CENTRO STUDI BANCA E FINANZA

www.cefin.unimore.it

CEFIN Working Papers
No 3

The Maximum Lq-Likelihood Method: an Application to Extreme Quantile Estimation in Finance

by Davide Ferrari e Sandra Paterlini

June 2007

CEFIN – Centro Studi di Banca e Finanza
Dipartimento di Economia Aziendale – Università di Modena e Reggio Emilia
Viale Jacopo Berengario 51, 41100 MODENA (Italy)
tel. 39-059.2056711 (Centralino) fax 059 205 6927

The Maximum Lq-Likelihood Method: an Application to Extreme Quantile Estimation in Finance

Davide Ferrari
School of Statistics
University of Minnesota
dferrari@stat.umn.edu

Sandra Paterlini
Department of Economics
CEFIN - Centro Studi Banca e Finanza
paterlini.sandra@unimore.it

June 28, 2007

Abstract

Estimating financial risk is a critical issue for banks and insurance companies. Recently, quantile estimation based on Extreme Value Theory (EVT) has found a successful domain of application in such a context, outperforming other approaches. Given a parametric model provided by EVT, a natural approach is Maximum Likelihood estimation. Although the resulting estimator is asymptotically efficient, often the number of observations available to estimate the parameters of the EVT models is too small in order to make the large sample property trustworthy. In this paper, we study a new estimator of the parameters, the Maximum Lq-Likelihood estimator (MLqE), introduced by Ferrari and Yang (2007). We show that the MLqE can outperform the standard MLE, when estimating tail probabilities and quantiles of the Generalized Extreme Value (GEV) and the Generalized Pareto (GP) distributions. First, we assess the relative efficiency between the the MLqE and the MLE for various sample sizes, using Monte Carlo simulations. Second, we analyze the performance of the MLqE for extreme quantile estimation using real-world financial data. The MLqE is characterized by a distortion parameter q and extends the traditional log-likelihood maximization procedure. When $q \to 1$, the new estimator approaches the traditional Maximum Likelihood Estimator (MLE), recovering its desirable asymptotic properties; when $q \neq 1$ and the sample size is moderate or small, the MLqE successfully trades bias for variance, resulting in an overall gain in terms of accuracy (Mean Squared Error).

1 Introduction

Recent financial crises and the new regulations for banks and insurance companies[1] have prompted intermediaries to regularly compute statistical tail-related measures of risk. One of the most popular measures of financial risk is the Value-at-Risk (VaR), usually defined as the α-th quantile of the distribution of losses (negative returns). Although the appropriateness of VaR as a risk

[1] Basel II for banks, Solvency II for insurance companies and IFRS 32 and 39 for all financial companies.

measure (Artzner et al. (1999)) has been recently questioned, it is still the most widely used for risk management, asset allocation and risk-adjusted performance evaluation. Various methods have been proposed to estimate VaR: historical approach, parametric quantile estimators (e.g., Normal or t-Student parametric models), variance-covariance models and Monte Carlo methods are the most commonly used techniques. Recently, Extreme Value Theory (EVT) has found extensive application in finance to estimate tail-related risk measures, as it has been shown that it can provide estimators that perform best overall in predicting Value-at-Risk (Brooks et al. (2005), Kuester et al. (2006)).

EVT is supported by a sound statistical theory and it relies on the asymptotic properties of the distributions of sample extrema. Specifically, the two prevailing parametric approaches for modelling extreme events are the Peaks-Over-Threshold (POT) and Block Maxima (BM) methods. The POT method exploits the Generalized Pareto (GP) distribution for modelling the exceedances over a certain threshold, while the BM method relies on the Generalized Extreme Value (GEV) distribution to model the maximum value that a variable takes in a given period of time (block).

Although maximum likelihood is the most popular estimation approach in this context, mainly due to its asymptotic properties and ease of implementation[2], often the number of observations available to estimate GEV and GPD parameters is too small to guarantee the desirable large sample properties of the Maximum Likelihood Estimator (MLE); thus, inference might not be trustworthy. Our investigation aims to address this issue by studying for the first time in the EVT context the performance of a new estimator of the parameters, the Maximum Lq-Likelihood Estimator (MLqE), which has been recently proposed by Ferrari and Yang (2007). The MLqE is based on the information measure introduced by Havrda and Charvát (1967) and generalizes the traditional log-likelihood maximization procedure: it preserves the desirable asymptotic properties of the traditional MLE, while it allows for a peculiar type of distortion introduced by the extra parameter q, resulting in a gain in terms of precision (Mean Squared Error) when the sample size is moderate or small.

The objective of this paper is to study the behavior of the new estimator on both simulated data and on real-world time series for extreme quantile estimation. First, we show that the new estimator is more efficient than the standard MLE when the goal is to estimate the tail probability of the GP and GEV distributions. The comparison is carried out through Monte Carlo simulations, where the performance of the two estimators is evaluated for different choices of the tail

[2]Other methods include the method of moments, the method of probability-weighted moments and the elemental percentile method. The reader is referred to Hosking and Wallis (1987), Grimshaw (1993), Castillo et al. (1997).

probability and sample size. We show that when the distortion parameter q is properly chosen, the Mean Squared Error of the MLqE is sensibly smaller than that of MLE. Second, we focus on extreme quantile estimation, assessing the performance of MLqE on a financial stock market index for both GEV and GP distributions. The comparison with the MLE indicates that choices of the distortion parameter q smaller than 1 can dramatically reduce the generalization error.

The paper is organized as follows. In section 2, we describe the two main parametric approaches for risk estimation based on EVT; in section 3 we introduce the Maximum Lq-Likelihood Estimator. In section 4 we present a Monte Carlo simulation study to explore the relative efficiency between the MLqE and the MLE in a finite-sample situation. Section 5 describes a hold-out validation procedure applied to real-world financial data and compares the generalization error of the new estimator with that of MLE. Finally, in section 6 we outline the conclusions.

2 Extreme Value Theory for tail-related risk measures

Extreme Value Theory has found numerous applications in various fields (e.g., Lazar (2004)), including finance. The reader is referred to Embrechts et al. (1997), and Reiss and Thomas (1997) for an overview of the main applications in finance, while a brief description of the two main approaches, namely the Peaks-Over-Threshold and the Block Maxima, is reported below.

2.1 Peaks-Over-Threshold

The POT approach considers exceedances over a certain threshold u. Let $\{X_i, 1 \leq i \leq n\}$ be a random sample from a distribution F with mean μ and variance σ^2. An *exceedance* occurs when $X_i > u$ and an *excess over* u is defined by $y = x - u$. The conditional distribution of the exceedances over u, taken at $X > u$ is

$$F_u(y) = P(X - u \leq y | X > u) = \frac{F(u+y) - F(u)}{1 - F(u)},\ y \geq 0. \tag{1}$$

Balkema and de Haan (1974) showed that for a large class of distributions, $F_u(y) \to G(y)$ as $u \to \infty$ where $G(y)$ is a *Generalized Pareto (GP)* distribution. A representation of the GP distribution is

$$G(y;\xi,\sigma) = \begin{cases} 1 - \left(1 + \dfrac{\xi}{\sigma} y\right)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp(-y/\sigma), & \xi = 0, \end{cases} \tag{2}$$

with

$$y \in \begin{cases} [0, \infty), & \xi \geq 0, \\ [0, -\sigma/\xi], & \xi < 0. \end{cases}$$

The probability density function g is obtained by differentiating with respect to x:

$$g(x;\xi,\sigma) = \begin{cases} \sigma^{-1}\left(1+\xi\frac{(x-u)}{\sigma}\right)^{-(1/\xi+1)}, & \xi \neq 0, \\ \sigma^{-1}\exp\left(-(x-u)/\sigma\right), & \xi = 0. \end{cases} \tag{3}$$

The shape parameter ξ can be positive, negative or zero and provides an indication on the heaviness of the tail. The GP can represent different distributions depending on the value taken by ξ. In particular, when $\xi > 0$, we obtain the ordinary Pareto distribution which is suitable for modelling heavy tailed distributions such as financial returns. When $\xi = 0$ and $\xi < 0$ we have respectively the exponential and the Pareto II type distributions.
From eq.(1) one can obtain the following equality for values of x larger than u:

$$1-F(x) = (1-F(u))(1-F_u(x-u)). \tag{4}$$

Given a sufficiently high threshold value, $F_u(x-u)$ can be estimated using the plug-in estimate based on GP distribution and $F(u)$ can be estimated using the sample proportion of observations. Thus, from eq.(4) one can write the tail estimator of $F(x)$. Inverting the expression for the tail gives the estimating equation of the Value-at-Risk[3]

$$\widehat{VaR}_{1-\alpha} = u + \frac{\hat{\sigma}}{\hat{\xi}}\left(\frac{n}{N_u}\alpha^{-\hat{\xi}} - 1\right), \tag{5}$$

where N_u denotes the observed number of exceedances over the threshold u. The reader is referred to McNeil et al. (2005) for a complete mathematical treatment of the POT approach for Value-at-Risk estimation.
Note that the asymptotic result poses some applicability constraints. In fact, the threshold u has to be large in order the Generalized Pareto approximation to hold; as a consequence, few exceedances would left. Thus, if an excessively high threshold is chosen, the plug-in estimator might be inaccurate with high variance. Furthermore, the asymptotic properties of the Maximum Likelihood estimator would hardly hold. Conversely, a low threshold would inevitably induce bias.

2.2 Block Maxima

The BM method models the maximum value that a variable takes in a given period of time (block). Consider a random variable X with cumulative distribution function $F(x)$ with mean

[3]The Value-at-Risk is usually defined as the α-th quantile of the distribution of losses, or the negative returns. Namely, $\text{VaR}_{1-\alpha} := \inf\{x \in \Re : P(X > x) \leq \alpha\}$ where X is a real-valued random variable representing losses or negative returns and $0 \leq \alpha \leq 1$. Typically, values of interest for α are 0.05 and 0.01.

μ and variance σ^2. Let $\{Y_i, 1 \le i \le n\}$ be a random sample from the standardized distribution $F\left(\frac{x-\mu}{\sigma}\right)$ and define[4]

$$Y_{n,n} = \max\{Y_1, Y_2, ..., Y_n\}.$$

In addition, let $\{a_n; n \ge 1\}$ and $\{b_n \ge 0; n \ge 1\}$ be sequences of numbers such that

$$P\left(\frac{Y_{n,n} - a_n}{b_n} \le x\right) \to G(y), \tag{6}$$

as $n \to \infty$ for some non-degenerate distribution G. Fisher and Tippett (1928), and Gnedenko (1943) showed that G belongs to one of the following three extreme value distributions:

Gumbel: $\Lambda(y) = \exp(-\exp(-y)), -\infty \le y \le \infty$

Fréchet: $\Phi(y;\alpha) = \begin{cases} 0, & y \le 0 \\ \exp(-y^{-\alpha}), & y > 0, \quad \alpha > 0 \end{cases}$

Weibull: $\Psi(y;\alpha) = \begin{cases} \exp(-(-y)^{\alpha}), & y \le 0 \quad \alpha > 0 \\ 1, & y > 0. \end{cases}$

Later, Jenkinson (1955) and von Mises (1954) suggested a re-parametrization of the above expressions by setting $\xi = \alpha^{-1}$ for the Fréchet distribution and $\xi = -\alpha^{-1}$ for the Weibull distribution. Thus, Gumbel, Fréchet and Weibull can be represented in a unified parametric model, known as the Generalized Extreme Value distribution (GEV), where ξ represents the shape parameter and gives an indication about the heaviness of the tail of the distribution.
The following characterization, which includes also the location and scale parameters μ and σ, is most commonly used:

$$H(x;\xi,\mu,\sigma) = \begin{cases} \exp\left[-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right], & \text{if } \xi \ne 0, \quad 1+\xi\frac{x-\mu}{\sigma} > 0 \\ \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right], & \text{if } \xi = 0. \end{cases} \tag{7}$$

The probability density function is then:

$$h(x;\xi,\mu,\sigma) = \begin{cases} \frac{1}{\sigma}\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi-1}\exp\left(-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{1/\xi}\right) & \text{if } \xi \ne 0, \\ \frac{1}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}\right)\exp\left(\exp\left(-\frac{x-\mu}{\sigma}\right)\right), & \text{if } \xi = 0, \end{cases} \tag{8}$$

[4]We could study as well the minimum rather than the maximum and the results for one of the two can be immediately transferred using the relationship $Y_{1,n} = -max\{-Y_1, -Y_2, ..., -Y_n\}$.

with $1+\xi\frac{x-\mu}{\sigma}>0$. We remark that the asymptotic results just described only guarantee that Y is *approximately* distributed according to a GEV distribution. Hence, the accuracy of such an approximation relies strongly on the size of the blocks from which the maxima are computed.

The block maxima approach allows to compute the so-called return-level, that is the level expected to be exceeded in one out of the k periods of length n. Given a block size large enough to hold the GEV approximation, the return level can be computed by inverting eq.(7) and thus obtaining

$$U_k = H^{-1}\left(1-\frac{1}{k};\xi,\sigma,\mu\right). \tag{9}$$

Substituting the parameter estimates, we have

$$\widehat{U}_k = \begin{cases} \hat{\mu}-\frac{\hat{\sigma}}{\hat{\xi}}\left(1-\left(-\log\left(1-\frac{1}{k}\right)\right)^{-\hat{\xi}}\right), & \text{if } \hat{\xi}\neq 0, \\ \hat{\mu}-\hat{\sigma}\log\left(-\log\left(1-\frac{1}{k}\right)\right), & \text{if } \hat{\xi}=0. \end{cases} \tag{10}$$

3 The Maximum L_q-Likelihood Method

Let $f(x;\boldsymbol{\theta_0})$ be the GP density in eq.(3) or the GEV density in eq.(8), where $\boldsymbol{\theta_0}=(\theta_{01},\ldots,\theta_{0p})\in\Theta$ denotes the vector of parameters to estimate (p=2 for GP and p=3 for GEV). Given a random sample $X_1,\ldots,X_n$ from $f(x;\boldsymbol{\theta_0})$, the Maximum Likelihood Estimator is

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta}\sum_{i=1}^{n}\log\left[f(X_i;\theta)\right]. \tag{11}$$

Maximum Likelihood is the standard approach in parametric estimation, mainly due to the desirable asymptotic properties of consistency, efficiency and normality. In particular, under some regularity conditions (e.g., see van der Vaart (1998), Ferguson (1996)), we have that $\sqrt{n}\left(\widehat{\theta}_n-\boldsymbol{\theta_0}\right)\xrightarrow{\mathscr{D}}N(0,V)$ as $n\to\infty$, where V represents the inverse of the Fisher information matrix.

Note that the asymptotic result is valid under the assumption that the underlying distribution is actually one of the extreme value distributions. However, the results presented in the previous section guarantee that the block maxima and the excesses over a threshold are only *approximately* from GEV and GP distributions. Thus, two contrasting sources of distortion characterize the estimation of the tail probability. The first concerns the limit results for the tail quantities. In the POT method we need an increasingly high threshold, u, in order to guarantee the convergence to the GP distribution; similarly, in the BM method a large block size is necessary in order to hold the GEV distribution. The second issue deals with the sample size necessary to make the asymptotic properties of MLE trustworthy, especially when the goal is to estimate small tail probabilities. Clearly, if we choose higher thresholds or larger block sizes, the number of available observations for ML estimation will be too small.

Recently, in order to handle the second issue, Ferrari and Yang (2007) introduced an estimator inspired to Havrda and Charvát (Havrda and Charvát (1967)) generalized information measure[5], the Maximum L_q-Likelihood Estimator (MLqE). The MLqE of $\boldsymbol{\theta_0}$ is defined as

$$\widetilde{\theta}_n = \arg\max_{\theta\in\Theta} \sum_{i=1}^{n} L_q\left[f(X_i;\theta)\right], \tag{12}$$

where

$$L_q(z) = \begin{cases} \dfrac{z^{1-q}-1}{1-q} & if\ q \neq 1, \\ \log z & if\ q = 1. \end{cases} \tag{13}$$

The function L_q represents a Box-Cox transformation in statistics and in other contexts it is often called deformed logarithm of order q. The estimates of the parameters are computed by solving the following system of equations:

$$\sum_{i=1}^{n} \frac{\partial}{\partial\theta_j} L_q\left[f(X_i;\theta)\right] = 0,\ j = 1,2,...,p. \tag{14}$$

When q is a fixed constant, $\widetilde{\theta}_n$ belongs to the class of *M-estimators*. Under some regularity conditions such estimators have well known asymptotic proprieties such as asymptotic normality (e.g., see van der Vaart (1998) and Huber (1981)).

The MLqE can be considered as a generalization of the traditional MLE. For values of q arbitrarily close to 1, we have that $L_q(\cdot) \to \log(\cdot)$ and the MLqE approaches the classical MLE. However, an advantage is obtained by having q slightly different from 1: in this situation the MLqE allows trading bias for variance and provides more accurate estimates when the sample size is small. A $q \neq 1$ corresponds to assign a different weight to the observations in the sample based on the rarity of their occurrence. In particular, when $q < 1$ the role played by extreme observations, which are the most influential on the estimates, is reduced. Consequently, when setting $q < 1$ the variability is reduced by increasing the bias, which can result in an overall gain in terms of Mean Squared Error, as we shall see. Conversely, if $q > 1$ the role of the observations corresponding to density values close to zero is accentuated (Ferrari and Yang (2007)).

[5] Such information measures, usually called α-order entropies (or q-entropies in physics), relax the additivity assumption that characterizes Shannon's information. In recent years α-order entropies have found successful applications in different fields, such as finance, biomedical sciences, environmental sciences and linguistics (e.g., see Gell-Mann (2004).

In the context of the class of distributions belonging to the exponential family, Ferrari and Yang (2007) derive the asymptotic properties of the MLqE. They show that the peculiar type of distortion introduced allows to gain in terms of precision (Mean Squared Error) by reducing the variance when both the sample size and the tail probability to be estimated are small. Conversely, when the sample size is large, reducing the amount of bias allows for the recovery of a number of desirable large sample properties such as efficiency and consistency. Hence, the MLq procedure extends the classic method resulting in a general inferential procedure that inherits most of the desirable features of traditional maximum likelihood methods and at the same time gains some new properties that can be usefully exploited in *ad hoc* estimation settings. The following sections report empirical results supporting the use of such estimator in the EVT framework.

4 Finite-sample efficiency of MLqE: Monte Carlo simulations

In this section we compare the relative efficiency between the MLqE and the MLE on simulated data from both GEV and GP distributions[6]. Our first aim is to investigate whether the MLqE can outperform, in terms of Mean Squared Error, the classical MLE when estimating small tail probabilities. The estimates of the tail probability are obtained by using the so-called *plug-in* approach, where the point estimate of the unknown parameter is substituted into the distribution of interest.

Let $F(x;\theta)$ be the cumulative distribution function for either GEV or GP distributions. The true parameter is denoted by $\theta_{\mathbf{0}}$ and the true tail probability by α (in particular, $\alpha = 1 - F(x;\theta_{\mathbf{0}})$ if the right tail is considered, and $\alpha = F(x;\theta_{\mathbf{0}})$ otherwise). Further, let $\widehat{\alpha}_n$ and $\widetilde{\alpha}_n$ be the plug-in estimates of α, obtained respectively via the ML and the MLq methods.

The relative performance of the two estimators is measured by taking the ratio between the two Mean Squared Errors:

$$R_n = \frac{MSE\,(\widehat{\alpha}_n)}{MSE\,(\widetilde{\alpha}_n)} = \frac{E(\widehat{\alpha}_n - \alpha)^2}{E(\widetilde{\alpha}_n - \alpha)^2} = \frac{(E(\widehat{\alpha}_n) - \alpha)^2 + Var(\widehat{\alpha}_n)}{(E(\widetilde{\alpha}_n) - \alpha)^2 + Var(\widetilde{\alpha}_n)}. \tag{15}$$

As pointed out by the error decomposition in the above expression, we are interested in the relative trade-off between bias and variance of the two estimators, for *a given sample size*. The simulations are then carried out as follows:

[6]The analyses presented in sections 4 and 5 are performed using the statistical computing envirnonment R (R Development Core Team (2006)). In the routines described we utilize functions from the Extreme Value Theory package *evir* (McNeil and Stephenson (2007))

- For any given sample size n, a number $B = 1000$ of random samples $X_1, ..., X_n$ are generated from either GEV or GP with parameter vector $\theta_{\mathbf{0}}$.

- For each sample, $\widehat{\alpha}_{n,b}$ and $\widetilde{\alpha}_{n,b}$, $b = 1, ..., B$, the ML and MLq estimates of the tail probability α are obtained. The estimates of the parameters for both estimators are computed by solving numerically the L_q-likelihood equations (14). The optimization is performed by using a variable metric algorithm (e.g., see Givens and Hoeting (2005)), where the MLE estimates $\widehat{\theta}_{n,b}$ are chosen as starting values.

- Finally, the relative performance between the two estimators is evaluated by the ratio

$$\widehat{R}_n = \frac{\widehat{\mu}}{\widetilde{\mu}} = \frac{\sum_{k=1}^{B}(\hat{\alpha}_{n,k} - \alpha)^2/B}{\sum_{k=1}^{B}(\widetilde{\alpha}_{n,k} - \alpha)^2/B}$$

  where $\widehat{\mu}$ and $\widetilde{\mu}$ represent the Monte Carlo estimates of the Mean Squared Error for MLE and MLqE, respectively. Furthermore, the standard error of $\widehat{R}_n$ is computed via the multivariate Delta Method as

$$se\left(\widehat{R}_n\right) = B^{-1/2}\left(\frac{\widehat{\sigma}_{11}}{\widetilde{\mu}^2} - 2\widehat{\sigma}_{12}\frac{\widehat{\mu}}{\widetilde{\mu}^3} + \widehat{\sigma}_{22}\frac{\widehat{\mu}^2}{\widetilde{\mu}^4}\right)^{1/2}$$

  where $\widehat{\sigma}_{11}$, $\widehat{\sigma}_{22}$ and $\widehat{\sigma}_{12}$ denote respectively the Monte Carlo estimates for the variances and the covariance of the squared errors (see Appendix 1 for the details of the calculation).

The procedure described above is repeated for several samples sizes (ranging from 5 to 200) and different choices of the true tail probability α and the distortion parameter q. The simulations discussed in the remainder of this section are obtained by sampling from a GEV distribution with parameters

$$\theta_{\mathbf{0}} = (\xi_0, \mu_0, \sigma_0) = (0.1, 0.05, 0.015),$$

and from a GP distribution with parameters

$$\theta_{\mathbf{0}} = (\xi_0, \sigma_0) = (0.5, 1).$$

We remark that the parameter values[7] are comparable in size to the estimates for various stock indexes computed by Gilli and Kellezi (2006) and McNeil et al. (2005). Nevertheless, we also performed simulations using other parameter settings, obtaining similar results.

[7]The value of the shape parameter ξ, which determines heaviness of the tail, is critical for both GEV and GP distributions. Since financial returns are have usually heavy tailed distributions (Cont (2001)), they can be suitably represented by considering $\xi > 0$.

Figures 1 and 2 show the results for the GP distribution. In particular, figure 1 shows the performance of the MLqE when q is 0.94 for different values of the tail probability α. For small and moderate sample sizes, we have that $\widehat{R}_n > 1$ and the MLqE is clearly more accurate than MLE. From figure 2 we can see that MLq estimates are more precise not only for small but even for larger sample sizes (up to 200). Moreover, for a given tail probability the gain is more accentuated when q is smaller.

Figures 3 and 4 present the case of the GEV distribution. Similarly to the GP distribution, figure 3 points out that MLqE is more accurate than the MLE for moderate or small sample sizes. Moreover, the gain appears to be more evident for smaller values of α. Actually, note that when α is 0.05, the MLqE outperforms the MLE in accuracy only for sample sizes smaller than 80, while this is not the case when α equal to 0.01. In figure 4 we can see that the relative performance of MLqE versus MLE improves when the tail size becomes smaller ($\alpha = 0.005$) and the parameter q decreases from 0.95 to 0.93. Recall that decreasing the distortion parameter q is equivalent to downweighting extreme observations that can be dramatically influential on the accuracy of the estimates when the size of α is small.

In general, if q is fixed, it is important to note that as the sample size gets larger, the bias component of the error becomes more relevant than the variance component and the MLE will always tend to dominate MLqE due to its asymptotic properties. This observation has suggested that a value of q closer to 1 should be preferred when the sample size increases.

5 Forecasting financial empirical quantiles

The simulation results have encouraged a further study on real-world financial data, where Extreme Value Theory plays a crucial role in forecasting the empirical quantiles. The analyses presented in the following sections have been carried out on publicly available financial data[8]: the daily log-returns of the Standard & Poor's 500 index (S&P500) from January 1960 to June 1993. Extreme value analysis on these data set has been previously discussed in literature (e.g., see McNeil and Frey (2000), and Knight et al. (2005)). The summary statistics for this data set are reported in table 1. This data set presents features that commonly characterize

Table 1: Descriptive statistics of the log-return series of S&P500 index.

Sample Size	Min	Max	Mean	St.Dev.	Skewness	Kurtosis
8414	-20.388	9.099	0.028	0.871	-1.510	44.300

[8]http://www.ma.hw.ac.uk/ mcneil/data.html

the distribution of financial log-returns. In particular, note that the distribution of returns for the S&P500 index is remarkably skewed. In the remainder of this section we consider the commonly employed hold-out procedure to estimate the generalization error of the estimates. We use such a measure (i) to compare the relative performance between MLqE and MLE when predicting empirical quantiles of one the extreme value distributions (GEV or GP); (ii) to study the performance of MLqE, relatively to the tail size α and the distortion parameter q.

5.1 Hold-out validation procedure

The comparison between the MLqE and the MLE is carried out using an estimate of the generalization error (Hastie et al. (2001)), obtained via a *repeated hold-out* procedure. First, from the original dataset of the log-returns we take the block maxima (for the BM model) or the exceedances over a certain threshold (for the POT model). Then, on the filtered data, the following steps are performed:

(i) The data are randomly divided into a training set of size $n^{(tr)}$ and a testing set of size $n^{(ts)} = n - n^{(tr)}$, where n is the size of the filtered sample. The training and testing samples are chosen such that $n^{(ts)} = n^{(tr)}$.

(ii) The ML and MLq estimates of the quantile τ, denoted by $\widehat{\tau}^{(tr)}$, are computed from the training set.

(iii) The sample quantile, $t^{(ts)}$, is computed from the testing set.

Steps (i),(ii) and (iii) are repeated for $B = 500$ times and then the performance of the estimator is evaluated by

$$\widetilde{\mathscr{E}} = B^{-1} \sum_{b=1}^{B} \left(\widehat{\tau}_b^{(tr)} - t_b^{(ts)} \right)^2. \tag{16}$$

Finally, the standard error of $\widetilde{\mathscr{E}}$ is calculated using nonparametric bootstrap, based on 2000 replications. The analysis is carried out for both left and right tails of the distributions of returns.

5.2 Empirical results on financial Data

In the filtering phase, 100 observations are extracted from the S&P500 log-return time series. In the BM model, the original sample is divided in $n = 100$ blocks, obtaining a block size reasonably large in order the GEV asymptotic approximation to apply (e.g., see Gilli and Kellezi (2006)). In the POT model, although some data driven procedures have been proposed (Lazar

(2004)), there seems to be no universal agreement on the choice of the threshold value to employ. However, Monte Carlo studies (e.g., see McNeil and Frey (2000)) have shown that for heavy tailed distributions a threshold corresponding to about $n = 100$ exceedances, performs well in terms of Mean Squared Error[9].

Tables 2 and 3 report the empirical results for the BM and POT approaches, for different quantiles and choices of the distortion parameter. Column 3 and 5 report the generalization error $\widetilde{\mathscr{E}}$ for the left tail and the right tail, while columns 4 and 6 report the percent gain (or loss) in terms of prediction error of the MLqE over that of MLE. The results for the MLE are reported in the row corresponding to $q = 1$, since the two estimators are the same for such a value.

For the BM method, a substantial improvement is obtained when $q < 1$. In all the cases, the improvement is relevant when the distortion parameter decreases to $q = 0.95$. Furthermore, we notice that the gain deriving from the MLq method is more evident on left tail, which is usually of major interest in risk analysis as it represents the losses. Actually, it is known that equity times series usually show a loss/gain asymmetry (Cont (2001)) with left-skewed distributions, as shown in table 1 for the data sets under exams. Finally, as expected, as the distortion parameter approaches 1, the usual MLE is recovered and the performance of the two estimators becomes similar.

Table 2 shows the results corresponding to the POT method. The analysis on the left tail confirms the considerations previously discussed for the BM method. However, the performance on the right tail shows only little or no improvement with respect the standard approach, when considering the 90th percentile. Nevertheless, the analysis clearly points out that the MLqE can be considered as a valid alternative to the MLE when computing the value at risk of financial losses, especially if interested in estimating extreme quantiles.

6 Discussion and Final Remarks

In this work, we have shown that the MLqE can be a valid alternative to the classical MLE when estimating a small tail probability or a large quantile in the context of Extreme Value Theory. The MLqE can be regarded as a natural extension of the classical MLE. Specifically, the distortion parameter q allows to adjust the relative weight of the information provided by each observation in the sample. If q is close to 1, the estimator preserves the large sample properties of the MLE, while for $q \neq 1$ the trade-off between bias and variance is modified, producing an overall gain in terms of accuracy (Mean Squared Error) when the sample size and/or the tail

[9]This choice is also confirmed by preliminary exploratory analyses carried out by using the graphical tools contained in the R package POT Ribatet (2006).

Table 2: Block Maxima method. The squared error, $\widetilde{\mathscr{E}}$, is computed for $q = 1, 0.995, 0.975$ and 0.95 (where $q = 1$ corresponds to the MLE) and considering two choices of the tail size. In parenthesis, the bootstrap standard error of $\widetilde{\mathscr{E}}$, computed from 2000 replicates. The percent gain is computed as $(\widetilde{\mathscr{E}}_{MLE}/\widetilde{\mathscr{E}}_{MLqE} - 1) \times 100$.

Percentile	q	Left Tail		Right Tail	
		$\widetilde{\mathscr{E}}$	% Gain	$\widetilde{\mathscr{E}}$	% Gain
	1.000	0.3836(0.0332)	/	0.2759(0.0248)	/
90th	0.995	0.3642(0.0320)	5.3155	0.2716(0.0244)	1.5981
	0.975	0.2981(0.0273)	28.6603	0.2565(0.0227)	7.5647
	0.950	0.2373(0.0231)	61.6476	0.2429(0.0210)	13.5677
	1.000	1.3706(0.1135)	/	0.5583(0.0531)	/
95th	0.995	1.3213(0.1128)	3.7305	0.5449(0.0523)	2.4618
	0.975	1.1594(0.1025)	18.2168	0.4971(0.0478)	12.3238
	0.950	1.0239(0.0953)	33.8575	0.4505(0.0432)	23.9352

probability to estimate are small. Such settings are typical in finance, where the attention is often on estimating very small probabilities with a small number of extrema. Although we have considered the MLqE for the specific purpose of Extreme Value Theory estimation, this stream of research seems to be very promising, due to the considerable flexibility of the new estimator to many classical estimation settings and its finite-sample variance reduction properties.

The simulation study has pointed out that the MLqE is more accurate than MLE in estimating tail probabilities for GEV and GP distributions for relatively small and moderate sample sizes. The gain from the MLqE appears to be more remarkable when the target tail probability is smaller. When the sample size is too large relative to the choice of the distortion parameter q, the bias component plays an increasingly relevant role and eventually we observe that the MLqE decreases its accuracy. This indicates that the distortion parameter should approach 1 as the sample size increases in order to preserve the efficiency gain. In addition, smaller values of the distortion parameter q enhance the accuracy attainable in small sample situation by reducing the role played by extreme (and more influential) observations. The findings from the simulation study are also confirmed by the empirical analysis on financial data. We show that for more extreme target quantiles, the MLqE achieves a superior performance in terms of

Table 3: Peaks-Over-Threshold method. Squared error, $\widetilde{\mathscr{E}}$, for $q = 1, .995, .975$ and $.95$ (when $q = 1$ we are computing the MLE) and two choices of the tail size. In parenthesis, the bootstrap standard error of $\widetilde{\mathscr{E}}$, computed from 2000 replicates. The percent gain is computed as $(\widetilde{\mathscr{E}}_{MLE}/\widetilde{\mathscr{E}}_{MLqE} - 1) \times 100$.

Percentile	q	Left Tail $\widetilde{\mathscr{E}}$	Left Tail % Gain	Right Tail $\widetilde{\mathscr{E}}$	Right Tail % Gain
90th	1.000	1.4190(0.1307)	/	0.522(0.0622)	/
	0.995	1.3627(0.1277)	4.1327	0.522(0.0618)	0.0094
	0.975	1.1630(0.1107)	22.0158	0.5233(0.0623)	-0.2430
	0.950	0.9671(0.0933)	46.7225	0.5285(0.0631)	-1.2296
95th	1.000	7.0898(0.6909)	/	1.7555(0.2685)	/
	0.995	6.7981(0.6763)	4.2903	1.7452(0.2666)	0.5887
	0.975	5.7576(0.5813)	23.1364	1.7103(0.2757)	2.6421
	0.950	4.7161(0.4806)	50.3295	1.6814(0.2883)	4.4060

generalization error, when the distortion parameter q is chosen to be smaller than 1.
Even if the arbitrariness of the choice of q could be one of the main critics of the new method, we believe that the main strength of the MLqE derives from the flexibility gained from the choice of such a parameter and further work need to be focused on this issue. Currently, two research directions are under investigation on the choice of q: (i) theoretical derivation of optimal values of q based on asymptotic theory and (ii) data-driven regularization procedures such as cross-validation.

Appendix: Delta Method Calculation

Consider α, $\widehat{\alpha}_{n,b}$ and $\widetilde{\alpha}_{n,b}$ defined as in section 4. Moreover, let $x_B = B^{-1}\sum_{b=1}^{B}(\hat{\alpha}_{n,b} - \alpha)^2$ and $y_B = B^{-1}\sum_{b=1}^{B}(\widetilde{\alpha}_{n,b} - \alpha)^2$. By the central limit theorem, for large values of B we have that

$$\sqrt{B}\begin{bmatrix} x_B \\ y_B \end{bmatrix} \xrightarrow{\mathscr{D}} N\left(\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}\right), \tag{17}$$

where $\mu_1 = MSE(\hat{\alpha}_n)$ and $\mu_2 = MSE(\widetilde{\alpha}_n)$. We are interested in the limiting distribution of $g(x_B, y_B) = x_B/y_B$ when $B \to \infty$. By the Delta Method (e.g., see Ferguson (1996)) we have that

$$\sqrt{B}\, g(x_B, y_B) \xrightarrow{\mathscr{D}} N\left(g(\mu), \overset{\bullet}{g}(\mu)^T \Sigma\, \overset{\bullet}{g}(\mu)\right), \text{ as } B \to \infty \tag{18}$$

where $\overset{\bullet}{g}(\cdot)$ is the gradient. In this case we have that

$$\overset{\bullet}{g}(\mu)^T = \left(\frac{\partial}{\partial \mu_1} g(\mu), \frac{\partial}{\partial \mu_2} g(\mu)\right)^T = \left(\frac{1}{\mu_1}, -\frac{\mu_1}{\mu_2^2}\right), \tag{19}$$

and

$$\begin{aligned} \overset{\bullet}{g}(\mu)^T \Sigma\, \overset{\bullet}{g}(\mu) &= \left(\frac{1}{\mu_1}, -\frac{\mu_1}{\mu_2^2}\right) \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \begin{pmatrix} 1/\mu_1 \\ -\mu_1/\mu_2^2 \end{pmatrix} \\ &= \frac{\sigma_{11}}{\mu_2^2} - 2\sigma_{12}\frac{\mu_1}{\mu_2^3} + \sigma_{22}\frac{\mu_1^2}{\mu_2^4}. \end{aligned}$$

Therefore, we obtained that

$$\sqrt{B}\left(\frac{x_B}{y_B}\right) \xrightarrow{\mathscr{D}} N\left(\frac{\mu_1}{\mu_2}, \frac{\sigma_{11}}{\mu_2^2} - 2\sigma_{12}\frac{\mu_1}{\mu_2^3} + \sigma_{22}\frac{\mu_1^2}{\mu_2^4}\right), \text{ as } B \to \infty. \tag{20}$$

References

P. Artzner, F. Delbaen, J. M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

A. A. Balkema and L. de Haan. Residual life time at great age. *The Annals of Probability*, 2: 792–804, 1974.

C. Brooks, J. Clare, J. Dalla Molle, and G. Persand. A comparison of extreme value theory approaches for determining value at risk. *Journal of Empirical Finance*, 22:1–22, 2005.

E. Castillo, J. María, and A. S. Hadi. Fitting continuous bivariate distributions to data. *The Statistician: Journal of the Institute of Statisticians*, 46:355–369, 1997.

R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1,2:223–236, 2001.

P. Embrechts, C. Kluppelberg, and T. Mikosch. *Modelling extremal events for Insurance and Finance*. Applications of Mathematics. Springer, 1997.

T. S. Ferguson. *A Course in Large Sample Theory*. Chapman & Hall Ltd, 1996.

D. Ferrari and Y. Yang. Estimation of tail probability via the maximum Lq-likelihood method. *Technical Report 659, School of Statistics, University of Minnesota*, 2007. URL `http//:www.stat.umn.edu/~dferrari/research/techrep659.pdf`.

R. Fisher and L. C. Tippett. Limiting forms of the frequency distribution of largest or smallest member of a sample. *Proceedings of the Cambridge philosophical society*, 24:180–190, 1928.

M. Gell-Mann, editor. *Nonextensive Entropy, Interdisciplinary Applications*. Oxford University Press, New York, 2004.

M. Gilli and E. Kellezi. An application of extreme value theory for measuring financial risk. *Computational Economics*, 2-3:207–228, 2006.

G. H. Givens and J. A. Hoeting. *Computational Statistics*. Wiley, New Jersey, 2005.

B. V. Gnedenko. Sur la distribution limite du terme d'une serie aleatoire. *Annals of Mathematics*, 44:423–453, 1943.

S. D. Grimshaw. Computing maximum likelihood estimates for the generalized Pareto distribution. *Technometrics*, 35:185–191, 1993.

T. Hastie, R. Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction: with 200 Full-color Illustrations*. Springer-Verlag Inc, 2001.

J. R. M. Hosking and J. R. Wallis. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, 29:339–349, 1987.

P. J. Huber. *Robust Statistics*. Wiley Series in Probability. John Wiley and Sons, 1981.

A. F. Jenkinson. The frequency distribution of the annual maximum (minimum) values of meteorological events. *Quarterly Journal of the Royal Meteorological Society*, 81:158–172, 1955.

J. Knight, S. Satchell, and G. Wang. Quantitative finance. 3:332–344, 2005.

K. Kuester, S. Mittnik, and M. Paolella. Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, 4,1:53–89, 2006.

N. A. Lazar. *Statistics of Extremes: Theory and Applications*. Wiley, England, 2004.

A. McNeil and R. Frey. Estimation of tail-related risk measures for heteroskedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7:271–300, 2000.

A. McNeil and A. Stephenson. *evir: Extreme Values in R*, 2007. URL `http://www.maths.lancs.ac.uk/~stephena/`. R package version 1.5. S original (EVIS) by Alexander McNeil and R port by Alec Stephenson.

A. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton Series in Finance, New Jersey, 2005.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006. URL `http://www.R-project.org`.

R. Reiss and M. Thomas. *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. Birkhauser Verlag, Basel, 1997.

M. A. Ribatet. A users guide to the pot package (version 1.0), 2006. URL `http://cran.r-project.org`.

A. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, New York, 1998.

R. von Mises. *La distribution de la plus grande de n valeurs, in Selected Papers, Volume II*, volume 44. American Mathematical Society, Providence, RI., 1954.

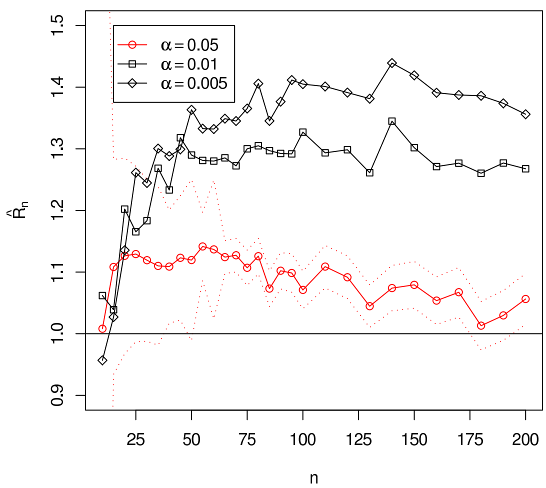

Figure 1: GP distribution. Monte Carlo Mean Squared Error ratio computed from $B = 1000$ samples of size n, for $\alpha = 0.05, 0.01, 0.005$ and $q = 0.94$. The dashed lines represent 95% confidence bands for the case when $\alpha = 0.05$.

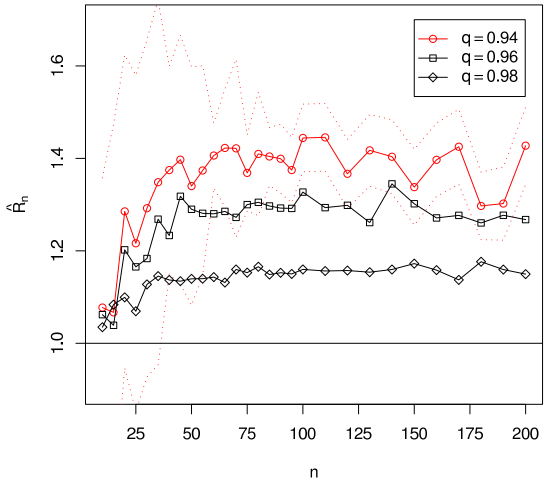

Figure 2: GP distribution. Monte Carlo Mean Squared Error ratio computed from $B = 1000$ samples of size n, for various values of the distortion parameter ($q = 0.94, 0.96, 0.98$) and true tail probability $\alpha = 0.01$.

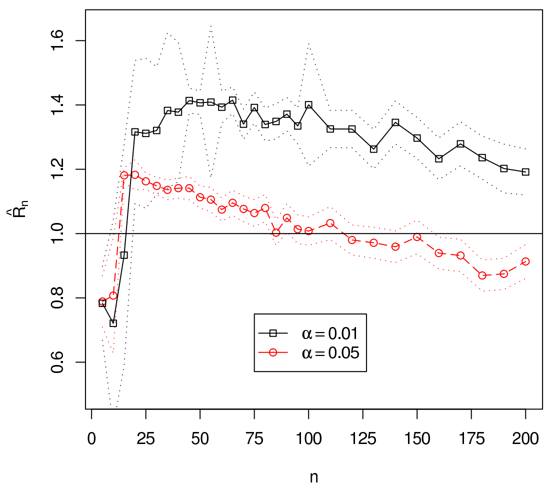

Figure 3: GEV distribution. Monte Carlo Mean Squared Error ratio computed from $B = 1000$ samples of size n, for two values of the true tail probability ($\alpha = 0.01, 0.05$) and distortion parameter $q = 0.95$. The dashed lines represent 95% confidence bands.

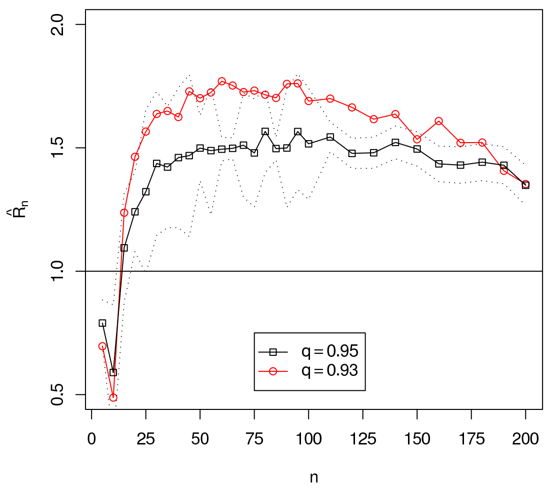

Figure 4: GEV distribution. Monte Carlo Mean Squared Error ratio computed from $B = 1000$ samples of size n, for two values of the distortion parameter ($q = 0.93, 0.95$) and true tail probability $\alpha = 0.005$. The dashed lines represent 95% confidence bands for the case when $q = 0.95$.