E(X | X > VaR_{cl}) = \left(\sum_{i=[ncl]}^n \hat{Z}_{n(i)} \right) / (n - [ncl]) \quad (25)$$

where $\hat{Z}_{n(1)} \leq \hat{Z}_{n(2)} \leq \dots \leq \hat{Z}_{n(n)}$ are order statistics from the volatility scaled bootstrapped series \hat{Z} . A similar approach applied to VaR estimation for volatile markets has yielded significant improvements over both parametric and nonparametric approaches (Žiković, 2007).

An overview of VaR and ES models analyzed in the empirical work that follows is given in Table 1.

4. Data and backtesting methodology

We have analyzed the performance of various VaR and ES models using the log of daily returns of equity indices from eight 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 - RTSIS; 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 June 30, 2008, which includes the beginning of the 2008 financial crisis.

VaR and ES figures were calculated for a one-day ahead horizon and 99 percent confidence level. VaR models are tested using Kupiec; Christoffersen Unconditional

Coverage (UC), Conditional Coverage (CC) and Independence (IND), 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 exceedences 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 skeptically since it automatically disadvantages VaR models that err on the conservative side.

Since more than one VaR model can often be accepted, the problem of ranking the models arises. Acceptable models can be ranked using their ability to forecast. Blanco and Ihle (1998) suggested evaluating forecasts according to a loss function equal to:

$$C_t = \begin{cases} \frac{L_t - VaR_t}{VaR_t} & \text{if } L_t > VaR_t \\ 0 & \text{if } L_t \leq VaR_t \end{cases} \quad (26)$$

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. We, however, can modify the Blanco-Ihle loss function to compare ES

with the actual value of the tail loss, a more meaningful comparison. The modified function equals:

$$C_t = \begin{cases} \frac{L_t - ES_t}{ES_t} & \text{if } L_t > VaR_t \\ 0 & \text{if } L_t \leq VaR_t \end{cases} \quad (27)$$

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 error statistics.

5. Findings and backtesting results

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 1,000 most recent observations from each stock index. 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 stock indices analyzed shows leptokurtosis, asymmetry and significant heteroskedasticity, with autoregression being especially pronounced in the emerging markets. An 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.⁵ 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 for the analyzed stock indices' threshold losses (losses surpassing the threshold value set by Hill estimator), and threshold innovations (innovations surpassing the threshold value set by Hill estimator) 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 modeling tail returns. For developed markets, the tail indexes vary between 0.006 (for the Nasdaq) and 0.1848 (for the DJIN). As expected, in emerging

⁵ For the BOVESPA, RTSIS, JALSH, KLCI and HENG SENG indices the asymmetric impact is not significantly different from zero. Results are available from the authors on request.

markets there is a greater difference between the minimum value of the tail index (0.009 for Mexbol) and the maximum value (0.247 for KLCI). The distribution of tail losses for the stock indices in South Africa and Malaysia may not have a finite fourth moment, since the estimated tail index is around 0.25⁶. The tails of the innovations from these time series are far heavier, ranging from 0.061 (for the CAC) to 0.386 (for the DAX) in developed markets and from 0 (for the Mexbol) to 0.197 (for the Bovespa) in emerging markets. High values of the estimated tail index for the left tail are an indication that these markets experienced severe crashes.

Backtesting results for VaR at the 99 percent confidence level are presented in Table 3. In all eight of the developed markets HHS and both EVT models (unconditional and conditional) satisfied both the Kupiec and Christoffersen independence test at the 5 percent significance level. Advanced non-parametric models (especially mirrored historical simulation (MHS 250) and BRW simulation with a decay factor of 0.99) exhibit good performance in comparison to other VaR models. MHS failed only for the DJIN and NIKKEI, while BRW failed for the S&P 500 and FTSE indices. Other models performed very poorly, with the parametric GARCH model passing the tests only with respect to the DJIN and RTY, and KHS 500 and HS 500 models only in the case of DAX and RTY. The purely parametric models, VCV and RiskMetrics, failed for all of the tested stock indexes. At the 10 percent significance level, the EVT models were the only ones that satisfied the Kupiec and Christoffersen IND tests for all indices. The HHS technique failed in the case of the DAX index, satisfying the Kupiec test only at a 10 percent significance level. The MHS 250 model was fourth ranked, failing for the DJIN, S&P500 and NIKKEI. These results are consistent

⁶ 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.

with those results obtained by McNeil and Frey (2000). The EVT-GARCH model performed far better than the unconditional EVT model, yielding far lower average VaR values that were much closer to the HHS estimates.

For the emerging markets, the HHS and unconditional EVT model were the only ones to satisfy the Kupiec and Christoffersen IND tests at a 5 percent or better significance level for all eight indexes. The conditional EVT-GARCH technique failed for the South African index (JALSH). A strong performance was also recorded by the MHS 250 and KHS 250, both of which failed only for the Sensex and Heng Seng index. BRW simulation with a decay factor of 0.99 failed for the Jalsh, Mexbol and Heng Seng index. Other models performed very poorly, with purely parametric models again being the worst performers. The GARCH model was acceptable only for the Brazilian Bovespa index and the VCV and RiskMetrics models failed for all eight indices. At the 10 percent significance level, the HHS and unconditional EVT models were the only ones that satisfied the Kupiec and Christoffersen IND tests for all of the indices.

Overall we find superb performance across both developed and emerging markets for extreme value based approaches as well as the newly-developed HHS model. 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-required Kupiec test as well as the Christoffersen IND criterion. Rankings according to the minimum average VaR value (provided the Basel II 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, EVT models provide the highest VaR estimates, meaning that they are the most conservative but also the most expensive in terms of capital requirements for financial institutions. For developed countries, the HHS model yielded the lowest average VaR five out of eight times (S&P 500, 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 HHS model yielded the lowest average VaR for three out of eight indices (JALSH, RTSIS and Heng Seng index) followed by the HS 500 model with two lowest VaRs (Mexbol and Taipei index). In summary, among VaR models that satisfy the Basel criteria, the HHS model provided the lowest average VaR in most cases, making it the model 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 root mean error (RMSE) and modified Blanco-Ihle error statistics at the 99 percent confidence level are presented in Tables 5 and 6. According to the RMSE statistic, in developed markets Bootstrapped HHS was the best performing ES model, yielding the lowest root mean square error in four out of eight markets. Following closely were the two mirrored historical simulations. The worst performing was the unconditional EVT model. In emerging markets, the Bootstrapped MHS 250 model was the best performing ES model, followed by the Bootstrapped HHS and MHS 500 models. The worst performing approach was kernel historical methods (KHS 250). Results according to the modified Blanco-Ihle statistic were similar. In developed markets, HHS was again the best performing ES model, again performing better than all other models for four out of eight markets. The second and third places were again shared by mirrored historical simulations.

The worst performing model was EVT-GARCH. In the emerging markets, the unconditional EVT model was the best performing ES model, performing best for six out of eight indices. Mirrored historical simulations were again close behind. The worst performing was again the kernel historical approach (KHS 250).

Thus, according to the backtesting results overall, mirrored historical simulation is the superior ES measure for emerging markets while Bootstrapped HHS is the best choice for developed markets. 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 exceedences over VaR often perform better than theoretically better-founded EVT models.

6. Conclusion

We developed a new hybrid approach for estimating VaR and ES, and a new loss statistic for comparing ES estimates with realized tail losses. This hybrid model's performance was compared to a wide array of VaR and ES models including unconditional and conditional EVT model. The results of VaR comparison for both developed and emerging markets are in line with the results reported by McNeil and Frey (2000). Regarding the performance of unconditional EVT models we support the findings of Mendes (2000), Gencay and Selcuk (2004) and Maghyreh and Al-Zoubi (2006). Backtesting shows that only the newly-proposed HHS and EVT-based VaR models provide adequate protection in both developed and emerging markets, but the hybrid approach does this at a significantly lower

cost in idle capital reserves. McNeil and Frey's (2000) conditional EVT-GARCH model performs far better than the unconditional EVT model and yielded far lower average VaR values. Among models that satisfied both conditional and unconditional coverage, our HHS yielded the lowest values of average VaR. Mirrored historical simulation, a simple extension of the historical simulation, also yielded surprisingly good risk coverage and satisfied the backtesting criteria for a great majority of stock indices. Backtest results also show that a kernel historical approach VaR estimator, although inferior to mirrored historical simulation, delivers significant variance and mean square error reduction in quantile estimation compared to plain historical simulation. For both emerging and developed markets, other VaR models consistently fail to produce satisfactory results and would, therefore provide risk managers with falsely optimistic data about the risk levels to which they have exposed their financial institutions. Purely parametric models, such as the VCV, RiskMetrics and GARCH models, were the worst performers for both developed and emerging markets. The wide use of such models may, indeed, have played a role in the recent under-appreciation of risks in financial markets.

Although extreme value theory has proven to be a powerful tool in risk management, both in emerging or developing countries, we have shown that other approaches can be equally or even more successful at a lower cost of capital. The results of ES backtesting are more mixed than the VaR results, with the same model sometimes being either the best or the worst for a given stock index depending on the error measure used. For developed markets, our new HHS ES model clearly yields the smallest error statistics in the developed markets, while mirrored historical simulation proved to be a superior ES measure for emerging markets. We again find no benefit to using a kernel approach instead of bootstrapped historical simulation. It is surprising, however, that although historical simulation models are clearly inferior to EVT models in VaR estimation, in ES estimation bootstrapping historical

exceedences 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. Our reported relative performance of specific models might be specific to the time period analyzed, although this seems less likely given that we ranked models for a large diverse set of developed and emerging stock market indices.

The results show that the strengths and weaknesses of every model are consistent between VaR and ES versions. Thus, recent advances in estimating VaR models can and should easily be adapted to apply to ES measures.

References:

- Acerbi C., Nordio C., Sirtori C. (2001), “Expected Shortfall as a Tool for Financial Risk Management”, Working Paper, <http://www.gloriamundi.org/var/wps.html>
- 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>
- Artzner P., Delbaen F., Eber J.M., Heath D. (1997), “Thinking coherently”, *Risk*, Vol. 10, No. 11, p. 68–71.
- Artzner P., Delbaen F., Eber J.M., Heath D. (1999), “Coherent measures of risk”, *Mathematical Finance*, Vol. 9, No. 3, p. 203–228.
- Blanco C., Ihle G. (1998), “How Good is Your VaR Using Backtesting to Assess System Performance”, *Financial Engineering News*, August, p. 1-2.
- 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.
- 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.
- Chen, X.C. (2008), “Nonparametric Estimation of Expected Shortfall”, *Journal of Financial Econometrics*, p. 87–107.
- Cotter, J. (2007), “Extreme risk in Asian equity markets”, MPRA Paper, <http://mpra.ub.uni-muenchen.de/3536/>
- Cotter, J. (2004), “Downside Risk for European Equity Markets”, *Applied Financial Economics*, Vol. 14, No. 10, p. 707-716.
- Danielsson, J., de Vries, C. (1997), “Tail Index and Quantile Estimation with Very High Frequency Data”, *Journal of Empirical Finance*, 4, p. 241-257.
- Danielsson, J., Hartmann, P., de Vries, C. (1998), “The cost of conservatism”, *Risk*, Vol. 1, No. 11, p. 101–103.
- 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.
- Dowd, K. (2005), “Measuring market risk”, New York, John Wiley & Sons
- Embrechts, P., Resnick, I. S., Samorodnitsky, G. (1997), “Extreme value theory as a risk management tool”, 28th International ASTIN Colloquium, Cairns

- Freedman, D.A., Peters, S.C. (1984), “Bootstrapping a regression equation: Some empirical results”, *Journal of American Statistical Association*, 79, p. 97-106.
- Gencay, R., Selcuk, F., Ulugulyagcı, A. (2003), “High volatility, thick tails and extreme value theory in Value-at-Risk estimations”, *Insurance: Mathematics and Economics*, p.337-356.
- 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.
- 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.
- Hull, J., White, A. (1998), “Incorporating volatility updating into the Historical Simulation Method for Value at Risk”, *Journal of Risk* (1), p.1-19
- Holton, A. G. (1998), “Simulating Value-at-Risk”, *Risk*, Vol. 11, No. 5, p. 60-63.
- 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.
- 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
- Kondor, I., Varga-Haszonits, I. (2008), “Feasibility of Portfolio Optimization under Coherent Risk Measures”, <http://xxx.tau.ac.il/abs/0803.2283>
- 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.
- McNeil, A. (1997), “Estimating the tails of loss severity distributions using extreme value theory”, *ASTIN Bulletin* 27, p. 117–137.
- 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.
- Mendes, B. (2000), “Computing robust risk measures in emerging equity markets using extreme value theory”, *Emerging Markets Quarterly*, Vol. 4, p. 25-41.
- Nyströmand, K., Skoglund, J. (2002), “Univariate Extreme Value Theory, GARCH and Measures of Risk”, Swedbank, Working Paper, Group Financial Risk Control, Sweden
- Silva, A., Mendes, B. (2003), “Value-at-risk and extreme returns in Asian stock markets”, *International Journal of Business*, Vol. 8, p. 17-40.

Silverman, B.W. (1986), "Density Estimation for Statistics and Data Analysis", London, Chapman and Hall

Yamai, Y., Yoshida, 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.

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

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

**Table 2: Maximum Likelihood Estimates of Shape and Scale Parameter for the GPD for
Negative Returns and Innovations
(1/1/2000 – 6/30/2008)**

	Returns			Innovations			Returns			Innovations		
	estimate	se	threshold value	estimate	se	threshold value	estimate	se	threshold value	estimate	se	threshold value
DJIN (12.07.2004 - 30.06.2008)						JALSH (30.06.2004 - 30.06.2008)						
Tail index	0,1848	0,1643	2,2742	0,1518	0,1597	2,0101	0,2277	0,2408	3,0713	0,0012	0,0817	1,4474
Sigma	0,6359	0,1357		0,5820	0,1225		0,7413	0,2278		0,5920	0,0684	
NASDAQ (12.07.2004 - 30.06.2008)						KLCI (15.06.2004 - 30.06.2008)						
Tail index	0,0064	0,1396	3,7934	0,2322	0,1709	1,9868	0,2465	0,1729	1,9166	0,1679	0,1620	2,0531
Sigma	1,2999	0,2558		0,3593	0,0782		0,9151	0,2004		0,7400	0,1568	
SP500 (12.07.2004 - 30.06.2008)						SENSEX (02.07.2004 - 30.06.2008)						
Tail index	0,0977	0,1394	2,3513	0,1841	0,1504	1,9777	0,0961	0,2451	4,8397	0,0598	0,1136	1,8330
Sigma	0,5487	0,1033		0,4479	0,0875		1,2232	0,4049		0,6160	0,0961	
RTY (12.07.2004 - 30.06.2008)						BOVESPA (11.06.2004 - 30.06.2008)						
Tail index	0,1828	0,1640	2,7208	0,2059	0,1672	1,9768	0,1008	0,15266	3,7364	0,1971	0,1660	2,0722
Sigma	0,5492	0,1172		0,3552	0,0765		0,9346	0,19231		0,3519	0,0755	
NIKKEI (07.06.2004 - 30.06.2008)						MEXBOL (15.07.2004 - 30.06.2008)						
Tail index	0,0144	0,1407	2,9648	0,1176	0,1550	2,0031	0,0087	0,1035	2,3247	0,0001	0,1021	1,7388
Sigma	0,9308	0,1839		0,4781	0,0991		1,0097	0,1471		0,5965	0,0864	
CAC (04.08.2004 - 30.06.2008)						RTSIS (21.06.2004 - 30.06.2008)						
Tail index	0,0846	0,2425	4,2138	0,0605	0,2371	2,4638	0,1206	0,1088	3,3341	0,0816	0,1051	1,6804
Sigma	0,8338	0,2746		0,5631	0,1834		1,4630	0,2127		0,6370	0,0910	
DAX (28.07.2004 - 30.06.2008)						HENG SENG (11.06.2004 - 30.06.2008)						
Tail index	0,0301	0,2660	5,0017	0,3855	0,1921	1,9833	0,1155	0,1547	2,9930	0,1058	0,1534	2,0348
Sigma	0,9356	0,3467		0,3420	0,0790		1,1308	0,2342		0,5630	0,1161	
FTSE (15.07.2004 - 30.06.2008)						TAIPEI (15.06.2004 - 30.6.2008)						
Tail index	0,0734	0,0980	1,8008	0,2098	0,1678	2,1291	0,0710	0,1360	3,2854	0,1921	0,1514	1,9550
Sigma	0,8221	0,1100		0,3820	0,0824		0,8760	0,1628		0,4926	0,0966	

Table 3: Number of VaR Model Successes According to Kupiec and Christoffersen IND Tests at 5 and 10 Percent Significance Level (1,000 observations, 99 percent confidence level)

Developed markets (8)							
	HS 250	HS 500	MHS 250	MHS 500	KHS 250	KHS 500	BRW $\lambda=0,97$
Kupiec test ($p>0.05$)	0	2	7	7	6	5	1
Kupiec test ($p>0.1$)	0	0	6	7	5	5	0
Christoffersen IND test	6	7	7	5	7	5	7
	BRW $\lambda=0,99$	VCV	Risk Metrics	GARCH	HHS	EVT GARCH	GPD
Kupiec test ($p>0.05$)	6	0	0	2	8	8	8
Kupiec test ($p>0.1$)	4	0	0	2	7	8	8
Christoffersen IND test	8	7	6	8	8	8	8
Emerging markets (8)							
	HS 250	HS 500	MHS 250	MHS 500	KHS 250	KHS 500	BRW $\lambda=0,97$
Kupiec test ($p>0.05$)	2	4	7	4	6	2	1
Kupiec test ($p>0.1$)	1	3	6	3	6	1	0
Christoffersen IND test	6	5	7	7	7	7	8
	BRW $\lambda=0,99$	VCV	Risk Metrics	GARCH	HHS	EVT GARCH	GPD
Kupiec test ($p>0.05$)	5	0	0	1	8	7	8
Kupiec test ($p>0.1$)	4	0	0	1	7	7	8
Christoffersen IND test	8	6	5	8	8	8	8

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; EVT-GARCH – McNeil, Frey (2002) conditional EVT model, GPD – unconditional EVT model using Generalized Pareto distribution; p – significance level;

**Table 4: Ranking According to Minimal Average VaR Values Provided Kupiec and
Christoffersen IND Test Are Satisfied
(at 5 percent significance level)**

	JALSH	BOVESPA	MEXBOL	KLCI	RTS\$	SENSEX	HENG SENG	TAIPEI
HS 250					-4,73%			-3,56%
HS 500		-4,07%	-3,00%**		-5,30%			-3,53%**
MHS 250	-3,38%	-4,42%	-3,52%	-2,24%	-5,30%			-3,78%
MHS 500	-3,43%	-4,63%	-3,58%		-5,66%			-4,05%
KHS 250	-3,04%	-4,28%	-3,27%	-2,14%	-5,26%			-3,73%
KHS 500	-3,02%	-4,22%	-3,11%		-5,38%			-3,66%
BRW $\lambda=0,97$						-3,76%**		
BRW $\lambda=0,99$		-4,16%		-2,11%**	-5,20%	-4,13%		-3,61%
Normal VCV								
Risk Metrics								
GARCH RM		-3,76%**						
HHS	-2,88%**	-4,18%	-3,10%	-2,14%	-4,62%**	-3,95%	-3,01%**	-3,56%
EVT GARCH		-4,30%	-3,22%	-3,94%	-5,43%	-4,12%	-4,05%	-4,24%
GPD	-4,79%	-5,40%	-4,69%	-7,75%	-6,99%	-6,32%	-6,80%	-4,71%
	S&P 500	DJIN	NASDAQ	RTY	NIKKEI	FTSE	DAX	CAC
HS 250								
HS 500				-2,64%			-2,81%	
MHS 250	-2,38%		-2,45%	-2,88%		-2,33%	-2,70%	-2,63%
MHS 500			-2,69%	-2,88%			-3,05%	
KHS 250			-2,38%	-2,80%		-2,29%	-2,70%	-2,63%
KHS 500				-2,74%			-2,91%	
BRW $\lambda=0,97$				-2,53%**				
BRW $\lambda=0,99$		-1,91%**	-2,32%**	-2,68%	-3,20%		-2,77%	-2,65%
Normal VCV								
Risk Metrics								
GARCH RM		-2,29%		-2,73%				
HHS	-2,34%**	-1,99%	-2,55%	-2,78%	-2,97%**	-2,05%**	-2,45%**	-2,42%**
EVT GARCH	-2,66%	-2,24%	-3,01%	-3,52%	-3,60%	-2,43%	-2,94%	-2,82%
GPD	-3,14%	-3,32%	-4,45%	-5,22%	-5,33%	-3,63%	-4,45%	-4,21%

** - lowest VaR value

Table 5: Ranking of Competing ES Models According to RMSE Error Statistic

(99 percent confidence level, 1,000 observations)

(1 – best, 9 – worst)

	JALSH	BOVESPA	MEXBOL	KLCI	RTSI\$	SENSEX	HENG SENG	TAIPEI	Total
Bootstrap HS250	7	2	3	5	5	8	8	7	6
Bootstrap HS500	5	5	8	4	8	6	5	8	7
Bootstrap MHS250	3	3	4	2	3	2	3	1	1
Bootstrap MHS500	4	8	5	3	4	3	4	2	3
KHS 250	8	4	6	8	7	9	7	6	9
KHS 500	6	7	7	7	6	5	6	9	8
Bootstrap HHS	1	6	2	6	2	4	2	5	2
EVT GARCH	2	9	9	1	1	1	9	3	4
GPD	9	1	1	9	9	7	1	4	5
	S&P500	DJIN	NASDAQ	RTY	NIKKEI	FTSE	DAX	CAC	Total
Bootstrap HS250	4	4	6	7	6	4	2	4	4
Bootstrap HS500	6	7	1	3	8	6	5	7	6
Bootstrap MHS250	2	2	4	1	2	3	4	2	2
Bootstrap MHS500	3	3	3	2	4	2	7	1	3
KHS 250	5	5	5	6	7	5	3	5	5
KHS 500	7	6	7	5	9	7	6	8	7
Bootstrap HHS	1	1	2	4	3	1	1	6	1
EVT GARCH	9	8	8	8	1	9	9	3	8
GPD	8	9	9	9	5	8	8	9	9

Table 6: Ranking of Competing ES Models According to Modified Blanco-Ihle Error

Statistic

(99 percent confidence level, 1,000 observations)

(1 – best, 9 – worst)

	JALSH	BOVESPA	MEXBOL	KLCI	RTSI\$	SENSEX	HENG SENG	TAIPEI	Total
Bootstrap HS250	9	2	6	6	8	9	8	8	7
Bootstrap HS500	6	4	9	9	6	5	5	5	6
Bootstrap MHS250	3	3	4	4	5	3	3	3	2
Bootstrap MHS500	4	5	2	5	4	4	4	2	3
KHS 250	7	7	7	7	9	8	9	7	9
KHS 500	8	8	8	8	7	6	6	6	8
Bootstrap HHS	2	6	5	2	3	7	2	9	5
EVT GARCH	5	9	3	1	2	1	7	4	4
GPD	1	1	1	3	1	2	1	1	1
	S&P500	DJIN	NASDAQ	RTY	NIKKEI	FTSE	DAX	CAC	Total
Bootstrap HS250	5	5	6	6	8	5	5	6	6
Bootstrap HS500	4	7	3	1	6	7	8	7	4
Bootstrap MHS250	2	3	1	5	3	2	2	2	2
Bootstrap MHS500	3	2	2	4	4	3	3	3	3
KHS 250	7	4	7	7	9	6	6	8	8
KHS 500	6	6	5	2	7	8	7	9	7
Bootstrap HHS	1	1	4	3	5	4	1	1	1
EVT GARCH	9	8	8	8	1	9	9	5	9
GPD	8	9	9	9	2	1	4	4	5

CESifo Working Paper Series

for full list see www.cesifo-group.org/wp

(address: Poschingerstr. 5, 81679 Munich, Germany, office@cesifo.de)

- 2757 Mirco Tonin and Michael Vlassopoulos, Disentangling the Sources of Pro-social Behavior in the Workplace: A Field Experiment, August 2009
- 2758 Nicole Grunewald and Inmaculada Martínez-Zarzoso, Driving Factors of Carbon Dioxide Emissions and the Impact from Kyoto Protocol, August 2009
- 2759 Yu-Fu Chen and Michael Funke, Booms, Recessions and Financial Turmoil: A Fresh Look at Investment Decisions under Cyclical Uncertainty, August 2009
- 2760 Jan-Egbert Sturm and Jakob de Haan, Does Central Bank Communication really Lead to better Forecasts of Policy Decisions? New Evidence Based on a Taylor Rule Model for the ECB, August 2009
- 2761 Larry Karp, Sacrifice, Discounting and Climate Policy: Five Questions, August 2009
- 2762 Marianna Belloc and Samuel Bowles, International Trade, Factor Mobility and the Persistence of Cultural-Institutional Diversity, August 2009
- 2763 Charles Noussair and Fangfang Tan, Voting on Punishment Systems within a Heterogeneous Group, August 2009
- 2764 Birgit Bednar-Friedl and Karl Farmer, Internationally Coordinated Emission Permit Policies: An Option for Withdrawers from the Kyoto Protocol?, August 2009
- 2765 Pierre M. Picard and David E. Wildasin, Labor Market Pooling, Outsourcing and Labor Contracts, August 2009
- 2766 Stefan Voigt and Lorenz Blume, The Economic Effects of Federalism and Decentralization – A Cross-Country Assessment, August 2009
- 2767 David S. Jacks, Christopher M. Meissner and Dennis Novy, Trade Booms, Trade Busts, and Trade Costs, August 2009
- 2768 Mario Jametti and Thomas von Ungern-Sternberg, Hurricane Insurance in Florida, August 2009
- 2769 Alessandro Balestrino, Kind of Black: The Musicians' Labour Market in Italy, August 2009
- 2770 Yosr Abid Fourati and Cathal O'Donoghue, Eliciting Individual Preferences for Pension Reform, August 2009
- 2771 Christian Breuer and Chang Woon Nam, VAT on Intra-Community Trade and Bilateral Micro Revenue Clearing in the EU, August 2009

- 2772 Choudhry Tanveer Shehzad, Jakob De Haan and Bert Scholtens, Growth and Earnings Persistence in Banking Firms: A Dynamic Panel Investigation, August 2009
- 2773 Erdal Yalcin, Uncertain Productivity Growth and the Choice between FDI and Export, August 2009
- 2774 Klaus Abberger, Wolfgang Nierhaus and Shynar Shaikh, Findings of the Signal Approach for Financial Monitoring in Kazakhstan, September 2009
- 2775 Sascha O. Becker, Francesco Cinnirella and Ludger Woessmann, The Trade-off between Fertility and Education: Evidence from before the Demographic Transition, September 2009
- 2776 Thomas Aronsson and Erkki Koskela, Optimal Income Taxation, Outsourcing and Policy Cooperation in a Dynamic Economy, September 2009
- 2777 Joel Slemrod, Old George Orwell Got it Backward: Some Thoughts on Behavioral Tax Economics, September 2009
- 2778 Cagri Seda Kumru and Athanasios C. Thanopoulos, Social Security Reform and Temptation, September 2009
- 2779 Alessandro Buccioli and Roel M. W. J. Beetsma, Inter- and Intra-generational Consequences of Pension Buffer Policy under Demographic, Financial and Economic Shocks, September 2009
- 2780 Eduardo Strube and Marcelo Resende, Complementarity of Innovation Policies in the Brazilian Industry: An Econometric Study, September 2009
- 2781 Henry Tulkens and Vincent van Steenberghe, "Mitigation, Adaptation, Suffering": In Search of the Right Mix in the Face of Climate Change, September 2009
- 2782 Maria L. Loureiro, Anna Sanz-de-Galdeano and Daniela Vuri, Smoking Habits: Like Father, Like Son, Like Mother, Like Daughter, September 2009
- 2783 Momi Dahan, Tehila Kogut and Moshe Shalem, Do Economic Policymakers Practice what they Preach? The Case of Pension Decisions, September 2009
- 2784 Eytan Sheshinski, Uncertain Longevity and Investment in Education, September 2009
- 2785 Nannette Lindenberg and Frank Westermann, How Strong is the Case for Dollarization in Costa Rica? A Note on the Business Cycle Comovements with the United States, September 2009
- 2786 Leif Danziger, Noncompliance and the Effects of the Minimum Wage on Hours and Welfare in Competitive Labor Markets, September 2009
- 2787 Gerlinde Fellner, Rupert Sausgruber and Christian Traxler, Testing Enforcement Strategies in the Field: Legal Threat, Moral Appeal and Social Information, September 2009

- 2788 Gabriel J. Felbermayr, Mario Larch and Wolfgang Lechthaler, Unemployment in an Interdependent World, September 2009
- 2789 Sebastian G. Kessing, Federalism and Accountability with Distorted Election Choices, September 2009
- 2790 Daniel Gros, Global Welfare Implications of Carbon Border Taxes, September 2009
- 2791 Louis N. Christofides, Michael Hoy and Ling Yang, The Gender Imbalance in Participation in Canadian Universities (1977-2005), September 2009
- 2792 Jan K. Brueckner and Robert W. Helsley, Sprawl and Blight, September 2009
- 2793 Vidar Christiansen and Stephen Smith, Externality-correcting Taxes and Regulation, September 2009
- 2794 John Beirne, Guglielmo Maria Caporale, Marianne Schulze-Ghattas and Nicola Spagnolo, Global and Regional Spillovers in Emerging Stock Markets: A Multivariate GARCH-in-mean Analysis, September 2009
- 2795 Rüdiger Pethig and Frieder Kolleß, Asymmetric Capital-Tax Competition, Unemployment and Losses from Capital Market Integration, September 2009
- 2796 Ngo Van Long, Horst Raff and Frank Stähler, Innovation and Trade with Heterogeneous Firms, September 2009
- 2797 Margit Osterloh and Bruno S. Frey, Research Governance in Academia: Are there Alternatives to Academic Rankings?, September 2009
- 2798 Thiess Buettner and Clemens Fuest, The Role of the Corporate Income Tax as an Automatic Stabilizer, September 2009
- 2799 Annette Alstadsæter, Measuring the Consumption Value of Higher Education, September 2009
- 2800 Peter Friedrich, Chang Woon Nam and Janno Reiljan, Local Fiscal Equalization in Estonia: Is a Reform Necessary?, September 2009
- 2801 Evžen Kočenda and Jan Hanousek, State Ownership and Control in the Czech Republic, September 2009
- 2802 Michael Stimmelmayer, Wage Inequality in Germany: Disentangling Demand and Supply Effects, September 2009
- 2803 Biswa N. Bhattacharyay, Towards a Macroprudential Surveillance and Remedial Policy Formulation System for Monitoring Financial Crisis, September 2009
- 2804 Margarita Katsimi, Sarantis Kalyvitis and Thomas Moutos, “Unwarranted” Wage Changes and the Return on Capital, September 2009

- 2805 Christian Lessmann and Gunther Markwardt, Aid, Growth and Devolution, September 2009
- 2806 Bas Jacobs and Dirk Schindler, On the Desirability of Taxing Capital Income to Reduce Moral Hazard in Social Insurance, September 2009
- 2807 Hans Gersbach and Noemi Hummel, Climate Policy and Development, September 2009
- 2808 David E. Wildasin, Fiscal Competition for Imperfectly-Mobile Labor and Capital: A Comparative Dynamic Analysis, September 2009
- 2809 Johan Eyckmans and Cathrine Hagem, The European Union's Potential for Strategic Emissions Trading through Minimal Permit Sale Contracts, September 2009
- 2810 Ruediger Bachmann and Christian Bayer, The Cross-section of Firms over the Business Cycle: New Facts and a DSGE Exploration, October 2009
- 2811 Slobodan Djajić and Michael S. Michael, Temporary Migration Policies and Welfare of the Host and Source Countries: A Game-Theoretic Approach, October 2009
- 2812 Devis Geron, Social Security Incidence under Uncertainty Assessing Italian Reforms, October 2009
- 2813 Max-Stephan Schulze and Nikolaus Wolf, Economic Nationalism and Economic Integration: The Austro-Hungarian Empire in the Late Nineteenth Century, October 2009
- 2814 Emilia Simeonova, Out of Sight, Out of Mind? The Impact of Natural Disasters on Pregnancy Outcomes, October 2009
- 2815 Dan Kovenock and Brian Roberson, Non-Partisan 'Get-Out-the-Vote' Efforts and Policy Outcomes, October 2009
- 2816 Sascha O. Becker, Erik Hornung and Ludger Woessmann, Catch Me If You Can: Education and Catch-up in the Industrial Revolution, October 2009
- 2817 Horst Raff and Nicolas Schmitt, Imports, Pass-Through, and the Structure of Retail Markets, October 2009
- 2818 Paul De Grauwe and Daniel Gros, A New Two-Pillar Strategy for the ECB, October 2009
- 2819 Guglielmo Maria Caporale, Thouraya Hadj Amor and Christophe Rault, International Financial Integration and Real Exchange Rate Long-Run Dynamics in Emerging Countries: Some Panel Evidence, October 2009
- 2820 Saša Žiković and Randall K. Filer, Hybrid Historical Simulation VaR and ES: Performance in Developed and Emerging Markets, October 2009