Extreme-quantile tracking for financial time series

V. Chavez-Demoulin[1*] , P. Embrechts[2] and S. Sardy[3]

[1] *Faculty of Business and Economics*
University of Lausanne
and
[2] *RiskLab, Department of Mathematics*
Swiss Finance Institute, ETH Zurich
and
[3] *Section of Mathematics*
University of Geneva

Abstract

Time series of financial asset values exhibit well known statistical features such as heavy tails and volatility clustering. We propose a nonparametric extension of the classical Peaks-Over-Threshold method from Extreme Value Theory to fit the time varying volatility in situations where the stationarity assumption may be violated by erratic changes of regime, say. As a result, we provide a method for estimating conditional risk measures applicable to both stationary and nonstationary series. A backtesting study for the UBS share price over the subprime crisis exemplifies our approach.

JEL classification: C.11; C.14; C.22; G.10; G.21. Keywords: Bayesian analysis; Conditional Risk Measures; Financial time series; Generalized Pareto distribution; Markov random field; Peaks-Over-Threshold; Quantile estimation; Regime switching; Statistics of extremes; Value-at-Risk.

[*]Corresponding author: valerie.chavez@unil.ch

1 Introduction

In numerous applications, one is interested in statistically estimating extremal risk measures based on time series observations. Typical examples include the estimation of record values, return periods and high-level crossings and this in applications ranging from climate research, over medicine and reliability to insurance and finance. Classical Extreme Value Theory (EVT), through the Peaks-Over-Threshold (POT) approach, yields a methodology for estimating such risk measures. From a mathematical point of view, the POT method is based on Balkema and de Haan [1974] and Pickands [1975]. The statistical theory was initially worked out by Davison and Smith [1990], Davison [1984a]. Although the techniques presented in this paper are applicable more widely, we focus attention on features present in financial (return) time series. Banking offers a unique example where an extreme risk measure, Value-at-Risk (VaR), is hardwired in the international regulatory framework referred to as the Basel Accords; for a history on these accords, see Tarullo [2008]. The so-called Basel I 1/2 Accord stipulated in the mid nineties that larger international banks have to hold regulatory (risk) capital for the trading book based on a 99% VaR over a 10-day holding period; for background reading, see Chapter 1 in McNeil et al. [2005]. The late-2000s financial crisis has shown to the extreme how inadequately the regulatory framework performed in times of crisis; for an early warning on this, see Danielsson et al. [2001] and Donnelly and Embrechts [2010] for a review of the crisis with an actuarial slant. In this paper we shall not discuss the issues underlying prudential regulation, nor question the wisdom of VaR-based risk capital calculation. Quantile based risk measures like VaR have been used in various fields of applied science, especially in the realm of rare event estimation, and this with great success. For the financial industry, they will further exist as an important tool in Quantitative Risk Management (QRM). As a consequence, we need methodological research not only on VaR's weaknesses (see McNeil et al. [2005]), but also on ways to improve the statistical estimation of quantiles in general for times series. The current paper addresses the latter, especially taking possible nonstationarity into account. We present a technique for the statistical estimation of conditional risk measures applicable to both stationary as well as nonstationary time series models. In order to achieve this, we adapt the method from Sardy and Tseng [2004] taking common features of financial time series into account. In particular, we combine the theory from the latter paper with quantile estimating techniques coming from EVT. The new methodology is illustrated in detail on UBS share price data and further exemplified on Nasdaq and equity portfolio data. It is not our aim in this paper to settle the issue to use specific stationary or nonstationary models for financial data, we side with Mikosch and Starica [2004] that frequent measurements over a long period of time are likely to be nonstationary.

More formally, let $\{V_t\}_{t\in\mathbb{N}}$ be a time series of values on a single financial asset or a portfolio and denote the negative log-returns $Z_{h,t} = -(\log V_{t+h} - \log V_t)$ at time t and horizon h. The conditional 100%VaR is defined as the α-quantile of the predictive distribution,

$$\mathrm{VaR}_\alpha(Z_{h,t}) = \inf\left\{z \in \mathbb{R} : F_{Z_{h,t}|\mathcal{H}_t}(z) \geq 1-\alpha\right\}; \tag{1}$$

here $F_{Z_{h,t}|\mathcal{H}_t}$ denotes the distribution function of $Z_{h,t}$ conditional on the entire history $\mathcal{H}_t$ of the underlying stochastic process up to time t (hence *conditional* risk measurement). For a discussion on the difference between a conditional and unconditional approach to the estimation of risk measures, see McNeil et al. [2005], Sections 2.1.2 and 2.3.6. The estimation of VaR is a topic of considerable interest to banking regulation for which many approaches of varying sophistication have been derived; see Jorion [2007]. A well-known method can be found in RiskMetrics (Mina [2001]) for the estimation of conditional quantiles, with the drawbacks of assuming Gaussianity and of modeling volatility with a convenient, but too simplistic, exponential-weighted moving average (EWMA) method (RiskMetrics [1996]). The Gaussianity assumption is known to often be strongly violated, particularly for a short horizon as the unconditional and conditional distributions of financial time series are known to be heavy–tailed. More accurate models have been developed for a better estimation of VaR. Two main approaches can be distinguished. The time series approach concentrates on modeling the temporal features of $\{V_t\}_{t\in\mathbb{N}}$ (e.g. volatility clustering and fat tails) with GARCH-type and stochastic volatility models; see Shephard [1996] for a review. Both the normal variance-covariance, as well as the Monte Carlo approach can be considered within this framework. The extreme value approach uses results from EVT to focus on the tail of the distribution in order to estimate the VaR, and this for instance in combination with the historical simulation approach. Pérignon and Smith [2010] find that 73% of the banks that report their VaR methodologies use historical simulation. This has led to interesting quantitative risk management methodology as for instance summarized in McNeil et al. [2005]. As an early example, Danielsson and de Vries [1997] tackle the unconditional quantile estimation problem for stationary time series. Numerous papers address the use of EVT for the estimation of the conditional distribution, examples are McNeil and Frey [2000], Brooks et al. [2005] and Chavez-Demoulin et al. [2005]. By now, there exists a huge literature on the topic; see for instance Danielsson [2011]. Finally, EVT offers a useful set of techniques for stress testing. A summary of one-dimensional EVT is to be found in Embrechts et al. [1997].

An early criticism on the use of EVT for financial time series can be found in Diebold et al. [1998] where nonstationarity, through regime switches, was stated as a salient feature of markets under distress. In Mikosch and Starica [2004], the authors show how parameter switches in a GARCH environment may fool standard test for long-range dependence. We want to show how a fairly straightforward adaption of the classical POT method in EVT can handle such nonstationarities. Classics on the topic of regime switching in finance are Hamilton [1990] and Hamilton [1989]. For a recent review, see Ang and Timmermann [2011]. In his thought-provoking, philosophical discussion on *The Market*, Ayache [2010], Appendix 1, p. 449, writes *"Recalibration means that any valuation model will virtually become stochastic and that the market price of any contingency claim, no matter how exotic or complex, will virtually act as a calibration input to the model. This imposes, as a pricing tool, a structure that must be at once open and invariant. We believe the regime-switching model to be such a one"*. A further early criticism by RM practitioners on EVT, as applied to financial risk management, concerns the possible slow-adaptiveness of the method to short-time shocks in the series, i.e. the VaR estimates may not sufficiently closely follow the dynamics of the series, especially in an after-shock phase. Our method adapts more quickly and at the same time improves on statistical accuracy as exemplified through backtesting.

As discussed in Danielsson [2011], volatility structures like in Figure 1 can be observed for wide classes of financial time series and this at various levels of sampling frequency. Despite a clear conceptual distinction between periods of volatility clustering of a stationary process and changing periods of high/low volatility due to nonstationarity, the nature of the resulting extremes may not always be easy to find out. In some cases, there is clear economic evidence for a regime switch; for instance at the announcement of a merger or takeover, or an economic policy decision concerning interest rates, or a rating update. For data related to economic crises, like the dot-com bubble or subprime crises, such a single causal event may not always be detectable. We therefore want to allow for a wider spectrum of models around the classical POT method and indeed develop an extreme-risk measurement tool applicable to a wide class of both stationary as well as nonstationary time series.

From a methodological point of view, we follow the POT method as described in Davison and Smith [1990] and briefly reviewed in Section 2 where we also point out the method's limitations. In Section 3 we propose a Bayesian nonparametric adaption of the POT method, relaxing the stationarity assumption. In the context of applications to financial risk management, we derive a time adaptive estimation and credible region construction of quantiles. The new methodology is backtested on UBS share price data and further exemplified on Nasdaq and portfolio equity data. We draw some conclusions in Section 4. Because of the potential relevance for finance and insurance, throughout the paper, we refer to quantiles as VaRs being well aware that there is much more to VaR-methodology than just quantile estimation. In the current discussion on new capital guidelines for market risk (the so-called Basel III Accord), the techniques of this paper belong to the EVT-VaR and stressed-VaR toolkit.

2 Review of the Peaks-Over-Threshold method

2.1 A useful decomposition

The classical POT method considers observations $Z_1, \ldots, Z_T$ that are independent and identically distributed (i.i.d.) from a distribution function F_Z that belongs to a wide class of continuous distributions; for a slightly wider class of model assumptions under which classical EVT works, see Embrechts et al. [1997]. In practice, the i.i.d. assumption is often violated for financial time series, because of dependence or nonstationarity. In Section 3 we relax these assumptions. For a given high threshold u, the POT method is based on the decomposition of the tail of F_Z as

$$1 - F_Z(z) = P(Z > u)\{1 - F_u(z)\}, z \geq u, \tag{2}$$

where the (conditional) excess distribution F_u is defined as

$$F_u(z) = \mathrm{P}(Z \leq z \mid Z > u), z \geq u.$$

Letting $E_u = \{t \in \{1, \dots, T\} : \ Z_t > u\}$ be the set of indexes t for which Z_t exceeds u, the POT method models the (random) number $N_u = \mathrm{card}(E_u)$ and the sizes $\{W_s = Z_{t_s} - u\}_{t_s \in E_u}$ of all the extreme observations $\{Z_{t_s}\}_{t_s \in E_u}$. For an appropriately chosen threshold u, mathematical theory supports the independent modeling of

- the number of exceedances $N_u \sim \mathrm{Poisson}(\lambda)$ as the limiting distribution of the sum of T Bernoulli random variables with success probability λ/T, and
- the excesses $W_1, \dots, W_{N_u}$, conditional on the number of exceedances $N_u = n_u$, as a sample from the generalized Pareto distribution (GPD)

$$G(w; \sigma, \kappa) = \begin{cases} 1 - (1 + \kappa w/\sigma)_+^{-1/\kappa}, & \kappa \neq 0, \\ 1 - \exp(-w/\sigma), & \kappa = 0, \end{cases} \tag{3}$$

  where $\sigma > 0$, and the support is $w \geq 0$ when $\kappa \geq 0$ and $0 \leq w \leq -\sigma/\kappa$ when $\kappa < 0$. The domain of attraction of the GPD includes many common distributions which can be classified based on their tail characteristics, or equivalently, on the shape parameter κ. The useful distributions for financial applications are for positive κ which corresponds to heavy-tailed (or power-tailed) distributions, e.g. Pareto, Student t or Fréchet. The case $\kappa = 0$ corresponds to distributions whose tail essentially decays exponentially, e.g. normal, gamma or lognormal, while $\kappa < 0$ is for short-tailed distributions with a finite right endpoint, like uniform or beta.

The choice of the threshold u is important, implying a balance between bias and variance. The threshold selection has a degree of arbitrariness in practice. On the one hand, the smaller u, the more observations are used for inference (small variance). On the other hand, mathematical theory (Leadbetter [1991]) suggests choosing a high u for limiting results to apply (small bias). Graphical techniques for the choice of u are for instance discussed in Smith [1987], Davison and Smith [1990] and Yang [1978]. Based on an extensive comparative simulation study, Chavez-Demoulin [1999] recommends choosing an initial threshold such that about 10% of the data are excesses. For this threshold selection, Chavez-Demoulin and Embrechts [2004] show that small variations in the value of the threshold typically have little impact on the estimation. A full POT application would include a sensitivity analysis across several threshold values u.

2.2 Parametric estimation for i.i.d. data

The parameters of the Poisson(λ) and the GPD(σ, κ) can be estimated from the number and size of exceedances in the time series. Letting g denote the density of the GPD, the log-likelihood

$$\begin{aligned}
l(\lambda, \sigma, \kappa; n_u, w_1, \ldots, w_{n_u}) &= \log\left(P\left(N_u = n_u\right) \prod_{n=1}^{n_u} g(w_n)\right) \\
&= n_u \log \lambda - \lambda - \log n_u! \\
&\quad - n_u \log \sigma - (1 + 1/\kappa) \sum_{n=1}^{n_u} \log(1 + \kappa w_n/\sigma)_+ \\
&= l(\lambda) + l(\sigma, \kappa) \qquad (4)
\end{aligned}$$

splits into two parts. Estimation of λ and (σ, κ) can therefore be performed separately by maximizing their respective likelihoods, leading to the estimates $\hat{\lambda} = n_u$ and $\hat{\sigma}, \hat{\kappa}$. To find the local maximum of the log-likelihood $l(\sigma, \kappa)$ requires numerical methods, for instance employing a Newton–Raphson method as discussed in Hosking and Wallis

[1987]. From the decomposition (2), the estimated distribution

$$\hat{F}_Z(z) = 1 - (1 - \exp(-n_u/T)) \left(1 + \hat{\kappa}\frac{z-u}{\hat{\sigma}}\right)_+^{-1/\hat{\kappa}} \tag{5}$$

leads to the estimated α-quantile or VaR at confidence level α, sufficiently close to 1:

$$\widehat{\mathrm{VaR}}_\alpha(Z) = \hat{F}_Z^{-1}(\alpha). \tag{6}$$

The estimation of the parameters κ and σ of a generalized Pareto distribution is regular if $\kappa \geq -1/2$ in the sense that the score statistic is asymptotically normal; see Davison [1984b], Davison [1984a] and Smith [1985]. Since heavy-tailedness is a feature of many financial time series, κ is typically positive, so that confidence intervals for VaR can be derived based on the asymptotic normality of the maximum likelihood estimators using the delta method or the profile likelihood approach; see pp. 501 and 284 in McNeil et al. [2005].

2.3 POT for time series

Within financial applications, the classical POT approach as explained under Sections 2.1 and 2.2 is mostly applied to (transformed) time series data ignoring the finer dependence structure. Temporal structures such as volatility clustering are however of crucial importance. Although the data declustering was already proposed in Leadbetter et al. [1983], automatic cluster identification remains a hard task and is often arbitrary; see for instance Coles [2001]. Declustering can also cause late detection of changes of regime when the process is nonstationary. To avoid declustering while taking volatility clustering into account, McNeil and Frey [2000] propose a two step procedure that combines a GARCH model to account for volatility and EVT to estimate the tail of the innovation distribution of the GARCH model. In Chavez-Demoulin et al. [2005] a marked point process model for the exceedances of a high threshold, with a self-exciting structure to describe the occurrence of clusters, is introduced. These and related approaches typically do not allow for nonstationarity.

2.4 An example: UBS data

Consider the daily UBS data of Figure 1 that shows the values of the UBS closing share prices (top) and the corresponding negative log-returns (bottom) from the 27th of June 2002 to the 18th of May 2010. The effect of the subprime crisis (2007-2009) can clearly be seen. It is worth noting that the backtesting results for UBS's own trading book (99%, 1-day VaR) most clearly emphasize the failure of its internal quantitative risk management tools. Rather than observing on average about 3 yearly violations of the above VaR limit, the following numbers were stated in the corresponding annual reports to the shareholders: 29 for 2007 and 50 for 2008 before returning to 4 for 2009.

For the UBS data, we apply the classical POT method to estimate the 99%VaR for the horizon $h = 1$ day, treating for now the time series of log–returns as i.i.d. To apply the POT method, at each time t we first determine the threshold u using the rule that around 10% of the data are excesses up to this date. We then use the tools of Section 2.2 to estimate the unconditional distribution (5), calculate the 99%VaR estimate (6), and compare it to the value realized the next day. A violation is said to occur when the realized negative log–return is higher than the estimated VaR. Repeating this operation from June 19, 2006 until May 17, 2010 is referred to as *backtesting*. For a more in depth discussion on VaR-backtesting, see Bontemps [2008] and Danielsson [2011], Chapter 8. Figure 2 plots these estimated 99% unconditional VaRs. Whereas we expect about 10 violations at the 99% level, we observe 35, giving evidence for an under–estimation of VaR till about 2009. The main reason for this discrepancy is the increasing volatility starting around mid 2007 and clearly visible in Figure 2; this feature is only marginally taken into account by the classical POT method. The under–estimation period is followed by an over–estimate curve that is not adaptive to the decreasing trend starting around 2009. This is due to the fact that for the classical POT method, the past two years of data used in the analysis are given the same weight in the estimation of the GPD parameters. It was this fact that was early on stated as a criticism on classical EVT-VaR by practitioners. In order to correct for this problem, numerous alternatives to classical EVT-VaR have been introduced in the literature, typically involving a detailed (for instance GARCH or stochastic volatility) modeling of the dynamics of the times series. Some of these approaches will be briefly exemplified on the data examples in Section 3.4. In Section 3 we offer an alternative approach, which is both flexible and straightforward to put into practice. The sampling period June 27, 2002 to May 18, 2010 for the UBS data contains at least three rather different periods, call them regimes. These are the higher volatility years 2003, 2008 and the rather persistent low-volatility period in between. In his comparative VaR analysis of the S&P 500 over this period, Danielsson [2011] p. 152, writes: "*This suggests that there was a structural break at the onset of the 2007 crisis causing difficulties for all methods*", and further "*The abrupt changes in volatilities seen in 2003 and 2008 are likely to cause problems for most VaR models. The models tried here clearly fail during those structural breaks*". From a VaR-based regulatory point of view, this is a rather sobering conclusion.

3 Nonparametric Peaks-Over-Threshold method

The discussion is Section 2.4 clearly calls for adapting the POT method to a wider class of models based on the conditional decomposition presented in Section 3.1 and mathematical foundations developed by Huesler [1986] and Leadbetter [1991]. At the end of Section 3.1 we briefly discuss alternative approaches.

To estimate the time varying conditional VaR in (1), our approach consists in developing a nonparametric POT method to fit the temporal evolution of the shape of the tail of the conditional distribution, as opposed to assuming a fixed shape throughout time. We develop in Section 3.2 a specific nonparametric fitting method capable of capturing erratic changes in the structure of the tail of the distribution. Using the smoothed estimates, we then predict future quantiles (VaR) and derive credible regions in Section 3.3.

3.1 A time dependent conditional decomposition

The conditional distribution $F_{Z_t|\mathcal{H}_t}$ of the negative, one-day log-returns Z_t (occasionally just referred to as "the returns") given the history $\mathcal{H}_t$ of the process up to time t, can be decomposed in its upper tail as

$$P\left(Z_t > z \mid \mathcal{H}_t\right) = P\left(Z_t > u \mid \mathcal{H}_t\right) P\left(Z_t - u > z - u \mid Z_t > u, \mathcal{H}_t\right), \tag{7}$$

for $z > u$ and a threshold u. This decomposition requires the estimation of two time–varying conditional probabilities on the right–hand side of (7). For an appropriately chosen threshold u, the POT approach described in Section 2 and the theory of Huesler [1986] support the independent modeling of

- the weekly counts

  $$N_u(l) \sim \text{Poisson}(\lambda_l^{\text{w}}) \tag{8}$$

  of returns above the threshold u during week l. The weekly Poisson parameters λ_l^{w} may vary from week to week. More formally, let $E_u(l) = \{t \in [5(l-1) + 1, \ldots, 5l] : \; Z_t > u\}$ be the set of time indexes during trading week l when Z_t exceeds the threshold u, then the weekly counts are $N_u(l) = \text{card}(E_u(l))$, and

- the sizes of exceedances

$$W_s \sim \mathrm{GPD}(\sigma_s, \kappa_s) \tag{9}$$

at times s when a negative log–return exceeds the threshold u.

Since we are particularly interested in estimating the conditional VaR near the end of the time series, we propose to choose the threshold u such that 10% of the data are excesses during the last year of recording only, as opposed to during the entire period.

A noticeable difference with the classical POT approach is that the parameters $\theta_t = (\lambda_t, \sigma_t, \kappa_t)$ of the Poisson and GPD distributions are allowed to vary with time. This was for instance done parametrically in Davison and Smith [1990] to adapt to seasonality, and nonparametrically with smoothing splines in Chavez-Demoulin and Davison [2005]. The latter paper contains an application to environmental data, fitting seasonal and slowly varying long-term climate factors.

3.2 Nonparametric Bayesian smoothing

To impose temporal smoothness in the evolution of the parameters of the Poisson and GPD models, we make the Bayesian assumption that the parameters are realizations of independent hidden processes, an idea similar to the Kalman filter or stochastic volatility models. Using Laplace innovations to model the hidden processes, along with maximum a posteriori estimation, leads to smooth coefficient estimates with occasional abrupt temporal changes, hence reflecting sudden changes of regimes; see Hamilton [1989] and Ang and Timmermann [2011]. The proposed estimator owes this characteristics to the ℓ_1-penalty induced by the Laplace prior in the same spirit as the lasso in Tibshirani [1996] for model selection; here the jumps will be selected by the estimator and will reflect regime switches. We discuss this idea of smoothing the daily information on the Poisson and GPD parameters in Sections 3.2.1 and 3.2.2, respectively.

3.2.1 Poisson parameters

As discussed in Section 3.1, the data $n_u(1), \ldots, n_u(L)$ are counts in $\{0, 1, 2, 3, 4, 5\}$ of the number of exceedances of returns in successive weeks $1, \ldots, L = \lfloor T/5 \rfloor$. From (8) they are assumed to be independent Poisson distributed with Poisson parameters $(\lambda_1^{\mathrm{w}}, \ldots, \lambda_L^{\mathrm{w}})$ that may change from week to week. The corresponding log-likelihood is

$$l(\lambda_1^{\mathrm{w}}, \ldots, \lambda_L^{\mathrm{w}}; n_u(1), \ldots, n_u(L)) = \sum_{l=1}^{L} \{n_u(l) \log \lambda_l^{\mathrm{w}} - \lambda_l^{\mathrm{w}}\}. \tag{10}$$

The maximum likelihood estimates are simply $\hat{\lambda}_l^{\mathrm{w}} = n_u(l)$, the weekly data, which exhibit a high variance. In order to achieve a better bias–variance trade–off, we smooth these estimates. Bayesian regularization consists in putting a prior distribution on the temporal smoothness of the Poisson parameters. As it is often done with Poisson data, see Nelder and Wedderburn [1972], we consider the parameters on a log-scale; hence we model $\eta = \log \lambda$ as the realization of a temporal first order Markov process with Laplace innovations $\eta_{l+1}^{\mathrm{w}} \mid \eta_l^{\mathrm{w}} = \mathrm{Laplace}(\eta_l^{\mathrm{w}}, \gamma_1)$, where $\mathrm{Laplace}(\eta, \gamma)$ is the Laplace distribution centered in η and the dispersion parameter $\gamma_1 > 0$ reflects whether the changes are abrupt and frequent (small γ_1) or mild and rare (large γ_1). The corresponding improper joint prior is

$$\pi(\eta_1^{\mathrm{w}}, \ldots, \eta_L^{\mathrm{w}}) = \left(\frac{\gamma_1}{2}\right)^{L-1} \exp(-\gamma_1 \sum_{l=2}^{L} |\eta_l^{\mathrm{w}} - \eta_{l-1}^{\mathrm{w}}|). \tag{11}$$

Using Bayes' Theorem, the log-posterior distribution of $\boldsymbol{\eta}^{\mathrm{w}}$, given the Poisson counts $n_u(1), \ldots, n_u(L)$, is

$$l(\eta_1^{\mathrm{w}}, \ldots, \eta_L^{\mathrm{w}}; n_u(1), \ldots, n_u(L)) = \sum_{l=1}^{L} (n_u(l)\ \eta_l^{\mathrm{w}} - \exp \eta_l^{\mathrm{w}}) - \gamma_1 \sum_{l=2}^{L} |\eta_l^{\mathrm{w}} - \eta_{l-1}^{\mathrm{w}}|. \tag{12}$$

The parameter $\gamma_1 > 0$ controls the amount of smoothing: choosing it as $\gamma_1 = +\infty$ leads to the parametric estimate of a constant Poisson parameter $\hat{\lambda}_1^{\mathrm{w}} = \ldots = \hat{\lambda}_L^{\mathrm{w}} = \sum_{l=1}^{L} n_u(l)/L$, while choosing it as $\gamma_1 = 0$ leads to the erratic maximum likelihood estimates $\hat{\lambda}_l^{\mathrm{w}} = n_u(l)$. A good bias–variance trade–off can be achieved by selecting the smoothing parameter γ_1 adaptively, for instance by minimizing the two–fold cross validation estimate of the Kullback-Leibler distance between the true and estimated Poisson likelihoods; see Sardy and Tseng [2004] for details. The information contained in $\hat{\gamma}_1$ is valuable as it reflects whether the time series is rather stable (large γ_1) or not (small γ_1). The estimated value of γ_1 is used in Section 3.3 to derive a measure of uncertainty of the estimated conditional VaR.

Computing the smooth maximum a-posteriori estimates of the temporal log-Poisson parameters $(\eta_1^{\mathrm{w}}, \ldots \eta_L^{\mathrm{w}})$ that minimize the negative penalized log-likelihood is not trivial because the function in (12), although convex, is not differentiable. The iterated dual mode (IDM) algorithm, which is easy to implement, is guaranteed to converge in this situation; see Sardy and Tseng [2004] for details. In the notation of Theorem 3 of Sardy and Tseng [2004], we have the negative log-likelihood function $g(\lambda) = -s\lambda + \exp\lambda$, its conjugate $g^{\star}(u) = (s+u)\{\log(y+u) - 1\}$ and the conjugate to the Laplace prior $h^{\star}(w) = 0$ if $\|w\|_\infty \leq \gamma_1$ and $+\infty$ otherwise.

3.2.2 Scale and shape parameters of the GPD

From (9), the sizes of exceedances $W_s \sim \mathrm{GPD}(\sigma_s, \kappa_s)$ have time varying parameters. Since the conditional distribution is typically heavy-tailed in finance, we expect the shape parameter κ_s to be positive. We assume here that the shape parameter $\kappa_s = \kappa > 0$ is not only positive but also constant over the entire period of the given time series. By doing so, the procedure loses flexibility but gains stability at the estimation stage for a better prediction of the conditional VaR. This assumption can be relaxed when deemed important.

Similarly to the Poisson parameters, the scale parameters σ_s can be estimated with some temporal smoothness. Assuming a first order Markov process with Laplace innovations for $\varphi_s = \log \sigma_s$ leads to the following log-posterior distribution of $\boldsymbol{\varphi}$:

$$\begin{aligned} l(\varphi_1, \ldots, \varphi_{n_u}, \kappa; w_1, \ldots w_{n_u}) &= \sum_{s=1}^{n_u} (-\varphi_s - (1 + 1/\kappa) \log\{1 + \kappa w_s \exp(-\varphi_s)\}) \\ & \quad -\gamma_2 \sum_{s=1}^{n_u - 1} |\varphi_{s+1} - \varphi_s|, \end{aligned} \tag{13}$$

where $n_u = \sum_{l=1}^{L} n_u(l)$ and γ_2 plays the role of a smoothing parameter. To compute the smooth maximum a-posteriori estimates of κ and $\varphi_1, \ldots, \varphi_{n_u}$, again the IDM algorithm can be employed.

3.3 Bayesian estimation of the conditional VaR

The nonparametric Bayesian smoothing of Section 3.2 leads to a Poisson parameter estimated every week, and GPD parameters estimated at times of exceedances. To obtain daily parameter estimates, we assume homogeneity within each week, so that the daily intensities are $\hat{\lambda}_t = \hat{\lambda}_l^{\mathrm{w}}/5$, for all days t in week l. For the daily scale parameter of the GPD, we assume piecewise constant interpolation between days of exceedances, i.e.,

$$\hat{\sigma}_t = \hat{\sigma}_{t_i} \quad \text{for} \quad t_i \leq t < t_{i+1}, \; i = 1, \ldots, n_u. \tag{14}$$

This interpolation scheme provides daily parameter estimates. So, given time series measurements $\{Z_t\}_{t=1,\ldots,T}$, the methodology described above leads to the model parameter estimates $\hat{\theta}_t = \{\hat{\lambda}_t, \hat{\sigma}_t, \hat{\kappa}\}_{t=1,\ldots,T}$, as well as two smoothing parameters $\hat{\gamma}_1$ and $\hat{\gamma}_2$. Moreover the Bayesian first order Markov random field priors provide not only point estimates up to time T, but also a measure of uncertainty by means of the predictive distributions of:

- the weekly log-Poisson parameter $\eta^{\mathrm{w}}_{L+1} \mid \eta^{\mathrm{w}}_{L} \sim \mathrm{Laplace}(\eta^{\mathrm{w}}_{L}, \gamma_1)$. Given the estimated value of the Poisson parameter $\hat{\eta}^{\mathrm{w}}_{L}$ at week L, the distribution of the Poisson parameter η^{w}_{L+1} at week $L+1$ can therefore be estimated by

$$\eta^{\mathrm{w}}_{L+1} \mid \eta^{\mathrm{w}}_{L} = \hat{\eta}^{\mathrm{w}}_{L} \sim \mathrm{Laplace}(\hat{\eta}^{\mathrm{w}}_{L}, \hat{\gamma}_1), \tag{15}$$

  and

- the daily log-scale parameter of the GPD

$$\varphi_{T+1} \mid \varphi_T \sim \begin{cases} \delta_{\varphi_T} & \text{with} \quad P(Z_T < u) \\ \mathrm{Laplace}(\varphi_T, \gamma_2) & \text{with} \quad 1 - P(Z_T < u) \end{cases}$$

  where δ_σ is the Dirac mass at σ. The decomposition between a Dirac and a Laplace stems from the fact that the process is assumed constant between times of exceedances (14), and has a Laplace innovation at a time of exceedance. Given the estimated value of the GPD parameter at time T, the distribution of the GPD parameter φ_{T+1} at time $T+1$ can therefore be estimated by

$$\varphi_{T+1} \mid \varphi_T = \hat{\varphi}_T \sim \begin{cases} \delta_{\hat{\varphi}_T} & \text{with} \quad \hat{P}(Z_T < u) \\ \mathrm{Laplace}(\hat{\varphi}_T, \hat{\gamma}_2) & \text{with} \quad 1 - \hat{P}(Z_T < u). \end{cases} \tag{16}$$

Taking the best bias–variance trade–off estimates $\hat{\lambda}_T$ and $\hat{\varphi}_T$ at time T obtained by smoothing the data up to time T, we insert from (15) and (16) the approximate distribution of $\theta_{T+1} \mid \theta_T = \hat{\theta}_T := (\hat{\lambda}_T, \hat{\sigma}_T, \hat{\kappa})$ in

$$\widehat{\mathrm{VaR}}_\alpha(Z_T) = F^{-1}_{\theta_{T+1}|\hat{\theta}_T}(\alpha), \tag{17}$$

where, with a slight misuse of notation,

$$F_{\theta_{T+1}|\hat{\theta}_T}(z) = 1 - \left\{1 - \exp(-\lambda_{T+1} \mid \hat{\lambda}_T)\right\} \left(1 + \hat{\kappa}\frac{z - u}{\sigma_{T+1} \mid \hat{\sigma}_T}\right)^{-1/\hat{\kappa}}. \qquad (18)$$

The conditional VaR at time T and horizon h is therefore a random variable whose distribution depends on that of $\theta_{T+1} \mid \hat{\theta}_T$. By generating $\theta_j \overset{\text{i.i.d.}}{\sim} \theta_{T+1} \mid \hat{\theta}_T$ from the predictive distributions, we can, by Monte Carlo simulation, estimate the distribution of the conditional VaR at time T and provide a credible region for VaR. Just like in the standard POT approach, alternative risk measures, like expected shortfall, can also be estimated within this conditional set-up.

3.4 UBS data: a nonparametric POT analysis

We now complete the preliminary backtesting study of Section 2.4 by including the Nonparametric POT method (NPOT) discussed above. We apply successively at each day T from June 19, 2006 until May 17, 2010 the NPOT on a past two-year window. Figure 3 shows the estimated parameters $\hat{\lambda}_T$ (upper), $\hat{\kappa}_T$ (middle), $\hat{\sigma}_T$ (bottom). Interestingly, smooth temporal trends and local bursts can be observed in the evolution of the parameters. The Poisson parameter λ for instance seems to follow a succession of bursts of varying intensities and lengths. The shape parameter κ is rather constant near zero with occasional bursts that can attain 0.3. The scale parameter σ reveals an interesting pattern with a noticeable increasing trend from 2008 till 2009 and a decreasing, somewhat stabilizing one from 2010, highlighting a higher variability during the subprime crisis.

The resulting estimated 99% conditional VaR with its 95% credible region (left of Figure 4) provides a time-dependent risk measure that is sensitive to short and large time scale volatility changes. This local adaptivity is beneficial for the nonparametric

POT method as revealed by the backtesting which no longer rejects the null hypothesis that the conditional VaRs are correctly estimated.

It is also possible to estimate the conditional expected shortfall (ES) and give a credible region for it by Monte Carlo simulation. The conditional ES at time T and horizon h, a random variable whose distribution depends on that of $\theta_{T+1} \mid \hat{\theta}_T$ in (18), can be estimated by

$$\widehat{\mathrm{ES}}_\alpha(Z_T) = \frac{\widehat{\mathrm{VaR}}_\alpha(Z_T)}{1-\hat{\kappa}} + \frac{\hat{\sigma}_{T+1} - \hat{\kappa} u}{1-\hat{\kappa}}.$$

Figure 4 plots the conditional 99% ES (right panel) and its 95% credible region for the period June 19, 2006 until May 17, 2010. The ES is always larger than the VaR as it defines the average loss when VaR is exceeded, also ES alleviates some conceptual problems (e.g. possible non-subadditivity) inherent to VaR; see McNeil et al. [2005].

For comparison, we also apply the two step method of McNeil and Frey [2000] extensively used by Fernandez [2004]: the two step method (CONDEVT) provides good results (according to several backtesting procedures) when the stationarity assumption seems to hold. The comparative backtesting results are presented for different confidence levels in Table 1 for the UBS data from June 19, 2006 to May 17, 2010. We have performed a similar NPOT analysis for the Nasdaq index over the period January 4, 1989 to April 10, 2003; see Table 2. The latter considers a longer time period and hence allows for backtesting at the higher confidence level of 99.9%. The NPOT method consistently provides observed numbers of violations closest to the expected number.

3.5 Portfolio of International Equity Indexes

We finally consider an application of NPOT to a hypothetical portfolio of international equity indexes analyzed in McNeil et al. [2005], Chapter 2. At any one day t the portfolio value V_t is standardized to have weights 30% FTSE100, 40% S&P 500 and 30% SMI. The portfolio is assumed to have domestic currency sterling (GBP) and consequently has currency exposure to US dollar (USD) and Swiss franc (CHF). The value of the portfolio is therefore influenced by five risk factors: three log–index values and two log-exchange rates.

We calculate VaR estimates at the 95% and 99% levels for all trading days in the years 1999 to 2003 using NPOT. From the results collected in the backtesting table (Table 3), we conclude that, also in this case, the NPOT method performs well.

4 Discussion

The Nonparametric Peaks-Over-Threshold method (NPOT) is an important extension of the classical POT model to situations where financial asset values may follow a nonstationary process. Comparing and contrasting the new method to existing ones on financial market data shows that the proposed method provides a realistic model for the extremal behavior of financial time series. Backtesting results confirm a rather precise and adapted estimation of high-quantile based risk measures (VaR, ES) for financial time series. Moreover, credible regions can be derived, which provide financial analysts with a valuable measure of uncertainty. The method proposed can be applied to other time series data for which high quantiles need to be tracked. This may be of particular interest to, for instance, climate change studies for environmental data. One may ask what would be the “value” for an investor or for a financial institution of the improved extreme-quantile (expected shortfall) estimates. Although we have written the paper more from a defensive regulatory framework point of view, clearly it would be possible to embed the new methodology into a portfolio tracking or optimization exercise. We have not done so yet. In the words of the Ang and Timmermann [2011] "*Regime switching models can match narrative stories of changing fundamentals that sometimes can only be interpreted ex-post, but in a way that can be used for ex-ante real-time forecasting, optimal portfolio choice, and other economic applications*". A combination of our results with those from the latter paper may prove useful in the context of an investor's optimal portfolio choice. We will return to these, and related issues in future publications.

Acknowledgment

The authors would like to thank the Associate Editor and anonymous referees for useful comments on earlier versions of the paper.

References

A. Ang and A. Timmermann. Regime changes and financial markets. *NBER Working Paper Series 17182, June 2011*, 2011.

E. Ayache. *The Blank Swan. The End of Probability.* Wiley, 2010.

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

C. Bontemps. Moment-based tests for discrete distributions. *Toulouse School of Economics*, Preprint, 2008.

C. Brooks, A.D. Clare, J.W. Dalle Molle, and G. Persand. A comparison of extreme value theory approaches for determining value at risk. *J. Empirical Finance*, 12: 339–352, 2005.

V. Chavez-Demoulin. Two problems in environmental statistics: Capture-recapture analysis and smooth extremal models. Ph.D. thesis. Department of Mathematics, Swiss Federal Institute of Technology, Lausanne, 1999.

V. Chavez-Demoulin and A. C. Davison. Generalized additive models for sample extremes. *Applied Statistics*, 54 (1):207–222, 2005.

V. Chavez-Demoulin and P. Embrechts. Smooth extremal models in finance. *Journal of Risk and Insurance*, 71 (2):183–199, 2004.

V. Chavez-Demoulin, A. C. Davison, and A.J. McNeil. A point process approach to value-at-risk estimation. *Quantitative Finance*, 5 (2):227–234, 2005.

S. Coles. *An Introduction to Statistical Modeling of Extreme Values.* Springer, 2001.

J. Danielsson. *Financial Risk Forecasting.* Wiley, 2011.

J. Danielsson and C.G. de Vries. Tail index and quantile estimation with very high frequency data. *Journal of Empirical Finance*, 4:241–257, 1997.

J. Danielsson, P. Embrechts, C. Goodhart, C. Keating, F. Muennich, O. Renault, and H. Song Shin. An Academic Response to Basel II. *Financial Markets Group, London School of Economics*, 2001.

A.C. Davison. Modelling excesses over high thresholds, with an application. In J.T. de Oliveira, editor, *Statistical Extremes and Applications*, pages 461–482. D. Reidel, 1984a.

A.C. Davison. A statistical model for contamination due to long-range atmospheric transport of radionuclides. Ph.D. thesis. Department of Mathematics, Imperial College of Science and Technology, London, 1984b.

A.C. Davison and R.L. Smith. Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society, Series B*, 52:393–442, 1990.

F.X. Diebold, T. Schuermann, and J. Stroughair. Pitfalls and opportunities in the use of extreme value theory in risk management. In J.D. Moody A.-P. N. Refenes and A.N. Burgess (eds.), editors, *Advances in Computational Finance*. Amsterdam: Kluwer Academic Publishers, 1998. Reprinted in *Journal of Risk Finance*, 1:2 (Winter 2000), 30–36.

C. Donnelly and P Embrechts. The devil is in the tails: actuarial mathematics and the subprime mortgage crisis. *ASTIN Bulletin*, 40(1):1–33, 2010.

P. Embrechts, C. Klueppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer, 1997.

V. Fernandez. Extreme value theory: value at risk and returns dependence around the world. Documentos de Trabajo 161, Centro de Economaa Aplicada, Universidad de Chile, 2004.

J.D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357–384, 1989.

J.D. Hamilton. Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45(1-2):39–70, 1990.

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

J. Huesler. Extreme values of non-stationary random sequences. *Journal of Applied Probability*, 23:937–950, 1986.

P. Jorion. *Value at Risk: the New Benchmark for Managing Financial Risk. 3rd ed.* McGraw-Hill, 2007.

M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Series.* Springer, 1983.

M.R. Leadbetter. On a basis for Peaks over Threshold modeling. *Statistics and Probability Letters*, 12:357–362, 1991.

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

A.J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques, Tools.* Princeton University Press, 2005.

T. Mikosch and C. Starica. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics*, 86 (1): 378–390, 2004.

J. Y. Mina, J. and Xiao. *Return to RiskMetrics: The Evolution of a Standard*. New York, 2001.

J. A. Nelder and R. W. M. Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society, Series B*, 135:370–384, 1972.

J. Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3:119–131, 1975.

C. Pérignon and D.R. Smith. The level and quality of value-at-risk disclosure by commercial banks. *J. of Banking & Finance*, 34:362–377, 2010.

RiskMetrics. *RiskMetrics Technical Document, 3rd ed.* New York, 1996.

S. Sardy and P. Tseng. On the statistical analysis of smoothing by maximizing dirty Markov random field posterior distributions. *Journal of the American Statistical Association*, 99:191–204, 2004.

N. Shephard. Statistical aspects of ARCH and stochastic volatility. In D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielsen, editors, *Time Series Models in Econometrics, Finance and Other Fields*, Monographs on Statistics and Applied Probability, pages 1–67. Chapman & Hall, 1996.

R.L. Smith. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72:67–92, 1985.

R.L. Smith. Estimating tails of probability distributions. *The Annals of Statistics*, 15: 1174–1207, 1987.

D.K. Tarullo. *Banking on Basel. The Future of International Financial Regulation.* Peterson Institute for International Economics, Washington DC, 2008.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B: Methodological*, 58:267–288, 1996.

G. L. Yang. Estimation of a biometric function. *Annals of Statistics*, 6:112–116, 1978.

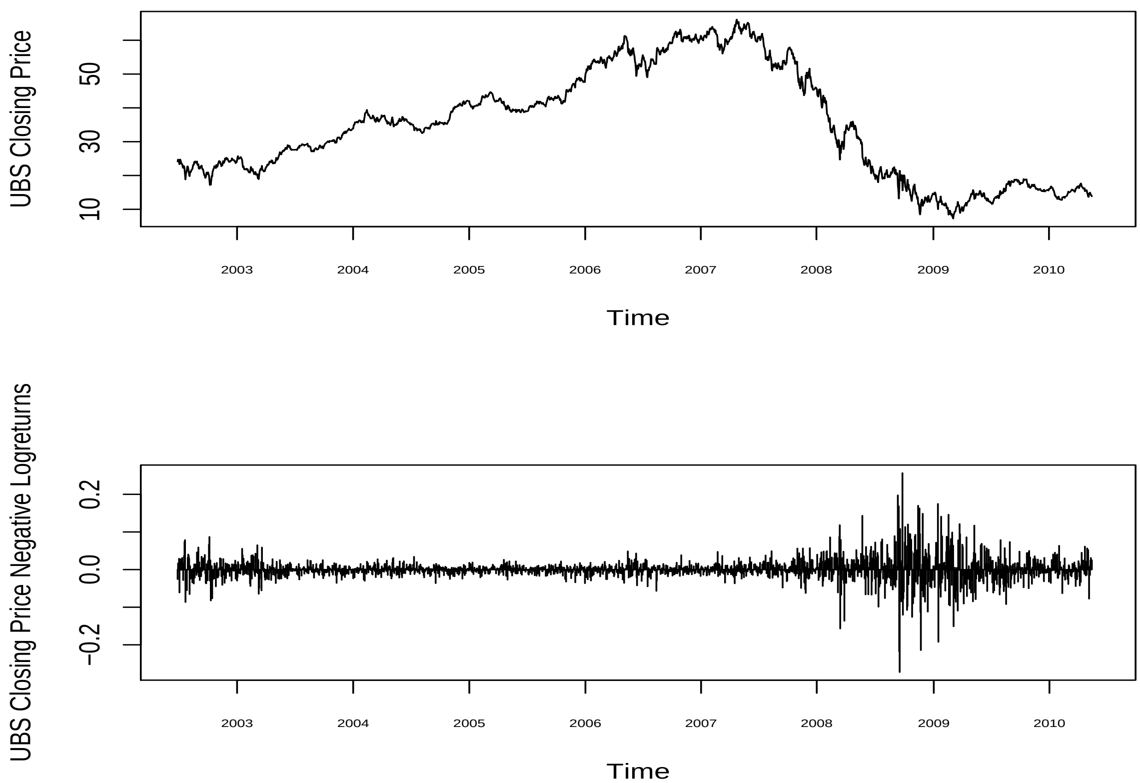

Figure 1: UBS data. The upper graph shows the UBS closing share prices from the 27th of June 2002 to the 18th of May 2010. The lower panel shows the negative log-returns.

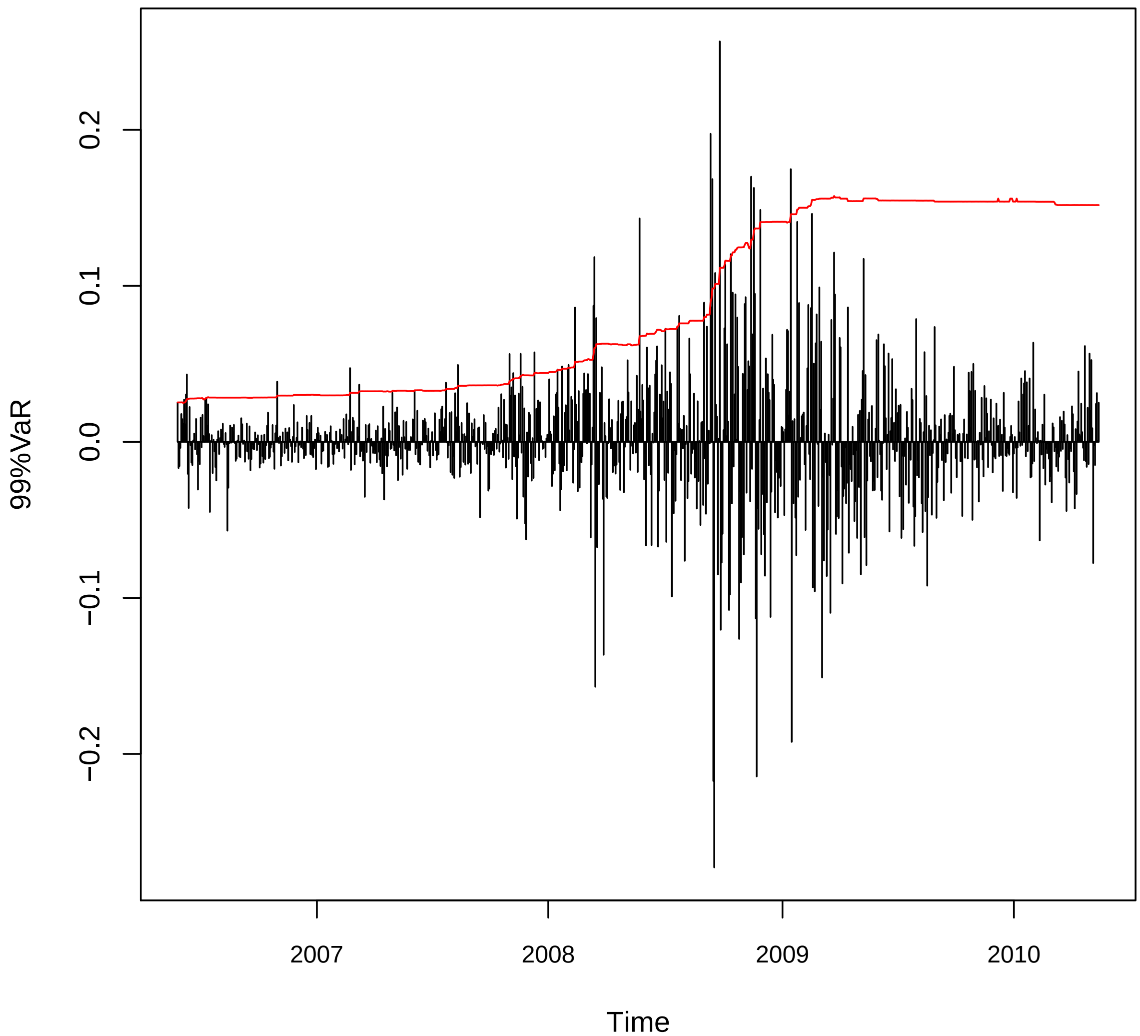

Figure 2: UBS data, negative log–returns. Estimated 99% Value-at-Risk (line) using the classical unconditional POT method from June 19, 2006 until May 17, 2010.

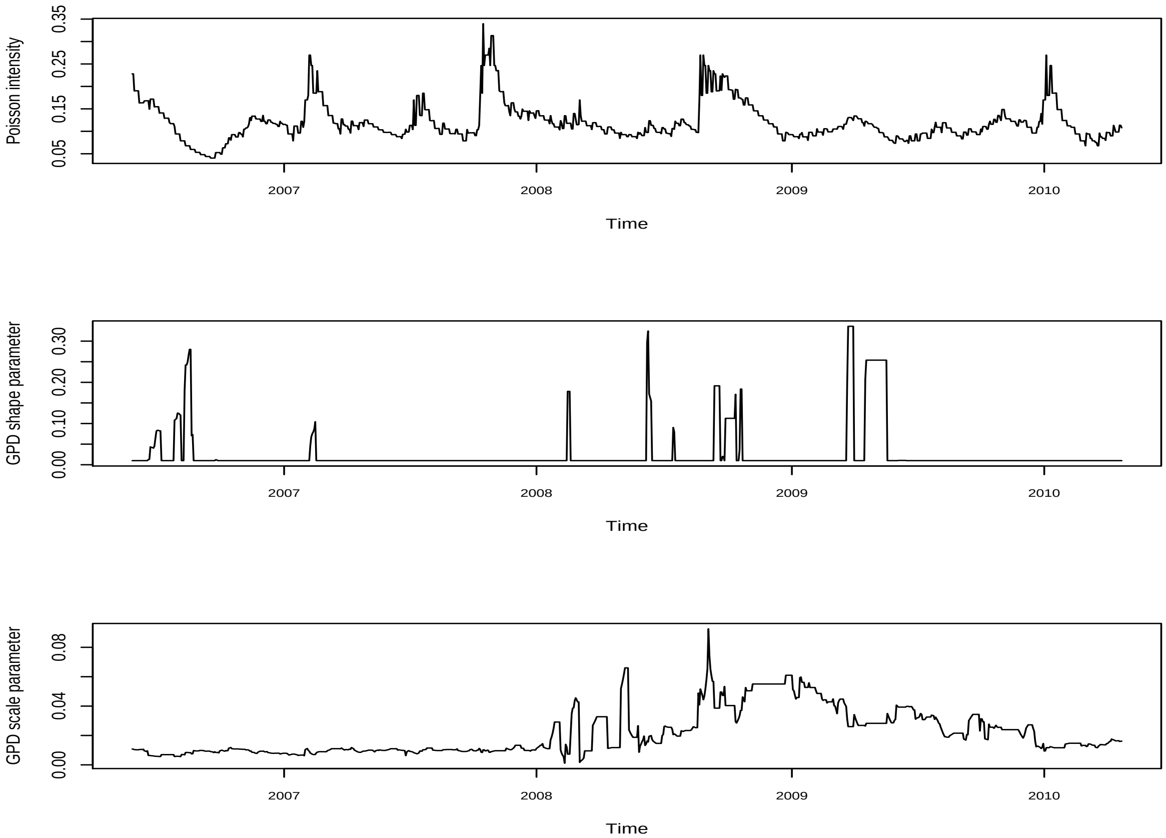

Figure 3: UBS data. Estimated parameters $\hat{\lambda}_t$ (upper), $\hat{\kappa}_t$ (middle), $\hat{\sigma}_t$ (bottom) from June 19, 2006 until May 17, 2010.

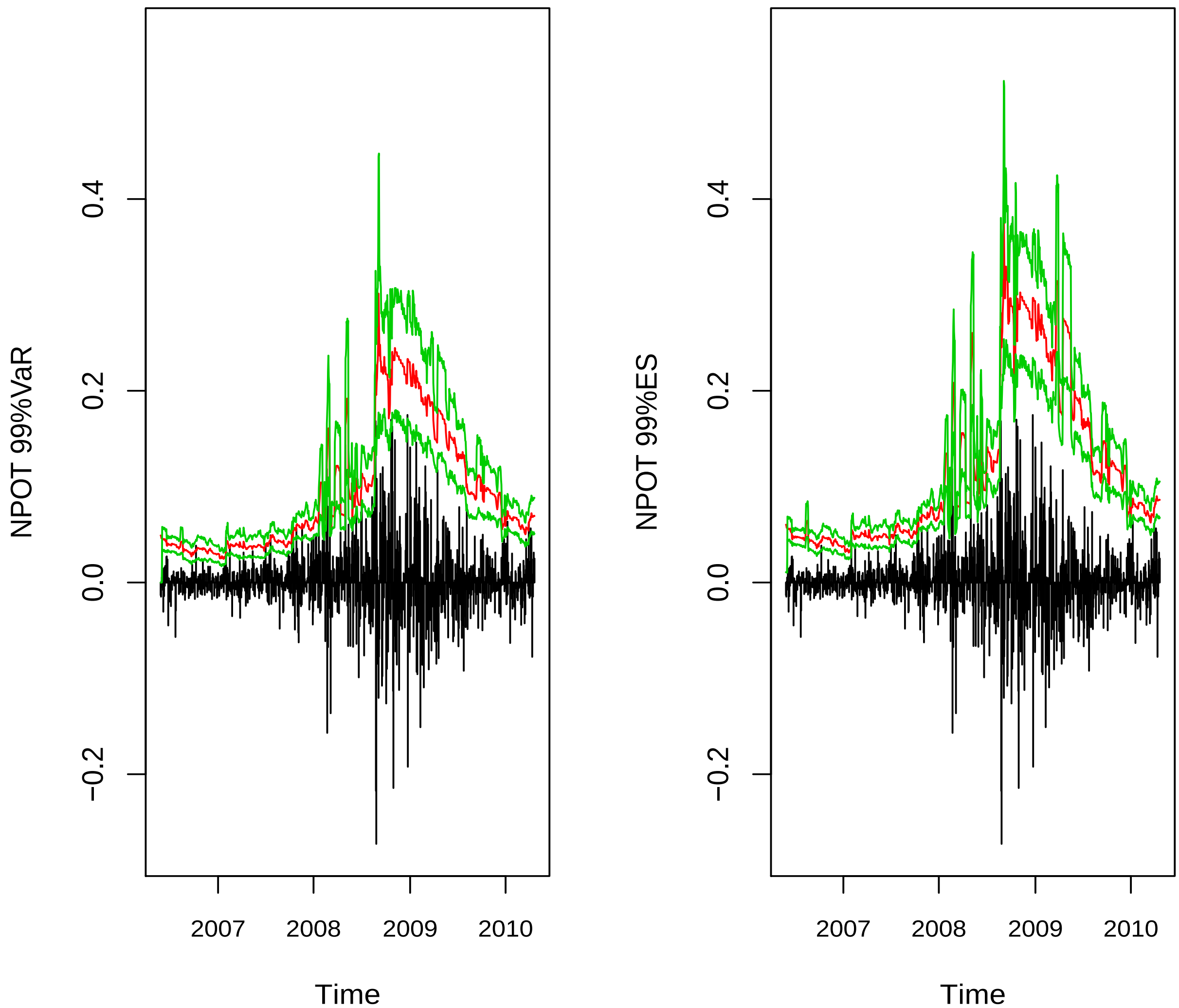

Figure 4: UBS data. Estimated 99% conditional VaR (left panel) using NPOT and its 95% credible region. Estimated 99% conditional ES (right panel) using NPOT and its 95% credible region.

	NPOT			CONDEVT		
	95%	**99%**	**99.5%**	**95%**	**99%**	**99.5%**
Expected	49.25	9.85	4.9	49.25	9.85	4.9
Observed	56	6	2	33	3	0
p-value	0.30	0.26	0.26	0.01	0.02	0.01

Table 1: UBS data. Backtesting results from June 19, 2006 until May 17, 2010. Expected number of violations, number of violations observed using NPOT and CONDEVT (McNeil and Frey [2000]), and p-values from the binomial test.

	NPOT			CONDEVT		
	99%	**99.5%**	**99.9%**	**99%**	**99.5%**	**99.9%**
Expected	36	18	3	36	18	3
Observed	35	12	1	19	9	9
p-value	0.93	0.19	0.28	0.01	0.04	0.01

Table 2: Nasdaq data. Backtesting results from January 4, 1989 to April 10, 2003. Expected number of violations, number of violations observed using NPOT and CONDEVT (McNeil and Frey [2000]), and p-values from the binomial test.

	1999	2000	2001	2002	2003
Trading days	260	259	260	260	260
95% cond. VaR					
Expected	13	12.95	13	13	13
Observed	10	16	17	20	11
p-value	0.47	0.38	0.25	0.06	0.67
99% cond. VaR					
Expected	2.6	2.59	2.6	2.6	2.6
Observed	3	2	1	3	1
p-value	0.74	1	0.52	0.74	0.52

Table 3: Portfolio of International Equity Indexes. Backtesting results. Expected and observed numbers of violations of the 95% and 99% conditional VaR from NPOT and corresponding binomial test p-values for the years 1999 to 2003.