An efficient method for market risk management under multivariate extreme value theory approach*

Miloš Božović †

Center for Investments and Finance, Krunska 7b, 11000 Belgrade, Serbia

This draft: October 3, 2010

Abstract

This paper develops an efficient multivariate extreme-value approach to calculating Value at Risk (VaR) and expected shortfall. It is based on the notion that some key results of the univariate extreme value theory can be applied separately to a set of orthogonal random variables, provided they are independent and identically distributed. Such random variables can be constructed from the principal components of ARMA-GARCH conditional residuals of a multivariate return series. The model's forecasting ability is then tested on a portfolio of foreign currencies. The results indicate that the generalized Pareto distribution of peeks over treshold of residuals performs well in capturing extreme events. In particular, model backtesting shows that the proposed multivariate approach yields more precise VaR forecasts than the usual methods based on conditional normality, conditional t-distribution or historical simulation, while maintaining the efficiency of conventional multivariate methods.

*I am grateful to Manuel Moreno, José M. Marín, Mats Hansson and participants of the European Financial Management Association 2010 Annual Conference for their useful comments and suggestions. The usual disclaimer applies.

†Email address: milos.bozovic@cif.co.rs

1 Introduction

Sometimes extreme times indeed call for extreme measures. Events like financial crises and market crashes have increased awareness of the need to quantify risk and assess the probability and extent of extremely large losses. Currently, the most popular tool used by financial institutions to measure and manage market risk is Value at Risk (VaR). VaR refers to the maximum potential loss over a given period at a certain confidence level. Originally intended as a reporting tool for senior management, it started becoming prevalent in the risk management world in 1994, when JPMorgan published the methodology behind its *RiskMetrics* system. Soon after, books by Jorion (1996) and Dowd (1998) introduced VaR to academic parlance and gave it more formal theoretical ground. VaR quickly entered other core areas of banking such as capital allocation, portfolio optimization or risk limitation. With its increasing importance, VaR was easily adopted by the regulators as well. In particular, the Basel II capital requirements for market risk are based on VaR.

In spite of being established an industry and regulatory standard, VaR is often criticized for not being a coherent risk measure.[1] Namely, VaR is not strictly sub-additive, since there might be situations in which $\mathrm{VaR}(X+Y) > \mathrm{VaR}(X) + \mathrm{VaR}(Y)$, as shown for example in Artzner *et al.* (1999), Acerbi & Tasche (2002) or Breuer *et al.* (2008). Furthermore, VaR completely ignores statistical properties of losses beyond the specified quantile of the profit-loss distribution, i.e. the tail risk. In order to overcome these drawbacks, Artzner *et al.* (1997) proposed the Expected Shortfall (ES) as an alternative risk measure. It is defined as the conditional expectation of loss beyond a fixed level of VaR. As such, ES takes into account tail risk and satisfies

[1] A coherent risk measure satisfies properties of monotonicity, sub-additivity, homogeneity and translational invariance.

the sub-aditivity property, which assures its coherence as a risk measure.

VaR and ES are usually estimated in analytical, simulation or historical framework. Analytical approach relies upon the assumption that returns or return innovations follow a known distribution, such as normal. Since financial time series commonly exhibit significant autocorrelation and heteroskedasticity, one typically models the conditional rather than the unconditional distribution of returns. However, many empirical results, such as McNeil (1997), da Silva & de Melo Mendez (2003) and Jondeau & Rockinger (2003), show that the normality assumption fails in explaining extreme events, even when autocorrelation and heteroskedasticity are taken into account. This follows from the fact that the high-frequency empirical returns are characterized by heavier tails than those implied by the normal distribution, as well as by a substantial skewness. In order to overcome these problems, a leptokurtic and/or skewed distribution, such as (standard or skewed) Student's t, may be used instead. However, empirical results based on the t-distribution have shown only a limited success. Alternatively, we can estimate VaR and ES via a simulation. The simulation method is quite useful, if not the only one available, when the underlying risk factors have non-linear payoffs, which is the case with options, for example. However, any simulation has to be based on a pre-specified model of dynamics of the underlying factors, thus the VaR and ES estimates will critically rely on a correct model specification with properly and precisely calibrated parameters.

To avoid ad-hoc assumptions of (un)conditional return distribution or dynamics of the underlying risk factors, the historical simulation (HS) is often used as an alternative. The HS employs historical data from recent past, thereby allowing for the presence of heavy tails without making assumptions about the probability distribution or dynamics of returns. This non-parametric approach is conceptually simple

as well as easy to implement. Moreover, it entirely overcomes the problem of model risk. Unfortunately, it suffers from some serious drawbacks. First, any extrapolation beyond past observations will be inaccurate, especially if the historical series is relatively short. If we try to mitigate this problem by considering longer samples, we will practically neglect the time-varying nature of volatility, as well as volatility clustering. In that case, the HS approach would not properly capture the risk in a sudden period of extremely high volatility – the VaR and ES estimates would change only marginally.

Beyond these traditional approaches, there is an alternative which uses the Extreme Value Theory (EVT) to characterize the tail behavior of the distribution of returns. By focusing on extreme losses, the EVT successfully avoids tying the analysis down to a single parametric family fitted to the whole distribution. Although there is a history of use of EVT in the insurance industry, its application to market risk calculations began about a decade ago. McNeil (1999), Bensalah (2000), Smith (2000), Nyström & Skoglund (2002b) and Embrechts *et al.* (2008) survey the mathematical foundations of EVT and discuss its applications to both financial and insurance risk management. The empirical results show that EVT-based models provide more accurate VaR estimates, especially in higher quantiles. For example, McNeil (1997), Nyström & Skoglund (2002b), Harmantzis *et al.* (2005) and Marinelli *et al.* (2007) show that EVT outperforms the estimates of VaR and ES based on analytical and historical methods.

EVT approach thus seems like a natural choice for risk measurement: its implementation is relatively easy and is based on a few assumptions required for the asymptotics to work. Regrettably, this elegance comes with a price, as the straightforwardness is limited to the univariate EVT. In practice, the number of assets in a typical portfolio

is large. We usually deal with a multitude of risk factors and hence our measurement method requires a multivariate approach. However, defining a multivariate model for the evolution of risk factors under extreme market conditions has so far been a daunting task. A seemingly obvious technique involves a multivariate version of the EVT, based on the multidimensional limiting relations (see Smith (2000)), but model complexity increases greatly with the number of risk factors. Alternatively, the joint distribution of returns can be seen as a product of marginal distributions and a copula. McNeil & Frey (2000) and Nyström & Skoglund (2002a), for example, describe the copula approach to assessment of the extreme co-dependence structure of risk factors. Not only that this technique introduces an additional model risk, inherent in the assumption of a specific analytical form of the co-dependence function, but it also becomes quite intractable with increase in dimensionality. Moreover, a typical copula method for multivariate EVT, such as the one described in Nyström & Skoglund (2002a), requires an additional simulation step in order to retrieve the innovations from the joint distribution, given the fitted marginals and parameters of the copula.

This paper introduces a multivariate EVT method for risk measurement that is based on separate estimations of the univariate model. A key assumption of the univariate EVT is that extreme returns are independent and identically distributed. Instead of estimating the joint n-dimensional distributions (using copulas or otherwise), the proposed method works with n orthogonal series of conditional residuals that are approximately independent and identically distributed. These residuals are obtained from the principal components of the joint return series that are free of any autocorrelation, heteroskedasticity and asymmetry. The latter is achieved by assuming that the joint return process follows a stationary n-dimensional model from the ARMA-GARCH family. To render the method free of any unnecessary distributional

assumption, the ARMA-GARCH parameters are estimated by a generalized method of moments.

As an illustration, the technique is applied to a sequence of daily interbank spot exchange rates of Euro, British Pound, Japanese Yen and Swiss Franc with respect to the U.S. Dollar. The VaR and ES estimates are compared to the actual losses. The results indicate that the method performs well in jointly capturing extreme events in all four series. It also yields more precise VaR and ES estimates and forecasts than the usual methods based on conditional normality, conditional t-distribution or historical simulation.

The remainder of the paper is organized as follows: Section 2 presents the theoretical background behind the EVT approach and the estimation methodology used in this paper. Section 3 describes the data and provides an example of estimation. Section 4 shows the back-tests of the model and its forecasting ability, and compares these results to the ones corresponding to the usual methods applied in risk modeling. Concluding remarks are given in Section 5.

2 Theoretical Framework and Estimation Methodology

2.1 Theoretical Framework

This subsection outlines some basic results of the univariate extreme value theory. First, I formally define the two risk measures used throughout the paper, VaR and

ES. Next, I present two most important results of EVT that concern the asymptotic distributions of the order statistics and of the exceedances over a given treshold.

Definition 1. *Let* $\{X_i\}_{i=1}^n$ *be a set of independent and identically distributed random variables with distribution function*

$$F(x) := \mathbb{P}\{X_i \leq x\}$$

for any i. Value at Risk *is the* q*-th quantile of the distribution* F*:*

$$\mathrm{VaR}_q := F^{-1}(q),$$

where $q \in (0,1)$ *and* F^{-1} *is the inverse of* F*. Similarly, the* Expected Shortfall *is the expected value of* X*, given that* VaR *is exceeded:*

$$\mathrm{ES}_q := \mathbb{E}[X|X > \mathrm{VaR}_q].$$

In order to compute VaR and ES we have to be able to assess the upper and lower tails of the distribution function F. Hence, it is natural to consider the order statistics

$$\begin{aligned} M_n &= \max\{X_1, X_2, \ldots, X_n\}, \\ m_n &= \min\{X_1, X_2, \ldots, X_n\}. \end{aligned}$$

Both M_n and m_n are random variables that depend on the length n of the sample. In analogy with the Central Limit Theorem, we will be interested in the asymptotic behavior of these random variables as $n \to \infty$. Since $m_n = -\max\{-X_1, -X_2, \ldots, -X_n\}$ it is sufficient to state all the results for M_n, that is, focus on observations in the upper tail of the underlying distribution. The results for the lower tail will be straightfor-

ward to generalize.

The following theorem is a limit law first derived heuristically by Fisher & Tippett (1928) and later from a rigorous standpoint by Gnedenko (1943).

Theorem 1. *Let $\{X_i\}_{i=1}^n$ be a set of n independent and identically distributed random variables with distribution function F and suppose that there are sequences of normalization constants, $\{a_n\}$ and $\{b_n\}$, such that, for some non-degenerated limit distribution F^*, we have*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} [F(a_n x + b_n)]^n = F^*(x), \quad x \in \mathbb{R}.$$

Then, there exist $\xi \in \mathbb{R}$, $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+$ such that $F^*(x) = \Gamma_{\xi,\mu,\sigma}(x)$ for any $x \in \mathbb{R}$, where

$$\Gamma_{\xi,\mu,\sigma}(x) := \exp\left[-\left(1 + \xi\frac{x-\mu}{\sigma}\right)_+^{-1/\xi}\right]$$

is the so-called generalized extreme value (GEV) *distribution.*

The GEV was first proposed by von Mises (1936) in this form. The $1/\xi$ is referred to as the tail index, as it indicates how heavy the upper tail of the underlying distribution F is. When $\xi \to 0$, the tail index tends to infinity and $\Gamma_{\xi,\mu,\sigma}(x) \to \exp[-\exp(-(x-\mu)/\sigma)]$.

The sign of ξ defines the three fundamental types of extreme value distributions:

• If $\xi = 0$, the distribution is called the Gumbel distribution. In this case, the distribution spreads out along the entire real axis.

• If $\xi > 0$, the distribution is called the Fréchet distribution. In this case, the distribution has a lower bound.

• If $\xi < 0$, the distribution is called the Weibull distribution. In this case, the

distribution has an upper bound.

Many of the well known distributions may be divided between these three classes of GEV distribution according to their behavior in the tail. For example, normal, gamma and log-normal distributions converge to Gumbell distribution ($\xi = 0$); Student's t, Pareto, log-gamma and Cauchy converge to Fréchet distribution ($\xi > 0$); uniform and beta converge to Weibull distribution ($\xi < 0$). The subset of all distributions F that converge to a given type of extreme value distribution is called the *domain of attraction* for that type. Some characterizations of a domain of attraction are given in Nyström & Skoglund (2002b). More details on GEV distribution and domains of attraction can be found, for example, in Embrechts *et al.* (2008).

EVT is sometimes applied directly – for example, by fitting GEV to the maxima of the series, see Smith (2000). An alternative approach is based on *exceedances over treshold*. The following theorem, first stated by Picklands (1975), gives the asymptotic form of conditional distribution beyond a very high treshold.

Theorem 2. *Let $\{X_i\}_{i=1}^n$ be a set of n independent and identically distributed random variables with distribution function F. Define*

$$F_u(y) := \mathbb{P}\left(X \leq u + y \mid X > u\right) = \frac{F(u+y) - F(u)}{1 - F(u)}, \quad y > 0$$

to be the distribution of excesses of X over the treshold u. Let x_F be the end of the upper tail of F, possibly a positive infinity. Then, if F is such that the limit given by Theorem 1 exists, there are constants $\xi \in \mathbb{R}$ and $\beta \in \mathbb{R}_+$ such that

$$\lim_{u \to x_F} \sup_{u < x < x_F} |F_u(x) - G_{\xi,\beta}(x-u)| = 0,$$

where

$$G_{\xi,\beta}(y) := 1 - \left(1 + \xi \frac{y}{\beta}\right)_{+}^{-1/\xi} \tag{1}$$

is known as the generalized Pareto (GP) *distribution.*

There is a close analogy between Theorems 1 and 2 because ξ is the same in both, and there is a one-to-one correspondence between GEV and GP distributions, given by

$$1 - G_{\xi,\beta}(x) = -\ln \Gamma_{\xi,0,\sigma}(x),$$

see Balkema & de Haan (1974), Davison & Smith (1990) and Nyström & Skoglund (2002b).

The application of EVT involves a number of challenges. First, the parameter estimates of the GEV and GP limit distributions will depend on the number of extreme observations used. Second, the choice of a treshold should be large enough to satisfy the conditions that permit the application of Theorem 2, i.e. $u \to x_F$, while at the same time leaving a sufficient number of observations to render the estimation feasible. There are different methods of making this choice, and some of them are examined in Bensalah (2000). Finally, Theorems 1 and 2 hold only if the extreme observations X are independent and identically distributed. Therefore, we cannot apply the results of EVT to returns on financial assets directly, since a typical financial time series exhibits autocorrelation and heteroskedasticity. Moreover, the EVT approach described in this subsection applies only to a single time series, whereas in practice we often deal with multidimensional series. The following subsection describes how to overcome these issues.

2.2 Estimation Methodology

Estimating Independent Univariate Excess Distributions

Theorem 2 states that for a large class of underlying excess distributions (namely, those satisfying Theorem 1), the distribution of exceedances over treshold converges to a generalized Pareto as the treshold is raised. Thus, the GP distribution is the natural model for the unknown excess distribution. The excess distribution above the threshold u may be therefore taken to be *exactly* GP for some ξ and β:

$$F_u(y) = G_{\xi,\beta}(y), \tag{2}$$

for any y satisfying $0 \leq y < x_F - u$.

Assuming that we have a set of realizations $\{z_{t,i}\}_{t=1}^{T}$, we can choose a sensible treshold u and estimate parameters ξ and β. If there are N_u out of a total of T data points that exceed the threshold, the GP will be fitted to the N_u exceedances. In the literature, several estimators have been used to fit the parameters of the GP distribution. Two most popular ones are the maximum likelihood (ML) and the Hill estimator. The ML estimator is based on the assumption that if the tail under consideration exactly follows a GP distribution, then the likelihood function can be written in a closed form. The estimators of the parameters ξ and β are then obtained using the standard ML approach. Provided that $\xi > -1/2$ the ML estimator of the parameters is consistent and asymptotically normal as the number of data points tends to infinity. The alternative is based on a combination of the ML method and the following semi-parametric result.

Theorem 3. *Suppose* $\{X_t\}_{t=1}^{T}$ *are independent and identically distributed random variables with distribution function* F*, and*

$$\lim_{k\to\infty} \frac{1-F(kx)}{1-F(k)} = x^{-1/\xi}, \quad x \in \mathbb{R}_+, \quad \xi > 0.$$

Then, for $x > 0$*,*

$$\lim_{T\to\infty} \mathbb{P}\left(\frac{M_T - b_T}{a_T} \leq x\right) = \Gamma_{\xi,0,1}(x),$$

where $b_T = 0$ *and* $a_T = F(1-1/T)$*.*

When estimating ξ one may, assuming a priori that $\xi > 0$, conjecture that the tail behaves as in Theorem 3 and obtain an ML estimator of the parameter ξ. This estimator is referred to as the Hill estimator, see Danielsson & de Vries (1997).

Nyström & Skoglund (2002b) have shown that ML typically performs better than the Hill estimator in terms of relative bias and relative standard deviation. In addition, ML has a useful property of being almost invariant to the choice of threshold. This is in sharp contrast to the Hill estimator which is very sensitive to this choice. Also, the Hill estimator is designed specifically for the heavy-tailed case whereas the ML method is applicable to light-tailed data as well.

Estimating Tails of Univariate Distributions

By setting $x = u + y$ and combining Theorem 2 and expression (2) we can write

$$F(x) = (1 - F(u))\, G_{\xi,\beta}(x-u) + F(u),$$

for $x > u$. This formula shows that we may easily interpret the model in terms of the tail of the underlying distribution $F(x)$ for $x > u$. Thus, the only additional element we require to construct a tail estimator is an estimate of $F(u)$. For this purpose, I use the method of historical simulation (HS) and take the obvious empirical estimator, $\widehat{F}(u) = 1 - N_u/T$. By setting a threshold at u, we are assuming that we have sufficient observations exceeding u for a reasonable HS estimate of $F(u)$, but for observations beyond u the historical method would be unreliable. Alternatively, we can find N_u that is closest to a predetermined $F(u)$. Thus, for example, in a sample of $T = 1000$ observations, $\widehat{F}(u) = 0.90$ will correspond to $N_u = 100$. The treshold is then set to $u = X_{900}$, if $\{X_t\}_{t=1}^{T}$ are ordered from the lowest to the highest.

Combining the HS estimate $\widehat{F}(u)$ with the ML estimates of the GP parameters, we obtain the tail estimator:

$$\widehat{F}(x) = 1 - \frac{N_u}{T}\left(1 + \widehat{\xi}\frac{x-u}{\widehat{\beta}}\right)^{-1/\widehat{\xi}}, \quad x > u. \tag{3}$$

Note that when the scale parameter β tends to infinity, $G_{\xi,\beta}(\cdot)$ vanishes and the tail estimator converges to the empirical one for any x. Thus, the tail estimator in (3) can be viewed as the HS estimator augmented by the tail behavior, which is captured by the GP distribution.

Estimating Univariate VaR and ES

For a given upper-tail probability $q > F(u)$ the VaR estimate is calculated by inverting the tail estimation formula (3) to get

$$\widehat{\mathrm{VaR}}_q = u + \frac{\widehat{\beta}}{\widehat{\xi}}\left[\left(\frac{T}{N_u}(1-q)\right)^{-\widehat{\xi}} - 1\right]. \tag{4}$$

This is a quantile estimate, where the quantile is an unknown parameter of an unknown underlying distribution. The confidence interval for $\widehat{\mathrm{VaR}}_q$ can be obtained using a method known as the *profile likelihood*.

Once we have $\widehat{\mathrm{VaR}}_q$, the point estimator of ES can be obtained from

$$\widehat{\mathrm{ES}}_q = \frac{1}{1-\widehat{\xi}}\left(\widehat{\mathrm{VaR}}_q + \widehat{\beta} - \widehat{\xi}u\right), \tag{5}$$

(see, for example, McNeil (1999)). As the tail index increases (equivalently, as $\widehat{\xi} \to 0$), the ES becomes progressively greater than VaR.

It is now easy to generalize the results for the VaR and ES such that they hold in the lower tail as well. Let $u_+ \equiv u$ be the upper-tail treshold, and let the lower-tail treshold u_- be defined symmetrically, that is by $F(u_-) = 1 - F(u_+)$. Then, for a given upper-tail probability $q_+ > F(u_+)$ or a given lower-tail probability $q_- < F(u_-)$ the general form of the VaR estimate is

$$\widehat{\mathrm{VaR}}_{q_\pm} = u_\pm \pm \frac{\widehat{\beta}_\pm}{\widehat{\xi}_\pm}\left[\left(\frac{T}{N_{u_\pm}}(1-q_\pm)\right)^{-\widehat{\xi}_\pm} - 1\right], \tag{6}$$

where the subscript $+$ $(-)$ refers to parameters in the upper (lower) tail. Similarly, the general form of the ES estimate is

$$\widehat{\mathrm{ES}}_{q_\pm} = \frac{1}{1-\widehat{\xi}_\pm}\left(\widehat{\mathrm{VaR}}_{q_\pm} \pm \widehat{\beta}_\pm - \widehat{\xi}_\pm u_\pm\right). \tag{7}$$

It is important to stress that the interpretation of VaR and ES may vary, depending on the meaning of the set of variables $\{X_t\}_{t=1}^T$. Usually, in the risk modeling context these variables represent profits and hence are expressed in monetary units.

Orthogonalization

The ultimate goal is to apply the EVT approach to a portfolio consisting of n assets. Before we can use any of the results of EVT outlined in Subsection 2.1, we have to construct a set of cross-sectionally uncorrelated random variables. A natural choice is to work with the principal components of the unconditional covariance matrix of the log returns.

Definition 2. *Define $\boldsymbol{\varepsilon}_t$ to be an n-dimensional random vector whose components $\varepsilon_{t,i}$ have zero mean for each $i = 1,\ 2,\ \ldots,\ n$. Let $\mathbf{V}_\infty = \mathbb{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t')$ be the n-by-n unconditional covariance matrix of $\boldsymbol{\varepsilon}_t$. Denote by $\boldsymbol{\Lambda}$ the diagonal matrix of the eigenvalues of $\mathbf{V}_\infty$,*

$$\boldsymbol{\Lambda} := \operatorname{diag}(\lambda_1,\ \lambda_2,\ \ldots,\ \lambda_n),$$

ordered by descending values, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. (The matrix $\mathbf{V}_\infty$ is positive definite, hence $\lambda_i > 0$ for any i.) Let $\mathbf{P}$ be the corresponding orthogonal matrix of normalized eigenvectors, so that the eigenvalue decomposition of $\mathbf{V}_\infty$ is given by

$$\mathbf{V}_\infty = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'.$$

Let further

$$\mathbf{L} := \mathbf{P}\boldsymbol{\Lambda}^{1/2}.$$

In other words, $\mathbf{L}$ is an n-by-n matrix whose singular value decomposition is given by the product of an orthogonal matrix $\mathbf{P}$, a diagonal matrix $\boldsymbol{\Lambda}^{1/2}$, and the n-by-n identity matrix $\mathbf{1}_n$. Then,

$$\mathbf{z}_t = \mathbf{L}^{-1}\boldsymbol{\varepsilon}_t, \tag{8}$$

is called the vector of principal components of $\boldsymbol{\varepsilon}_t$, *for any t. The i-th element of the vector $\mathbf{z}_t$ is called the i-*th principal component of $\boldsymbol{\varepsilon}_t$.

Note that

$$\mathbb{E}\left(\mathbf{z}_t\right) = \mathbf{L}^{-1}\mathbb{E}\left(\boldsymbol{\varepsilon}_t\right) = \mathbf{0}$$

and

$$\begin{aligned}
\operatorname{var}\left(\mathbf{z}_t\right) &= \mathbb{E}\left(\mathbf{z}_t\mathbf{z}_t'\right) \\
&= \mathbf{L}^{-1}\mathbb{E}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)\mathbf{L}^{-1\prime} \\
&= \mathbf{L}^{-1}\mathbf{V}_\infty\mathbf{L}^{-1\prime} \\
&= \mathbf{1}_n, \qquad (9)
\end{aligned}$$

since $\mathbf{V}_\infty = \mathbf{L}\mathbf{L}'$. Hence, $\mathbf{z}_t$ are cross-sectionally uncorrelated and each component has a unit variance.

Since $\boldsymbol{\varepsilon}_t = \mathbf{L}\mathbf{z}_t$, each coordinate of $\boldsymbol{\varepsilon}_t$ can be written as a linear combination of the principal components,

$$\varepsilon_{t,i} = \sum_{j=1}^{n} L_{ij} z_{t,j}, \quad i = 1, 2, \ldots, n,$$

where L_{ij} are the elements of $\mathbf{L}$. The fraction of total variation in $\boldsymbol{\varepsilon}_t$ explained by the j-th principal component is

$$\frac{\lambda_j}{\sum_{k=1}^{n}\lambda_l}.$$

This property leads to another convenient feature of the principal component approach. Namely, if low-ranked components do not add much to the overall explained variance, which is often the case in financial time series, we can work with a reduced number of m principal components, where $m < n$. The first m components will then

explain

$$\frac{\sum_{j=1}^{m} \lambda_j}{\sum_{k=1}^{n} \lambda_l} \lesssim 1$$

of the variation in $\boldsymbol{\varepsilon}_t$. In that case, $\mathbf{L}$ is replaced by a n-by-m matrix $\mathbf{L}_m$, where

$$\mathbf{L}_m := \mathbf{P}_m \mathbf{\Lambda}_m^{1/2}, \tag{10}$$

$\mathbf{P}_m$ is a n-by-m matrix of the first m normalized eigenvectors, and

$$\mathbf{\Lambda} := \text{diag}(\lambda_1,\ \lambda_2,\ \ldots,\ \lambda_m)$$

is a diagonal matrix of the first m eigenvalues. The m-dimensional vector of the first m principal components of $\boldsymbol{\varepsilon}_t$ is then given by

$$\mathbf{z}_t = \mathbf{L}_m^{-1} \boldsymbol{\varepsilon}_t, \tag{11}$$

for any t.

Filtering

Orthogonalization transforms a cross-sectionally correlated series into a set of uncorrelated ones. We also have to filter out any serial correlation and volatility clustering. As a net result we will obtain sequences of orthogonal, serially uncorrelated and identically distributed conditional residuals.

Specifically, I will assume that for each asset $i = 1, 2, \ldots, n$ the log returns $y_{t,i} :=$

$\ln(S_{t,i}/S_{t-1,i})$ at time t follow an ARMA(r,m) process

$$y_{t,i} = \mu_i + \sum_{s=1}^{r} b_{s,i} y_{t-s,i} + \varepsilon_{t,i} + \sum_{s=1}^{m} \theta_{s,i} \varepsilon_{t-s,i}. \tag{12}$$

For each t and i, the residuals $\varepsilon_{t,i}$ are serially uncorrelated random variables with a continuous density function of zero mean. Conditionally on the information available at $t-1$, the vector of residuals,

$$\boldsymbol{\varepsilon}_t := [\varepsilon_{t,1} \quad \varepsilon_{t,2} \quad \ldots \quad \varepsilon_{t,n}]',$$

has a zero mean and a covariance matrix $\mathbf{V}_t$. That is,

$$\mathbb{E}\left(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}\right) = \mathbb{E}\left(\boldsymbol{\varepsilon}_t\right) = [0 \quad 0 \quad \ldots \quad 0]' =: \mathbf{0}, \tag{13}$$

$$\operatorname{var}\left(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}\right) = \mathbb{E}\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' | \mathcal{F}_{t-1}\right) =: \mathbf{V}_t, \tag{14}$$

where, for any t, the matrix $\mathbf{V}_t$ is positive definite and measurable with respect to the information set $\mathcal{F}_{t-1}$, a σ-field generated by the past residuals $\{\boldsymbol{\varepsilon}_{t-1},\ \boldsymbol{\varepsilon}_{t-2},\ \ldots,\ \boldsymbol{\varepsilon}_1\}$. Note that the vector form of the ARMA process given by equation (12) then reads

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{s=1}^{r} \mathbf{b}_s \mathbf{y}_{t-s} + \boldsymbol{\varepsilon}_t + \sum_{s=1}^{m} \boldsymbol{\theta}_s \boldsymbol{\varepsilon}_{t-s}, \tag{15}$$

where $\mathbf{y}_t$ and $\boldsymbol{\mu}$ are vectors with elements indexed by $i = 1, 2, \ldots, n$, while

$$\begin{aligned} \mathbf{b}_s &:= \operatorname{diag}\left(b_{s,1},\ b_{s,2},\ \ldots,\ b_{s,n}\right) \\ \boldsymbol{\theta}_s &:= \operatorname{diag}\left(\theta_{s,1},\ \theta_{s,2},\ \ldots,\ \theta_{s,n}\right) \end{aligned}$$

are n-by-n diagonal matrices of ARMA coefficients.

To capture the volatility clustering, I will assume that the conditional covariance

matrix follows a model from the GARCH family. The standard GARCH(p,q) model is sufficient to capture most of the clustering, and – to some extent – excess kurtosis. However, it has a drawback of being symmetric, in the sense that negative and positive shocks have the same impact on volatility. There is a strong empirical evidence that the positive and negative innovations to returns exhibit different correlations with innovations to volatility. This asymmetry can be captured, for example, by assuming that the conditional residuals follow an asymmetric distribution, such as skewed Student's t. Alternatively, we can model the asymmetry explicitly in the equation followed by the conditional covariance matrix. In order to keep the estimation method free of any distributional assumptions I opt for the alternative approach. As Glosten *et al.* (1993), I will assume that the conditional covariance $\mathbf{V}_t$ follows a multivariate asymmetric GARCH(p,q), also known as multivariate GJR-GARCH(p,q):

$$\mathbf{V}_t = \mathbf{\Omega} + \sum_{s=1}^{p} \mathbf{A}_s \mathbf{E}_{t-s} + \sum_{s=1}^{p} \mathbf{\Theta}_s \mathbf{I}_{t-s} \mathbf{E}_{t-s} + \sum_{s=1}^{q} \mathbf{B}_s \mathbf{V}_{t-s}, \tag{16}$$

where $\mathbf{\Omega}$, $\mathbf{A}_1$, $\ldots$, $\mathbf{A}_p$, $\mathbf{\Theta}_1$, $\ldots$, $\mathbf{\Theta}_p$, $\mathbf{B}_1$, $\ldots$, $\mathbf{B}_q$ are constant, positive semidefinite n-by-n matrices,

$$\mathbf{E}_t := \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t',$$

and

$$\mathbf{I}_t := \operatorname{diag}(\operatorname{sgn}(-\varepsilon_{t,1})_+,\ \operatorname{sgn}(-\varepsilon_{t,2})_+,\ \ldots,\ \operatorname{sgn}(-\varepsilon_{t,n})_+),$$

for any t. As usual, the coefficients in matrices $\mathbf{A}_s$ in (16) measure the extent to which volatility shocks in previous periods affect the current volatility, while $\mathbf{A}_s + \mathbf{B}_s$ measure the rate at which this effect fades away. The terms proportional to matrices $\mathbf{\Theta}_s$ capture the impact of asymmetric return shocks to volatility. For any t, the

unconditional covariance matrix of $\boldsymbol{\varepsilon}_t$ is given by

$$\mathbf{V}_\infty := \left(\mathbf{1}_n - \sum_{s=1}^{p} \left(\mathbf{A}_s + \frac{1}{2}\boldsymbol{\Theta}_s \right) - \sum_{s=1}^{q} \mathbf{B}_s \right)^{-1} \boldsymbol{\Omega}.$$

Hence, covariance stationarity of the GJR-GARCH(p,q) process (16) is assured by setting the matrix

$$\mathbf{1}_n - \sum_{s=1}^{p} \left(\mathbf{A}_s + \frac{1}{2}\boldsymbol{\Theta}_s \right) - \sum_{s=1}^{q} \mathbf{B}_s$$

to be positive definite.

It is worth noting that there are many plausible and often implemented alternatives to asymmetric GARCH model of Glosten *et al.* (1993), such as EGARCH model of Nelson (1991). I have chosen to work with the Glosten *et al.* (1993) specification for the sake of simplicity.

Cross-sectional correlations are reflected in the off-diagonal terms of matrices $\mathbf{V}_t$ and $\mathbf{E}_t$. This in turn makes the matrices $\boldsymbol{\Omega}$, $\mathbf{A}_1, \ \ldots, \ \mathbf{A}_p$, $\boldsymbol{\Theta}_1, \ \ldots, \ \boldsymbol{\Theta}_p$, $\mathbf{B}_1, \ \ldots, \ \mathbf{B}_q$ non-diagonal. In total, one would have to estimate $(1 + 2p + q)(n + 1)n/2$ different parameters. Clearly, this number explodes as we increase the number of assets in the portfolio. However, this is only one facet of the problem. The other is that we cannot apply the results of univariate EVT to conditional residuals $\varepsilon_{t,i}$ directly.

For that matter, we can work in the orthonormal basis of principal components by applying the linear transformation (8) to the conditional residuals $\boldsymbol{\varepsilon}_t$. In the orthonormal basis of principal components, equation (15) and (16) for the ARMA(r, m)–GJR-GARCH(p, q) process then read:

$$\widehat{\mathbf{y}}_t = \widehat{\boldsymbol{\mu}} + \sum_{s=1}^{r} \widehat{\mathbf{b}}_s \widehat{\mathbf{y}}_{t-s} + \mathbf{z}_t + \sum_{s=1}^{m} \widehat{\boldsymbol{\theta}}_s \mathbf{z}_{t-s} \tag{17}$$

and

$$\widehat{\mathbf{V}}_t = \widehat{\boldsymbol{\Omega}} + \sum_{s=1}^{p} \widehat{\mathbf{A}}_s \widehat{\mathbf{E}}_{t-s} + \sum_{s=1}^{p} \widehat{\boldsymbol{\Theta}}_s \widehat{\mathbf{I}}_{t-s} \widehat{\mathbf{E}}_{t-s} + \sum_{s=1}^{q} \widehat{\mathbf{B}}_s \widehat{\mathbf{V}}_{t-s}, \tag{18}$$

where $\widehat{\mathbf{y}}_t := \mathbf{L}^{-1}\mathbf{y}_t$ for any t, $\widehat{\boldsymbol{\mu}} := \mathbf{L}^{-1}\boldsymbol{\mu}$, and

$$\widehat{\mathbf{M}} := \mathbf{L}^{-1}\mathbf{M}\mathbf{L}^{-1\prime}$$

for any $\mathbf{M} \in \{\widehat{\mathbf{b}};\ \boldsymbol{\Omega},\ \mathbf{A}_1,\ \ldots,\ \mathbf{A}_p,\ \boldsymbol{\Theta}_1,\ \ldots,\ \boldsymbol{\Theta}_p,\ \mathbf{B}_1,\ \ldots,\ \mathbf{B}_q\}$ and any $\mathbf{M} \in \{\boldsymbol{V}_t,\ \mathbf{E}_t,\ \mathbf{I}_t\}_{t \geq \max\{p,q\}}$. In particular,

$$\widehat{\mathbf{E}}_t := \mathbf{L}^{-1}\mathbf{E}_t\mathbf{L}^{-1\prime} = \mathbf{z}_t\mathbf{z}_t'$$

and

$$\widehat{\mathbf{I}}_t := \mathbf{L}^{-1}\mathbf{I}_t\mathbf{L}^{-1\prime} = \mathrm{diag}(\mathrm{sgn}(-z_{t,1})_+,\ \mathrm{sgn}(-z_{t,2})_+,\ \ldots,\ \mathrm{sgn}(-z_{t,n})_+).$$

Equation (13) implies

$$\mathbb{E}\left(\mathbf{z}_t|\mathcal{F}_{t-1}\right) = \mathbf{L}^{-1}\mathbb{E}\left(\boldsymbol{\varepsilon}_t\right) = \mathbf{0}. \tag{19}$$

On the other hand, let

$$\begin{aligned} \widehat{\mathbf{V}}_t &:= \mathrm{var}\left(\mathbf{z}_t|\mathcal{F}_{t-1}\right) \\ &= \mathbb{E}\left(\mathbf{z}_t\mathbf{z}_t'|\mathcal{F}_{t-1}\right) \\ &= \mathbf{L}^{-1}\mathbf{V}_t\mathbf{L}^{-1\prime} \end{aligned}$$

be the conditional covariance matrix of principal components. Since the principal components $\mathbf{z}_t$ are orthogonal, it is reasonable to assume that the matrix $\widehat{\mathbf{V}}_t$ is diagonal (see, for example, Alexander (2001)). Then, the process given by equation

(18) can be estimated separately for each principal component. This gives a set of n independent scalar equations of the form

$$\widehat{V}_{t,i} = \widehat{\Omega}_i + \sum_{s=1}^{p} \widehat{A}_{s,i}\widehat{E}_{t-s,i} + \sum_{s=1}^{p} \widehat{\Theta}_{s,i}\widehat{I}_{t-s,i}\widehat{E}_{t-s,i} + \sum_{s=1}^{q} \widehat{B}_{s,i}\widehat{V}_{t-s,i}, \tag{20}$$

where, in general, $\widehat{M}_i := \widehat{M}_{ii}$ is the i-th diagonal element of the matrix $\widehat{\mathbf{M}}$, i being $1, 2, \ldots, n$ for the first, second, ..., n-th principal component, respectively.

Once we estimate the set of parameters $\{\widehat{\mathbf{\Omega}},\ \widehat{\mathbf{A}}_1,\ \ldots,\ \widehat{\mathbf{A}}_p,\ \widehat{\mathbf{\Theta}}_1,\ \ldots,\ \widehat{\mathbf{\Theta}}_p,\ \widehat{\mathbf{B}}_1,\ \ldots,\ \widehat{\mathbf{B}}_q\}$ we can apply the inverse transformation

$$\mathbf{V}_t := \mathbf{L}\widehat{\mathbf{V}}_t\mathbf{L}' \tag{21}$$

for $t \geq \max\{p, q\}$, to retrieve the series of conditional covariance matrices in the original basis of log returns. This allows us to estimate VaR and ES in a multivariate framework, for an arbitrary portfolio.

Note that it is straightforward to generalize the above approach to the case of $m < n$ principal components. Using definition (10), we can transform any n-by-n matrix $\mathbf{M}$ into the basis of the first m principal components via transformation

$$\widehat{\mathbf{M}} := \mathbf{L}_m^{-1}\mathbf{M}\mathbf{L}_m^{-1\prime},$$

yielding an m-by-m matrix $\widehat{\mathbf{M}}$. Equations (18) and (20) maintain the same form.

GMM Estimation

Estimation of the GJR-GARCH(p, q) parameters in the basis of principal components can be performed in several ways. Let us focus on the set of scalar equations (20). Under the additional assumption of a known conditional distribution for the residuals, it is straightforward to set up the likelihood function for the entire ARMA(r, m)–GJR-GARCH(p, q) model. This gives the ML estimator for the set of parameters

$$\left\{\widehat{\boldsymbol{\mu}},\ \widehat{\mathbf{b}}_1,\ \ldots,\ \widehat{\mathbf{b}}_r,\ \widehat{\boldsymbol{\theta}}_1,\ \ldots,\ \widehat{\boldsymbol{\theta}}_m;\ \widehat{\boldsymbol{\Omega}},\ \widehat{\mathbf{A}}_1,\ \ldots,\ \widehat{\mathbf{A}}_p,\ \widehat{\boldsymbol{\Theta}}_1,\ \ldots,\ \widehat{\boldsymbol{\Theta}}_p,\ \widehat{\mathbf{B}}_1,\ \ldots,\ \widehat{\mathbf{B}}_q\right\}.$$

In the principal component framework, there are $(1+r+m+1+2p+q)n$ parameters in total to be estimated.

However, as indicated earlier, it is desirable to have an estimator which avoids specific assumptions about the conditional distribution, while maintaining the efficiency of the ML (or quasi-ML) estimator. Such an estimator is based on the Generalized Method of Moments (GMM). Instead of making distributional assumptions, it proceeds by postulating conditional moments. Here, I will briefly outline its implementation. The details of the GMM approach to ARMA-GARCH models can be found, for example, in Skoglund (2001).

For a fixed t and any principal component i define

$$\mathbf{e}_t := \left[z_t \quad z_t^2 - \widehat{V}_t\right]',$$

where, with a slight abuse of notation, I use $z_t = z_{t,i}$ and $\widehat{V}_t = \widehat{V}_{t,i}$. Let the score be given by

$$\mathbf{g}_t(\psi) := \mathbf{F}_t'(\psi)\mathbf{e}_t,$$

where $\mathbf{F}_t$ is an instrumental variable function. The GMM estimator of univariate ARMA(r, m)–GJR-GARCH(p, q) parameters ψ is defined as

$$\widehat{\psi} = \arg\min_{\psi} \ \mathbf{m}(\psi)' \ \mathbf{W} \ \mathbf{m}(\psi), \tag{22}$$

where

$$\mathbf{m}(\psi) := \frac{1}{T} \sum_{t=1}^{T} \mathbf{g}_t(\psi)$$

is the sample analog of the expected score, while the weighting matrix $\mathbf{W}$ is a consistent estimate of the inverse asymptotic covariance matrix of the score. The set of moment conditions is given by

$$\mathbb{E}\left[\mathbf{g}_t(\psi)\right] = \mathbf{0}.$$

An efficient choice of instrumental variable function and weighting matrix corresponds to setting

$$\mathbf{F}_t(\psi) = \mathbf{\Sigma}_t^{-1} \mathbf{J}_t(\psi),$$

where $\mathbf{\Sigma}_t := \mathrm{var}\left(\mathbf{e}_t | \mathcal{F}_{t-1}\right)$,

$$\mathbf{J}_t(\psi) := \frac{\partial \mathbf{e}_t}{\partial \psi'}$$

is the Jacobian matrix, and

$$\mathbf{W} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{g}_t(\psi) \mathbf{g}_t(\psi)',$$

see Newey & McFadden (1994). Standard errors can be calculated in a usual way from a consistent estimate of the Fisher information matrix. A recursive semi-closed form solution for $\mathbf{g}_t$ can be found in Skoglund (2001), Nyström & Skoglund (2002b) and Nyström & Skoglund (2002b), for a particular (and common) case of ARMA$(1, 0)$–GJR-GARCH$(1, 1)$ process.

It is worth noting that the application of the GMM estimator requires an initial guess on the third and fourth moments of z_t. Therefore, in order to obtain an initial estimator of the set of parameters ψ, it is convenient to use the quasi-ML estimator to obtain the initial consistent estimates (i.e., to assume conditional normality of residuals).

Forecasting

A one-step-ahead forecast of the transformed log return vector can be obtained from (17):

$$\mathbb{E}\left(\widehat{\mathbf{y}}_{t+1}|\mathcal{F}_t\right) = \widehat{\boldsymbol{\mu}} + \sum_{s=1}^{r} \widehat{\mathbf{b}}_s \widehat{\mathbf{y}}_{t-s+1}. \tag{23}$$

Using the fact that principal components $\mathbf{z}_t$ are independent, we can write the forecast for an arbitrary time horizon $h \geq 1$:

$$\begin{aligned} \mathbb{E}\left(\widehat{\mathbf{y}}_{t+h}|\mathcal{F}_t\right) &= \widehat{\boldsymbol{\mu}} + \sum_{s=1}^{r} \widehat{\mathbf{b}}_s \mathbb{E}\left(\widehat{\mathbf{y}}_{t-s+h}|\mathcal{F}_t\right) \\ &= \widehat{\boldsymbol{\mu}} + \sum_{s=1}^{h} \widehat{\mathbf{b}}_s \widehat{\mathbf{y}}_{t-s+h} + \sum_{s=h+1}^{r} \widehat{\mathbf{b}}_s \mathbb{E}\left(\widehat{\mathbf{y}}_{t-s+h}|\mathcal{F}_t\right). \end{aligned} \tag{24}$$

Equation (24) is recursive and the last term contains the forecasts for $1,\ 2,\ \ldots,\ h-1$ steps ahead.

Next, from equation (18), it follows that a one-step-ahead forecast of conditional covariance in the basis of principal components is given by

$$\mathbb{E}\left(\widehat{\mathbf{V}}_{t+1}|\mathcal{F}_t\right) = \widehat{\mathbf{V}}_{t+1}$$

$$= \widehat{\mathbf{\Omega}} + \sum_{s=1}^{p} \widehat{\mathbf{A}}_s \widehat{\mathbf{E}}_{t-s+1} + \sum_{s=1}^{p} \widehat{\mathbf{\Theta}}_s \widehat{\mathbf{I}}_{t-s+1} \widehat{\mathbf{E}}_{t-s+1} + \sum_{s=1}^{q} \widehat{\mathbf{B}}_s \widehat{\mathbf{V}}_{t-s+1},$$

since $\widehat{\mathbf{V}}_{t+1}$ is measurable with respect to the information available at t. A two-steps-ahead forecast is

$$\mathbb{E}\left(\widehat{\mathbf{V}}_{t+2}|\mathcal{F}_t\right) = \widehat{\mathbf{\Omega}} + \sum_{s=1}^{p} \left(\widehat{\mathbf{A}}_s + \frac{1}{2}\widehat{\mathbf{\Theta}}_s\right) \widehat{\mathbf{V}}_{t-s+1} + \sum_{s=1}^{q} \widehat{\mathbf{B}}_s \widehat{\mathbf{V}}_{t-s+1},$$

which can be obtained by substituting the matrices known up until and including time t. Iteratively, we can derive a covariance forecast for an arbitrary horizon. Applying the inverse transformation (21), we can obtain the covariance forecast in the original basis of log returns.

Given the upper- and lower-tail quantiles $q_{\pm}$, the confidence interval $[z^{-}_{t+h,i},\ z^{+}_{t+h,i}]$ for the forecast of the value of i-th principal component h steps ahead is given by

$$z^{\pm}_{t+h,i} = F_i^{-1}(q_{\pm})\sqrt{\widehat{V}_{t+h,i}}, \tag{25}$$

where, $F_i^{-1}(\cdot)$ is the inverse of the univariate probability function followed by the set of random variables $\{z_{t,i}\}_{t=1}^{T}$. It can be obtained by inverting the tail estimator (3). As before, $\widehat{V}_{t+h,i}$ stands for the i-th diagonal element of the matrix $\widehat{\mathbf{V}}_{t+h}$.

Estimating Multivariate VaR and ES

Our final goal is to estimate VaR and ES for a portfolio of n assets. Denote by $\mathbf{a}$ the vector of portfolio positions, in monetary units.[2] Then, h-steps-ahead portfolio VaR

[2]This, among other things, facilitates the treatment of short positions, when portfolio weights may not be well defined.

is defined by

$$\mathrm{VaR}_{q_\pm} = \mathbf{a}'\mathbf{L}\left[\mathbb{E}\left(\widehat{\mathbf{y}}_{t+h}|\mathcal{F}_t\right) + \mathbf{z}^{\pm}_{t+h}\right], \tag{26}$$

where $\mathbf{z}^{\pm}_{t+h}$ is the vector whose i-th component is given by (25). The intuition behind formula (26) is the following. The first term,

$$\mathbf{a}'\mathbf{L}\ \mathbb{E}\left(\widehat{\mathbf{y}}_{t+h}|\mathcal{F}_t\right) = \mathbf{a}'\mathbb{E}\left(\mathbf{y}_{t+h}|\mathcal{F}_t\right),$$

represents the expected return on the portfolio for h steps ahead. The second term is determined by the vector $\mathbf{L}\mathbf{z}^{\pm}_{t+h}$, which defines the confidence intervals in the n-dimensional space of log returns. Hence, the second term $\mathbf{a}'\mathbf{L}\mathbf{z}^{\pm}_{t+h}$ is the confidence interval for portfolio returns around their mean, for h steps ahead and at a confidence level defined by $q_\pm$.

In analogy with equations (26) and (5), the h-steps-ahead portfolio ES is given by

$$\mathrm{ES}_{q_\pm} = \mathbf{a}'\mathbf{L}\widetilde{\mathbf{z}}^{\pm}_{t+h}, \tag{27}$$

where

$$\widetilde{z}^{\pm}_{t+h,i} = \widetilde{F}_i^{-1}(q_\pm)\sqrt{\widehat{V}_{t+h,i}}$$

and

$$\widetilde{F}_i^{-1}(q_\pm) = \frac{1}{1-\xi_\pm}\left[F_i^{-1}(q_\pm) \pm \beta_\pm - \xi_\pm u_\pm\right].$$

3 Data and Empirical Results

3.1 Data

The empirical results that follow are based on average daily interbank spot exchange rates of Euro, British Pound, Japanese Yen and Swiss Franc with respect to the U.S. Dollar, from January 4, 1999 to September 30, 2008, a sample of 2542 observations. The four time series were obtained from Thomson Financial's Datastream. Table 1 provides summary statistics for the exchange rate levels and the corresponding daily log returns (in percent), computed as $y_t = 100 \ln(S_t/S_{t-1})$. Daily sampling is chosen in order to capture high-frequency fluctuations in return processes that may be critical for identification of rare events in the tails of distribution, while avoiding to model the intraday return dynamics, abundant with spurious market microstructure distortions and trading frictions.

I perform several preliminary test on the data. The values of skewness and kurtosis in Table 1 indicate that both the levels and returns deviate from normality. This is also confirmed by Jarque-Bera and Kolmogorov-Smirnov tests (not reported), whose p-values are at most of the order of 10^{-3}. Table 2 shows the results of Ljung-Box Q-statistics for the autocorrelation of returns, up to order 10 (Panel A). The null hypotheses of no autocorrelation in returns cannot be rejected. The absence of a significant short-run return predictability is consistent with high efficiency of the currency market. The autocorrelation in the squared returns is, on the other hand, highly significant in all four series, indicating the presence of heteroskedasticity (Panel B). The correlation coefficients between squared returns and their lags (not reported) are all positive, confirming the notion of clustering – the periods of high volatility are

Table 1: **Summary Statistics**

Daily interbank spot exchange rates of Euro, British Pound, Japanese Yen and Swiss Franc with respect to the U.S. Dollar, from January 4, 1999 to September 30, 2008 (2542 observations).

Panel A: Daily exchange rate levels				
Currency	Mean	Variance	Skewness	Kurtosis
EUR	1.1511	0.0376	0.2234	2.1992
GBP	1.7103	0.0376	0.0661	1.7385
JPY	0.8774	0.0030	−0.0339	2.2511
CHF	0.7380	0.0121	0.1375	2.1955

Panel B: Daily returns (percent)				
Currency	Mean	Variance	Skewness	Kurtosis
EUR	0.0084	0.3539	−0.0267	4.5420
GBP	0.0040	0.2338	0.0757	4.1778
JPY	0.0026	0.3493	0.2267	4.8656
CHF	0.0088	0.4012	0.1411	4.2532

likely to be followed by high volatility.

Table 3 reports the results of the unit root tests. Both Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) statistics indicate that the unit root hypothesis is convincingly rejected in favor of stationary returns (the critical values of ADF and PP statistics at 5 and 1 percent confidence are −3.41 and −3.96, respectively).

Table 2: **Autocorrelation**

Ljung-Box test for autocorrelation of returns and squared returns up to 10^{th} lag.

Panel A: Autocorrelation of returns		
Currency	Q statistic	p-value
EUR	3.9867	0.9479
GBP	9.4858	0.4867
JPY	6.8611	0.7385
CHF	12.7326	0.2390

Panel B: Autocorrelation of squared returns		
Currency	Q statistic	p-value
EUR	111.5435	$< 10^{-5}$
GBP	105.7946	$< 10^{-5}$
JPY	81.5108	$< 10^{-5}$
CHF	42.7107	$< 10^{-5}$

Table 3: **Stationarity**

Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests for the presence of unit roots, based on the regression

$$y_t = c + \delta t + \phi y_{t-1} + \sum_{L=1}^{10} b_L \Delta y_{t-L} + \varepsilon_t,$$
$$\mathrm{H}_0 : \phi = 1,\ \delta = 0.$$

Currency	ADF statistic	PP statistic
EUR	−15.8257	−50.4768
GBP	−15.2648	−48.0508
JPY	−14.6802	−49.2301
CHF	−15.6345	−50.9751
5% crit. value	−3.41	−3.41
1% crit. value	−3.96	−3.96

3.2 Empirical Results

I apply the method described in Section 2 to the exchange rate data. The ex-post analysis of autocorrelations in principal components and squared principal components have shown that it is sufficient to use an ARMA$(1,0)$–GJR-GARCH$(1,1)$ model to obtain independent and identically distributed residuals. Hence, the estimation steps are the following. First, estimate

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{b}\mathbf{y}_{t-s} + \boldsymbol{\varepsilon}_t. \tag{28}$$

Then, calculate the unconditional variance matrix $\mathbf{V}_\infty$ of the residuals $\boldsymbol{\varepsilon}_t$ and apply the eigenvalue decomposition following Definition 2 to obtain the principal components $\mathbf{z}_t$. The conditional covariance matrix of the principal components then follows

$$\widehat{\mathbf{V}}_t = \widehat{\boldsymbol{\Omega}} + \widehat{\mathbf{A}}\widehat{\mathbf{E}}_{t-1} + \widehat{\boldsymbol{\Theta}}\widehat{\mathbf{I}}_{t-1}\widehat{\mathbf{E}}_{t-1} + \widehat{\mathbf{B}}\widehat{\mathbf{V}}_{t-1}. \tag{29}$$

Next, to obtain the GJR-GARCH$(1,1)$ parameters

$$\boldsymbol{\psi} := \left\{\widehat{\boldsymbol{\Omega}},\ \widehat{\mathbf{A}},\ \widehat{\boldsymbol{\Theta}},\ \widehat{\mathbf{B}}\right\},$$

run the GMM estimation (22) separately for each principal component. Since there are four exchange rates in the sample, there are $4 \times 4 = 16$ parameters in total to be estimated from the GMM step if we work with a full set of four principal components.

Once we have the ARMA$(1,0)$–GJR-GARCH$(1,1)$ parameters, we can compute the forecasts, as well as VaR and ES for an arbitrary portfolio, following formulas (26) and (27). Note that, in general, parameter estimates change in time as we move through the time series. Hence, a proper dynamic method for VaR and ES forecasting

would involve regular updating of parameters.

I illustrate the method by running a dynamic estimation over the sample period. To have sufficient observations for the estimation runs, I start from January 1, 2004. For each of the remaining 1239 daily observations, I calculate one-step-ahead forecasts of VaR and ES. As an example, the estimation details are shown in Tables 4–7 and Figures 1–4, for January 1, 2008.

Table 4 shows the summary of the principal component analysis. The first principal component explains almost 70 percent of joint variations in the four exchange rates.

Table 4: **Principal Components**

Variance explained by each of the four principal components (PCs). Estimation period is January 4, 1999 – December 31, 2007.

	PC 1	PC 2	PC 3	PC 4
Eigenvalue	0.9283	0.2682	0.1036	0.0269
Variance explained	0.6995	0.2021	0.0781	0.0202
Cumulative	0.6995	0.9017	0.9798	1.0000

Table 5 summarizes the results of the univariate parameter estimation. For each principal component i, the table shows the values of ARMA(1,0)–GJR-GARCH(1,1) parameters obtained by the GMM estimation, along with their standard errors. Clearly, mean stationarity is satisfied, since $|\widehat{b}_i| < 1$ for every i. Also, it is easy to check that the GARCH parameters are very close but still within the bounds of covariance-stationary regime, as $\widehat{A}_i + \widehat{\Theta}_i/2 + \widehat{B}_i < 1$. The constant terms $\widehat{\Omega}_i$ appear to be insignificant; however, the corresponding unconditional variances $\widehat{V}_{\infty,i} = \widehat{\Omega}_i/[1 - (\widehat{A}_i + \widehat{\Theta}_i/2 + \widehat{B}_i)]$ are significant.

Table 6 displays the estimates of the upper- and lower-tail parameters of the uni-

Table 5: **ARMA-GARCH Estimates**

Parameter estimates in the ARMA(1,0)–GJR-GARCH(1,1) model, in the basis of principal components. Estimation period is January 4, 1999 – December 31, 2007.

Parameter	PC 1	PC 2	PC 3	PC 4
$\widehat{\mu}_i$	0.0018	0.0043	0.0034	−0.0002
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$\widehat{b}_i$	0.0212	0.0777	0.0910	0.0093
	(0.0077)	(0.0112)	(0.0136)	(0.0443)
$\widehat{\Omega}_i$	0.0005	0.0039	0.0016	0.0016
	(0.0092)	(0.0215)	(0.0179)	(0.0407)
$\widehat{A}_i$	0.0216	0.0446	0.0361	0.1593
	(0.0069)	(0.0238)	(0.0210)	(0.0594)
$\widehat{\Theta}_i$	0.0030	0.0162	0.0007	−0.0942
	(0.0002)	(0.0001)	(0.0001)	(0.0001)
$\widehat{B}_i$	0.9751	0.9218	0.9350	0.6337
	(0.0201)	(0.0185)	(0.0196)	(0.0139)

(Standard errors in parentheses.)

variate GP distribution, $\xi_\pm$ and $\beta_\pm$. Following the procedure described in Subsection 2.2, these parameters are estimated separately for each of the principal components, as the exceedances can be assumed to be not only independent and identically distributed, but also orthogonal. The upper and lower tresholds, u_+ and u_-, are determined by $F(u_+) = 0.90$ and $F(u_-) = 0.10$, respectively. This gives a sufficient number of observations in the tails to render the ML estimation of the parameters possible. The inverse of the tail index, $\xi_\pm$, is significant and negative for the first principal component, which corresponds to Weibull distribution. The other values of $\xi_\pm$ are statistically not different from zero, with an exception of the upper tail of the fourth principal component, where ξ_+ is significant and positive. The scale parameter $\beta_\pm$ has values that range between 0.52 and 0.70. Also, the asymmetry between the upper and the lower tail implied by the parameters is apparent.

Using formula (1) and the values in Table 6, we can plot the function $G_{\xi,\beta}(\cdot)$ for the distribution of excesses of $z_{t,i}$ over the upper and lower tresholds, u_+ and u_-. Figures 1–4 show the graphs for the tail behavior of each of the four principal components. I compare the empirical with the GP distribution function, as well as with the normal and Student's t distributions calibrated across the sample, using the parameter estimates prior to January 1, 2008. Clearly, GP distribution drastically outperforms the alternatives in explaining the tail behavior.

Table 6: **Parameters of the Univariate GP Distribution**

Upper- and lower-tail parameters of the univariate generalized Pareto distribution, estimated separately for each of the standardized ARMA-GARCH orthogonal residuals. The upper and lower tresholds are determined by the quantiles corresponding to probabilities of 0.90 and 0.10, respectively. Estimation period is January 4, 1999 – December 31, 2007.

Upper tail				
Parameter	PC 1	PC 2	PC 3	PC 4
$\widehat{\xi}_+$	−0.1096	0.0544	0.0058	0.1804
	(0.0612)	(0.0703)	(0.0755)	(0.0752)
$\widehat{\beta}_+$	0.6397	0.5153	0.5724	0.5386
	(0.0573)	(0.0497)	(0.0575)	(0.0536)
Lower tail				
Parameter	PC 1	PC 2	PC 3	PC 4
$\widehat{\xi}_-$	−0.2030	0.0570	−0.0293	0.0239
	(0.0575)	(0.0714)	(0.0572)	(0.0625)
$\widehat{\beta}_-$	0.7013	0.6765	0.6379	0.6031
	(0.0605)	(0.0658)	(0.0556)	(0.0548)

(Standard errors in parentheses.)

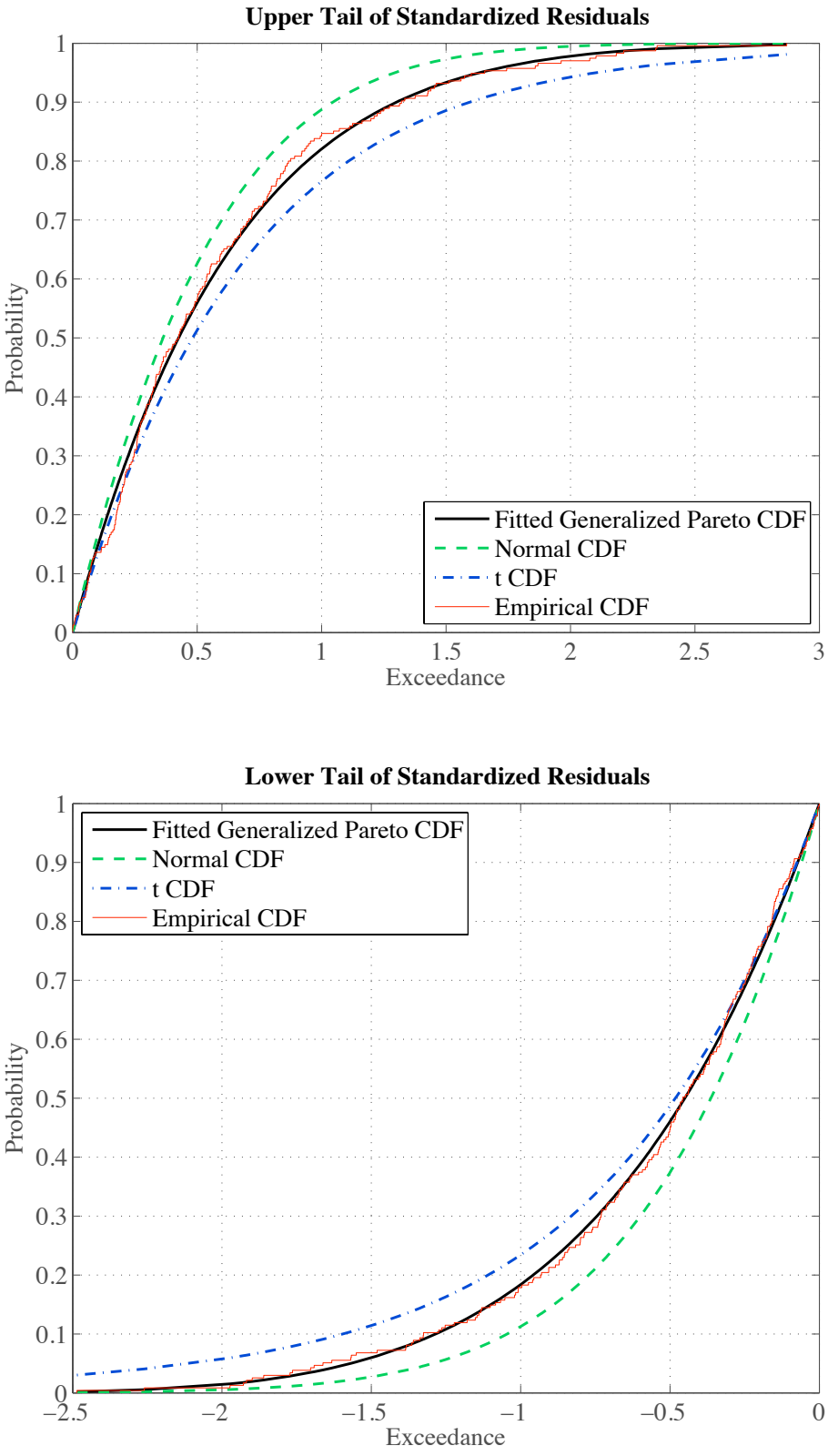

Figure 1: **First principal component.** Upper and lower tails of standardized residuals.

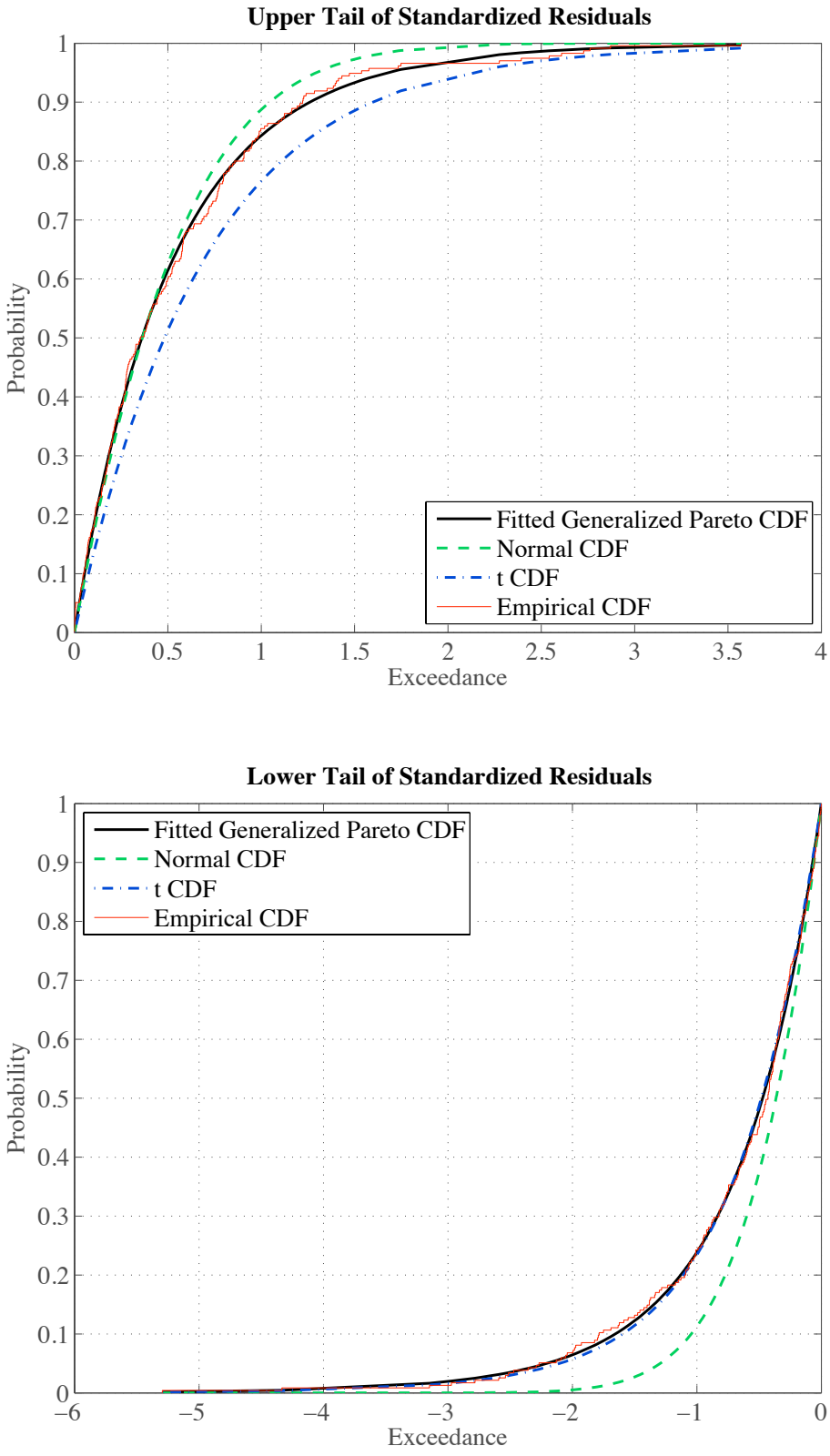

Figure 2: **Second principal component.** Upper and lower tails of standardized residuals.

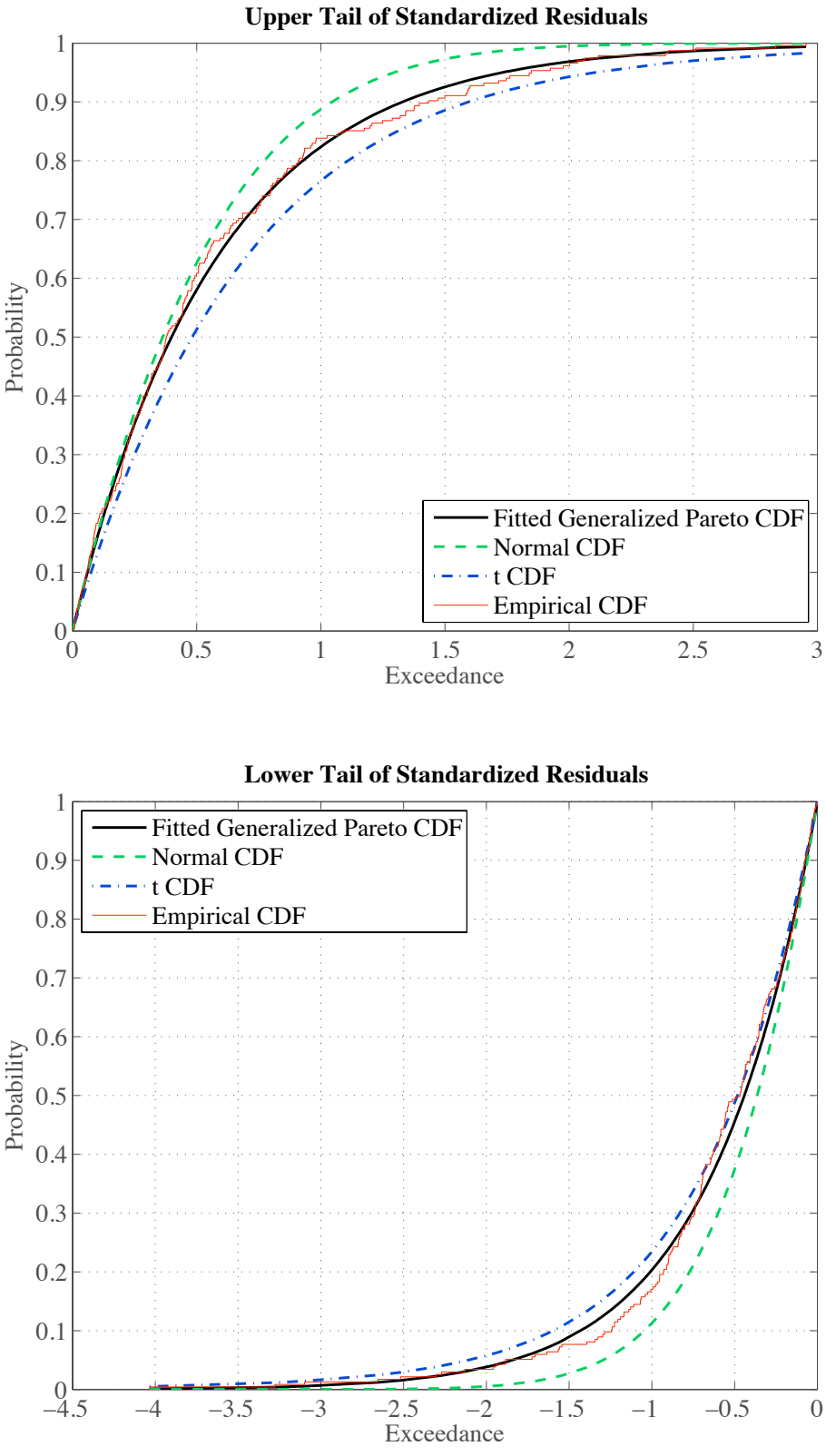

Figure 3: **Third principal component.** Upper and lower tails of standardized residuals.

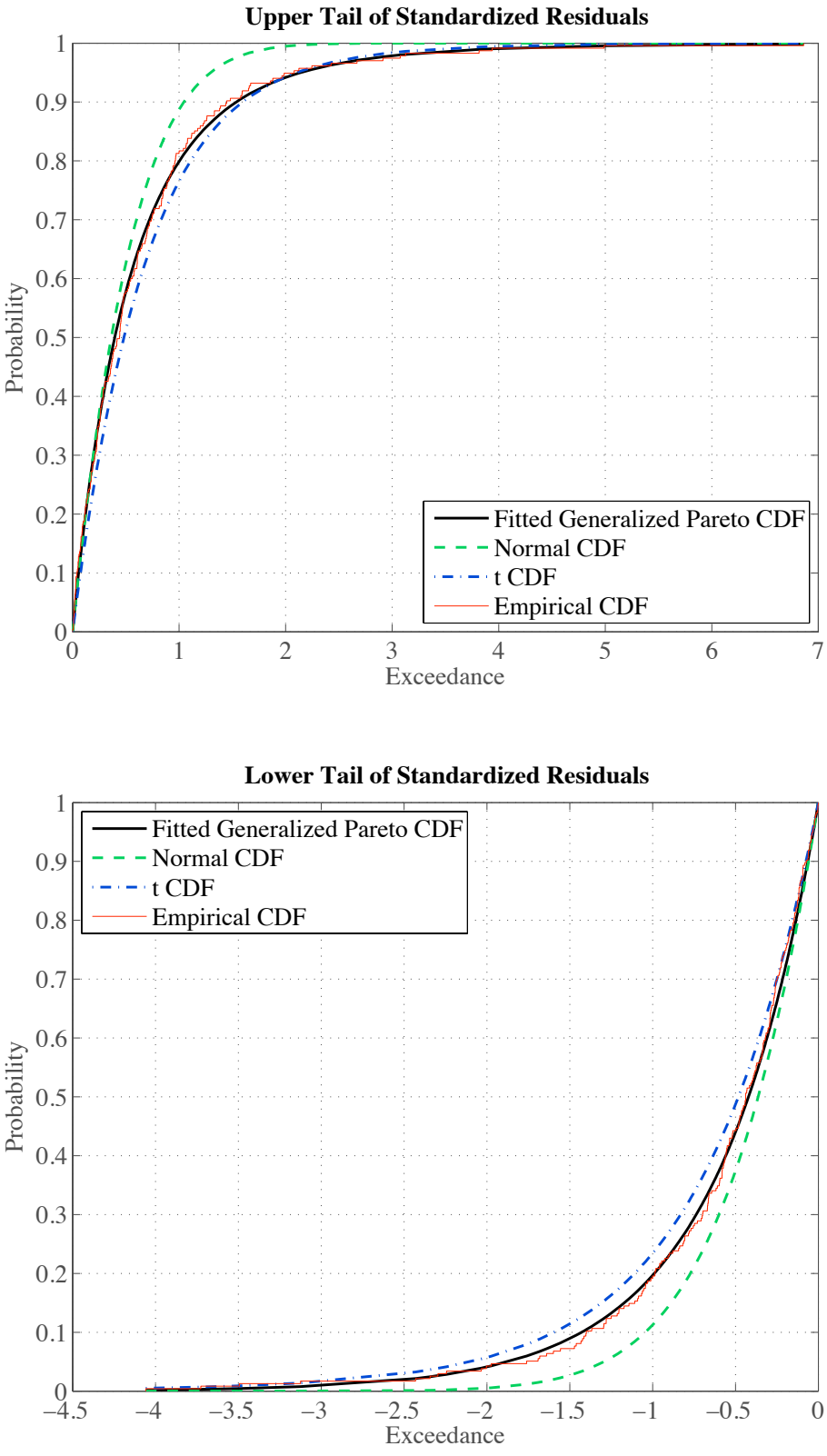

Figure 4: **Fourth principal component.** Upper and lower tails of standardized residuals.

Finally, I compute the one-day VaR and ES forecasts for an equally weighted portfolio of currencies using formulas (26) and (27), for January 1, 2008. The results are summarized in Table 7. The forecasts are given for the confidence levels of 90, 95, 99 and 99.9 percent. The values for VaR and ES are reported in percent. The lower- (upper-) tail values are applicable for the losses associated with holding a long (short) position in the portfolio. Evidently, the distributional asymmetry is reflected in pronounced differences between the risk measures in the upper and lower tail.

Table 7: **VaR and ES Forecasts**

One-day upper- and lower-tail VaR and ES forecasts for January 1, 2008, for an equally weighted portfolio of currencies and for several confidence levels.

Upper tail				
CL	0.90	0.95	0.99	0.999
VaR	0.5294	0.6947	1.0750	1.6101
ES	0.7284	0.8924	1.2694	1.7994

Lower tail				
CL	0.90	0.95	0.99	0.999
VaR	−0.4212	−0.5923	−0.9203	−1.2495
ES	−0.6790	−0.8251	−1.1026	−1.3743

4 Backtesting

Any risk management model needs to be tested before we can successfully apply it in practice. A variety of tests has been proposed to evaluate the accuracy of a VaR model. These tests are constructed to give an assessment of adequacy of the proposed models in predicting the size and frequency of losses. The standard backtests of VaR models compare the VaR forecasts for a given horizon with the actual portfolio losses. In its simplest form, the backtesting procedure consists of calculating the absolute

or relative number of times that the actual portfolio returns fall outside the VaR estimate, and comparing that number to the confidence level used.

Model backtesting is also important for financial institutions that are subjected to regulatory requirements. Since the late 1990s, regulatory guidelines require that banks with substantial trading activity have to set aside capital to insure against extreme portfolio losses. The size of the set-aside, or market risk capital requirement, is directly related to a measure of portfolio risk. In most of developed markets, the present regulatory framework follows the recommendations of Basel II, the second of the accords issued by the Basel Committee on Banking Supervision. The purpose of Basel II (initially published in June 2004) and its subsequent amendments was to create an international standard that can be used by national banking regulators. Currently, there are two general methodologies for assessment of market risk capital requirements under Basel II. The first one is the so-called Standardized Approach (SA), and is based on a set of simple rules on how to calculate minimum capital requirements using basic cross-sectional information about the assets in the bank's trading book. The more advanced approach is the Internal Models Method (IMM), which is based on VaR, and – being more precise – typically yields lower capital requirements. Specifically, the minimum capital requirement under IMM is defined as

$$\mathrm{MCR}_t := \max\left\{\mathrm{VaR}_t,\ (M+P_t)\overline{\mathrm{VaR}}\right\} + \mathrm{SRC}_t,$$

where VaR_t is the ten-days-ahead VaR forecast at 99 percent confidence level, $\overline{\mathrm{VaR}}$ is the average of these forecasts over the past 60 trading days, M is a multiplication factor set by national regulators (usually equal to 3), P_t is the penalty associated with the backtesting results, while SRC_t is the specific risk capital charge. The penalty P_t is determined by classifying the number of violations I of one-day 99-percent VaR in the previous 250 trading days into three distinct categories:

- $P_t = 0$, if $I \leq 4$ (green zone);
- $P_t = (I - 4)/5$, if $5 \leq I \leq 9$ (yellow zone);
- $P_t = 1$, if $10 \leq I$ (red zone).

Hence, a VaR model with more violations leads to a greater capital requirement.

The Basel II "traffic-light" approach to backtesting represents the only assessment of VaR accuracy prescribed in the current regulatory framework. Although its simple implementation is suitable for informational purposes, this approach merely counts the breaches of the 99-percent confidence level and fails to discard any model that, for example, overestimates the risk, or performs poorly when compared to other confidence levels. The ability of a backtest to discard all the models that systematically overstate as well as understate the risk is known as the unconditional coverage property. Christoffersen (1998) points out that the problem of determining the accuracy of a VaR model can be reduced to the problem of determining whether the sequence of breach counts satisfies both the unconditional coverage and independence. The latter property refers to intuition that the previous history of VaR violations must not convey any information about the future violations.

Some of the earliest VaR backtests proposed in the literature focused on the property of unconditional coverage, that is, whether or not the reported VaR is violated more or less than α percent of the time, where $1 - \alpha$ is the confidence level. Kupiec (1995), for example, proposed a proportion of failures (POF) test that examines how many times VaR forecasts are violated over a given span of time. If the number of violations differs significantly from α times the size of the sample, then the accuracy of the underlying risk model is called into question. Using a sample of T observations,

Kupiec (1995) test statistic takes the form,

$$\text{POF} := 2\ln\left[\left(\frac{1-\widehat{\alpha}}{1-\alpha}\right)^{T-I(\alpha)}\left(\frac{\widehat{\alpha}}{\alpha}\right)^{I(\alpha)}\right], \tag{30}$$

where

$$\begin{aligned}\widehat{\alpha} &:= \frac{I(\alpha)}{T},\\ I(\alpha) &:= \sum_{t=1}^{T} I_t(\alpha),\end{aligned}$$

where $I_t(\alpha)$ is an indicator function taking the value one if the actual return at t breaches the forecasted value of VaR for the confidence level determined by α, and zero if it stays within the VaR bounds. Hence, if the proportion of VaR violations, $\widehat{\alpha}$, is exactly equal to α then the POF statistic takes the value zero, indicating no evidence of any inadequacy in the underlying VaR measure. As the proportion of VaR violations differs from α, the POF statistic grows, indicating that the proposed VaR measure either systematically understates or overstates the underlying level of risk. The POF statistic given by (30) is a likelihood ratio and hence converges in distribution to a χ^2 with $I(\alpha)$ degrees of freedom.

Figure 5 shows the comparison between actual returns (dots) and VaR forecasts (continuous lines) for different confidence levels for an equally weighted portfolio of four currencies, for the period January 1, 2004 – September 30, 2008. The forecasts are computed using the formula (26), both for the lower and the upper tail of the return distribution in order to take into account losses both of a long and a short position, respectively. Table 8 summarizes the backtesting results, comparing the expected number of violations with the actual ones. The actual violations were compared across different multivariate models (EVT, conditional normality and conditional t-

distribution), as well as the univariate historical simulation. The multivariate normal and t models applied here follow the orthogonal GARCH approach of Alexander (2001). In other words, these forecasts were also obtained using the ARMA(1,0)–GJR-GARCH(1,1) filtering of principal components, except that the estimation of the covariance matrices was performed across the entire sample (thereby including both the center and the tails of the distribution) via ML method assuming normally- or t-distributed conditional residuals. The number of violations by quantiles clearly shows that HS markedly deviates from the expected values. The multivariate normal underestimates, while the multivariate t model overestimates the tail risk. At the same time, the multivariate EVT appears to yield much better forecasts. This is also verified formally by means of the Kupiec (1995) test, see Table 9. Clearly, all the models give predictions that are within statistically significant bounds for confidence levels of 90 and 95 percent, except for the normal model in the upper tail at 95 percent confidence level. However, the HS model performs poorly at all higher confidence levels, the multivariate t at 99 percent confidence, while the multivariate normal falls short in explaining the upper-tail returns above 99 percent level, and both the upper- and lower-tail extreme returns above 99.9 percent level. On the other hand, VaR forecasts based on the proposed multivariate EVT method violate the corresponding confidence bounds by a number of times that is not statistically different from the expected one. The only exception is perhaps the extreme confidence interval of 99.9 percent, where we observe no violations in the upper tail and one violation in the lower tail, compared to the expectation of 1.239, so for an appropriate sense of statistical significance at these extreme return regions we might need an even longer backtesting sample.

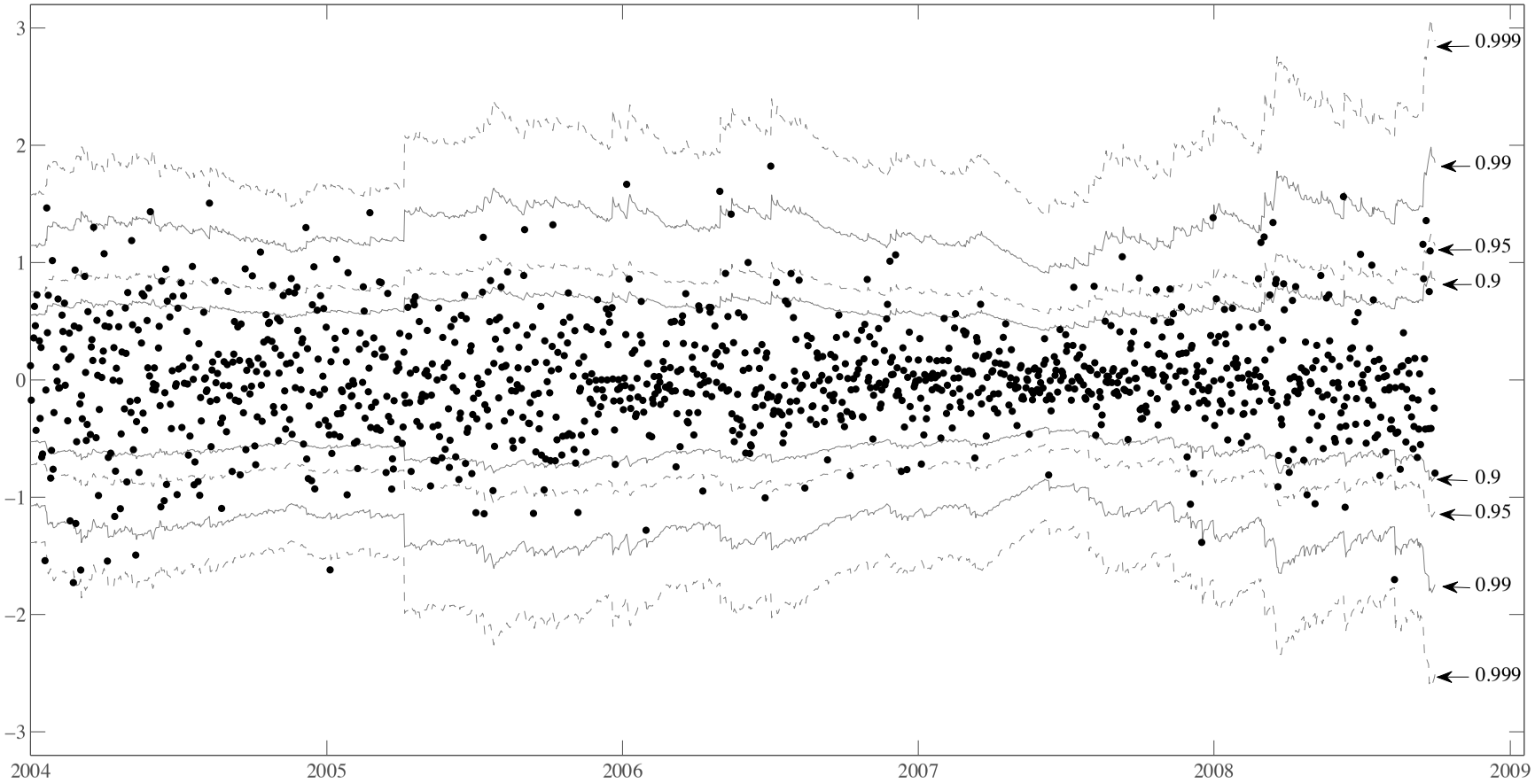

Figure 5: **VaR Backtesting of the Multivariate EVT Method.** Actual returns (dots) and VaR forecasts in the upper- and lower-tail at different confidence levels (continuous lines), for an equally weighted portfolio of currencies between January 1, 2004 and September 30, 2008.

Table 8: **VaR Backtesting: Violations by Quantiles**

Expected versus the actual number of violations obtained by several models for an equally weighted portfolio of currencies, between January 1, 2004 and September 30, 2008 (a total of 1239 observations).

Upper tail				
Method	Number of violations			
	CL = 0.90	CL = 0.95	CL = 0.99	CL = 0.999
EVT	109	56	12	0
Normal	144	89	31	16
t	121	49	11	0
HS	235	166	70	21
Expected	123.9	61.95	12.39	1.239

Lower tail				
Method	Number of violations			
	CL = 0.90	CL = 0.95	CL = 0.99	CL = 0.999
EVT	88	48	9	1
Normal	118	67	19	8
t	86	40	7	0
HS	219	142	54	28
Expected	123.9	61.95	12.39	1.239

(p-values in parentheses.)

Table 9: **VaR Backtesting: Kupiec Test**

The Kupiec (1995) POF test statistics and p-values obtained by several models for an equally weighted portfolio of currencies, between January 1, 2004 and September 30, 2008 (a total of 1239 observations).

Upper tail				
Method	Number of violations			
	CL = 0.90	CL = 0.95	CL = 0.99	CL = 0.999
EVT	2.0665	0.6207	0.0125	2.4792
	($< 10^{-4}$)	($< 10^{-4}$)	($< 10^{-4}$)	$--$
Normal	3.4620	11.0174	19.9238	52.5198
	($< 10^{-4}$)	(0.4564)	(0.0626)	(~ 1.0)
t	0.0759	3.0602	0.1637	2.4792
	($< 10^{-4}$)	($< 10^{-4}$)	(0.2022)	$--$
HS	90.1083	128.6208	129.9539	79.6643
	($< 10^{-4}$)	(0.0142)	(~ 1.0)	(~ 1.0)

Lower tail				
Method	Number of violations			
	CL = 0.90	CL = 0.95	CL = 0.99	CL = 0.999
EVT	12.7273	3.5725	1.0354	0.0494
	($< 10^{-4}$)	($< 10^{-4}$)	(0.0006)	(0.1760)
Normal	0.3167	0.4226	3.0626	16.3572
	($< 10^{-4}$)	($< 10^{-4}$)	($< 10^{-4}$)	(0.9625)
t	14.2719	9.3110	2.8099	2.4792
	($< 10^{-4}$)	($< 10^{-4}$)	(0.0980)	$--$
HS	67.6349	81.0498	77.1940	121.6632
	($< 10^{-4}$)	($< 10^{-4}$)	(0.9791)	(~ 1.0)

(p-values in parentheses.)

By examining a variety of different quantiles instead of a single one, some types of backtests can detect violations of the independence across a range of different VaR levels, while satisfying the unconditional coverage property. A variety of such tests has been proposed during the past decade, and Campbell (2005) gives a good review of these and other backtesting methods. An example of such a test is Pearson's test for goodness of fit. This test is based upon the number of observed violations at a variety of different VaR levels, separated into bins on the unit interval. The Pearson's test statistic is given by

$$Q := \sum_{k=1}^{K} \frac{\left(N_k^{\text{obs}} - N_k^{\text{exp}}\right)^2}{N_k^{\text{exp}}}, \tag{31}$$

where N_k^{obs} and N_k^{exp} are, respectively, the observed and the expected number of violations in the k-th bin. The Q statistic converges in distribution to a χ^2 with $K-1$ degrees of freedom, K being the number of bins. The results of the Pearson's test for the currency portfolio are summarized in Table 10, for the set of bins given by $\alpha \in [0.00, 0.001) \cup [0.001, 0.01) \cup [0.01, 0.05) \cup [0.05, 0.10) \cup [0.10, 1.00]$. They show that models based on conditionally normal or t-distributed residuals, as well as the HS model, can be rejected in favor of the proposed multivariate EVT alternative.

Table 10: **VaR Backtesting: Pearson's Test**

Pearson's test statistics and p-values obtained using VaR forecasts for an equally weighted portfolio of currencies between January 1, 2004 and September 30, 2008 (a total of 1239 observations). The partition of the unit interval used was $\alpha \in [0.00, 0.001) \cup [0.001, 0.01) \cup [0.01, 0.05) \cup [0.05, 0.10) \cup [0.10, 1.00]$.

Method	Lower tail	Upper tail
EVT	0.4170	1.3142
	(0.0189)	(0.1410)
Normal	38.5252	7.5773
	(~ 1.0)	(0.8917)
t	2.2298	2.0934
	(0.3064)	(0.2814)
HS	123.1067	146.7974
	(~ 1.0)	(~ 1.0)

(p-values in parentheses.)

5 Conclusion

This paper develops an efficient procedure for estimation of Value at Risk and expected shortfall based on a multivariate extreme value theory approach. The method is based on separate estimations of the univariate EVT model. It works with a set of orthogonal conditional residuals, obtained from the principal components of the joint return series. Autocorrelation, heteroskedasticity and asymmetry that are inherent in the original return series can be removed by assuming an ARMA process for the conditional mean and an asymmetric GARCH process for the conditional variance of the principal components. In this way, we can obtain a set of independent and identically distributed random variables, which is a prerequisite for any univariate EVT approach. The tails of the univariate distributions are modeled by a generalized Pareto distribution of peeks over treshold, while the interiors are fitted with an empirical distribution function. Furthermore, the method can be free of any unnecessary distributional assumption since the estimation of the ARMA-GARCH parameters can be performed via a generalized method of moments. Also, the method is free of estimation of a joint multivariate distribution, which would require a technique such as copula approach with simulations.

As an illustration, the method is applied to a sequence of daily interbank spot exchange rates of Euro, British Pound, Japanese Yen and Swiss Franc with respect to the U.S. Dollar. The forecasts of VaR and ES are backtested through a comparison with the actual losses over an out-of-the-sample period of four years and three quarters. The backtesting results indicate that the proposed multivariate EVT method performs well in forecasting the risk of a portfolio of four currencies. It certainly gives more precise estimate of VaR than the usual methods based on conditional normality, conditional t-distribution or historical simulation, while having the efficiency of an

orthogonal GARCH method.

References

Acerbi, C., & Tasche, D. 2002. Expected Shortfall: A Natural Coherent Alternative to Value at Risk. *Economic Notes*, **31**(2), 1–10.

Alexander, C. 2001. Orthogonal GARCH. *Pages 21–38 of:* Alexander, C. (ed), *Mastering Risk*, vol. 2. Pearson.

Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. 1997. Thinking Correctly. *Risk*, **10**(11), 68–71.

Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. 1999. Coherent Measures of Risk. *Mathematical Finance*, **9**(3), 203–228.

Balkema, A. A., & de Haan, L. 1974. Residual Life Time at Great Age. *Annals of Probability*, **2**, 792–804.

Bensalah, Y. 2000. *Steps in Applying Extreme Value Theory to Finance: A Review.* Working Paper, Bank of Canada.

Breuer, T., Jandacka, M., Rheinberger, K., & Summer, M. 2008. *Regulatory Capital for Market and Credit Risk Interaction: Is Current Regulation Always Conservative?* Working Paper.

Campbell, S. D. 2005. *A Review of Backtesting and Backtesting Procedures.* Finance and Economics Discussion Series, Federal Reserve Board, Washington, D.C.

Christoffersen, P. 1998. Evaluating Interval Forecasts. *International Economic Review*, **39**, 841–862.

da Silva, L. C. A., & de Melo Mendez, B. V. 2003. Value-at-Risk and Extreme Returns in Asian Stock Markets. *International Journal of Business*, **8**(1), 17–40.

Danielsson, J., & de Vries, C. 1997. *Value at Risk and Extreme Returns.* FMG Discussion paper No. 273, Financial Markets Group, London School of Economics.

Davison, A. C., & Smith, R. L. 1990. Models for Exceedances over High Tresholds. *Journal of the Royal Statistical Society B*, **52**, 393–442.

Dowd, K. 1998. *Beyond Value at Risk: The New Science of Risk Management.* Chichester: John Wiley & Sons.

Embrechts, P., Klüppelberg, C., & Mikosch, T. 2008. *Modeling Extremal Events for Insurance and Finance.* New York: Springer.

Fisher, R. A., & Tippett, L. H. C. 1928. Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. *Proceedings of the Cambridge Philosophical Society*, **24**, 180–190.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. 1993. On the Relation between Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance*, **48**, 1779–1801.

Gnedenko, B. V. 1943. Sur la Distribution limite du Terme maximum d'une Série aléatoir. *Annals of Mathematics*, **44**, 423–453.

Harmantzis, F., Chien, Y., & Miao, L. 2005. Empirical Study of Value-at-Risk and Expected Shortfall as a Coherent Measure of Risk. *Journal of Risk Finance*, **7**(2), 117–135.

Jondeau, E., & Rockinger, M. 2003. Testing for Differences in the Tails of Stock Market Returns. *Journal of Empirical Finance*, **10**, 559–581.

Jorion, P. 1996. *Value at Risk: The New Benchmark for Controlling Market Risk*. Chicago: Irwin.

Kupiec, P. 1995. Techniques for Verifying the Accuracy of Risk Management Models. *Journal of Derivatives*, **3**, 73–84.

Marinelli, C., d'Addona, S., & Rachev, S. T. 2007. A Comparison of Some Univariate Models for Value-at-Risk and Expected Shortfall. *International Journal of Theoretical and Applied Finance*, **10**(6), 1–33.

McNeil, A., & Frey, R. 2000. Estimation of Tail Related Risk Measure for Heteroscedastic Financial Time Series: An Extreme Value Approach. *Journal of Empirical Finance*, **7**, 271–300.

McNeil, A. J. 1997. Estimating the Tails of Loss Severity Distribution Using Extreme Value Theory. *Theory ASTIN Bulletin*, **27**(1), 1117–1137.

McNeil, A. J. 1999. *Extreme Value Theory for Risk Managers*. Working Paper, ETH Zürich.

Nelson, D. B. 1991. Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, **59**, 347–370.

Newey, W. K., & McFadden, D. 1994. Large Sample Estimation and Hypothesis Testing. *Chap. 36, pages 2111–2245 of:* Engle, R. F., & McFadden, D. (eds), *Handbook of Econometrics*, vol. 4. Amsterdam: North Holland.

Nyström, K., & Skoglund, J. 2002a. *A Framework for Scenario-Based Risk Management*. Working Paper, Swedbank.

Nyström, K., & Skoglund, J. 2002b. *Univariate Extreme Value Theory, GARCH and Measures of Risk*. Working Paper, Swedbank.

Picklands, J. 1975. Statistical Inference Using Extreme Order Statistics. *Annals of Statistics*, **3**, 119–131.

Skoglund, J. 2001. *A Simple Efficient GMM Estimator of GARCH Models*. Working Paper Series in Economics and Finance. Stockholm School of Economics.

Smith, R. L. 2000. Measuring Risk with Extreme Value Theory. *In: Risk Management: Theory and Practice*. Cambridge University Press.

von Mises, R. 1936. La Distribution de la plus grande de n Valeurs. *Pages 271–294 of: Selected Papers II, 1954*. Providence, RI: American Mathematical Society.