UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL

# Excesses, Durations and Forecasting Value-at-Risk

Paulo José Araújo dos Santos

Doutoramento em Estatística e Investigação Operacional
(Especialidade de Probabilidades e Estatística)

2011

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL

# Excesses, Durations and Forecasting Value-at-Risk

Paulo José Araújo dos Santos

Doutoramento em Estatística e Investigação Operacional
(Especialidade de Probabilidades e Estatística)

Sob Orientação de:
Professora Doutora Maria Isabel Fraga Alves

2011

## Abstract

The first part of this thesis is devoted to the semiparametric estimation of high quantiles. The classic estimators do not enjoy a desirable property in the presence of linear transformation of the data. To solve this problem, the Peaks Over a Random Threshold (PORT) methodology and PORT estimators are proposed. The consistency and asymptotic normality of the estimators are demonstrated. The finite sample behaviour of the proposed PORT estimators is studied and compared with some competitors.

Under the context of financial time series and forecasting Value-at-Risk (VaR), the tendency to clustering of violations problem arises. To deal with this, a new class of independence tests for interval forecasts evaluation is proposed and the choice of one test is addressed. The exact and the asymptotic distributions of the corresponding test statistic are derived. In simulation studies, the proposed test revealed to be more powerful than the other tests under study, with few exceptions.

The tendency to clustering of violations problem is related with the discrete Weibull distribution, through the shape parameter. A new estimator for this parameter is proposed. The conditional distribution function and the moments are derived.

In order to solve the tendency to clustering of violations problem, a new risk model based on durations between excesses over a high threshold (DPOT) is proposed and compared with state-of-the art models under the probability 0.01, established in the Basel Accords.

Under the context of extremal quantiles and using one of the oldest financial time series, the DPOT model and a risk model that uses an PORT estimator are compared with other risk models.

In the empirical studies presented, to predict the VaR at a level 0.01 or lower, these models revealed more accuracy than the conditional parametric models widely used by the econometricians.



**Keywords:** extreme value theory; quantitative risk management; financial time series; clustering of violations; discrete Weibull distribution; backtesting.

## Resumo em português

As principais contribuições desta tese enquadram-se nas áreas da estatística de valores extremos e gestão quantitativa do risco, com exemplos de aplicações em finanças. O desenvolvimento deste trabalho apoia-se num ambiente probabilista subjacente da Teoria de Valores Extremos (EVT do inglês "Extreme Value Theory"). Na primeira parte assumimos que os dados são realizações de variáveis aleatórias (va´s) independentes e identicamente distribuídas (iid). A distribuição do máximo de $n$ observações de uma amostra de va´s iid, após a normalização adequada, converge para um dos três tipos possiveis de distribuições de valores extremos (Fréchet, Weibull ou Gumbel), representadas na forma unificada pela distribuição generalizada de valores extremos (GEV do inglês "Generalized Extreme Value"). O parâmetro de forma da distribuição GEV é designado por índice de cauda, sendo a sua estimação de grande importâcia na estimação de outros parâmetros de acontecimentos raros como por exemplo quantis elevados. Estão disponíveis na literatura, referida no Capítulo 1, muitos exemplos da importância da estimação destes parâmetros num vasto leque de aplicações em hidrologia, engenharia sísmica, ciências do ambiente, modelação de tráfico de redes, gestão de riscos em seguros, finanças, entre outras áreas. A primeira parte desta tese é dedicada à estimação semiparamétrica do índice de cauda e de quantis elevados. Os estimadores clássicos não gozam de uma propriedade desejável na presença de transformações lineares dos dados. Esta propriedade consiste nas estimativas não serem perturbadas por mudanças de localização. Para resolver este problema, propomos uma metodologia que foi designada por metodologia PORT (do inglês "Peaks Over a Random Threshold") e estimadores PORT para o índice de cauda e para quantis elevados. A consistência e a normalidade assintótica destes estimadores é demonstrada. É estudado o comportamento de estimadores PORT em amostras finitas e o seu desempenho comparado com outros estimadores usados na literatura da especialidade.

Outro tema central tratado nesta tese é a quantificação do risco. O Value-at-Risk (VaR), que de forma simplista não é mais do que um quantil extremo, tem vindo a substituir a volatilidade e o desvio padrão, sendo actualmente a medida de risco mais utilizada pelos profissionais na área financeira. Desde os Acordos de Basileia que grande parte das instituições financeiras utilizam diáriamente o VaR para cálculo dos requisitos de capital. Num contexto de séries temporais financeiras e previsão do VaR, surge o problema de violações em grupos (*clusters*). Christoffersen (1998) mostrou que o problema de determinar a precisão de um modelo de previsão intervalar, como o VaR, pode ser reduzido

ao estudo das propriedades de cobertura não condicional (UC) e independência (IND) da sucessão "hit", indicadora de violação. Uma infracção problemática da hipótese IND é a que surge associada a violações que ocorrem em *clusters*. Este tipo de violações sinalizam um modelo que não reage atempadamente à mudança de condições e num contexto de mercados financeiros, corresponde à ocorrência de um grande número de perdas elevadas num espaço curto de tempo. Nesta tese propomos uma nova classe de testes para a hipótese IND e uma definição para tendência para violações em *clusters*. Estes testes utilizam a duração até à primeira violação e as durações entre as violações. As distribuições exacta e assintótica da correspondente estatística de teste são deduzidas e estudamos o problema da escolha de um teste pertencente a esta classe. Este teste é adequado para detectar modelos com tendência para produzirem violações em *clusters* e apresenta várias vantagens em relação às alternativas presentes na literatura, designadamente, é baseado numa distribuição exacta, é baseado numa estatística cuja distribuição não depende de um parâmetro perturbador e estudos por simulação mostram que tem um desempenho muito superior em termos de potência relativamente aos testes em estudo, com poucas excepções, apresentando mais do dobro da potência em muitos casos. Este teste também é aplicado a dados reais que abrangem a recente crise financeira global de 2008. A análise destes dados reais proporciona evidência que ignorar a propriedade IND foi uma importante razão para o fraco desempenho, durante a crise de 2008, do procedimento de "backtesting" definido nos Acordos de Basileia. Neste caso estudado, no qual usámos um modelo que produz violações em clusters, o teste que propomos rejeita a hipótese IND antes de todos os outros testes.

A problemática de violações em *clusters*, motivou o estudo da distribuição Weibull discreta. Sob a hipótese IND, as durações entre violações são va's com distribuição geométrica que é um caso particular da distribuição Weibull discreta. Com violações em clusters, verifica-se um excessivo número de durações muito curtas e um excessivo número de durações muito longas. A distribuição Weibull discreta com o parâmetro de forma inferior a 1, gera este padrão de durações e por isso a estimativa deste parâmetro pode ser usada para detectar modelos que violam a hipótese IND desta forma. Nesta tese, um novo estimador para o parâmetro de forma da distribuição Weibull discreta é proposto, sendo deduzidas a função distribuição condicional e os momentos deste estimador. Utilizando as expressões teóricas deduzidas e um estudo por simulação, o estimador proposto é comparado com outros estimadores em termos de viés e erro quadrátic médio, verificando-se que o novo estimador tem um desempenho muito superior quando o parâmetro de forma é inferior a 1. Recorrendo a uma série temporal financeira, a utilização deste estimador para identificar modelos de risco que produzem violações em *clusters* é ilustrada. A distribuição Weibull discreta tem muitas aplicações para além da apresentada no âmbito da gestão quantitativa do risco e por isso a aplicabilidade do estimador proposto também se estende para além desta área.

No Capítulo 1, é apresentada evidência que o pressuposto iid não é realista num contexto de séries temporais financeiras. Nestas séries é usual obter forte evidência de ocorrência de clusters de volatilidade e dependência serial não linear. Isto pode levar a um desempenho não adequado de modelos VaR que assumem o pressuposto iid para os retornos. Para lidar com esta dependência, surgiu na literatua uma metodologia híbrida que combina modelos de volatilidade tipo-GARCH (do inglês "Generalized AutoRegressive Conditional Heteroskedasticity") com a EVT. Nesta tese propomos uma metodologia alternativa com base apenas na EVT e que não necessita de assumir um modelo paramétrico para toda a distribuição dos retornos mas apenas na cauda e com base em sólida teoria assintótica. A metodologia proposta utiliza os excessos acima de um nível elevado e as durações entre estes excessos como covariáveis, tendo sido designada por DPOT (do inglês "Duration based Peaks Over Threshold"). Na literatura, métodos baseados no ajustamento de um modelo estocástico aos excessos acima de um nível elevado $u$ foram desenvolvidos sob a designação POT (do inglês "Peaks Over Threshold"). Um dos mais importantes Teoremas da EVT estabelece que os excessos acima de $u$ seguem aproximadamente uma distribuição de Pareto Generalizada, quando a distribuição subjacente pertence ao domínio de atracção de uma distribuição GEV. A metodologia DPOT recorre ao método POT e à modelação do parâmetro de forma utilizando as durações como covariáveis. Com base neste método, três modelos DPOT, para previsão do VaR para o dia seguinte, foram comparados com outros modelos utilizando os retornos históricos de índices de acções. Os resultados empíricos mostram que, para a probabilidade 0.01 estipulada nos Acordos de Basileia e com os índices estudados, estes modelos têm um desempenho muito bom em termos de cobertura não condicional e em termos de independência. Os modelos DPOT revelaram um melhor desempenho "out-of-sample" que os modelos que constituem o estado-da-arte de modelos de risco. Em comparação com o popular modelo RiskMetrics desenvolvido pela J.P. Morgan, o desempenho é muito superior quer em termos de cobertura condicional, quer em termos de independência. Adicionalmente, tendo em conta o cálculo dos requisitos de capital no âmbito dos Acordos de Basileia, no período estudado e para os índices considerados, os modelos DPOT propostos conduziram a requisitos de capital médio inferiores aos dos outros modelos, mas antecipando melhor os momentos de elevada volatilidade.

Finalmente, importa notar que o pressuposto iid não é necessáriamente uma limitação. Na última parte desta tese, recorrendo a dados reais, mostramos que no caso do VaR com níveis de probabilidade muito baixos, i.e., no caso de quantis elevados, um estimador de variância mínima e viés reduzido (MVRB do inglês ”Minimum Variance Reduced Bias”), introduzido recentemente na literatura, que incorpora um dos estimadores PORT sugerido nesta tese e que é baseado no pressuposto iid, pode ser extraordináriamente preciso. Nesta última parte, é comparado o desempenho “out-of-sample” de modelos VaR baseados neste estimador, de modelos DPOT e de outros modelos VaR. Neste contexto, os níveis de probabilidade muito baixos utilizados foram $p = 0.001$ e $p = 0.0005$, que correspondem a alterações adversas de preços que se espera ocorrerem em média uma vez de quatro em quatro anos ou em média uma vez de oito em oito anos. O VaR com estes níveis de probabilidade pode ter interesse no desenvolvimento dos “testes de stress”.

Nos estudos empíricos apresentados, para previsão do VaR com um nível de probabilidade igual ou inferior a 0.01, os modelos DPOT e os modelos baseados num estimador MVRB que incorpora um estimador PORT, revelaram mais precisão que os modelos paramétricos condicionais que são muito utilizados pelos econometristas, verificando-se em alguns casos uma enorme diferença no desempenho ”out-of-sample”.

## Acknowledgements

First of all I would like to thank my supervisor Prof. Isabel Fraga Alves, for her guidance, advices, for pushing me, believing in me and for all the time and patience, both during difficult periods and during periods of enthusiasm when, after hard work, the objectives are achieved. Her help and knowledge was essential for the accomplishment of this thesis. My deep gratitude.

I would also like to thank Prof. Maria Ivette Gomes for the privilege of having two chapters as joint work with her.

I thank Markus Haas, Mark Paollela, Juan-Ángel Jiménez-Martin, Pilar Abad and Sonia Benito for useful suggestions and comments.

I would like to thank the Center of Statistics and Applications of University of Lisbon (CEAUL) for financial support and the Fundação para a Ciência e Tecnologia (FCT) that through the PROTEC program provides financial support and the last two years with only six hours of classes per week. I also thank Escola Superior de Gestão e Tecnologia de Santarém (ESGTS) and Instituto Politécnico de Santarém (IPS) for the institutional support.

I thank to my friends for the support and for understanding my absences.

Finally I would like to express my deep gratitude to all my family, and especially to Antonieta, Beatriz and Carolina, for their understanding, their support and for the joy of their presence during the development of this work.

To Antonieta, Beatriz and Carolina

## Abbreviations

| | |
|---|---|
| ADI | Authorized Deposit-taking Institution |
| APARCH | Asymmetric Power Autoregressive Conditional Heteroscedasticity |
| AR | Autoregressive |
| CC | Conditional Coverage |
| cdf | Cumulative distribution function |
| CEVT | Conditional Extreme Value Theory |
| DPOT | Duration based POT method |
| df | Density function |
| ecdf | Empirical cumulative distribution function |
| EVT | Extreme Value Theory |
| FES | Frequency of excluded samples |
| iid | independent and identically distributed |
| IND | Independence |
| GARCH | Generalized AutoRegressive Conditional Heteroskedasticity |
| GMM | Generalised Method of Moments |
| GEV | Generalized Extreme Value Distribution |
| GPD | Generalized Pareto Distribution |
| HS | Historical Simulation |
| ML | Maximum Likelihood |
| MM | Maximum to Median test |
| MSE | Mean Square Error |
| MVRB | Minimum Variance Reduced Bias |
| o.s. | Order statistic |
| POT | Peaks Over Threshold |
| PORT | Peaks Over a Random Threshold |
| pmf | Probability mass function |
| REFF | Relative Efficiency Indicator |
| RMSE | Root of Mean Square Error |
| rv | Random variable |
| UC | Unconditional Coverage |
| VaR | Value-at-Risk |

# Contents

# 1 Introduction

## 1.1 Extreme value theory

Extreme value theory (EVT) is the theory of modeling and measuring events which occur with very small probability. In this sense, EVT gives a probabilistic framework to model what is often called extreme or rare events. One classic example of the need of EVT is the problem raised in the Netherlands with the assessment of dikes height needed to protect the land (de Haan, 1990). Starting from the first edition in 1958 of Gumbel (2004), universally acknowledged as the classic text about statistics of extremes, many examples of application have been provided over the last few decades and an extensive literature is available, on applications in various areas such as traffic analysis, hydrology, earthquake engineering, environmental science, finance, insurance, among others. Reference Books in the field of real world applications of EVT are Kotz and Nadarajah (2000), Coles (2001), Embrechts et al. (2001), Beirlant et al. (2004), Castillo et al. (2005), Reiss and Thomas (2007), Falk et al. (2010). In this thesis, the main contributions are within the fields of EVT and quantitative risk management with examples of applications to finance. The chapters 2 and 3 were developed under the assumption that $X_1, X_2, ..., X_n$ is a sample of independent and identically distributed (iid) random variables (rv) from some cumulative distribution function (cdf) $F$. Section 1.3 shows evidence that in the field of finance and in the context of log returns the iid assumption is unrealistic and later on, using durations, we will take into account the dependence and the non identical distribution of the returns.

Under the iid assumption the cdf of the maximum, denoted by $X_{n:n}$, is

$$P[X_{n:n} \leq x] = P[X_1 \leq x, X_2 \leq x, \cdots, X_n \leq x] = F^n(x).$$

Usually, in practical applications, the cdf $F$ is unknown and the previous result is insufficient to obtain the cdf of $X_{n:n}$. To overcome this, the convergence of $(X_{n:n} - b_n)/a_n$ with $n \to \infty$, was studied, where $a_n > 0$ and $b_n$ are sequences of real numbers known as the normalizing constants. The following theorem (Fisher and Tippett, 1928; Gnedenko, 1943) states that $X_{n:n}$ after the proper normalization, converges in distribution to one of three possible distributions.

**Theorem 1.1.1** (Fisher-Tippett-Gnedenko theorem)**.** *If a sequences of real numbers $a_n > 0$ and $b_n$ exists such that*

$$\lim_{n\to\infty} P\Big(\frac{X_{n:n} - b_n}{a_n} \leq x\Big) = F^n(a_n x + b_n) = G(x),$$

*then if $G$ is a non degenerate distribution function, it belongs to one of the following types*

$$\Phi_\alpha(x) = \exp(-x^{-\alpha}), \qquad x > 0, \alpha > 0 \qquad \textit{Fréchet,}$$

$$\Psi_\alpha(x) = \exp(-(-x^{\alpha})), \qquad x < 0, \alpha > 0 \qquad \textit{Weibull}$$

$$\Lambda_\alpha(x) = \exp(-\exp(-x)), \qquad x \in \mathbb{R} \qquad \textit{Gumbel.}$$

The three distributions of the theorem 1.1.1 are particular cases of the Generalized Extreme Value (GEV)

$$G_\gamma(x) = \begin{cases} \exp\left\{-(1+\gamma x)^{-1/\gamma}\right\}, & 1+\gamma x > 0,\ \gamma \neq 0 \\ \exp\left(-e^{-x}\right), & x \in \mathbb{R},\ \gamma = 0, \end{cases} \tag{1.1.1}$$

where the parameter $\gamma$ is known as the *tail index* and is related with the tail weight of the distribution. This unified model is due to Von Mises (1936) and reduces to the Fréchet, Weibull and Gumbel, respectively, for $\gamma > 0$, $\gamma < 0$ and $\gamma = 0$. If the result in theorem 1.1.1 holds for a distribution $F$ it is said that $F$ belongs to the max-domain of attraction of the cdf $G_\gamma$ and this condition is denoted by $F \in D(G_\gamma)$. The case $\gamma < 0$ corresponds to a cdf $F$ with finite endpoint, such as the uniform and the beta distributions. The case $\gamma = 0$ corresponds to a cdf $F$ with exponentially decaying tail such as normal and lognormal distributions. The case $\gamma > 0$ corresponds to a cdf $F$ with polynomially decaying tail such as the Pareto, Burr, Student's t, among others, and these distributions are referred as being *heavy-tailed*. Recently, Neves and Fraga Alves (2008) proposed the concept of *super heavy-tailed* distributions. Heavy-tailed distributions are accepted in the literature as realistic distributions for several phenomena (e.g., Embrechts et al., 1997; Resnick, 2007). In the next Section, with the NASDAQ index returns, we will show empirical evidence about the need for distributions with heavier tails than the normal distribution, to model the unconditional distribution of the returns.

For small values of p, we want to extrapolate beyond the sample, estimating a high quantile

$$\chi_p(X) := F^{\leftarrow}(1-p),\ p = p_n \to 0,\ \text{as } n \to \infty,\ np_n \to c \geq 0. \tag{1.1.2}$$

Here $F^{\leftarrow}(t) := \inf\{x : F(x) \geq t\}$ denotes the generalized inverse function of $F$. As far as we know, semi-parametric high quantile-estimators in the literature prior to 2006, do not enjoy the adequate behavior in the presence of linear transformations of the data, related with the theoretical linearity of a quantile

$$\chi_p(\delta X + \lambda) = \delta\chi_p(X) + \lambda,\ \text{for any real } \lambda \text{ and real positive } \delta. \tag{1.1.3}$$

In chapter 2 we suggest a class of semi-parametric high quantile-estimators for which the empirical counterpart of the theoretical linear property (1.1.3) holds. In chapters 2 and 3, a new methodology for tail index and high quantiles estimation, based on the *excesses* over a random threshold - PORT methodology - is suggested and studied, under the iid assumption. This allow us, for example, to study the behavior of the tails of the unconditional distribution of the returns and to estimate the unconditional Value-at-Risk with small values of $p$. In chapter 3, among other results, for symmetric distributions with infinite left endpoint, we also prove the non-consistency of the classic Hill estimator (Hill, 1975) when a practical statistical methodology of transforming the original data through the subtraction of the minimum is used.

Another very important result in EVT is the Pickands-Balkema-de Haan theorem, which we review in chapter 6. This theorem specifies the form of the limit distribution of *excesses* over a high threshold. A new risk model, suggested in chapter 6, uses the excesses over a high threshold and is based on this theorem.

## 1.2 Returns and forecasting Value-at-Risk

Throughout this work, one of the main themes is the measurement of risk and the risk will be measured in terms of price changes. These changes can take the form of absolute price change, simple return, simple gross return and, log return. Using a similar notation as Campbell, Lo, and Mackinlay (1997), we recall the definitions of these forms of price changes. Let $P_t$ be the price of an asset at time $t$, the *absolute price change* of the asset between day $t$ and $t-1$ is defined as

$$D_t = P_t - P_{t-1},$$

the *simple return* on the asset, for the same period, is

$$R_t^* = \frac{P_t - P_{t-1}}{P_{t-1}},$$

and the *simple gross return* on the asset is

$$1 + R_t^* = \frac{P_t}{P_{t-1}}.$$

Note that the asset's $k$-period simple gross return, written as $1+R_t^*(k)$, is the product of the $k$ one period simple gross returns involved

$$1 + R_t^*(k) = \frac{P_t}{P_{t-k}} = \frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}} \cdots \times \frac{P_{t-k+1}}{P_{t-k}} = \prod_{j=0}^{k-1}(1 + R_{t-j}^*),$$

and the simple return over the $k$ periods, written $R_t^*(k)$, is equal to the $k$-period simple gross return minus one.

Let $C_0$ be the initial deposit in a Bank and $C_1$ the capital at the end of the period. Assume that the interest rate of the bank is $r \times 100\%$ per period and the bank pays interest $m$ times during the period. The final capital is $C_1 = C_0(1 + r/m)^m$. With continuous compounding, i.e., with $m \to \infty$, the final capital is $C_1 = C_0 \exp(r)$, and the continuously compounded interest rate is equal to $r$. Note that $r = \log(C_1/C_0)$. In a similar way, the continuously compounded return or *log return* of an asset is defined to be the natural logarithm of the simple gross return

$$R_t = \log(1 + R_t^*) = \log \frac{P_t}{P_{t-1}}. \tag{1.2.1}$$

Returns (simple and log returns) have more attractive statistical properties than absolute price changes and the latter do not measure change in terms of the initial price level. These are the reasons for working with the returns rather than the absolute price changes. Furthermore, for the log returns the following property holds

$$R_t(k) = \log(1 + R_t^*(k)) = \log\Big(\prod_{j=0}^{k-1}(1 + R_{t-j}^*)\Big) = R_t + R_{t-1} + \cdots + R_{t-k+1}.$$

Thus, the $k$-period log return is simple the sum of the one-period log returns involved. This property gives the log returns some advantages over the simple returns. For modeling the statistical behavior of returns over time, it is easier to derive the time series properties of additive processes than of multiplicative processes. Note also that $R_t = \log P_t - \log P_{t-1}$, is the first difference of $\log P_t$. The logarithm transformation and the first difference are very important transformations used in time series analysis for achieving stationarity.

To illustrate the differences between the forms of price changes, Table 1.1 presents daily close prices for the S&P 500 index for the period October 13, 1987 through October 19, 1987, and the corresponding daily absolute price changes, simple returns and log returns. The web site http://finance.yahoo.com was the source of the data. This period includes Monday, 19 October 1987, known as the Black Monday, when the stock markets around the world crashed. In this day the S&P 500 index lost more than 20% of its value with a price change equal to -57.86 usd, a simple return equal to -0.2047 and a log return equal to -0.229. As expected, the simple return and the log return series are similar to one another for small changes in the prices, but even with the largest change in the history of the index, the difference between the simple return and the log return is small.

Table 1.1
Absolute price changes, simple returns and log returns for the S&P 500 index.

| Date | Price S&P500, $P_t$ | Absolute price change, $D_t$ | Simple return $R_t^* \times 100$ | Log return $R_t \times 100$ |
|---|---|---|---|---|
| 1987-10-13 | 314.52 | 5.13 | 1.66 | 1,645 |
| 1987-10-14 | 305.23 | -9.29 | -2.95 | -2,998 |
| 1987-10-15 | 298.08 | -7.15 | -2.34 | -2,370 |
| 1987-10-16 | 282.70 | -15.38 | -5.16 | -5,298 |
| 1987-10-19 | 224.84 | -57.86 | -20.47 | -22,900 |

One of the main subjects treated in this thesis is Value-at-Risk (VaR), which allow us to measure the size of the risk. This measure is replacing the standard deviation or volatility as the most widely used measure of risk. VaR give us a monetary value (or a return) that we risk losing during a time horizon and with a confidence level. For example, a -2% one-day-ahead VaR(0.05) for a portfolio means that during the next day we can be 95 percent certain that the value of the portfolio will not decrease by more than 2%. For a detailed discussion of VaR, see Jorion (2000), the reference Holton (2003) provides details about the history of VaR and the reference McNeil et. al. (2006) is a general text about quantitative risk management. The need for a risk measure for setting of capital adequacy limits for financial institutions justify the emergence of VaR. Since the Basel II Accord, forecast at day $t$ the Value-at-Risk (VaR) for day $t+1$, become a daily task for many financial institutions. More formally, considering time-series of daily log returns (1.2.1), the VaR for time $t+1$, $\mathrm{VaR}_{t+1}(p)$, is defined by

$$P[R_{t+1} \leq \mathrm{VaR}_{t+1}(p)] = p, \tag{1.2.2}$$

where $p$ is the coverage rate or probability level. $\mathrm{VaR}_{t+1}(p)$ is a quantile $p$ of the return $R_{t+1}$ distribution. In Fig. 1.2.1, this risk measure is illustrated assuming a standard normal distribution. In this case, $\mathrm{VaR}_{t+1}(0.05) = -1.645$ and $[\mathrm{VaR}_{t+1}(0.05); +\infty[$ is a one sided interval forecast for $R_{t+1}$, with a confidence level equal to 0.95. We can write $P[\mathrm{VaR}_{t+1}(0.05) < R_{t+1} < +\infty] = 1 - p = 0.95$.

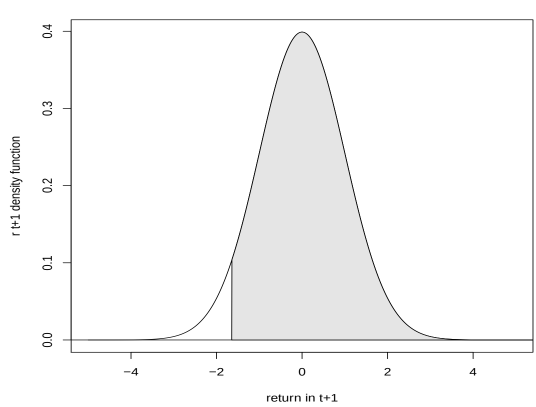



Figure 1.2.1: Area below the density curve and between $\mathrm{VaR}_{t+1|t}(0.05) = -1.645$ and $+\infty$, assuming a standard normal distribution for $R_{t+1}$.

When we forecast the $\text{VaR}_{t+1}(p)$ at time $t$, this one-day-ahead VaR forecast is denoted by $\text{VaR}_{t+1|t}(p)$. The hit function is defined as

$$I_{t+1}(p) = \begin{cases} 1 & \text{if } R_{t+1} < \text{VaR}_{t+1|t}(p) \\ 0 & \text{if } R_{t+1} \geq \text{VaR}_{t+1|t}(p). \end{cases} \tag{1.2.3}$$

To evaluate interval forecasts the hit function (1.2.3) is considered. This function produces a sequence of zeros and ones. The hit function, at time $t+1$, assumes the value one when the return falls outside the interval $[\text{VaR}_{t+1|t}(p), +\infty[$ and zero otherwise. When the return falls outside the interval we say that a *violation* occurs and usually this corresponds to a day with a large loss. Christoffersen (1998) showed that evaluating interval forecasts can be reduced to examining whether the hit sequence, $\{I_t\}_{t=1}^{T}$, satisfies the unconditional coverage (UC) and independence (IND) properties. These properties are explained in Chapter 4, where a new class of tests for the IND property is proposed.

Let us now suppose that the true distribution of $R_{t+1}$ is heavy-tailed, i.e., $F \in D(G_\gamma)$ with $\gamma > 0$. For example, let us suppose that the true distribution is $Student - t$ with 2 degrees of freedom (d.f.) $(\gamma = 1/2)$. In this case, if the standard normal model is assumed $(\gamma = 0)$ to forecast the $\text{VaR}_{t+1}(0.05)$, providing the value $-1.645$, we are underestimating the absolute value of $\text{VaR}_{t+1}(0.05)$ such that the probability of a violation is not 0.05 (the area below the density curve and between $-\infty$ and $-1.645$ in Fig. 1.2.1) but 0.120852 (the area below the density curve and between $-\infty$ and $-1.645$ in Fig. 1.2.2), which represents more than the double of the correct probability. We will show in Chapters 6 and 7, that something similar to this risk underestimation happens when the widely used RiskMetrics model (RiskMetrics, 1996), based on the normal distribution, is applied to portfolios that replicate stock indexes.

Empirical properties of returns are well documented in the literature (see for example Tsay, 2010). In Fig. 1.2.3, the histogram obtained with all the returns from the NASDAQ index until March 25, 2011, is presented. The normal density with mean equal to 0.0327 (the empirical mean) and standard deviation equal to 1.258 (the empirical standard deviation) is also ploted in the same Figure. The histogram has a higher peak around the mean, but heavier tails than that of the normal distribution. We also note some asymmetry with a heavier left tail than the right tail, which we will study in detail for this index in Chapter 3. The empirical evidence clearly indicates that the normality assumption is not appropriate to model the unconditional distribution of the NASDAQ index returns. This histogram give us an approximation to the unconditional distribution of $R_{t+1}$ but it is very important to note that it is possible to use the recent information until time $t$ to estimate the conditional distribution and improve the one-day-ahead

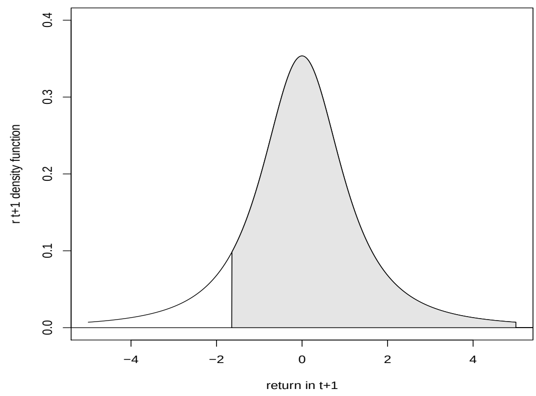



Figure 1.2.2: Area below the density curve and between −1.645 and +∞, assuming a Student-t distribution with 2 d.f. for $r_{t+1}$.

VaR forecast, $\mathrm{VaR}_{t+1|t}(p)$. The conditional distribution instead of the unconditional distribution is used by the parametric conditional models and by the conditional EVT model reviewed in Chapter 6. In the new model, proposed in Chapter 6, a conditional distribution is also used, but only for the tail and not for the entire distribution of the return. In the next section we discuss the concept of volatility, important to model the conditional distribution. There are several statistical approaches to VaR estimation; see, e.g., Kuester et al. (2005) and the references therein for a survey. Some of these approaches are reviewed in Chapter 6. Several studies conclude that conditional models based on EVT provide better out-of-sample performance to forecast one-day-ahead VaR; see, e.g., McNeil and Frey (2000), Byström (2004), Bekiros and Georgoutsos (2005), Kuester et al. (2006), Ghorbel and Trabelsi (2008), Ozun et al. (2010), Araújo Santos and Fraga Alves (2011).

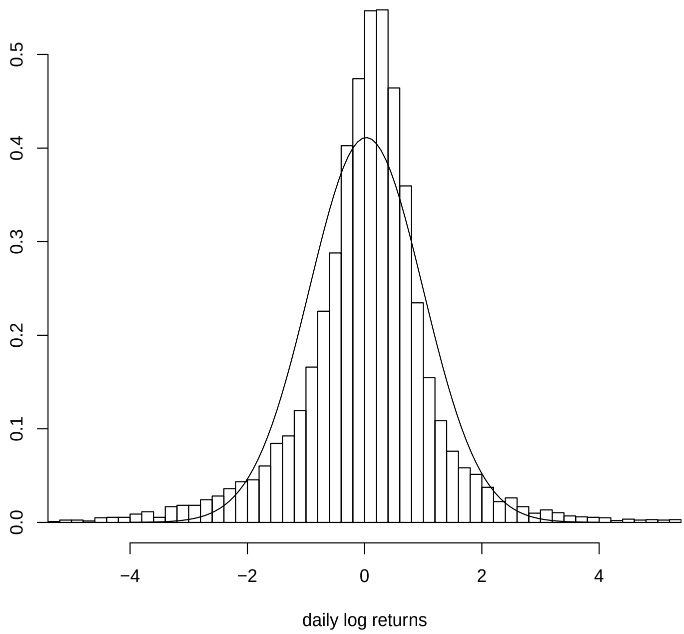



Figure 1.2.3: Histogram of 10123 returns for NASDAQ index from February 8, 1971 through March 25, 2011, with a normal density adjusted.

## 1.3 Volatility clustering, clusters of violations and durations

Fully parametric models in the location-scale class assumes for the returns,

$$R_t = \mu_t + \varepsilon_t = \mu_t + Z_t\sigma_t, \tag{1.3.1}$$

where $Z_t$ are a sequence of iid rv's with zero mean and unit variance, $\mu_t$ the conditional mean and $\sigma_t$ the conditional standard deviation. Unconditional parametric models set $\mu_t \equiv \mu$ and $\sigma_t \equiv \sigma$, conditionally homoskedastic parametric models set only $\sigma_t \equiv \sigma$ and conditionally heteroskedastic parametric models allow both the mean and standard deviation to be functions of past information. In light of the following evidence, to model realistically the returns, the chosen model must allow for a conditional standard deviation $\sigma_t$.

Volatility is usually defined as the conditional standard deviation $\sigma_t$ of the asset return. Volatility plays an important role in the field of finance. Two main applications are options pricing and risk management. Volatility is not directly observable but it has some characteristics that are frequently seen in asset returns. A very important characteristic is recognized in the literature at least since Mandelbroat (1963), which noted that large changes tend to be followed by large changes and small changes tend to be followed by small changes. This phenomenon of high volatility for certain time periods and low for other periods is known as *volatility clustering*. Fig. 1.3.1 shows the time series plot of log returns for the S&P 500 index where we observe that periods of large returns are clustered and distinct from periods of small returns, which are also clustered. The variabilities of returns vary over time and appear in clusters. If we measure the volatility in terms of standard deviation, then we have clearly evidence that the standard deviations change with time. Changing standard deviations (or variances) are denoted by the term heteroscedasticity. With time-varying standard deviations, the distribution of the returns is not constant over time, i.e., the returns are not identically distributed rv's. Moreover, the following evidence shows that returns are not independent.

Considering time series of returns, the autocorrelations are usually very weak, however the autocorrelations of non linear transformations as the square of returns or the absolute values of returns are highly significative until a large umber of lags as we can observe in Fig. 1.3.2 with the returns of S&P 500 index. The autocorrelations with lags between 1 and 41 are all much higher than the 95% confidence bands (dashed lines) and we have strong evidence that all the autocorrelations are higher than zero. This means that the past and present behavior of returns can help to predict the future behavior of returns. For log return series, usually strong evidence of non-linear serial dependence is found. With the strong evidence against the independence of returns over time and against the assumption of identically distributed returns, VaR models which assume iid returns, can suffer from a severe drawback. Diebold et al. discuss this drawback and to deal with it they suggest a hybrid method combining a volatility model with the EVT approach. In Chapter 6, we propose other approach which address this issue only under the EVT framework. However, the iid assumption is not necessarily a limitation. In Chapter 7 we show with real data that if the VaR at very low levels is considered, an high quantile estimator recently introduced in the literature, which uses one tail index estimator proposed in Chapter 2 and based on the iid assumption, can be extraordinarily accurate.

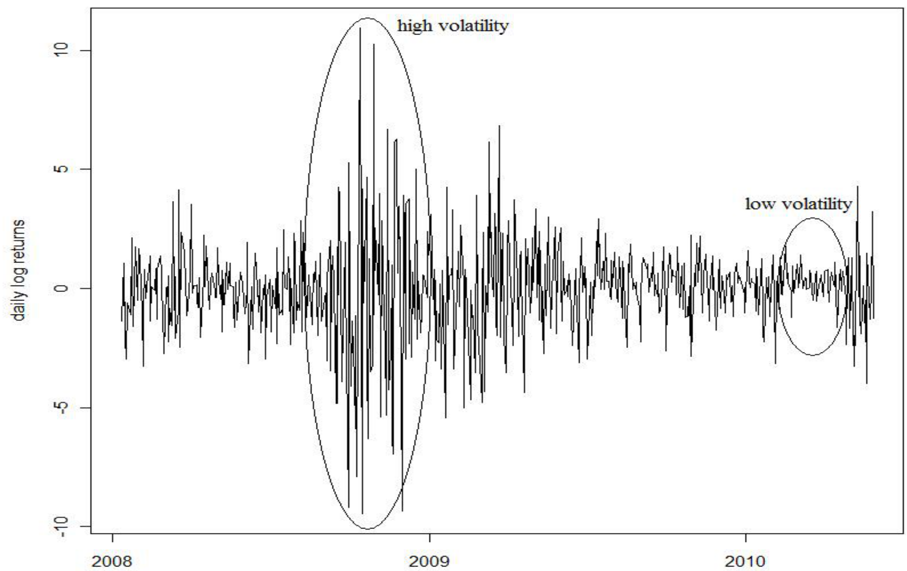



Figure 1.3.1: S&P 500 index returns.

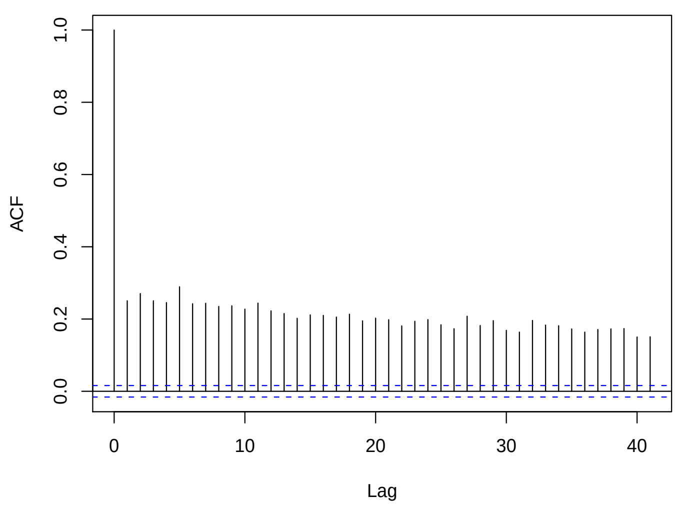



Figure 1.3.2: Sample autocorrelation coefficients for the absolute value of S&P 500 index returns.

Models that do not account for the volatility clustering phenomenon tend to produce clusters of violations. To illustrate this we choose the Historical Simulation (HS) method,

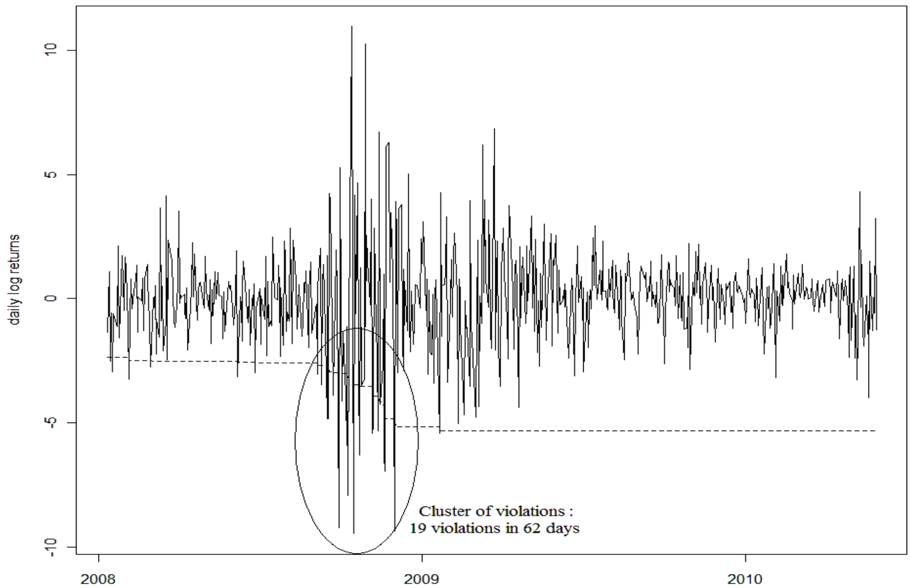



Figure 1.3.3: S&P 500 index returns (solid line) and HS VaR(0.01) forecasts (dashed lines).

a popular unconditional method to forecast VaR which easily generates clusters of violations when applied to heteroscedastic processes, especially with large samples. The most popular variety of this method compute VaR as an empirical quantile of a moving window of $n_w$ observations up to day $t$. For a detailed discussion of the HS method see Dowd (2002) and Christoffersen (2003). In Fig. 1.3.3 for the S&P 500 index and the same period of Fig. 1.3.1, we plot the returns and one-day-ahead 1% VaR computed with the HS method using $n_w = 1000$. The cluster of high volatility marked in Fig. 1.3.1, which corresponds to the 2008 financial crises, produced a cluster of violations with nineteen violations in only sixty-two trading days, when the expected with an accurate risk model is one violation each one hundred days. In Chapter 6, using an EVT method, under unconditional setup, which also do not account for the volatility clustering phenomenon, the same problem of clusters of violations is illustrated. One of the main objectives of this thesis is to suggest a conditional model to deal with this problem and one, based on EVT, is suggested in Chapter 6. Clusters of violations correspond to several large losses occurring in short periods of time and this constitutes one problematic infraction to the IND hypothesis of the hit sequence. In Chapter 4, a new independence test for the hit sequence and a definition for tendency to clustering of violations are proposed. This test can be used to backtesting VaR models and will help to identify models that suffer from the problem illustrated in Fig. 1.3.3; moreover, this test is more general and can be used in any context of interval forecasts evaluation.

The clusters of violations problematic has motivated the study of the discrete Weibull distribution. Let us define the duration between two consecutive violations as

$$D_i := t_i - t_{i-1}, \tag{1.3.2}$$

where $t_i$ denotes the day of violation number $i$ and $t_0 = 0$, which implies that $D_1$ is the time until the first violation. We denote a sequence of $N$ durations by $\{D_i\}_{i=1}^{N}$. We will show in Chapter 4 that the IND hypothesis of the hit sequence can be written as

$$D_i \overset{\text{iid}}{\sim} D \sim \text{Geometric}(\pi), \text{ with } 0 < \pi < 1.$$

The geometric distribution is a particular case of the discrete Weibull distribution, presented in Chapter 5, with the shape parameter $\theta$ equal to 1 and this allows us to write the IND hypothesis of the hit sequence as

$$D_i \overset{\text{iid}}{\sim} D \sim \text{discrete Weibull}(\theta = 1).$$

With clusters of violations, we have an excessive number of very short durations and an excessive number of very long durations. The discrete Weibull with $\theta < 1$ will generate this pattern and the estimate of $\theta$ can be used to identify a model that violates IND in this way. In Chapter 5, a new shape parameter estimator for the discrete Weibull distribution is suggested and their use to identify risk models that suffers from the tendency to clustering of violations problem is exemplified. In addition, the discrete Weibull distribution has many applications outside the field of quantitative risk management, some of them are referred in the introduction of Chapter 5, therefore the applicability of the proposed estimator also goes beyond this field.

In Chapter 6, the dependence and the non identical distribution of the returns is considered trough the use of durations as covariates. Here, the durations used are between excesses over a high threshold. We show that for the very important value of $p = 0.01$, established in the Basel II accord, the new conditional model proposed can perform better than state-of-the art VaR models, both in terms of out-of-sample accuracy and in terms of minimization of capital requirements under the Basel II accord. Finally, in Chapter 7 the application of the model proposed in Chapter 6 is extended to high quantiles. An high quantile estimator which uses one tail index estimator proposed in Chapter 2, is also included in the out-of-sample study realized. VaR at very small levels, with for example $p = 0.001$, may have interest in the development of stress tests (e.g., Longin, 2001;Tsay, 2010). In the appendix, R programs are provided to implement the suggested and used tests and models.

*Main reference:*
*Paulo Araújo Santos, Isabel Fraga Alves and Maria Ivette Gomes*
*REVSTAT - Statistical Journal*
*Volume 4, Number 3, November 2006, 227-247.*

# 2

# Peaks Over Random Threshold Methodology for Tail Index and High Quantile Estimation

A class of semi-parametric *high quantile* estimators which enjoy a desirable property in the presence of linear transformations of the data are presented. Such a feature is in accordance with the empirical counterpart of the theoretical linearity of a quantile $\chi_p$: $\chi_p(\delta X+\lambda) = \delta\chi_p(X)+\lambda$, for any real $\lambda$ and positive $\delta$. This class of estimators is based on the sample of excesses over a random threshold, originating what we denominate *PORT* (*Peaks Over Random Threshold*) methodology. We prove consistency and asymptotic normality of two high quantile estimators in this class, associated with the *PORT*-estimators for the tail index. The exact performance of the new tail index and quantile *PORT*-estimators is compared with the original semi-parametric estimators, through a simulation study.

## 2.1 Introduction

In this Chapter we deal with semi-parametric estimators of the *tail index* $\gamma$ and *high quantiles* $\chi_p$, which enjoy desirable properties in the presence of linear transformations of the available data. We recall that a *high quantile* is a value exceeded with a small probability. Formally, we denote by $F$ the heavy-tailed distribution function (cdf) of a random variable (rv) $X$, the common cdf of the i.i.d. sample $\underline{X} := \{X_i\}_{i=1}^{n}$, for which the high quantile (1.1.2) has to be estimated.

We consider estimators based on the $k+1$ top order statistics (o.s.), $X_{n:n} \geq \cdots \geq X_{n-k:n}$, where $X_{n-k:n}$ is an intermediate o.s., i.e., $k$ is an intermediate sequence of integers such that

$$k = k_n \to \infty, \quad k_n/n \to 0, \quad \text{as } n \to \infty. \tag{2.1.1}$$

We assume that we are working in a context of heavy tails, i.e., $\gamma > 0$ in the extreme value distribution (1.1.1), the non-degenerate cdf to which the maximum $X_{n:n}$ is attracted, after a suitable linear normalization. When this happens we say that the cdf $F$ is in the Fréchet domain of attraction and we write $F \in D(G_\gamma)_{\gamma>0}$.

The Chapter is developed under the first order regular variation condition, which allows the extension of the empirical cdf beyond the range of the available data, assuming a polynomial decay of the tail. This condition can be expressed by

$$F \in D(G_\gamma)_{\gamma>0} \quad \text{iff} \quad \overline{F} := 1 - F \in RV_{-1/\gamma} \quad \text{iff} \quad U \in RV_\gamma, \tag{2.1.2}$$

where $U$ is the quantile function defined as $U(t) := F^{\leftarrow}(1 - 1/t)$, $t \geq 1$; the notation $RV_\alpha$ stands for the class of regularly functions at infinity with index of regular variation $\alpha$, i.e., positive measurable functions $h$ such that $\lim_{t\to\infty} h(tx)/h(t) = x^\alpha$, for all $x > 0$.

It is interesting to note that the $p$-quantile can be expressed as $\chi_{p_n} = U\left(1/p_n\right)$.

To get asymptotic normality of estimators of parameters of extreme events, it is usual to assume the following extra second regular variation condition, that involves a non-positive parameter $\rho$:

$$\lim_{t\to\infty} \frac{U(tx)/U(t) - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho}, \tag{2.1.3}$$

for all $x > 0$, where $A$ is a suitably chosen function of constant sign near infinity. Then, $|A| \in RV_\rho$ and $\rho$ is called the second order parameter (Geluk and de Haan, 1987). For the strict Pareto model, with tail function $\overline{F}(x) = (x/C)^{-1/\gamma}$ and quantile function $U(t) = Ct^\gamma$, $U(tx)/U(t) - x^\gamma \equiv 0$. We then consider that (2.1.3) holds with $A(t) \equiv 0$.

More restrictively, we might consider that $F$ belonged to the wide class of Hall (Hall, 1982), that is, the associated quantile function $U$ satisfies

$$U(t) = \delta t^\gamma (1 + \gamma\beta t^\rho/\rho + o(t^\rho)), \ \rho < 0, \ \gamma, \delta > 0, \ \beta \in \mathbb{R}, \quad \text{as } t \to \infty, \tag{2.1.4}$$

or equivalently, (2.1.3) holds, with $A(t) = \gamma\beta t^\rho$. The strict Pareto model appears when both $\beta$ and the remainder term $o(t^\rho)$ are null.

Returning to the problem of high quantile estimation, we recall the classical semi-parametric Weissman-type estimator of $\chi_{p_n}$ (Weissman, 1978),

$$\widehat{\chi}_{p_n} = \widehat{\chi}_{p_n}(\underline{X}) = X_{n-k_n:n}\left(\frac{k_n}{np_n}\right)^{\hat{\gamma}_n}, \tag{2.1.5}$$

with $\hat{\gamma}_n = \hat{\gamma}_n(\underline{X})$ some consistent estimator of the tail parameter $\gamma$.

In the classical approach one considers for $\hat{\gamma}_n$ the well known Hill estimator (Hill, 1975),

$$\hat{\gamma}_n^H = \hat{\gamma}_n^H(\underline{X}) = \frac{1}{k_n}\sum_{j=1}^{k_n}\log\frac{X_{n-j+1:n}}{X_{n-k_n:n}}, \tag{2.1.6}$$

or the Moment estimator (Dekkers et al., 1989),

$$\hat{\gamma}_n^M = \hat{\gamma}_n^M(\underline{X}) = M_n^{(1)} + 1 - \frac{1}{2}\left\{1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}}\right\}^{-1}, \tag{2.1.7}$$

with $M_n^{(r)}$, the $r$-Moment of the log-excesses, defined by

$$M_n^{(r)} = M_n^{(r)}(\underline{X}) = \frac{1}{k_n}\sum_{j=1}^{k_n}\left(\log\frac{X_{n-j+1:n}}{X_{n-k_n:n}}\right)^r,\ r = 1, 2. \tag{2.1.8}$$

We use the following notation:

$$\widehat{\chi}_{p_n}^H = X_{n-k_n:n}\left(\frac{k_n}{np_n}\right)^{\hat{\gamma}_n^H}, \qquad \widehat{\chi}_{p_n}^M = X_{n-k_n:n}\left(\frac{k_n}{np_n}\right)^{\hat{\gamma}_n^M}. \tag{2.1.9}$$

Finally, we explain the question that motivated this chapter. It is well known that scale transformations to the data do not interfere with the stochastic behaviour of the tail index estimators (2.1.6) and (2.1.7), i.e., we can say that they enjoy scale invariance. The incorporation of (2.1.6) or (2.1.7) in the Weissman-type estimator in (2.1.5), allows us to obtain the following desirable exact property for quantile estimators: for any real positive $\delta$,

$$\widehat{\chi}_{p_n}(\delta\underline{X}) = \delta X_{n-k_n:n}\left(\frac{k_n}{np_n}\right)^{\hat{\gamma}_n} = \delta\widehat{\chi}_{p_n}(\underline{X}). \tag{2.1.10}$$

But we want a similar linear property in the case of location transformations to the data, $Z_j := X_j + \lambda,\ j = 1, \cdots, n$, for any real $\lambda$. That is, our main goal is that, for the transformed data $\underline{Z} := \{Z_j\}_{j=1}^n$, the quantile estimator satisfies

$$\widehat{\chi}_{p_n}(\underline{Z}) = \widehat{\chi}_{p_n}(\underline{X}) + \lambda. \tag{2.1.11}$$

Altogether, this represents the empirical counterpart of the theoretical linear property for quantiles (1.1.3).

Here we present a class of high quantile-estimators for which (2.1.10) and (2.1.11) hold exactly, pursuing the empirical counterpart of the theoretical linear property (1.1.3). For a simple modification of (2.1.5) that enjoys (2.1.11) approximately, see Fraga Alves and Araújo Santos (2004). For the use of reduced bias tail index estimation in high quantile estimation for heavy tails, see Gomes and Figueiredo (2003), Matthys and Beirlant (2003) and Gomes and Pestana (2005), where the second order reduced bias tail index estimator in Caeiro *et al.* (2005) is used for the estimation of the *Value at Risk*.

### 2.1.1 The class of high quantile estimators under study

The class of estimators suggested here is function of a sample of excesses over a random threshold $X_{n_q:n}$,

$$\underline{X}^{(q)} := \left(X_{n:n} - X_{n_q:n}, X_{n-1:n} - X_{n_q:n}, \cdots, X_{n_q+1:n} - X_{n_q:n}\right), \tag{2.1.12}$$

where $n_q := [nq] + 1$, with:

- $0 < q < 1$, for cdf's with finite or infinite left endpoint $x_F := \inf\{x : F(x) > 0\}$ (*the random threshold is an empirical quantile*);
- $q = 0$, for cdf's with finite left endpoint $x_F$ (*the random threshold is the minimum*).

A statistical inference method based on the sample of excesses $\underline{X}^{(q)}$ defined in (2.1.12) will be called a *PORT*-methodology, with *PORT* standing for*Peaks Over Random Threshold*. We propose the following *PORT-Weissman* estimators:

$$\widehat{\chi}^{(q)}_{p_n} = (X_{n-k_n:n} - X_{n_q:n}) \left(\frac{k_n}{np_n}\right)^{\hat{\gamma}_n^{(q)}} + X_{n_q:n}, \tag{2.1.13}$$

where $\hat{\gamma}_n^{(q)}$ is any consistent estimator of the tail parameter $\gamma$, made location/scale invariant by using the transformed sample $\underline{X}^{(q)}$. Indeed, the incorporation in the Adapted-Weissman estimator in (2.1.13), of tail index estimators, as function of the sample of excesses, allows us to obtain exactly the linear property (2.1.11).

### 2.1.2 Shifts in a Pareto model

To illustrate the behaviour of the new quantile estimators in (2.1.13), we shall first consider a parent $X$ from a Pareto$(\gamma, \lambda, \delta)$,

$$F_{\gamma,\lambda,\delta}(z) = 1 - \left(\frac{z-\lambda}{\delta}\right)^{-1/\gamma},\ z > \lambda + \delta,\ \delta > 0, \tag{2.1.14}$$

with $\lambda = 0$ and $\gamma = \delta = 1$. Let us assume that we want to estimate an upper $p = p_n = \frac{1}{n}$-quantile in a sample of size $n = 500$. Then, we want to estimate the parameter $\chi_p(X) = 500$. If we induce a shift $\lambda = 100$ to our data, we would obviously like our estimates to approach $\chi_p(X + 100) = 600$.

In Figure 2.1.1 we plot, for the Pareto$(\lambda, 1, 1)$ parents, with $\lambda = 0$ and $\lambda = 100$ and for $q = 0$ in (2.1.12), the simulated mean values of the *Weissman* and *PORT-Weissman* quantile estimators based on the Hill, denoted $\hat{\chi}_p^H$ and $\hat{\chi}_p^{H(q)}$, respectively. These mean values are based on $N = 500$ replications, for each value $k$, $5 \leq k \leq 500$, from the above mentioned models.

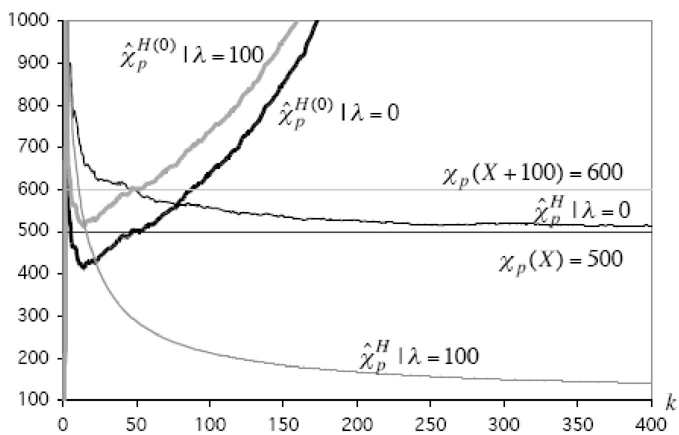



Figure 2.1.1: Mean values of $\widehat{\chi}_{p_n}^H$ and $\widehat{\chi}_{p_n}^{H(0)}$, $p_n = 0.002$ for samples of size $n = 500$ from a *Pareto*$(1, 0, 1)$ parent (target quantile $\chi_{p_n} = 500$) and from the *Pareto*$(1, 100, 1)$ (target quantile $\chi_{p_n} = 600$).

Similarly to the Hill horror plots (Resnick, 2004), associated to slowly varying functions $L_U(t) = t^{-\gamma}U(t)$, we also obtain here Weissman-Hill horror plots whenever we induce a shift in the simple standard Pareto model. Indeed, for a standard Pareto model ($\lambda = 0$ in (2.1.14)), Weissman type estimators in (2.1.5) perform reasonably well, with $\hat{\gamma}_n = \hat{\gamma}_n^H$. However, a small shift in the data may lead to disastrous results, even in this simple and specific case. For the *PORT-Weissman* estimates, the shift in the quantile estimates is equal to the shift induced in the data, a sensible property of quantile estimates. Figure 2.1.1 also illustrates how serious can be the consequences to the sample paths of the classical high quantile estimators, when we induce a shift in the data, as suggested in Drees (2003). We may indeed be led to dangerous misleading conclusions, like a systematic underestimation, for instance, mainly due to "stable zones" far away of the target quantile to be estimated.

### 2.1.3 Scope of the chapter

As far as we know, no systematic study has been done concerning asymptotic and exact properties of semi-parametric methodologies for tail index and high quantile estimation, using the transformed sample in (2.1.12). Somehow related with this subject, Gomes and Oliveira (2003), in a context of regularly varying tails, suggested a simple generalization of the classical Hill estimator associated to artificially shifted data. The shift imposed to the data is deterministic, with the aim of reducing the main component of the bias of Hill's estimator, getting thus estimates with stable sample paths around the target value. A preliminary study has also been carried out, by the same authors, replacing the artificial deterministic shift by a random shift, which in practice represents a transformation of the original data through the subtraction of the smallest observation, added by one, whenever we are aware that the underlying heavy-tailed model has a finite left endpoint.

With the purpose of tail index and high quantile estimation there is, in our opinion, a gap in the literature regarding classical semi-parametric estimation methodologies adapted for shifted data, the main topic of this paper.

In Section 2.2, we derive asymptotic properties for the adapted Hill and Moment estimators, as functions of the sample of excesses (2.1.12). In Section 2.3, we propose two estimators for $\chi_p$ that belong to the class (2.1.13) and prove their asymptotic normality. In Section 2.4, and through simulation experiments, we compare the performance of the new estimators with the classical ones. Finally, in Section 2.5, we draw some concluding remarks.

## 2.2 Asymptotic Behavior of Tail Index PORT-Estimators

For the classical Hill and Moment estimators, we know that for any intermediate sequence $k$ as in (2.1.1) and under the validity of the second order condition in (2.1.3),

$$\hat{\gamma}_n^H \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}} P_k^H + \frac{A(n/k)}{1-\rho}\big(1 + o_p(1)\big) \tag{2.2.1}$$

and

$$\hat{\gamma}_n^M \stackrel{d}{=} \gamma + \frac{\sqrt{\gamma^2+1}}{\sqrt{k}} P_k^M + \frac{(\gamma(1-\rho)+\rho)A(n/k)}{\gamma(1-\rho)^2}\big(1 + o_p(1)\big), \tag{2.2.2}$$

where $P_k^H$ and $P_k^M$ are asymptotically standard normal rv's.

In this Section we present asymptotic results for the classical Hill estimator in (2.1.6) and the Moment estimator in (2.1.7), both based on the sample of excesses $\underline{X}^{(q)}$ in (2.1.12), which will be denoted respectively, by

$$\hat{\gamma}_n^{H(q)} := \hat{\gamma}_n^H(\underline{X}^{(q)}) \quad \text{and} \quad \hat{\gamma}_n^{M(q)} := \hat{\gamma}_n^M(\underline{X}^{(q)}), \quad 0 \le q < 1. \tag{2.2.3}$$

In the following, $\chi_q^*$ denotes the $q$-quantile of $F$: $F(\chi_q^*) = q$ (by convention $\chi_0^* := x_F$), so that

$$X_{n_q:n} \overset{p}{\longrightarrow} \chi_q^*, \qquad \text{as } n \to \infty, \qquad \text{for } 0 \le q < 1.$$

For the estimators in (2.2.3) we have the asymptotic distributional representations expressed in Theorem 2.2.1.

**Theorem 2.2.1 (PORT-Hill and PORT-Moment).** *For any intermediate sequence $k$ as in* (2.1.1)*, under the validity of the second order condition in* (2.1.3)*, for any real $q$, $0 \le q < 1$, and with $T$ generally denoting either $H$ or $M$, the asymptotic distributional representation*

$$\hat{\gamma}_n^{T(q)} \overset{d}{=} \gamma + \frac{\sigma_T}{\sqrt{k}} P_k^T + \left( c_T A(n/k) + d_T \frac{\chi_q^*}{U(n/k)} \right) \Big(1 + o_p(1)\Big) \tag{2.2.4}$$

*holds, where $P_k^T$ is an asymptotically standard normal rv,*

$$\sigma_H^2 := \gamma^2, \quad c_H := \frac{1}{1-\rho}, \quad d_H := \frac{\gamma}{\gamma+1}, \tag{2.2.5}$$

$$\sigma_M^2 := \gamma^2 + 1, \quad c_M := \frac{\gamma(1-\rho)+\rho}{\gamma(1-\rho)^2} \quad \textit{and} \quad d_M := \left(\frac{\gamma}{\gamma+1}\right)^2. \tag{2.2.6}$$

**Remark 2.2.1.** Notice that $\sigma_M^2 = \sigma_H^2 + 1$, $c_M = c_H + \frac{\rho}{\gamma(1-\rho)^2}$ and $d_M = (d_H)^2$. Consequently, $\sigma_M > \sigma_H$, $c_M \le c_H$ and $d_M < d_H$.

The proof of Theorem 2.2.1 relies on the the following Lemmas 2.2.1 and 2.2.2.

**Lemma 2.2.1.** *Let $F$ be the cdf of $X$, and assume that the associated $U$-quantile function satisfies the second order condition* (2.1.3)*. Consider a deterministic shift transformation to $X$, defining the r.v. $X_q := X - \chi_q^*$ with cdf $F_q(x) = F(x + \chi_q^*)$ and associated $U_q$-quantile function given by $U_q(t) := F_q^{\leftarrow}(1 - 1/t) = U(t) - \chi_q^*$.*
*Then $U_q$ satisfies a second order condition similar to* (2.1.3)*, that is*

$$\lim_{t\to\infty} \frac{U_q(tx)/U_q(t) - x^\gamma}{A_q(t)} = x^\gamma \left(\frac{x^{\rho_q} - 1}{\rho_q}\right), \qquad \textit{for } x > 0, \quad \rho_q \le 0, \tag{2.2.7}$$

*with*

$$\left(A_q(t), \rho_q\right) := \begin{cases} (A(t)\,,\,\rho) & \textit{if} \quad \rho > -\gamma; \\[2ex] \left(A(t) + \dfrac{\gamma \chi_q^*}{U(t)}\,,\,-\gamma\right) & \textit{if} \quad \rho = -\gamma; \\[2ex] \left(\dfrac{\gamma \chi_q^*}{U(t)}\,,\,-\gamma\right) & \textit{if} \quad \rho < -\gamma. \end{cases} \tag{2.2.8}$$

**Proof:** Under (2.1.3), for $x > 0$,

$$\begin{aligned} \frac{U_q(tx)}{U_q(t)} &= \frac{U(tx) - \chi_q^*}{U(t) - \chi_q^*} \\ &= \frac{U(tx)}{U(t)} \left\{ \frac{1 - \chi_q^*/U(tx)}{1 - \chi_q^*/U(t)} \right\} \\ &= \frac{U(tx)}{U(t)} \left\{ 1 + \chi_q^* \frac{1/U(t) - 1/U(tx)}{1 - \chi_q^*/U(t)} \right\} \\ &= \frac{U(tx)}{U(t)} \left\{ 1 + \frac{\chi_q^*}{U(t)} \left[ 1 - \frac{U(t)}{U(tx)} \right] (1 + o(1)) \right\} \\ &= x^\gamma \left\{ 1 + \frac{x^\rho - 1}{\rho} A(t)(1 + o(1)) \right\} \left\{ 1 + \frac{\gamma \chi_q^*}{U(t)} \frac{x^{-\gamma} - 1}{-\gamma} (1 + o(1)) \right\} \\ &= x^\gamma \left\{ 1 + \frac{x^\rho - 1}{\rho} A(t) + \frac{\gamma \chi_q^*}{U(t)} \frac{x^{-\gamma} - 1}{-\gamma} + o(A(t)) + o(1/U(t)) \right\}. \end{aligned}$$

Then $U_q$ satisfies (2.2.7), for $A_q$ and $\rho_q$ defined in (2.2.8) and the result follows. □

**Lemma 2.2.2.** *Denote by* $M_n^{(r,q)}$ *the* $M_n^{(r)}$ *statistics in* (5.3.10)*, as functions of the transformed sample* $\underline{X}^{(q)}$*,* $0 \le q < 1$ *in* (2.1.12)*; that is,*

$$M_n^{(r,q)} := M_n^{(r)}(\underline{X}^{(q)}) = \frac{1}{k} \sum_{j=1}^{k} \left( \log \frac{X_{n-j+1:n} - X_{n_q:n}}{X_{n-k:n} - X_{n_q:n}} \right)^r, \quad r = 1, 2.$$

*Then, for any intermediate sequence* $k$ *as in* (2.1.1)*, under the validity of the second order condition in* (2.1.3) *and for any real* $q$*,* $0 \le q < 1$*,*

$$M_n^{(r,q)} - \frac{1}{k} \sum_{j=1}^{k} \left( \log \frac{X_{n-j+1:n} - \chi_q^*}{X_{n-k:n} - \chi_q^*} \right)^r = o_p\left( \frac{1}{U(n/k)} \right), \quad r = 1, 2.$$

**Proof:** We will consider $r=1$. Using the first order approximation $\ln(1+x) \sim x$, as $x \to 0$, together with the fact that $X_{n_q:n} = \chi_q^*(1+o_p(1))$, we will have successively

$$
\begin{aligned}
M_n^{(1,q)} \quad & - \quad \frac{1}{k}\sum_{j=1}^{k} \log \frac{X_{n-j+1:n}-\chi_q^*}{X_{n-k:n}-\chi_q^*} = \\
& = \quad \frac{1}{k}\sum_{j=1}^{k} \log \frac{X_{n-j+1:n}-X_{n_q:n}}{X_{n-k:n}-X_{n_q:n}} - \log \frac{X_{n-j+1:n}-\chi_q^*}{X_{n-k:n}-\chi_q^*} \\
& = \quad \frac{1}{k}\sum_{j=1}^{k} \log \frac{1-X_{n_q:n}/X_{n-j+1:n}}{1-X_{n_q:n}/X_{n-k:n}} - \log \frac{1-\chi_q^*/X_{n-j+1:n}}{1-\chi_q^*/X_{n-k:n}} \\
& = \quad \frac{1}{k}\sum_{j=1}^{k} \left( \frac{X_{n_q:n}}{X_{n-k:n}} - \frac{X_{n_q:n}}{X_{n-j+1:n}} + \frac{\chi_q^*}{X_{n-j+1:n}} - \frac{\chi_q^*}{X_{n-k:n}} \right)(1+o_p(1)) \\
& = \quad \frac{X_{n_q:n}-\chi_q^*}{X_{n-k:n}} \frac{1}{k}\sum_{j=1}^{k} \left(1 - \frac{X_{n-k:n}}{X_{n-j+1:n}}\right)(1+o_p(1)) \qquad (2.2.9) \\
& = \quad \frac{o_p(1)}{X_{n-k:n}} \frac{1}{k}\sum_{j=1}^{k} \left(1 - \frac{X_{n-k:n}}{X_{n-j+1:n}}\right)(1+o_p(1)) .
\end{aligned}
$$

Denote by $\{Y_j\}_{j=1}^k$ iid $Y$ standard Pareto rv's, with cdf $F_Y(y) = 1 - y^{-1}$, for $y > 1$ and $\{Y_{j:k}\}_{j=1}^k$ the associated o.s.'s.
Since $X_{n-k:n} \stackrel{d}{=} U(Y_{n-k:n})$, with $Y_{n-k:n}$ the $(n-k)$-th o.s. associated to an iid standard Pareto sample of size $n$ and $\left(\frac{k}{n}\right) Y_{n-k:n} \stackrel{p}{\longrightarrow} 1$, for any intermediate sequence $k$, then $\frac{X_{n-k:n}}{U(n/k)} \stackrel{p}{\longrightarrow} 1$; this together with the fact that $\left\{\frac{Y_{n-j+1:n}}{Y_{n-k:n}}\right\}_{j=1}^{k} \stackrel{d}{=} \{Y_{k-j+1:k}\}_{j=1}^k$ , allow us to write

$$\begin{aligned} M_n^{(1,q)} &- \frac{1}{k}\sum_{j=1}^{k} \log \frac{X_{n-j+1:n}-\chi_q^*}{X_{n-k:n}-\chi_q^*} = \\ &= \frac{o_p(1)}{U(Y_{n-k:n})}\frac{1}{k}\sum_{j=1}^{k}\left(1-\frac{U(Y_{n-k:n})}{U(\frac{Y_{n-j+1:n}}{Y_{n-k:n}}Y_{n-k:n})}\right)(1+o_p(1)) \\ &= \frac{1}{k}\sum_{j=1}^{k}\left(1-Y_{k-j+1:k}^{-\gamma}\right) o_p\left(\frac{1}{U(n/k)}\right)(1+o_p(1)) \\ &= \frac{1}{k}\sum_{j=1}^{k}\left(1-Y_{j}^{-\gamma}\right) o_p\left(\frac{1}{U(n/k)}\right)(1+o_p(1)). \end{aligned}$$

Now $E\big[Y^{-\gamma}\big] = \frac{1}{\gamma+1}$ and by the weak law of large numbers we obtain

$$\begin{aligned} M_n^{(1,q)} &- \frac{1}{k}\sum_{j=1}^{k} \log \frac{X_{n-j+1:n}-\chi_q^*}{X_{n-k:n}-\chi_q^*} = \\ &= \frac{\gamma}{\gamma+1}\left(1+o_p(1/\sqrt{k}\right) o_p\left(\frac{1}{U(n/k)}\right) \\ &= o_p\left(\frac{1}{U(n/k)}\right). \end{aligned}$$

For $r=2$ steps similar to the previous ones lead us to the result. □

**Remark 2.2.2.** *Note that if $q \in (0,1)$, $X_{n_q:n} - \chi_q^* = O_p(1/\sqrt{n})$ and from (2.2.9), for $r = 1, 2$, $\sqrt{k}\left[M_n^{(r,q)} - \frac{1}{k}\sum_{j=1}^{k}\left\{\log\frac{X_{n-j+1:n}-\chi_q^*}{X_{n-k:n}-\chi_q^*}\right\}^r\right] = O_p\left(\sqrt{k/n}\frac{1}{U(n/k)}\right) = o_p(1)$ holds.*

**Proof:** (*Theorem 2.2.1* ) Taking into account Lemma 2.2.2

$$\hat{\gamma}_n^{H(q)} = \frac{1}{k}\sum_{j=1}^{k}\log\frac{X_{n-j+1:n}-\chi_q^*}{X_{n-k:n}-\chi_q^*} + o_p\left(\frac{1}{U(n/k)}\right).$$

Now, considering the result in Lemma 2.2.1 and representation (2.2.1) adapted for the deterministic shift data from $X_q := X - \chi_q^*$ model, we obtain the following representation for PORT-Hill estimator

$$\hat{\gamma}_n^{H(q)} \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}P_k^H + \frac{A_q(n/k)}{1-\rho_q}\big(1+o_p(1)\big) + o_p\left(\frac{1}{U(n/k)}\right),$$

with $A_q(t)$ provided in (2.2.8), and the result (2.2.4) follows with $T = H$.

Similarly, considering Lemmas 2.2.1 and 2.2.2 and the representation (2.2.2) adapted for the deterministic shift data from $X_q := X - \chi_q^*$ model, we obtain for the PORT-Moment estimator the representation

$$\hat{\gamma}_n^{M(q)} \stackrel{d}{=} \gamma + \frac{\sqrt{\gamma^2+1}}{\sqrt{k}} P_k^M + \frac{(\gamma(1-\rho_{_q}) + \rho_{_q}) A_q(n/k)}{\gamma(1-\rho_{_q})^2} \big(1 + o_p(1)\big)$$
$$+ o_p\left(\frac{1}{U(n/k)}\right),$$

and result (2.2.4) follows with $T = M$. □

**Remark 2.2.3.** *Note that if we induce a deterministic shift $\lambda$ to data $X$ from a model $F =: F_0$, i.e., if we work with the new model $F_\lambda(x) := F_0(x - \lambda)$, the associated $U$-quantile function changes to $U_\lambda(t) = \lambda + \delta U_0(t)$. Then, as expected,* (2.2.4) *holds whenever we replace $\hat{\gamma}_n^{H(q)}$ by $\hat{\gamma}_n^H|\lambda$ (the Hill estimator associated with the shifted population with shift $\lambda$) provided that we replace $\chi_q^*$ by $-\lambda$. This topic has been handled in Gomes and Oliveira* (2003)*, where the shift $\lambda$ is regarded as a tuning parameter of the statistical procedure that leads to the tail index estimates. The same comments apply to the classical Moment estimator.*

**Corollary 2.2.1.** *For the strict Pareto model, i.e., the model in* (2.1.14) *with $\lambda = 0$ and $\gamma = \delta = 1$, the distributional representations* (2.2.4) *holds with $A(t)$ replaced by 0.*

Under the conditions of Theorems 2.2.1 and with the notations defined in (2.2.5) and (2.2.6), the following results hold:

**Corollary 2.2.2.** *Let $\mu_1$ and $\mu_2$ be finite constants and let $T$ generically denote either $H$ or $M$.*

**i)** *For $\gamma > -\rho$,*

$$\hat{\gamma}_n^{T(q)} \stackrel{d}{=} \gamma + \frac{\sigma_{_T}}{\sqrt{k}} P_k^T + c_{_T} A(n/k)\Big(1 + o_p(1)\Big).$$

*If $\sqrt{k}\ A\,(n/k) \to \mu_1$, then*

$$\sqrt{k}\left(\hat{\gamma}_n^{T(q)} - \gamma\right) \underset{n\to\infty}{\stackrel{d}{\longrightarrow}} Normal\left(\mu_1 c_{_T}, \sigma_{_T}^2\right).$$

**ii)** *For $\gamma < -\rho$,*

$$\hat{\gamma}_n^{T(q)} \stackrel{d}{=} \gamma + \frac{\sigma_{_T}}{\sqrt{k}} P_k^T + d_{_T} \frac{\chi_q^*}{U(n/k)}\Big(1 + o_p(1)\Big).$$

*If* $\sqrt{k}/U\left(n/k\right) \to \mu_2$, *then*

$$\sqrt{k}\left(\hat{\gamma}_n^{T(q)} - \gamma\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} Normal\left(\mu_2 d_{\scriptscriptstyle T}\chi_q^*, \sigma_{\scriptscriptstyle T}^2\right).$$

**iii)** *For* $\gamma = -\rho$,

$$\hat{\gamma}_n^{T(q)} \stackrel{d}{=} \gamma + \frac{\sigma_{\scriptscriptstyle T}}{\sqrt{k}}P_k^T + \left[c_{\scriptscriptstyle T} A(n/k) + d_{\scriptscriptstyle T}\frac{\chi_q^*}{U(n/k)}\right]\Big(1 + o_p(1)\Big).$$

*If* $\sqrt{k}A\left(n/k\right) \to \mu_1$ *and* $\sqrt{k}/U\left(n/k\right) \to \mu_2$, *then*

$$\sqrt{k}\left(\hat{\gamma}_n^{T(q)} - \gamma\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} Normal\left(\mu_1 c_{\scriptscriptstyle T} + \mu_2 d_{\scriptscriptstyle T}\chi_q^*, \sigma_{\scriptscriptstyle T}^2\right).$$

## 2.3 High Quantile PORT-Estimators

On the basis of (2.1.13), we shall now consider the following estimators of $\chi_{p_n}$, functions of the sample of excesses over $X_{n_q:n}$, i.e., of the sample $\underline{X}^{(q)}$ in (2.1.12):

$$\widehat{\chi}_{p_n}^{H(q)} := \left(X_{n-k_n:n} - X_{n_q:n}\right)\left(\frac{k_n}{np_n}\right)^{\hat{\gamma}_n^{H(q)}} + X_{n_q:n}, \quad 0 \le q < 1, \tag{2.3.1}$$

$$\widehat{\chi}_{p_n}^{M(q)} := \left(X_{n-k_n:n} - X_{n_q:n}\right)\left(\frac{k_n}{np_n}\right)^{\hat{\gamma}_n^{M(q)}} + X_{n_q:n}, \quad 0 \le q < 1. \tag{2.3.2}$$

For these estimators we have the asymptotic distributional representations presented in Theorem 2.3.1.

**Theorem 2.3.1.** *In Hall's class* (2.1.4), *for intermediate sequences* $k_n$ *that satisfy*

$$\log\left(np_n\right)/\sqrt{k_n} \to 0, \;\; \text{as } n \to \infty, \tag{2.3.3}$$

*with* $p_n$ *such that* (1.1.2) *holds, then, with* $T$ *denoting either* $H$ *or* $M$, $(c_{\scriptscriptstyle H}, d_{\scriptscriptstyle H}, \sigma_{\scriptscriptstyle H})$ *and* $(c_{\scriptscriptstyle M}, d_{\scriptscriptstyle M}, \sigma_{\scriptscriptstyle M})$ *defined in* (2.2.5) *and* (2.2.6), *respectively, and for any real* $q$, $0 \le q < 1$,

$$\frac{\sqrt{k_n}}{\sigma_T \log(k_n/(np_n))}\left(\frac{\widehat{\chi}_{p_n}^{T(q)}}{\chi_{p_n}} - 1\right) = P_k^T + \sqrt{k_n}\left(c_{\scriptscriptstyle T} A(n/k) + d_{\scriptscriptstyle T}\frac{\chi_q^*}{U(n/k)}\right)\Big(1 + o_p(1)\Big),$$

*where* $P_k^T$ *is an asymptotically standard normal rv*

**Proof:** From now on, we denote $a_n := \frac{k_n}{np_n}$. With the underlying conditions in (1.1.2), $a_n$ tends to infinity, as $n \to \infty$, and the quantile to be estimated can be expressed as

$$\chi_{p_n} = U\left(\frac{1}{p_n}\right) = U\left(\frac{na_n}{k_n}\right).$$

We will present the proof for $T = H$, since for $T = M$ the proof follows the same steps. First notice that

$$\begin{aligned}\widehat{\chi}_{p_n}^{H(q)} &= \left(X_{n-k_n:n} - X_{n_q:n}\right) a_n^{\hat{\gamma}_n^{H(q)}} + X_{n_q:n} \\ &= X_{n-k_n:n}\left[\left(1 - \frac{X_{n_q:n}}{X_{n-k_n:n}}\right) a_n^{\hat{\gamma}_n^{H(q)}} + \frac{X_{n_q:n}}{X_{n-k_n:n}}\right].\end{aligned}$$

Now, since $X_{n_q:n} \xrightarrow{p} \chi_q^*$, we have $\frac{X_{n_q:n}}{X_{n-k_n:n}} = o_p(1)$. Then

$$\widehat{\chi}_{p_n}^{H(q)} = X_{n-k_n:n}\left[a_n^{\hat{\gamma}_n^{H(q)}}(1 + o_p(1))\right],$$

which means that the proposed estimator $\widehat{\chi}_{p_n}^{H(q)}$ is asymptotically equivalent to the Weissman type estimator (2.1.5), whenever we use the consistent estimator $\hat{\gamma}_n \equiv \hat{\gamma}_n^{H(q)}$.

Consider now a convenient representation for the difference,

$$\widehat{\chi}_{p_n}^{H(q)} - \chi_{p_n} = X_{n-k_n:n}\left\{a_n^{\hat{\gamma}_n^{H(q)}} - a_n^{\hat{\gamma}_n^{H(q)}}\left(\frac{X_{n_q:n}}{X_{n-k_n:n}}\right) + \frac{X_{n_q:n}}{X_{n-k_n:n}} - \frac{\chi_{p_n}}{X_{n-k_n:n}}\right\},$$

and recall that we may write

$$\frac{\chi_{p_n}}{X_{n-k_n:n}} = \frac{U(\frac{n}{k_n}a_n)}{U(\frac{n}{k_n})}\frac{U(\frac{n}{k_n})}{U(Y_{n-k_n:n})}.$$

According to (2.1.3), for $\rho < 0$, $U(\frac{n}{k_n}a_n)/U(\frac{n}{k_n}) = a_n^{\gamma}\left(1 - A(n/k_n)/\rho\right)(1 + o_p(1))$.

Considering that for the estimator $\hat{\gamma}_n^{H(q)}$, the representation (2.2.4) holds, we get successively, for sequences $k_n$ that verify (2.3.3),

$$a_n^{\hat{\gamma}_n^{H(q)}} = a_n^{\gamma}\Big(1 + \log a_n\left(\hat{\gamma}_n^{H(q)} - \gamma\right)\Big)(1 + o_p(1))$$

and

$$\begin{aligned}\widehat{\chi}_{p_n}^{H(q)} - \chi_{p_n} &= a_n^{\gamma} X_{n-k_n:n}\Big\{1 + \log a_n\left(\hat{\gamma}_n^{H(q)} - \gamma\right)(1 + o_p(1)) \\ &\qquad - \left(1 - A(n/k_n)/\rho\right)(1 + o_p(1))\Big\} \\ &= a_n^{\gamma} X_{n-k_n:n}\Big\{\log a_n\left(\hat{\gamma}_n^{H(q)} - \gamma\right) + A(n/k_n)/\rho\Big\}(1 + o_p(1)).\end{aligned}$$

Now, we consider the following representation for intermediate statistics, proved in Ferreira et al. (2003),

$$X_{n-k_n:n} \stackrel{d}{=} U\left(\frac{n}{k_n}\right)\left(1 + \frac{\gamma B_k}{\sqrt{k_n}} + o_p\left(\frac{1}{\sqrt{k_n}}\right) + o_p(A(n/k_n))\right), \qquad (2.3.4)$$

with $B_k$ an asymptotically standard normal rv

Using (2.2.4) and (2.3.4) , we may write

$$\widehat{\chi}_{p_n}^{H(q)} - \chi_{p_n} = U\left(\frac{n}{k_n}\right) a_n^{\gamma} \left(1 + O_p(1/\sqrt{k_n}\right) \left\{W_n + A\left(\frac{n}{k_n}\right)/\rho\right\}(1 + o_p(1)),$$

where

$$\begin{aligned} W_n &= \log a_n \left(\hat{\gamma}_n^{H(q)} - \gamma\right) \\ &= \log a_n \left(\frac{\sigma_H}{\sqrt{k_n}} P_k^H + \left(c_{_H} A(n/k) + d_{_H} \frac{\chi_q^*}{U(n/k)}\right)\left(1 + o_p(1)\right)\right), \end{aligned}$$

with $P_k^H$ an asymptotically standard normal rv.

$$\frac{\widehat{\chi}_{p_n}^{H(q)} - \chi_{p_n}}{a_n^{\gamma} U(\frac{n}{k_n})} = \{W_n + A(n/k)/\rho\}(1 + o_p(1))$$

and

$$\frac{\sqrt{k_n}}{\sigma_H \log a_n}\left(\frac{\widehat{\chi}_{p_n}^{H(q)}}{\chi_{p_n}} - 1\right) = P_k^H + \sqrt{k_n}\left(c_{_H} A(n/k) + d_{_H}\frac{\chi_q^*}{U(n/k)}\right)\left(1 + o_p(1)\right).$$

□

The following result is a direct consequence of Corollary 2.2.2 and Theorem 2.3.1.

**Corollary 2.3.1.** *Under the same conditions of Theorem 2.3.1, then, with $T$ replaced by $H$ or $M$, and $(c_{_H}, d_{_H}, \sigma_{_H})$ and $(c_{_M}, d_{_M}, \sigma_{_M})$ defined in* (2.2.5) *and* (2.2.6)*, respectively, the following results hold.*

**i)** *For $\gamma > -\rho$,*

$$\frac{\sqrt{k_n}}{\sigma_{_T} \log(k_n/(np_n))}\left(\frac{\widehat{\chi}_{p_n}^{T(q)}}{\chi_{p_n}} - 1\right) = P_k^T + \sqrt{k_n}\left(c_{_T} A(n/k)\right)\left(1 + o_p(1)\right),$$

*If $\sqrt{k_n}A(n/k_n) \to \mu_1$, finite, as $n \to \infty$, then the mean value is $\mu_1 c_{_T}$.*

**ii)** *For $\gamma < -\rho$,*

$$\frac{\sqrt{k_n}}{\sigma_{_T} \log(k_n/(np_n))}\left(\frac{\widehat{\chi}_{p_n}^{T(q)}}{\chi_{p_n}} - 1\right) = P_k^T + \sqrt{k_n}\left(d_{_T}\frac{\chi_q^*}{U(n/k_n)}\right)\left(1 + o_p(1)\right),$$

*If $\sqrt{k_n}/U(n/k_n) \to \mu_2$, finite, as $n \to \infty$, then the mean values is $\mu_2 d_{_T} \chi_q^*$ .*

**iii)** *For $\rho = -\gamma$,*

$$\frac{\sqrt{k_n}}{\sigma_T \log(k_n/(np_n))}\left(\frac{\widehat{\chi}^{T(q)}_{p_n}}{\chi_{p_n}} - 1\right) = P^T_k + \sqrt{k_n}\Big(c_T A(n/k) + d_T \frac{\chi^*_q}{U(n/k_n)}\Big)\Big(1 + o_p(1)\Big),$$

*If $\sqrt{k_n}A(n/k_n) \to \mu_1$, finite, and $\sqrt{k_n}/U(n/k_n) \to \mu_2$, finite, as $n \to \infty$, then the mean value is $\mu_1 c_T + \mu_2 d_T \chi^*_q$.*

## 2.4 Simulations

Here, we compare the finite sample behavior of the proposed high quantile estimators $\widehat{\chi}^{H(q)}_{p_n}$ in (2.3.1) and $\widehat{\chi}^{M(q)}_{p_n}$ in (2.3.2) with the classical semi-parametric estimators $\widehat{\chi}^{H}_{p_n}$ and $\widehat{\chi}^{M}_{p_n}$ in (2.1.9). We have generated $N = 200$ independent replicates of sample size $n = 1000$ from the following models:

- Burr Model: $X \frown \text{Burr}(\gamma, \rho)$, $\gamma = 1$, $\rho = -2, -0.5$, with cdf

$$F(x) = 1 - (1 + x^{-\rho/\gamma})^{1/\rho}, \quad x \geq 0.$$

- Cauchy Model: $X \frown \text{Cauchy}$, $\gamma = 1$, $\rho = -2$, with cdf

$$F(x) = \frac{1}{2} + \frac{1}{\pi}\text{arctang } x, \quad x \in \mathbb{R}.$$

At a first stage, we generate samples from the standard models $F_0 := F$. At a second stage, we introduce a positive shift $\lambda = \chi_{0.01}$, i.e., a new location chosen in a comparable basis as the percentile 99% of the starting point distribution $F_0$. This defines a new model $F_\lambda(x) := F_0(x - \lambda)$ from the same family. We estimate the high quantile $\chi_{0.001}$, for each model $F_0$ or $F_\lambda$ from the referred Burr and Cauchy families, and we present patterns of Mean Values and Root of Mean Squared Errors, plotted against $k = 6, \cdots, 800$. The simulations illustrate the dramatic disturbance on the behavior of the classical quantile estimators in (2.1.9), when a shift is introduced. We, again, enhance that the flat stable zones achieved with these estimators, in the presence of shifts, could lead us to dangerous misleading conclusions, unless we are aware of the suitable threshold $k$ or of specific properties of the underlying model.

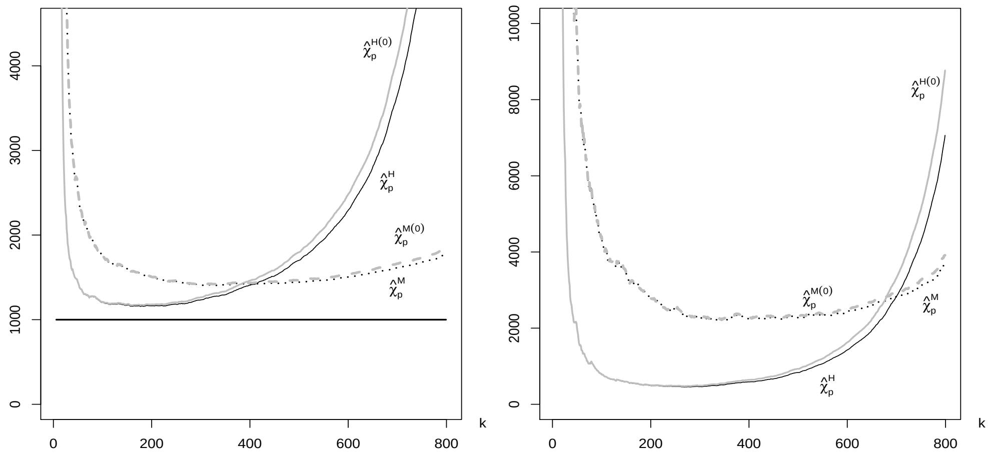



Figure 2.4.1: Mean values (left) and root mean squared errors (right), of $\widehat{\chi}_{p_n}^{H(0)}$, $\widehat{\chi}_{p_n}^{M(0)}$, $\widehat{\chi}_{p_n}^{H}$ and $\widehat{\chi}_{p_n}^{M}$, for a sample size $n = 1000$, from a Burr model with $\gamma = 1$, $\rho = -2$ and $\lambda = 0$ (target quantile $\chi_{0.001} = 1000$).

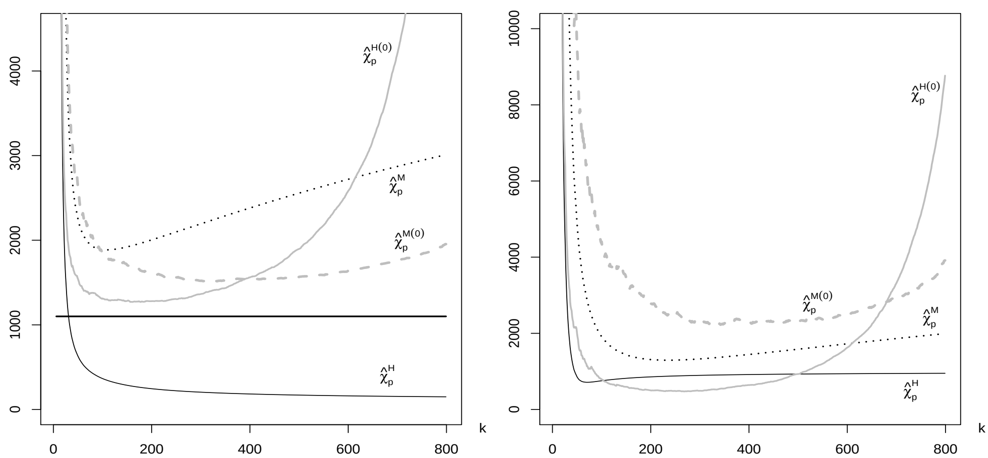



Figure 2.4.2: Mean values (left) and root mean squared errors (right), of $\widehat{\chi}_{p_n}^{H(0)}$, $\widehat{\chi}_{p_n}^{M(0)}$, $\widehat{\chi}_{p_n}^{H}$ and $\widehat{\chi}_{p_n}^{M}$, for a sample size $n = 1000$, from a Burr model with $\gamma = 1$, $\rho = -2$ and $\lambda = 99.99$ (target quantile $\chi_{0.001} = 1099.99$).

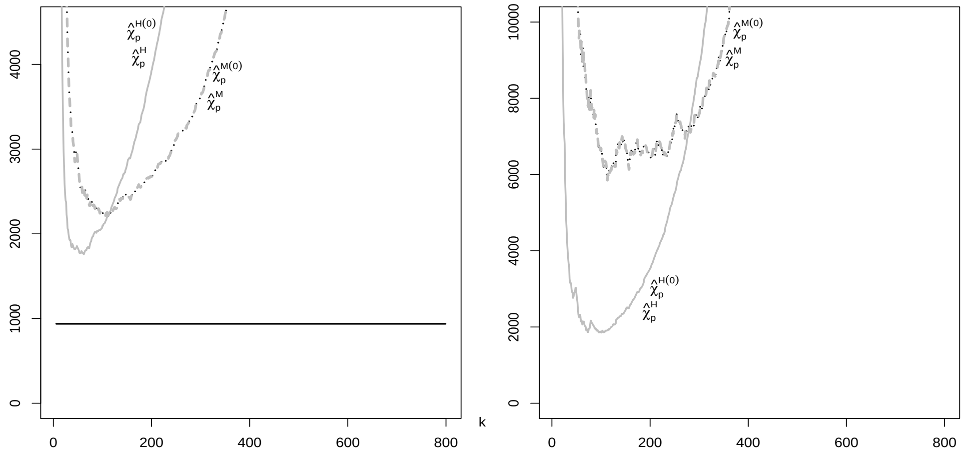



Figure 2.4.3: Mean values (left) and root mean squared errors (right), of $\widehat{\chi}_{p_n}^{H(0)}$, $\widehat{\chi}_{p_n}^{M(0)}$, $\widehat{\chi}_{p_n}^{H}$ and $\widehat{\chi}_{p_n}^{M}$, for a sample size $n = 1000$, from a Burr model with $\gamma = 1$, $\rho = -0.5$ and $\lambda = 0$ (target quantile $\chi_{0.001} = 937.731$).

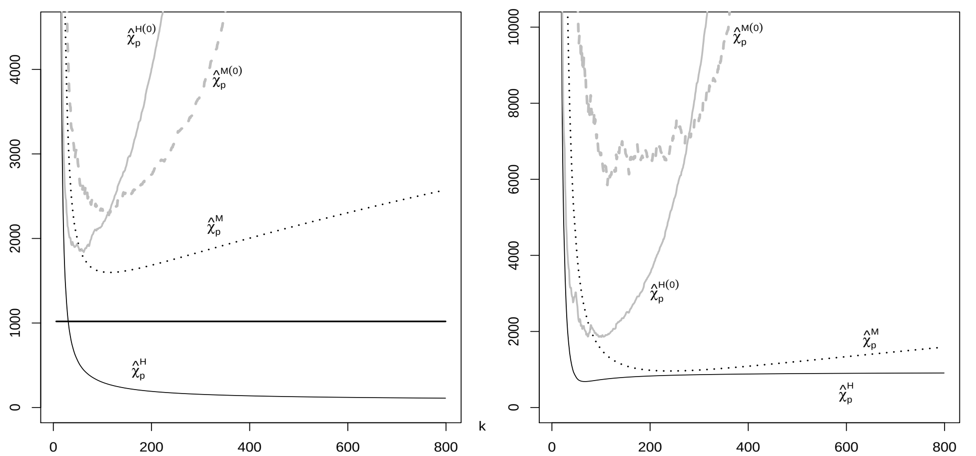



Figure 2.4.4: Mean values (left) and root mean squared errors (right), of $\widehat{\chi}_{p_n}^{H(0)}$, $\widehat{\chi}_{p_n}^{M(0)}$, $\widehat{\chi}_{p_n}^{H}$ and $\widehat{\chi}_{p_n}^{M}$, for a sample size $n = 1000$, from a Burr model with $\gamma = 1$, $\rho = -0.5$ and $\lambda = 81.023$ (target quantile $\chi_{0.001} = 1018.754$).

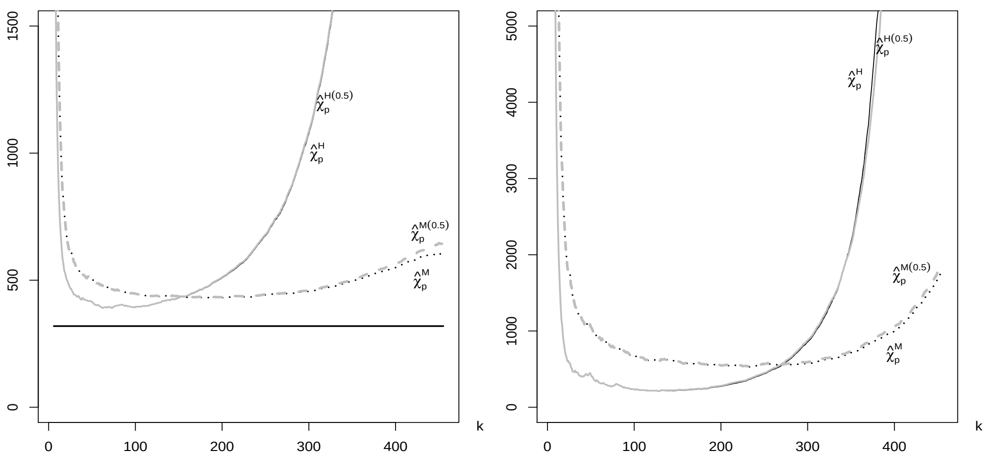



Figure 2.4.5: Mean values (left) and root mean squared errors (right), of $\widehat{\chi}_{p_n}^{H(0.5)}$, $\widehat{\chi}_{p_n}^{M(0.5)}$, $\widehat{\chi}_{p_n}^{H}$ and $\widehat{\chi}_{p_n}^{M}$, for a sample size $n = 1000$, from a Cauchy model with $\gamma = 1$, $\rho = -2$ and $\lambda = 0$ (target quantile $\chi_{0.001} = 319.309$).

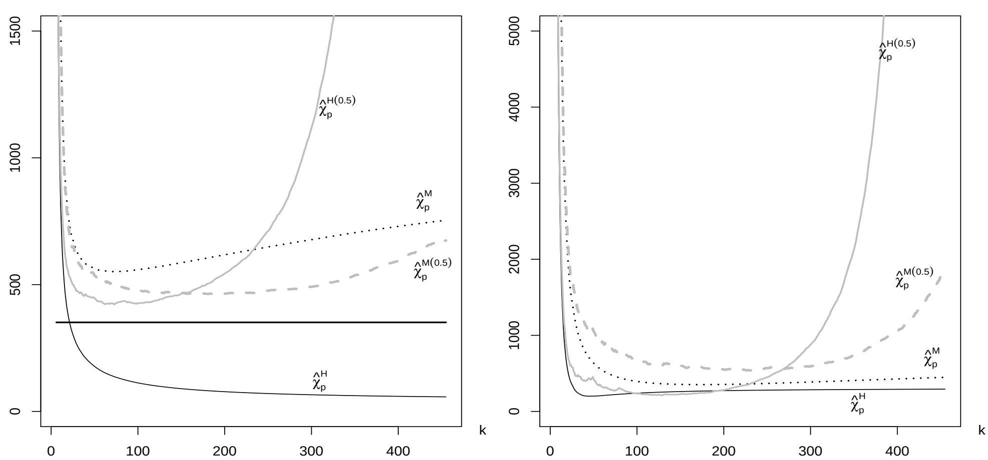



Figure 2.4.6: Mean values (left) and root mean squared errors (right), of $\widehat{\chi}_{p_n}^{H(0.5)}$, $\widehat{\chi}_{p_n}^{M(0.5)}$, $\widehat{\chi}_{p_n}^{H}$ and $\widehat{\chi}_{p_n}^{M}$, for a sample size $n = 1000$, from a Cauchy model with $\gamma = 1$, $\rho = -2$ and $\lambda = 31.821$ (target quantile $\chi_{0.001} = 351.13$).

From the figures, in this section, we observe that the classical quantile estimators diverge a lot from the important linear property (2.1.11). On the other hand, the estimators we propose, (2.3.1) and (2.3.2), enjoy exactly this property.

## 2.5 Concluding Remarks

- The *PORT* tail index and quantile estimators, based on the sample of excesses, $\underline{X}^{(q)}$, in (2.1.12), provide us with interesting classes of estimators, invariant for changes in location, as well as scale, a property also common to the classical estimators.

- In practice, whenever we use a *tuning parameter* $q$ in $(0, 1)$, we are always safe. Indeed, in such a case, the new estimators may or may not behave better than the classical ones, but they are consistent and asymptotically normal for the same type of $k$-values.

- A *tuning parameter* $q = 0$ is appealing but should be used carefully. Indeed, if the underlying parent has not a finite left endpoint, we are led to non-consistent estimators, with sample paths that may be erroneously flat around a value quite far away from the real target. This topic will be object of further study in the next Chapter.

*Main reference:*
*Maria Ivette Gomes, Isabel Fraga Alves and Paulo Araújo Santos*
*Communications in Statistics - Simulation and Computation*
*Volume 37, Number 7, 2008, 1281-1306.*

# 3

# PORT Hill and Moment Estimators for Heavy-Tailed Models

In this Chapter, we use the peaks over random threshold (PORT)-methodology, and consider Hill and moment PORT-classes of extreme value index estimators. These classes of estimators are invariant not only to changes in scale, like the classical Hill and moment estimators, but also to changes in location. They are based on the sample of excesses over a random threshold, the order statistic $X_{[nq]+1:n}$, $0 \leq q < 1$, being $q$ a tuning parameter, which makes them highly flexible. Under convenient restrictions on the underlying model, these classes of estimators are consistent and asymptotically normal for adequate values of $k$, the number of top order statistics used in the semi-parametric estimation of the extreme value index $\gamma$. In practice, there may however appear a stability around a value distant from the target $\gamma$ when the minimum is chosen for the random threshold, and attention is drawn for the danger of transforming the original data through the subtraction of the minimum. A new bias-corrected moment estimator is also introduced. The exact performance of the new extreme value index PORT-estimators is compared, through a large-scale Monte-Carlo simulation study, with the original Hill and moment estimators, the bias-corrected moment estimator, and one of the minimum-variance reduced-bias (MVRB) extreme value index estimators recently introduced in the literature. As an empirical example we estimate the tail index associated to a set of real data from the field of finance.

## 3.1 Introduction

The *extreme value index* (or *tail index*) $\gamma$ is the shape parameter in GEV (1.1.1). This cdf appears as the limiting cdf, as $n \to \infty$, of the linearly normalised maximum $X_{n:n}$ of an independent, identically distributed (iid), or even weakly dependent stationary sample of size $n$, $(X_1, ..., X_n)$. We shall work in a context of heavy-tailed models, i.e., we shall consider that $\gamma > 0$ in (1.1.1). Let us denote $F^{\leftarrow}(t) := \inf\{x : F(x) \geq t\}$, the generalized inverse function of F, $U(t) := F^{\leftarrow}(1 - 1/t)$, and $RV_\alpha$ the class of regularly functions at infinity with an index of regular variation $\gamma$, i.e., positive measurable functions $h$ such that $\lim_{t\to\infty} h(tx)/h(t) = x^\alpha$, for all $x > 0$. We shall work with models $F$ that are in the domain of attraction for maxima of $EV$ with $\gamma > 0$, denoted $\mathcal{D}_{\mathcal{M}}(EV_{\gamma>0})$, i.e., with models F such that:

$$1 - F \in RV_{-1/\gamma} \quad \text{or equivalently} \quad U \in RV_\gamma. \tag{3.1.1}$$

For the estimation of the right tail we consider two classical estimators of the *extreme value index* $\gamma$ based on the $k+1$ top order statistics (o.s.), denoted $\underline{X}_k := (X_{n:n} \geq \cdots \geq X_{n-k:n})$, where $X_{n-k:n}$ is an intermediate o.s. (2.1.1). Those estimators are the Hill estimator (Hill, 1975), with the functional expression

$$\hat{\gamma}^H_{n,k} = \hat{\gamma}^H_n(\underline{X}_k) := \frac{1}{k}\sum_{j=1}^{k} V_{jk}, \quad V_{jk} := \ln X_{n-j+1:n} - \ln X_{n-k:n}, \tag{3.1.2}$$

and the moment estimator (Dekkers et al., 1989),

$$\hat{\gamma}^M_{n,k} = \hat{\gamma}^M_n(\underline{X}_k) := M^{(1)}_{n,k} + 1 - \frac{1}{2}\{1 - (M^{(1)}_{n,k})^2/M^{(2)}_{n,k}\}^{-1} \tag{3.1.3}$$

with

$$M^{(r)}_{n,k} = M^{(r)}_n(\underline{X}_k) = \frac{1}{k}\sum_{j=1}^{k}\{V_{jk}\}^r, \quad r = 1, 2. \tag{3.1.4}$$

It is a well-known result in the field of *statistics of extremes* that the estimator in (3.1.2) is valid only for $\gamma \geq 0$, whereas the estimator in (3.1.3) is valid for all $\gamma \in \mathbb{R}$. They are both scale invariant, but not location invariant. Indeed, the associated estimates, particularly the Hill estimates, may suffer drastic changes when we induce an arbitrary shift in the data. Apart from the classical Hill and moment estimators, often simply denoted $H$ and $M$, respectively, we shall also consider one of the three classes of second-order reduced-bias extreme value index estimators recently introduced in Caeiro et al. (2005) and Gomes et al. (2007b, 2008b). These classes are based on the adequate estimation of a "scale" and a "shape" second order parameters, $\beta$ and $\rho$, respectively, are valid for a large class of heavy-tailed models and are appealing in the sense that we are able to reduce the asymptotic bias of the Hill estimator in (3.1.2) without increasing the asymptotic variance, which is kept at the value $\gamma^2$, the asymptotic variance of Hill's estimator. We shall call these estimators "minimum-variance reduced-bias" (MVRB) estimators. These MVRB-estimators are also non invariant for changes in location. However, they are much less sensitive to changes in location than the classical Hill estimator in (3.1.2). The simplest one, and the one we use, is the class provided in Caeiro et al. (2005) and further studied in Caeiro and Gomes (2008) here denoted $H$ for the sake of simplicity. Such a class has the functional form:

$$\hat{\gamma}^{\overline{H}}_{n,k,\hat{\beta},\hat{\rho}} = \hat{\gamma}^{\overline{H}}_{n,k,\hat{\beta},\hat{\rho}}(\underline{X}) := \hat{\gamma}^{H}_{n}\Big(1 - \frac{\hat{\beta}}{1-\hat{\rho}}\Big(\frac{n}{k}\Big)^{\hat{\rho}}\Big), \tag{3.1.5}$$

where $\hat{\beta}$ and $\hat{\rho}$ are adequate consistent estimators of the second-order parameters $\beta$ and $\rho$, respectively, to be specified later on in Section 3.2. We shall also consider a bias-corrected moment estimator, given by:

$$\hat{\gamma}^{\overline{M}}_{n,k,\hat{\beta},\hat{\rho}} = \hat{\gamma}^{\overline{M}}_{n,k,\hat{\beta},\hat{\rho}}(\underline{X}) := \hat{\gamma}^{M}_{n}\Big(1 - \frac{\hat{\beta}}{1-\hat{\rho}}\Big(\frac{n}{k}\Big)^{\hat{\rho}}\Big) - \frac{\hat{\beta}\hat{\rho}}{(1-\hat{\rho})^2}\Big(\frac{n}{k}\Big)^{\hat{\rho}}. \tag{3.1.6}$$

However, the main classes of estimators considered in this thesis are, just as the quantile estimators in Araújo Santos et al. (2006), functionals of a sample of excesses over a random threshold $X_{[nq]+1:n}$, i.e., functionals of (2.1.12). These new classes of extreme value index estimators are the so-called PORT-Hill estimators, also denoted $H(q)$, and the PORT-moment estimators, also denoted $M(q)$, theoretically studied, for heavy tails, in Araújo Santos et al. (2006). They are denoted here by

$$\hat{\gamma}^{T(q)}_{n,k} := \hat{\gamma}^{T}_{n}(\underline{X}^{(q)}_{k}) \quad 0 \le p < 1, \quad \text{with} \quad T = H \,\text{or}\, M, \tag{3.1.7}$$

where $\hat{\gamma}^{H}_{n,k}$, $\hat{\gamma}^{M}_{n,k}$, and $\underline{X}^{(q)}_{k}$ are provided in (3.1.2), (3.1.3), and (2.1.12), respectively. The estimators in (3.1.7) are now invariant for both changes of scale and location in the data, and depend on the *tuning parameter* $q$, that provides a highly flexible class of extreme value index estimators, which may even compare favorably with the MVRB extreme value index estimators, provided that we adequately choose the tuning parameter $q$. The choice $q = 0$ is appealing in practice, but should be used with care, as it can induce a problem of sub-estimation.

In Section 2.1.2 with the study of the behaviour of the classical high quantile estimators when we induce a shift in the data, we gave a motivation to the need of new estimation procedures like the above mentioned PORT methodology. This is also valid for the tail index estimators, since with the classical tail index estimators we achieve a similar behaviour. In Section 3.2, we provide the asymptotic properties of the estimators under study, we show the non-consistency of the PORT-Hill estimator $H(0)$, for symmetric models with infinite left endpoint and, through simulation experiments, we compare the exact performance of the new estimators in (3.1.7) with the classical Hill and moment estimators in (3.1.2) and (3.1.3), respectively, as well as with the reduced-bias extreme value index estimators in (3.1.5) and in (3.1.6). Finally, in Section 3.3 we provide an illustration of the behaviour of the estimators for a set of real data in the field of finance.

## 3.2 Distributional Behaviour of the Estimators Under Comparison

### 3.2.1 A Brief Reference to Their Asymptotic Behaviour

In order to obtain a non degenerate behavior for any extreme value index estimator, under a semi-parametric framework, it is convenient to assume the second-order condition (2.1.3). Here, and mainly because of the reduced-bias estimators in (3.1.5) and (3.1.6), we shall more restrictively assume that F belongs to the wide class of Hall (1982) presented in (2.1.4). For the classical $H$ and $M$ estimators, generally denoted $T$, we know that for any intermediate sequence $k$ as in (2.1.1) and under the validity of the second-order condition in (2.1.3):

$$\hat{\gamma}^T_{n,k} \stackrel{d}{=} \gamma + \frac{\sigma_T P^T_k}{\sqrt{k}} + c_T A(n/k)(1 + o_P(1)), \tag{3.2.1}$$

where

$$\sigma_H = \gamma, \quad c_H = \frac{1}{1-\rho}, \tag{3.2.2}$$

$$\sigma_M = \sqrt{\sigma^2 + 1}, \quad c_M = \frac{1}{1-\rho} + \frac{\rho}{\gamma(1-\rho)^2}, \tag{3.2.3}$$

being $P^T_k$ ($T = H$ or $M$) asymptotically standard normal rv's (de Haan and Peng, 1998).

We may now generalize Theorem 3.1 in Caeiro et al. (2005), where it is possible to find a proof of the following theorem for the estimator $\hat{\gamma}^{\overline{H}}_{n,k,\hat{\beta},\hat{\rho}}$ in (3.1.5). Let $T$ generically denote either $H$ or $M$.

**Theorem 3.2.1.** *For any intermediate sequence $k$ as in (2.1.1), for models in (2.1.4), for any $(\hat{\beta}, \hat{\rho})$, consistent for the estimation of $(\beta, \rho)$ and such that $(\hat{\rho} - \rho)\ln(n/k) = o_p(1)$, the asymptotic distributional representation*

$$\hat{\gamma}^{\overline{T}}_{n,k,\hat{\beta},\hat{\rho}} \stackrel{d}{=} \gamma + \frac{\sigma_T P^T_k}{\sqrt{k}} + o_p(A(n/k)),$$

*holds both for $\hat{\gamma}^{\overline{H}}_n$ in (3.1.5) as well as for $\hat{\gamma}^{\overline{M}}_n$ in (3.1.6), where $(P^T_k, \sigma_T)$ with $T = H$ or $T = M$ are given in (3.2.2) and (3.2.3).*

**Proof:** If we estimate consistently $\beta$ and $\rho$ through the estimators $\hat{\beta}$ and $\hat{\rho}$ in the conditions of the theorem, we may use Cramer's delta-method, and write:

$$\begin{aligned}
&\hat{\gamma}^{\overline{M}}_{n,k,\hat{\beta},\hat{\rho}}(k)\\
&= \hat{\gamma}^{M}_{n,k}(k) \times \Big(1 - \frac{\beta}{1-\rho}\Big(\frac{n}{k}\Big)^{\rho} - (\hat{\beta}-\beta)\frac{1}{1-\rho}\Big(\frac{n}{k}\Big)^{\rho}(1+o_p(1))\\
&\quad - \frac{\beta}{1-\rho}(\hat{\rho}-\rho)\Big(\frac{n}{k}\Big)^{\rho}\Big(\frac{1}{1-\rho} + \ln(n/k)\Big)(1+o_p(1))\Big) - \frac{\beta\rho}{(1-\rho)^2}\Big(\frac{n}{k}\Big)^{\rho}\\
&\quad - \Big\{(\hat{\beta}-\beta)\frac{\rho}{(1-\rho)^2}\Big(\frac{n}{k}\Big)^{\rho} + \frac{\beta(\hat{\rho}-\rho)}{1-\rho}\Big(\frac{n}{k}\Big)^{\rho}\Big(\frac{\rho\ln(n/k)}{1-\rho} + 3 - \rho\Big)\Big\}(1+o_p(1))\\
&\overset{d}{=} \quad \hat{\gamma}^{\overline{M}}_{n,k,\beta,\rho}(k) - \frac{A(n/k)}{1-\rho}\Big(\hat{\gamma}^{M}_{n,k}(k) - \frac{\rho}{1-\rho}\Big)\Big(\frac{\hat{\beta}-\beta}{\beta} + (\hat{\rho}-\rho)\ln(n/k)\Big)(1+o_p(1)).
\end{aligned}$$

The reasoning is then quite similar to the one used in Caeiro et al. (2005) for the $\overline{H}$-estimator. Since $\hat{\beta}$ and $\hat{\rho}$ are consistent for the estimation of $\beta$ and $\rho$, respectively, and $(\hat{\rho}-\rho)\ln(n/k) = o_p(1)$, the last summand is $o_p(A(n/k))$, and the result in the theorem, related to the $\overline{M}$-estimator, follows immediately. □

Finally, for the PORT-Hill and PORT-moment estimators in (3.1.7), we have the asymptotic distributional representation given by Theorem 2.2.1.

**Remark 3.2.1.** *Note that as both $d_H$ and $d_M$ in (2.2.5) and (2.2.6), as well as $U(t)$, are positive, the dominant component of the bias of $\hat{\gamma}^{T(q)}_{n,k}$, given in (2.2.4), is increasing as a function of $q$.*

**Remark 3.2.2.** *Note also that if we induce a deterministic shift $\lambda$ to data $X$, considering $X+\lambda$, i.e., if instead of working with data from a model $F := F_0$, we work with the new model $F_\lambda(x) := F_0(x-\lambda)$, the associated $U$-quantile function changes to $U_\lambda(t) = \lambda + U_0(t) \equiv \lambda + U(t)$. Then, if the second-order condition (2.1.3) holds for $F \equiv F_0$, with an auxiliary function $A(t) \equiv A_0(t)$, we straightforwardly get*

$$\frac{U_\lambda(tx)}{U_\lambda(t)} = \frac{U(tx)}{U(t)}\Big\{1 - \frac{\lambda\gamma}{U(t)}\Big(\frac{x^{-\gamma}-1}{-\gamma}\Big) + o\Big(\frac{1}{U(t)}\Big)\Big\}.$$

*Consequently,*

$$\frac{U_\lambda(tx)}{U_\lambda(t)} - x^{\gamma} = x^{\gamma}\Big(A(t)\Big(\frac{x^{\rho}-1}{\rho}\Big) - \frac{\lambda\gamma}{U(t)}\Big(\frac{x^{-\gamma}-1}{-\gamma}\Big) + o(A(t)) + o(1/U(t))\Big),$$

*and we get, for instance, for the Hill estimator associated to this shift $\lambda$, denoted $\hat{\gamma}^{H|\lambda}_{n,k}$ or $H|\lambda$ for the sake of simplicity, the distributional representation*

$$\hat{\gamma}^{H|\lambda}_{n,k} \overset{d}{=} \gamma + \frac{\sigma_H}{\sqrt{k}}P^H_k + \Big(c_H A(n/k) - d_H\frac{\lambda}{U(n/k)}\Big)(1+o_p(1)), \tag{3.2.4}$$

*i.e., as expected, (2.2.4) holds whenever we replace* $\hat{\gamma}_n^{H(q)}$ *by* $\hat{\gamma}_n^{H|\lambda}$*, provided that we replace* $\chi_q$ *by* $-\lambda$*. For details, see Gomes and Oliveira (2003), where the shift* $\lambda$ *is regarded as a tuning parameter of the statistical procedure that leads to the tail index estimates. On the basis of the bias term associated with the Hill functional applied to shifted data, these authors have found easily a justification for some kind of "magic numbers", like* $\lambda = 0.5$*, appearing for a Fréchet model, with tail function* $1 - F(x) = 1 - exp(-x^{-1/\gamma})$*,* $x > 0$*, and* $\lambda = 1/\gamma$*, appearing for a generalized Pareto (GPD) distribution, with tail function* $1 - F(x) = (1 + \gamma x)^{-1/\gamma}$*,* $x > 0$ *(*$\gamma > 0$*). Indeed, from a theoretical point of view, let us assume we are working in Hall's class of distributions, where*

$$1 - F(x) = Cx^{-1/\gamma}(1 + Dx^{\rho/\gamma}(1 + o(1))), \quad as\, x \to \infty.$$

*Then, regular variation theory (Bingham et al., 1987) enables us to obtain the asymptotic inverse of* $F$*,*

$$U(t) := F^{\leftarrow}(1 - 1/t) = (Ct)^{\gamma}(1 + \gamma D(Ct)^{\rho}(1 + o(1))), \quad as\, t \to \infty,$$

*and we may choose any A function, such that* $A(t) \sim \gamma\rho D(Ct)^{\rho}$*, as* $t \to \infty$*.*
*Whenever* $\rho = -\gamma$*, we may thus choose* $A(t)$ *such that:*

$$A(t)U(t) = -\gamma^2 D, \quad i.e., \, 1/U(t) = -A(t)/(\gamma^2 D).$$

*If we look at (3.2.4) we see that the dominant component of asymptotic bias is then given by* $(A(n/k) - \lambda\gamma/U(n/k))/(1 + \gamma) = A(n/k)(1 + \lambda/(\gamma D))/(1 + \gamma)$*. Such a component is thus null whenever* $\lambda = -\gamma D$*.*
*The Fréchet model belongs to Hall's class, with* $C = 1$*,* $D = -1/2$*, and* $\rho = -1$*. Then, for* $\gamma = 1$*,* $\lambda = 0.5$ *enables us to remove the main component of asymptotic bias. If we think on a GPD model, we are again in Hall's class of models with* $C = \gamma^{-1/\gamma}$*,* $D = -1/\gamma^2$*, and* $\rho = -\gamma$*. Then, for every* $\gamma$ *if we induce in the data a shift* $\lambda = -\gamma D = 1/\gamma = -1/\rho$ *we are able to remove the dominant component of asymptotic bias.*

**Remark 3.2.3.** *The comments in Remark 3.2.2 are also true for the classical moment estimator, i.e., if we induce a shift* $\lambda$ *to the data, (2.2.4) holds whenever we replace* $\hat{\gamma}_n^{M(q)}$ *by* $\hat{\gamma}_n^{M|\lambda}$*, provided that we replace* $\chi_q$ *by* $-\lambda$*. Moreover, also for the moment estimator the dominant component of asymptotic bias is null whenever in Hall's class of models, we have* $\rho = -\gamma$ *and we induce a shift* $\lambda = \rho D = -\gamma D$*.*

We still add the following.

**Remark 3.2.4.** *Let us now consider the general $EV_\gamma$ model in (1.1.1). Then, we may write*

$$1-F(x)=(\gamma x)^{-1/\gamma}.\begin{cases}\Big(1-\frac{1}{\gamma^2 x}+o(x^{-1})\Big) & \text{if } 0<\gamma<1\\ \Big(1-\frac{3}{2x}+o(x^{-1})\Big) & \text{if } \gamma=1\\ \Big(1-\frac{(\gamma x)^{-1/\gamma}}{2}+o(x^{-1/\gamma})\Big) & \text{if } \gamma>1,\end{cases}$$

*i.e.,*

$$C=\gamma^{-1/\gamma},\quad \rho=\begin{cases}-\gamma & \text{if } 0<\gamma\le 1\\ -1 & \text{if } \gamma>1\end{cases},\quad D=\begin{cases}-1/\gamma^2 & \text{if } 0<\gamma<1\\ -3/2 & \text{if } \gamma=1\\ -\gamma^{-1/\gamma}/2 & \text{if } \gamma>1.\end{cases}$$

*For the $EV_\gamma$ model, with $\gamma\le 1$, we may thus get a second-order reduced-bias extreme value index estimator, on the basis of both the Hill and the moment functionals, in (3.1.2) and (3.1.3), respectively, provided that we induce the deterministic shift*

$$\lambda=-\gamma D=\begin{cases}1/\gamma & \text{if } 0<\gamma<1\\ 3/2 & \text{if } \gamma=1.\end{cases}$$

*Note, however, that with a deterministic shift, as suggested in Gomes and Oliveira (2003), the estimators lose even the scale invariance property.*

### 3.2.2 The non-consistency of $H(0)$ for symmetric models with infinite left endpoint

In this subsection we show that for heavy-tailed models symmetric around any real value and with $x_F=-\infty$, the Hill estimator, adapted to the sample of excesses over the minimum, can be non consistent for $\gamma$.

**Theorem 3.2.2 (Non-consistency of PORT-Hill).** *For any intermediate sequence $k$ as in* (2.1.1), *under the validity of the first order condition in* (2.1.2) *for a symmetric cdf $F$ :*

1. *if $\gamma\ge\log 2$,* $\quad\hat\gamma_n^{H(0)}\underset{n\to\infty}{\overset{p}{\not\to}}\gamma$;
2. *if $\gamma<1$,* $\quad\hat\gamma_n^{H(0)}\underset{n\to\infty}{\overset{p}{\longrightarrow}}0.$

**Proof:** (1) The proof relies on the representation of $H(0)$ as a function of the extremal quotient defined by $Q_n:=-X_{1:n}/X_{n:n}$, which converges to 1 in probability, for symmetric cdf's. Details on the asymptotic properties of this extremal quotient can be found in Gumbel and Keeney (1950). Consider the representation of the o.s.'s

$\{X_{i:n} = U(Y_{i:n})\}_{i=1}^n$, with $Y_{i:n}$ the $i$-th increasing o.s. associated with a random sample from a standard Pareto cdf, $F_Y(y) = 1 - 1/y$, $y > 1$. In $\mathcal{D}_{\mathcal{M}}(EV_{\gamma>0})$ the scaled intermediate o.s. $\frac{k}{n} Y_{n-k:n} \underset{n\to\infty}{\overset{p}{\longrightarrow}} 1$ and the scaled maximum $\frac{1}{n} Y_{n:n} \underset{n\to\infty}{\overset{p}{\longrightarrow}} 1$. Consequently, by (2.1.2), the ratio $Q_{n,k} := \frac{X_{n-k:n}}{X_{n:n}} = O_p(k^{-\gamma})$. Since for the extremal quotient $Q_n = 1 + o_p(1)$, we have successively

$$\begin{aligned}
0 < \hat{\gamma}_n^{H(0)} &= \frac{1}{k}\sum_{i=1}^k \ln \frac{X_{n-i+1:n} - X_{1:n}}{X_{n-k:n} - X_{1:n}} = \frac{1}{k}\sum_{i=1}^k \ln \frac{Q_{n,i-1} + Q_n}{Q_{n,k} + Q_n} \\
&= \frac{1}{k}\sum_{i=1}^k \ln\{Q_{n,i-1} + Q_n\} - \ln\{Q_{n,k} + Q_n\} \\
&= \frac{1}{k}\sum_{i=1}^k \ln\left\{Q_n\left(1 + \frac{Q_{n,i-1}}{Q_n}\right)\right\} - \ln\{Q_{n,k} + Q_n\} \\
&< \ln Q_n + \ln\left(1 + Q_n^{-1}\right) - \ln\{Q_{n,k} + Q_n\} \\
&= \ln(1 + o_p(1)) + \ln(2 + o_p(1)) - \ln\{1 + o_p(1)\} \underset{n\to\infty}{\overset{p}{\longrightarrow}} \ln 2
\end{aligned}$$

which assures that, for $\gamma \geq \ln 2$, $\quad \hat{\gamma}_n^{H(0)} \underset{n\to\infty}{\overset{p}{\nrightarrow}} \gamma$.

(2) Consider now $0 < \gamma < 1$, in first order condition (2.1.2), and write

$$\hat{\gamma}_n^{H(0)} = \frac{1}{k}\sum_{i=1}^k \ln\{Q_{n,i-1} + Q_n\} - \ln\{Q_{n,k} + Q_n\} =: A_n + B_n \tag{3.2.5}$$

Since the extremal quotient $Q_n = 1 + o_p(1)$ and $Q_{n,k} = O_p(k^{-\gamma})$ the second term in (3.2.5) $B_n = \ln\{1 + o_p(1)\} = o_p(1)$. For the first term in (3.2.5) we have

$$\begin{aligned}
A_n &= \ln Q_n + \frac{1}{k}\sum_{i=1}^k \ln\left(1 + \frac{Q_{n,i-1}}{Q_n}\right) \\
&= \ln Q_n + \frac{1}{k}\sum_{i=1}^k \ln\left(1 + Q_n^{-1} Q_{n,k} \frac{X_{n-i+1:n}}{X_{n-k:n}}\right) \\
&\overset{d}{=} \ln Q_n + \frac{1}{k}\sum_{i=1}^k \ln\left(1 + Q_n^{-1} Q_{n,k} \frac{U(\frac{Y_{n-i+1:n}}{Y_{n-k:n}} Y_{n-k:n})}{U(Y_{n-k:n})}\right) \\
&\overset{d}{=} \ln Q_n + \frac{1}{k}\sum_{i=1}^k \ln\left(1 + Q_n^{-1} Q_{n,k} \frac{U(Y'_{k-i+1:k} Y_{n-k:n})}{U(Y_{n-k:n})}\right) \\
&\overset{d}{=} \ln Q_n + \frac{1}{k}\sum_{i=1}^k \ln\left(1 + Q_n^{-1} Q_{n,k} \frac{U(Y'_i Y_{n-k:n})}{U(Y_{n-k:n})}\right),
\end{aligned} \tag{3.2.6}$$

since Rényi's representation (Rényi, 1953) enables us to write

$$\left\{\frac{Y_{n-i+1:n}}{Y_{n-k:n}}\right\}_{i=1}^{k} \stackrel{d}{=} \left\{Y'_{k-i+1:k}\right\}_{i=1}^{k} \text{ and } \sum_{i=1}^{k} g\left(Y'_{k-i+1:k}\right) = \sum_{i=1}^{k} g(Y'_i),$$

for any measurable function $g$, with $Y'_{i:k}$, $i = 1, \cdots, k$ denoting the o.s.'s associated with a unit Pareto random sample $\{Y'_i\}_{i=1}^{k}$.

We will show now that $\frac{1}{k}\sum_{i=1}^{k} \ln\left(1 + Q_n^{-1} Q_{n,k} \frac{U(Y'_i Y_{n-k:n})}{U(Y_{n-k:n})}\right) \underset{n\to\infty}{\overset{p}{\longrightarrow}} 0$, for $\gamma < 1$. We do this by using Potter's inequalities (Bingham, Goldie and Teugels, 1987): since $U \in RV_\gamma$, $\gamma > 0$, for any $\epsilon > 0$ there exists $t_0 = t_0(\epsilon)$ such that for $t \geq t_0$ and $x \geq 1$

$$(1-\epsilon)x^{\gamma-\epsilon} \leq \frac{U(tx)}{U(t)} \leq (1+\epsilon)x^{\gamma+\epsilon}. \qquad (3.2.7)$$

The use of (3.2.7) enables us to get an upper bound for the second summand in (3.2.6)

$$\frac{1}{k}\sum_{i=1}^{k} \ln\left(1 + (1+\epsilon) Q_n^{-1} Q_{n,k} {Y'_i}^{\gamma+\epsilon}\right)$$

which is asymptotic equivalent to

$$\frac{1+\epsilon}{k} Q_n^{-1} Q_{n,k} \sum_{i=1}^{k} {Y'_i}^{\gamma+\epsilon} = (1+\epsilon)(1+o_p(1)) O_p(k^{-\gamma}) \frac{1}{k}\sum_{i=1}^{k} {Y'_i}^{\gamma+\epsilon}.$$

Since $\frac{1}{k}\sum_{i=1}^{k} {Y'_i}^{\gamma}$ converges to $E[Y^\gamma] = 1/(1-\gamma)$ in probability, for $\gamma < 1$, assured by the law of large numbers. A similar reasoning leads us to a lower bound for the second summand in (3.2.6). We get $\frac{1}{k}\sum_{i=1}^{k} \ln\left(1 + Q_n^{-1} Q_{n,k} \frac{U(Y'_i Y_{n-k:n})}{U(Y_{n-k:n})}\right) \underset{n\to\infty}{\overset{p}{\longrightarrow}} 0$, for $\gamma < 1$; consequently,

$$\frac{1}{k}\sum_{i=1}^{k} \ln\left\{\frac{X_{n-i+1:n}}{X_{n:n}} - \frac{X_{1:n}}{X_{n:n}}\right\} = o_p(1)$$

and the result in 2. follows. □

**Remark 3.2.5.** *This result constitutes an alert to the practical statistical methodology of transforming the original data through the subtraction of the minimum of the sample. In the view of tail index estimation, this is not assured as a consistent inference procedure taking the example of Hill estimator, with consequently practical misleading conclusions.*

### 3.2.3 The Estimation of Second-Order Parameters

For the estimation of the second-order parameters, needed for the estimators in (3.1.5) and in (3.1.6), we suggest here an algorithm similar to the ones in Gomes and Pestana (2007a,c) and Gomes et al. (2008b,c):

1. Given a sample $(X_1, X_2, ..., X_n)$, with the notation $a^{b\tau} = b \ln a$ whenever $\tau = 0$, and $M_{n,k}^{(r)}$ given in (3.1.4), plot, for $\tau = 0, 1$, the estimates

$$\hat{\rho}_\tau(k) := -\Big|\frac{3(T_{n,k}^{(\tau)} - 1)}{(T_{n,k}^{(\tau)} - 3)}\Big|, \text{ with } T_{n,k}^{(\tau)} := \frac{(M_{n,k}^{(1)})^\tau - (M_{n,k}^{(2)}/2)^{\tau/2}}{(M_{n,k}^{(2)}/2)^{\tau/2} - (M_{n,k}^{(3)}/6)^{\tau/3}}. \quad (3.2.8)$$

2. Consider $\{\hat{\rho}_\tau(k)\}_{k\in\mathcal{K}}$, for integer values $k \in \mathcal{K} = ([n^{0.995}], [n^{0.999}])$, and compute their median, denoted $\chi_\tau$, $\tau = 0, 1$. Choose

$$\tau^* := \begin{cases} 0 & \text{if } \sum_{k\in\mathcal{K}}(\hat{\rho}_0(k) - \chi_0)^2 \le \sum_{k\in\mathcal{K}}(\hat{\rho}_1(k) - \chi_1)^2 \\ 1 & \text{otherwise.} \end{cases}$$

3. Compute, for $k_1 = [n^{0.995}]$, $\hat{\rho}^* = \hat{\rho}(k_1; \tau^*)$ and $\hat{\beta}^* := \hat{\beta}(k_1; \hat{\rho}^*)$,

$$\hat{\beta}(k; r) := \left(\frac{k}{n}\right)^r \frac{d_k(-r) \times D_k(0) - D_k(-r)}{d_k(r) \times D_k(-r) - D_k(-2r)} \quad (3.2.9)$$

where for any $\alpha \le 0$, and with $W_i := i\{\ln X_{n-i+1,n} - \ln X_{n-i,n}\}$, $1 \le i \le k$,

$$D_k(\alpha) := \frac{1}{k}\sum_{i=1}^{k}(i/k)^\alpha W_i, \quad d_k(\alpha) := \frac{1}{k}\sum_{i=1}^{k}(i/k)^\alpha W_i. \quad (3.2.10)$$

**Remark 3.2.6.** *The implementation of this algorithm in practice leads often to $\tau^* = 0$ whenever $\rho \le 1$ and $\tau^* = 1$ whenever $\rho > 1$ (see Gomes and Pestana, 2007c). This is the reason why we are going to use such a rule in the simulations. The choices of $\mathcal{K}$ in Step 2 and $k_1$ in Step 3 are not crucial, provided that we restrict ourselves to reasonably large values of $k$, the number of o.s. used.*

Regarding the reduced-bias extreme value index estimators in (3.1.5) and (3.1.6), the estimators $(\hat{\beta}_\tau, \hat{\rho}_\tau)$ of $(\beta, \rho)$, $\tau = 0, 1$, have been used, leading to:

$$\overline{H}_\tau \equiv \overline{H}_\tau(k) \equiv \hat{\gamma}^{\overline{H}}_{n,k,\hat{\beta}_\tau,\hat{\rho}_\tau}, \quad \overline{M}_\tau \equiv \overline{M}_\tau(k) \equiv \hat{\gamma}^{\overline{M}}_{n,k,\hat{\beta}_\tau,\hat{\rho}_\tau}, \quad \tau = 0, 1.$$

The simulations in Caeiro et al. (2005) and Gomes and Pestana (2007c) show that the tail index estimators $\overline{H}_\tau$, with $\tau$ equal to either 0 or 1, according as $|\rho| \le 1$ or $|\rho| > 1$, work quite well. The use of $\tau = 1$ always enables us to achieve a better performance than the one we get with the Hill estimator $H$. In a "blind way", we might thus advise such a choice, and we shall do it for the reduced-bias moment estimator $\overline{M}_\tau$. But for $\overline{H}_\tau$, $\tau = 0$ provides much better results than $\tau = 1$ whenever $|\rho|$, unknown, is smaller than or equal to 1.

### 3.2.4 Simulated Behaviour of the Tail Index Estimators

We have implemented multi-sample Monte Carlo simulation experiments of size $5,000\times 10$ for the extreme value index estimators under study.

*3.3.1. Mean Values and Mean Squared Error Patterns of the Tail Index Estimators.* In Fig 3.2.1 , for samples of size $n = 1,000$ from a Fréchet($\gamma$), with $\gamma = 1$, we show the simulated patterns of the mean values, $E[\bullet]$, and mean squared errors, $MSE[\bullet]$, of the Hill estimator $H$ in (3.1.2) and its location invariant versions $H(p)$, $p = 0, 0.25$, and 0.5, in (3.1.7), together with the ones of the MVRB estimators $\overline{H}_0$ in (3.1.5). Figure 3.2.2 is similar to Fig. 3.2.1, but for the moment estimator $M$ in (3.1.3), its location invariant versions $M(p)$, $p = 0, 0.25$, and 0.5, in (3.1.7) and the MRVB estimator $\overline{M}_1$ in (3.1.6). The mean values and mean squared errors of the estimators are based on the first replicate, with a run of size 5,000. Figures 3.2.3 and 3.2.4 are equivalent to Figs. 3.2.1 and 3.2.2, respectively, but for the $EV_\gamma$ model in (1.1.1), with $\gamma = 0.25$. Similar comment applies to Figs. 3.2.5 and 3.2.6, where we consider the underlying parent $EV_\gamma$, with $\gamma = 1$. Finally, the pairs of Figs. 3.2.7, 3.2.8 and Figs. 3.2.9, 3.2.10 are equivalent to the pair of Figs. 3.2.1, 3.2.2, but for Student $t_v$, with $v = 4$ and $v = 2$, respectively. The Student $t_v$ probability density function (df) is:

$$f_v(x) = \Gamma((v+1)/2)[1 + x^2/(v-2)]^{-(v+1)/2}/(\sqrt{\pi(v-2)}\Gamma(v/2)), \quad x \in \mathbf{R}.$$

For the Student-$t_v$ model, we get $\gamma = 1/v$ and $\rho = -2/v$.
We may draw the following specific comments:

- As expected, on the basis of Remark 3.2.1, $H(q)$ and $M(q)$ are increasing in $q$. However, and with $T$ generally denoting either $H$ or $M$, we expect to have $T < T(0)$ if the left endpoint $x_F$ of the underlying model $F$ is zero, but things work the other way round, i.e., $T(0) < T$ if $x_F \neq 0$.

- For a Fréchet model, and perhaps as expected more generally, if we induce a shift (random shift) through a central o.s. (or even the minimum, equal to 0), applying the Hill or the moment functionals to $X_i - X_{[nq]+1:n}$, $1 \le i \le n$, $0 \le q < 1$, we get worse results than when we work with either the Hill or the moment estimators, respectively. This result is not astonishing in the sense that we are replacing estimators that are only scale invariant by scale and location invariant estimators. Indeed, from the results in Gomes and Oliveira (2003), we know that, concerning the Hill estimator, we should shift our data from $X$ towards $X + 0.5$ in order to remove the dominant component of bias of the Hill estimator, and $-0.5 < x_F = 0$. But then, we are working with estimators that are neither invariant for changes in scale nor location.

- As mentioned before, for the $EV_\gamma$ model, with $0 < \gamma < 1$, we have $\rho = -\gamma$, and with a shift $\lambda = 1/\gamma = -1/\rho$ we would remove the dominant component of bias of Hill's as well as moment's estimators. This means that we should apply the Hill or the moment functionals to $X - x_F = X + 1/\gamma$. Given that $X_{1:n} \to x_F = -1/\gamma$, we expect to be reasonably close to a reduced-bias extreme value index estimator whenever we apply Hill's or moment's functionals to $X_i - X_{1:n}$, $1 \le i \le n$. If we look at Figs. 3.2.3 and 3.2.4, we see that $H(0)$ and $M(0)$ behave even better than the corresponding MVRB-estimators. For the $EV_1$ model, the shift that would reduce the dominant component of bias would be induced by $\lambda = 3/2$. We should thus go below the minimum, given that $x_F = -1$, and our estimator would no longer be location invariant (nor scale invariant). The statistics $H(0)$ and $M(0)$ are the best ones among the non reduced-bias estimators, but the corresponding MVRB estimators behave better than either $H(0)$ and $M(0)$. For the $EV_\gamma$ model with $\gamma > 1$, although we have $\rho \neq -\gamma$, the relative behaviour of the PORT-estimators is quite similar to the one appearing when $\gamma = 1$. The location invariant estimators $H(q)$, $q \le 0.25$, behave better than the Hill, although not better than the MVRB-estimator $\overline{H}_0$.

- We have decided to consider also Student $t_v$ parents with $v$ degrees of freedom. Then, we have $\rho \neq -\gamma$. These parents have infinite left and right endpoints, and consequently, it is no longer sensible to consider $q = 0$ in the PORT-estimators, because of the possible non-consistency of the associated PORT-statistics. We did it merely to draw the attention for the erroneous conclusions we may take from a quite common behavior in data analysis practice. Indeed, a usual solution to take care of the Pareto approximation $U(t) \sim \delta t^\gamma$ is to make statistical inference only after a suitable shift of the data. In the literature, it has been sometimes suggested to subtract a random quantity, usually the minimum of the sample. This shifted data set has the advantage of working out with usually more non negative values, a desirable property for classical semi-parametric estimators of a positive tail index. An extensive discussion about this type of shifted procedures can be found for instance in Drees (2003). Therein, it is studied the effect of subtracting the minimum of the sample, previously to the subsequent analysis of the Nasdaq

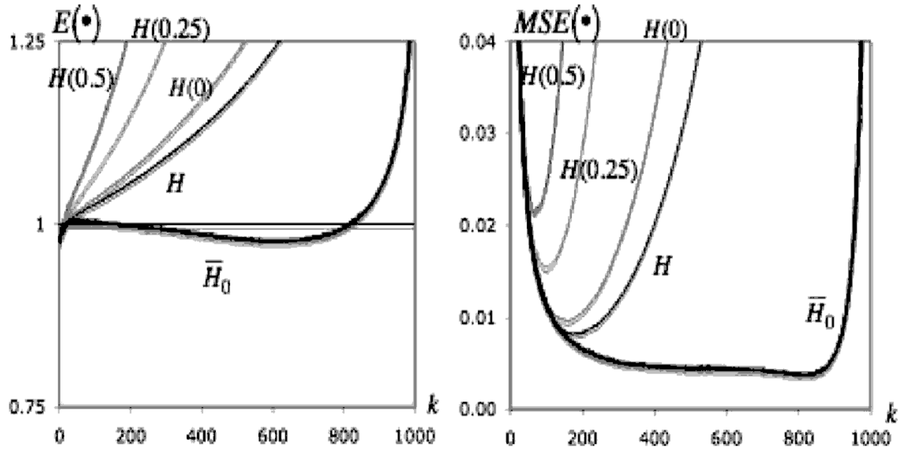



Figure 3.2.1: Simulated Mean values (left) and root mean squared errors (right) of the Hill estimator $H$ in (3.1.2) and $H(q)$, $q = 0, 0.25$ and $0.5$ in (3.1.7), together with $\overline{H}_0$ in (3.1.5), for samples of size $n = 1,000$ from a Fréchet parent with $\gamma = 1$ $\rho = -1$.

Composite index log-returns data set, in the context of VaR estimation. In fact, for that particular data, it is therein observed that this procedure constitutes a considerable improvement, arising for the Hill $\gamma$-estimates a larger flat zone in the associated sample path, after transforming the original data through the subtraction of the smallest observation. However, if we look at Figs. 3.2.7 and 3.2.9, we easily see that "flat" zones in the sample path of the shifted-Hill (by the minimum) estimator can lead to serious underestimation of the extreme value index.

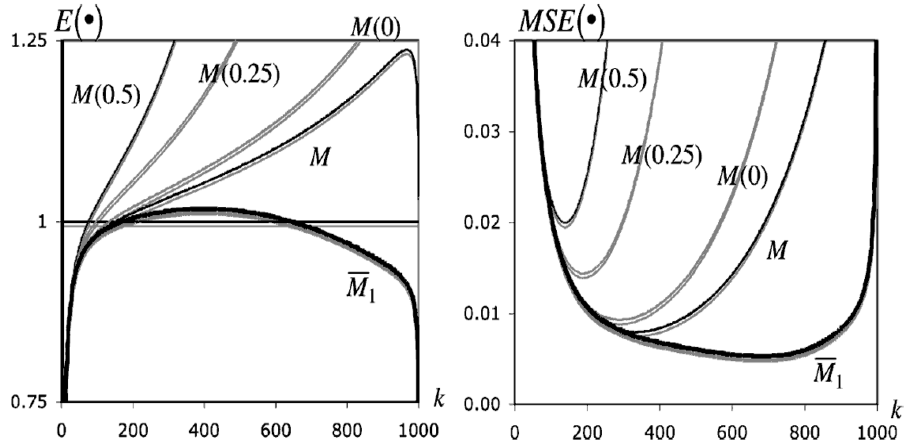



Figure 3.2.2: Simulated Mean values (left) and root mean squared errors (right) of $M$ and $M(q)$, $q = 0, 0.25$ and 0.5 in (3.1.7), together with $\overline{M}_1$ in (3.1.6), for samples of size $n = 1,000$ from a Fréchet parent with $\gamma = 1$ $\rho = -1$.

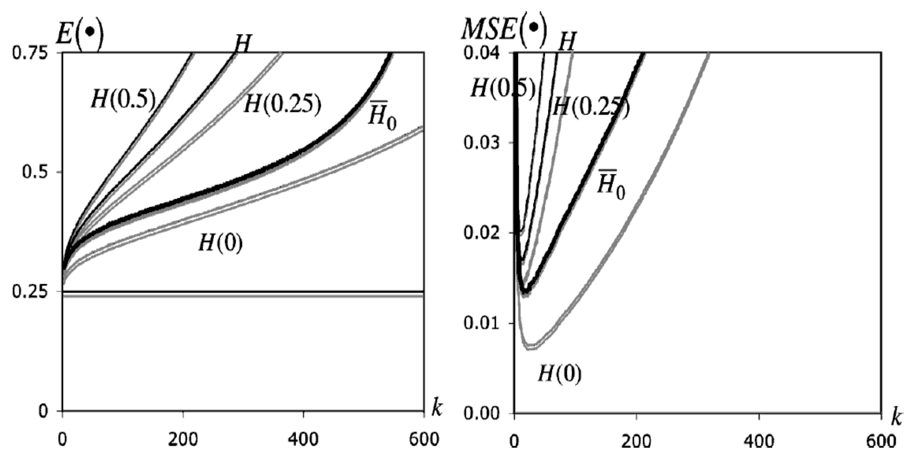



Figure 3.2.3: Simulated Mean values (left) and root mean squared errors (right) of $H$ and $H(q)$, $q = 0, 0.25$ and 0.5, together with $\overline{H}_0$, for samples of size $n = 1,000$ from a $EV_\gamma$ parent with $\gamma = 0.25$ $\rho = -0.25$.

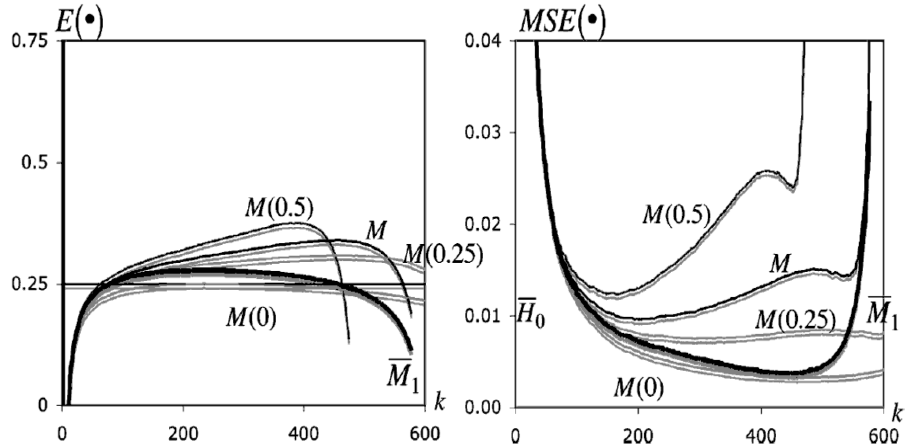



Figure 3.2.4: Simulated Mean values (left) and root mean squared errors (right) of $M$ and $M(q)$, $q = 0, 0.25$ and $0.5$, together with $\overline{M}_1$, for samples of size $n = 1,000$ from a $EV_\gamma$ parent with $\gamma = 0.25$ $\rho = -0.25$.

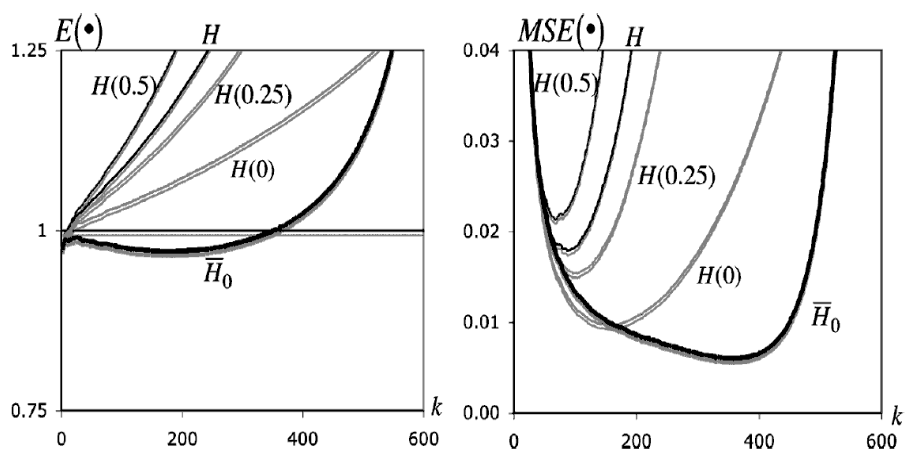



Figure 3.2.5: Simulated Mean values (left) and root mean squared errors (right) of $H$ and $H(q)$, $q = 0, 0.25$ and $0.5$, together with $\overline{H}_0$, for samples of size $n = 1,000$ from a $EV_\gamma$ parent with $\gamma = 1$ $\rho = -1$.

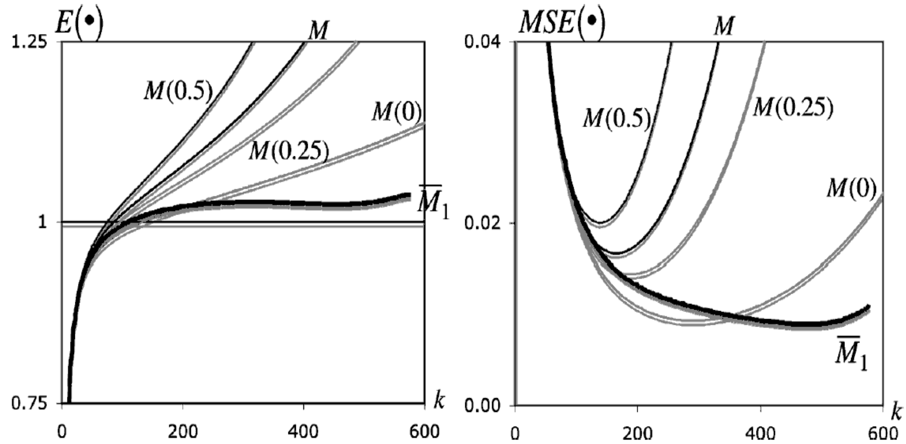



Figure 3.2.6: Simulated Mean values (left) and root mean squared errors (right) of $M$ and $M(q)$, $q = 0, 0.25$ and $0.5$, together with $\overline{M}_1$, for samples of size $n = 1,000$ from a $EV_\gamma$ parent with $\gamma = 1$ $\rho = -1$.

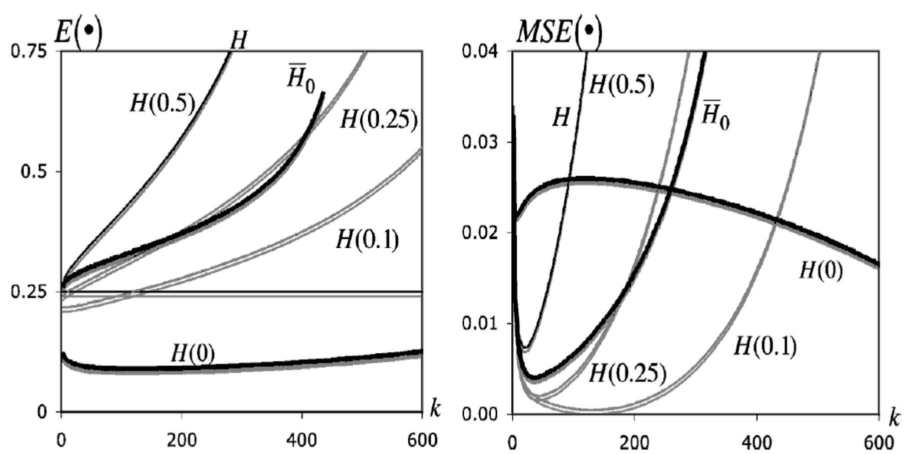



Figure 3.2.7: Simulated Mean values (left) and root mean squared errors (right) of $H$ and $H(q)$, $q = 0, 0.25$ and $0.5$, together with $\overline{H}_0$, for samples of size $n = 1,000$ from a $t_4$ parent with $\gamma = 0.25$ $\rho = -0.5$.

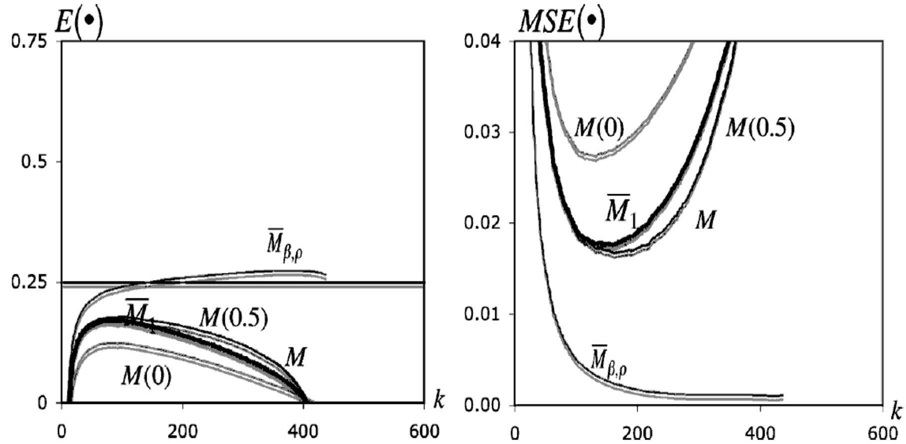



Figure 3.2.8: Simulated Mean values (left) and root mean squared errors (right) of $M$ and $M(q)$, $q = 0, 0.25$ and 0.5, together with $\overline{M}_1$ and the rv $\overline{M}_{\beta,\rho}$, for samples of size $n = 1,000$ from a $t_4$ parent with $\gamma = 0.25$ $\rho = -0.5$.

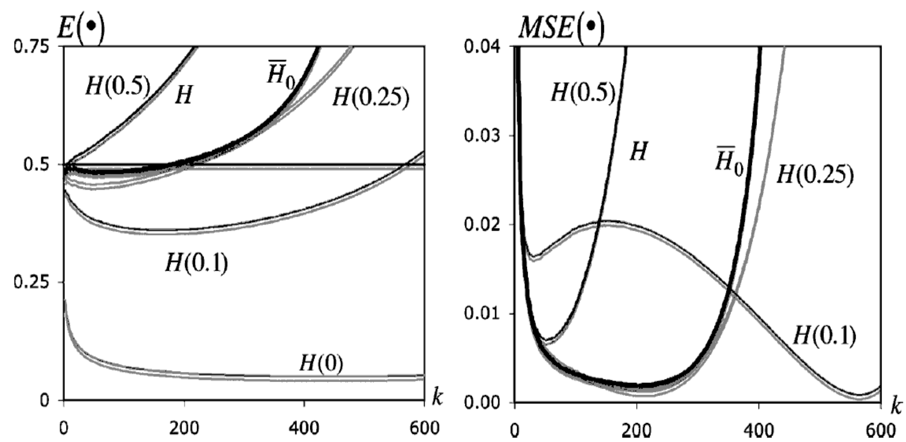



Figure 3.2.9: Simulated Mean values (left) and root mean squared errors (right) of $H$ and $H(q)$, $q = 0, 0.25$ and 0.5, together with $\overline{H}_0$, for samples of size $n = 1,000$ from a $t_2$ parent with $\gamma = 0.5$ $\rho = -1$.

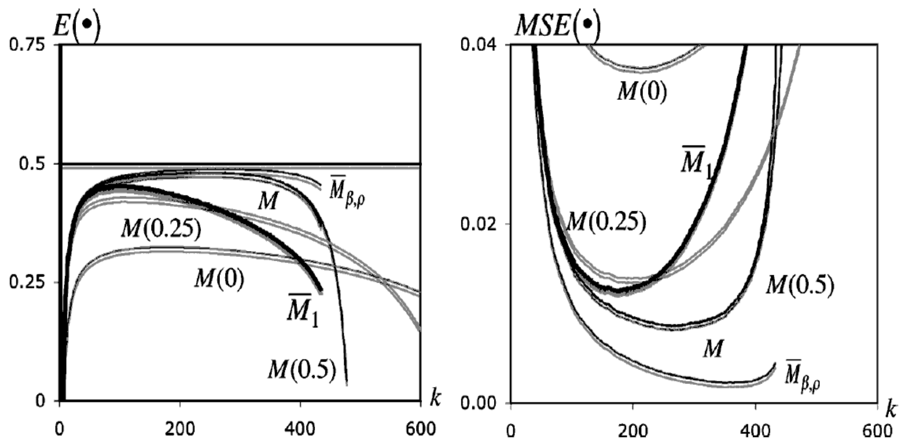



Figure 3.2.10: Simulated Mean values (left) and root mean squared errors (right) of $M$ and $M(q)$, $q = 0, 0.25$ and 0.5, together with $\overline{M}_1$ and the rv $\overline{M}_{\beta,\rho}$, for samples of size $n = 1,000$ from a $t_2$ parent with $\gamma = 0.5$ $\rho = -1$.

*3.3.2. Mean Values of the Tail Index Estimators at Optimal Levels.* Tables 3.1, 3.2 and 3.3 are related to underlying models with $|\rho| < 1$, $\rho < -1$ and $|\rho| > 1$, respectively. We shall there present, for $n = 200, 500, 1,000, 2,000$ and $5,000$, the simulated mean values at optimal levels (levels where mean squared errors are minima as functions of $k$) of the Hill estimator $H$ in (3.1.2), the moment estimator $M$ in (3.1.3), the MVRB-estimators, $\overline{H}_0$, $\overline{M}_1$, in (3.1.5), (3.1.6), respectively, and the PORT-Hill and moment estimators in (3.1.7) associated with $q = 0, 0.1, 0.25,$ and 0.5. Information on 95% confidence intervals, computed on the basis of the 10 replicates with 5,000 runs each, is also provided. Among the estimators considered, the one providing the smallest squared bias is underlined and in **bold**.

*3.3.3. Mean Squared Errors and Relative Efficiency Indicators at Optimal Levels.* We shall compute Hill's estimator at the simulated value of $k_o^H := \arg\min_k MSE[\hat{\gamma}_{n,k}^H]$, the simulated optimal $k$ in the sense of minimum mean squared error, not relevant in practice, but providing an indication of the best possible performance of Hill' s estimator. Such an estimator will be denoted $H_0$. Let us generically denote $T$ any of the extreme value index estimators under study. We shall now compute $T_0$, the estimator $T$ computed at its simulated optimal level, again in the sense of minimum mean squared error. The simulated indicators are:

$$REFF_{T|H} := \sqrt{\frac{MSE[H_0]}{MSE[T_0]}} \tag{3.2.11}$$

**Remark 3.2.7.** *An indicator higher than one means a better performance than the Hill estimator. Consequently, the higher these indicators are, the better the $T_0$-estimators perform, comparatively to $H_0$.*

In Tables 3.4 - 3.12, we present in the first row, the mean squared error of $H_0$, so that we can easily recover the mean squared errors of all other estimators $T_0$. The following rows provide the $REFF$ indicators, $REFF_{T|H}$ in (3.2.11), for the different extreme value index estimators under study. Again, the estimator providing the highest $REFF$ indicator (minimum mean squared error at optimal level) is underlined and in **bold**.

*Some Comments Regarding the $REFF$ Indicators:*

- For Fréchet parents and regarding $REFF$ indicators, the reduced-bias estimator $\overline{H}_0$ is the one exhibiting the better behaviour (higher $REFF$). The moment estimator, at the optimal level, slightly overpasses the Hill estimator, also at its optimal level, for all $n$. Whenever we consider the PORT-estimators, the $REFF$ indicators are always smaller than 1, and they decrease as $q$ increases. For the same $q$, $M(q)$, and $H(q)$ have $REFF$ indicators close together, with a slightly better performance of the $M(q)$ estimator.

- For the $EV_\gamma$, $\gamma = 0.25$, and regarding $REFF$ indicators, only $H(0.5)$ exhibits a REFF measure smaller than one for all $n$. The reduced-bias estimator $\overline{H}_0$ behaves better than the Hill and quite close to $H(0.25)$, but not so high as for Fréchet parents. Both for $H(q)$, and $M(q)$ the $REFF$ indicators increase as $q$ decreases, with the moment estimator behaving better than the Hill estimator, for the same $q$. The estimator with the highest $REFF$ indicator, among the ones considered is $M(0)$. However, $H(0)$ provides a $REFF$ indicator quite close to 1.5 for all $n$. For the $EV_\gamma$ with $\gamma = 1$ the main difference lies in the fact that now the reduced-bias indicator $\overline{H}_0$ provides the highest $REFF$ indicators for all $n \geq 500$. The relative behavior of the $REFF$ indicators for $H(q)$ and $M(q)$ follows a pattern similar to the one associated to an $EV_{0.25}$, but both $H(0.5)$ and $M(0.5)$ have $REFF$ indicators smaller than one for all $n$.

- For all Student models, and as expected due to the symmetry of the model around 0, $H(0.5)$ is almost coincident with $H$, as well as $M(0.5)$ almost equals $M$. For the Student model with $v = 4$ degrees of freedom, the reduced-bias estimator $\overline{H}_0$ behaves quite well, even for small values of $n$, but $H(0.25)$ overpasses it, being $H(0.1)$ the best estimator among the ones considered. All $M(q)$ estimators behave worse than the Hill estimator at optimal levels when $\rho$ approaches 0, but for $v = 2$ the moment estimator $M$ behaves slightly better than the Hill for large $n$. As mentioned before, $H(0)$, possibly not even consistent for the estimation of $\gamma$, as well as $M(0)$, behave really very badly, with sample paths quite stable, but around a value a long way from the target.

Table 3.1
Simulated mean values, at optimal levels, of $H$, $M$, $\overline{H}_0$, $\overline{M}_1$, $H(q)$, and $M(q)$, $q = 0.1, 0.25$, and 0.5, for parents with $|\rho| = -1$

| $n$ | 200 | 500 | 1000 | 2000 | 5000 |
|---|---|---|---|---|---|
| | | | Student $t_4$ $(\gamma = 0.25)(\rho = -0.5)$ | | |
| $H$ | 0.3402±0.0792 | 0.3409±0.0636 | 0.3205±0.0357 | 0.3062±0.0486 | 0.2856±0.0253 |
| $M$ | 0.0845±0.1027 | 0.1538±0.0855 | 0.1630±0.0495 | 0.1821±0.0496 | 0.1916±0.0301 |
| $\overline{H}_0$ | 0.3231±0.0521 | 0.3010±0.0453 | 0.2876±0.0276 | 0.2919±0.0350 | 0.2862±0.0205 |
| $\overline{M}_1$ | 0.0799±0.0722 | 0.1313±0.0774 | 0.1595±0.0592 | 0.1704±0.0533 | 0.2019±0.0223 |
| $H(0)$ | 0.2735±0.0504 | 0.1892±0.0569 | 0.1889±0.0298 | 0.1534±0.0227 | 0.1040±0.0145 |
| $H(0.1)$ | **0.2639**±0.0216 | **0.2645**±0.0145 | **0.2561**±0.0136 | **0.2608**±0.0060 | **0.2576**±0.0070 |
| $H(0.25)$ | 0.2937±0.0328 | 0.2766±0.0345 | 0.2645±0.0195 | 0.2721±0.0241 | 0.2633±0.0146 |
| $H(0.5)$ | 0.3450±0.0814 | 0.3410±0.0664 | 0.3186±0.0370 | 0.3059±0.0477 | 0.2853±0.0253 |
| $M(0)$ | 0.0389±0.0899 | 0.0978±0.0720 | 0.1154±0.0558 | 0.1385±0.0603 | 0.1643±0.0315 |
| $M(0.1)$ | 0.0474±0.0908 | 0.1208±0.0714 | 0.1311±0.0586 | 0.1589±0.0606 | 0.1816±0.0302 |
| $M(0.25)$ | 0.0635±0.0949 | 0.1210±0.0794 | 0.1497±0.0652 | 0.1593±0.0666 | 0.1943±0.0249 |
| $M(0.5)$ | 0.0888±0.1075 | 0.1549±0.0872 | 0.1623±0.0496 | 0.1816±0.0498 | 0.1914±0.0299 |
| | | | $EV_\gamma$ $(\gamma = 0.25)(\rho = -0.25)$ | | |
| $H$ | 0.3754±0.0806 | 0.3910±0.0951 | 0.3370±0.0585 | 0.3909±0.0801 | 0.3237±0.0333 |
| $M$ | 0.3473±0.0957 | 0.2489±0.0956 | 0.2923±0.0718 | 0.3077±0.0499 | 0.2957±0.0350 |
| $\overline{H}_0$ | 0.4026±0.0903 | 0.3396±0.0522 | 0.3648±0.0597 | 0.3884±0.0768 | 0.3230±0.0394 |
| $\overline{M}_1$ | 0.2449±0.0722 | 0.2012±0.0955 | 0.2618±0.0411 | 0.2834±0.0338 | 0.2617±0.0182 |
| $H(0)$ | 0.3710±0.0692 | 0.3120±0.0553 | 0.3242±0.0479 | 0.3434±0.0431 | 0.2990±0.0290 |
| $H(0.1)$ | 0.3808±0.0750 | 0.3335±0.0716 | 0.3606±0.0576 | 0.3772±0.0760 | 0.3218±0.0326 |
| $H(0.25)$ | 0.3842±0.0806 | 0.3739±0.0870 | 0.3562±0.0617 | 0.3904±0.0849 | 0.3206±0.0337 |
| $H(0.5)$ | 0.3847±0.0765 | 0.4274±0.1251 | 0.3722±0.0745 | 0.3848±0.0691 | 0.3255±0.0496 |
| $M(0)$ | 0.2088±0.0581 | 0.2223±0.0595 | **0.2471**±0.0423 | **0.2650**±0.0342 | **0.2514**±0.0107 |
| $M(0.1)$ | **0.2406**±0.0649 | 0.2700±0.0628 | 0.2946±0.0453 | 0.2875±0.0429 | 0.2652±0.0191 |
| $M(0.25)$ | 0.3136±0.0742 | **0.2469**±0.1100 | 0.3042±0.0633 | 0.3064±0.0545 | 0.2732±0.0273 |
| $M(0.5)$ | 0.3820±0.0993 | 0.2822±0.0750 | 0.3066±0.0729 | 0.3226±0.0597 | 0.3067±0.0405 |
| | | | $GPD_\gamma$ $(\gamma = 0.5)(\rho = -0.5)$ | | |
| $H$ | 0.5938±0.1056 | 0.6289±0.0777 | 0.5993±0.0553 | 0.5690±0.0392 | 0.5366±0.0463 |
| $M$ | 0.5559±0.1575 | 0.5814±0.0916 | 0.5805±0.0551 | 0.5693±0.0371 | 0.5245±0.0384 |
| $\overline{H}_0$ | 0.5889±0.0805 | 0.6004±0.0648 | 0.5897±0.0458 | 0.5864±0.0227 | 0.5339±0.0251 |
| $\overline{M}_1$ | 0.5769±0.1488 | 0.5908±0.0877 | 0.5891±0.0491 | 0.5794±0.0319 | 0.5270±0.0367 |
| $H(0)$ | 0.5941±0.1057 | 0.6290±0.0777 | 0.5993±0.0553 | 0.5690±0.0391 | 0.5366±0.0463 |
| $H(0.1)$ | 0.5928±0.1163 | 0.6344±0.0846 | 0.5991±0.0533 | 0.5745±0.0421 | 0.5420±0.0467 |
| $H(0.25)$ | 0.6266±0.1150 | 0.6270±0.0919 | 0.6148±0.0547 | 0.5643±0.0329 | 0.5452±0.0679 |
| $H(0.5)$ | 0.6474±0.1651 | 0.6280±0.0854 | 0.6003±0.0663 | 0.5910±0.0529 | 0.5466±0.0798 |
| $M(0)$ | 0.5564±0.1577 | **0.5814**±0.0916 | **0.5805**±0.0551 | **0.5681**±0.0381 | 0.5245±0.0384 |
| $M(0.1)$ | 0.5518±0.1502 | 0.5861±0.0889 | 0.5836±0.0552 | 0.5722±0.0385 | 0.5220±0.0409 |
| $M(0.25)$ | **0.5266**±0.1700 | 0.5924±0.8900 | 0.5828±0.0593 | 0.5773±0.0379 | 0.5213±0.0466 |
| $M(0.5)$ | 0.5302±0.1998 | 0.6043±0.0946 | 0.5829±0.0523 | 0.5803±0.0502 | **0.5152**±0.0612 |

Table 3.2
Simulated mean values, at optimal levels, of $H$, $M$, $\overline{H}_0$, $\overline{M}_1$, $H(q)$, and $M(q)$, $q = 0.1, 0.25$, and 0.5, for parents with $|\rho| = -1$

| $n$ | 200 | 500 | 1000 | 2000 | 5000 |
|---|---|---|---|---|---|
| | | | Frechet $(\gamma = 1)(\rho = -1)$ | | |
| $H$ | 1.0498±0.1085 | 1.0750±0.0624 | 1.0657±0.0463 | 1.0775±0.0487 | 1.0356±0.0268 |
| $M$ | 1.0612±0.1197 | 1.0709±0.0809 | 1.0656±0.0489 | 1.0697±0.0650 | 1.0385±0.0257 |
| $\overline{H}_0$ | **1.0296**±0.1034 | 1.0353±0.0813 | 1.0226±0.0398 | **1.0286**±0.0506 | **1.0033**±0.0209 |
| $\overline{M}_1$ | 0.9607±0.1274 | **1.0091**±0.0824 | **1.0136**±0.0447 | 1.0300±0.0558 | 1.0099±0.0211 |
| $H(0)$ | 1.0540±0.1302 | 1.1040±0.0740 | 1.0626±0.0486 | 1.0697±0.0480 | 1.0373±0.0282 |
| $H(0.1)$ | 1.0714±0.1430 | 1.0832±0.0683 | 1.0743±0.0499 | 1.0770±0.0604 | 1.0504±0.0155 |
| $H(0.25)$ | 1.0471±0.1752 | 1.0863±0.0867 | 1.1003±0.0662 | 1.0793±0.0656 | 1.0495±0.0234 |
| $H(0.5)$ | 1.0763±0.1877 | 1.1428±0.1132 | 1.1070±0.0824 | 1.0693±0.0606 | 1.0445±0.0302 |
| $M(0)$ | 1.0719±0.1262 | 1.0761±0.0869 | 1.0687±0.0530 | 1.0736±0.0677 | 1.0414±0.0275 |
| $M(0.1)$ | 1.0574±0.1572 | 1.0836±0.1007 | 1.0768±0.0573 | 1.0845±0.0747 | 1.0467±0.0310 |
| $M(0.25)$ | 1.0811±0.1663 | 1.0904±0.1134 | 1.0847±0.0637 | 1.0889±0.0835 | 1.0497±0.0335 |
| $M(0.5)$ | 1.1069±0.1832 | 1.1002±0.1386 | 1.0978±0.0729 | 1.0976±0.0957 | 1.0571±0.0418 |
| | | | Student $t_2$ $(\gamma = 0.5)(\rho = -1)$ | | |
| $H$ | 0.5599±0.1079 | 0.6104±0.0698 | 0.5548±0.0325 | 0.5271±0.0328 | 0.5246±0.0270 |
| $M$ | 0.4063±0.1195 | 0.5123±0.1031 | 0.4924±0.0285 | 0.4754±0.0472 | 0.4969±0.0299 |
| $\overline{H}_0$ | 0.4714±0.0607 | 0.5233±0.0454 | 0.5023±0.0213 | **0.5019**±0.0215 | **0.4988**±0.0155 |
| $\overline{M}_1$ | 0.3427±0.0009 | 0.4655±0.0005 | 0.4703±0.0002 | 0.4483±0.0001 | 0.4781±0.0001 |
| $H(0)$ | 0.1270±0.0488 | 0.2613±0.1874 | 0.2535±0.1125 | 0.2457±0.1554 | 0.1645±0.0793 |
| $H(0.1)$ | 0.4761±0.0423 | 0.4849±0.0349 | 0.4940±0.0225 | 0.4941±0.0137 | 0.4962±0.0087 |
| $H(0.25)$ | **0.5111**±0.0680 | 0.4891±0.0224 | **0.5000**±0.0179 | 0.4968±0.0150 | 0.4964±0.0093 |
| $H(0.5)$ | 0.5595±0.1090 | 0.6074±0.0675 | 0.5561±0.0316 | 0.5241±0.0326 | 0.5245±0.0269 |
| $M(0)$ | 0.2290±0.0781 | 0.3323±0.0780 | 0.3571±0.0454 | 0.3436±0.0459 | 0.3681±0.0322 |
| $M(0.1)$ | 0.2979±0.0950 | 0.4028±0.0797 | 0.4248±0.0371 | 0.4121±0.0444 | 0.4360±0.0388 |
| $M(0.25)$ | 0.3521±0.1068 | 0.4406±0.0768 | 0.4497±0.0358 | 0.4285±0.0394 | 0.4539±0.0356 |
| $M(0.5)$ | 0.4153±0.1188 | **0.5107**±0.1015 | 0.4990±0.0315 | 0.4760±0.0468 | 0.4950±0.0289 |
| | | | $EV_\gamma$ $(\gamma = 1)(\rho = -1)$ | | |
| $H$ | 1.0430±0.1501 | 1.0668±0.1310 | 1.1578±0.0625 | 1.0805±0.0600 | 1.0478±0.0287 |
| $M$ | 0.9902±0.1870 | 1.0006±0.1000 | 1.1287±0.0812 | 1.0931±0.0883 | 1.0527±0.0373 |
| $\overline{H}_0$ | 0.8708±0.0823 | 0.9667±0.0638 | **1.0495**±0.0579 | **1.0377**±0.0507 | **1.0146**±0.0214 |
| $\overline{M}_1$ | 0.8123±0.1108 | 0.9220±0.0651 | 1.0759±0.0703 | 1.0874±0.0645 | 1.0518±0.0315 |
| $H(0)$ | 1.0918±0.1355 | **1.0037**±0.0717 | 1.0909±0.0765 | 1.0697±0.0480 | 1.0373±0.0282 |
| $H(0.1)$ | 1.0938±0.1606 | 1.0267±0.0930 | 1.0973±0.0646 | 1.0770±0.0604 | 1.0504±0.0155 |
| $H(0.25)$ | 1.0911±0.1505 | 1.0475±0.1138 | 1.1369±0.0740 | 1.0793±0.0656 | 1.0495±0.0234 |
| $H(0.5)$ | 1.0854±0.1723 | 1.0581±0.1348 | 1.1379±0.0838 | 1.0693±0.0606 | 1.0445±0.0302 |
| $M(0)$ | 1.0082±0.1261 | 0.9948±0.0659 | 1.1000±0.0684 | 1.0736±0.0677 | 1.0414±0.0275 |
| $M(0.1)$ | **1.0022**±0.1561 | 0.9904±0.0806 | 1.1133±0.0763 | 1.0845±0.0747 | 1.0467±0.0310 |
| $M(0.25)$ | 1.0162±0.1653 | 0.9932±0.0916 | 1.1224±0.0803 | 1.0889±0.0835 | 1.0497±0.0335 |
| $M(0.5)$ | 0.9944±0.1805 | 1.0133±0.1102 | 1.1399±0.0817 | 1.0976±0.0957 | 1.0571±0.0418 |

Table 3.3
Simulated mean values, at optimal levels, of $H$, $M$, $\overline{H}_0$, $\overline{M}_1$, $H(q)$, and $M(q)$, $q = 0.1, 0.25$, and 0.5, for parents with $|\rho| > 1$

| $n$ | 200 | 500 | 1000 | 2000 | 5000 |
|---|---|---|---|---|---|
| | | | Student $t_1$ $(\gamma = 1)(\rho = -2)$ | | |
| $H$ | 1.1198±0.1513 | 1.1329±0.0991 | 1.0744±0.0868 | 1.0401±0.0452 | 1.0258±0.0490 |
| $M$ | 1.0550±0.1661 | 1.0937±0.1257 | 1.0349±0.0829 | 1.0392±0.0509 | 1.0164±0.0470 |
| $\overline{H}_0$ | 1.0228±0.1191 | 1.1255±0.1113 | 1.0356±0.0766 | 1.0344±0.0267 | 1.0071±0.0466 |
| $\overline{M}_1$ | 1.0767±0.1144 | 1.0719±0.0661 | 1.0647±0.0471 | 1.0739±0.0451 | 1.0371±0.0269 |
| $H(0)$ | 0.3105±0.1960 | 0.3997±0.2939 | 0.2189±0.1124 | 0.4788±0.3238 | 0.4629±0.2564 |
| $H(0.1)$ | 0.7925±0.1710 | 0.8584±0.1290 | 0.8289±0.1213 | 0.9069±0.0813 | 0.9355±0.0764 |
| $H(0.25)$ | **1.0161**±0.0839 | **0.9735**±0.0395 | **1.0138**±0.0306 | **1.0167**±0.0133 | **1.0000**±0.0134 |
| $H(0.5)$ | 1.1086±0.1493 | 1.1348±0.1006 | 1.0799±0.0842 | 1.0367±0.0473 | 1.0256±0.0484 |
| $M(0)$ | 0.4514±0.0753 | 0.4716±0.0317 | 0.4753±0.0197 | 0.4948±0.0097 | 0.4967±0.0064 |
| $M(0.1)$ | 0.7012±0.1303 | 0.8056±0.1272 | 0.8139±0.0903 | 0.9015±0.0732 | 0.9004±0.0702 |
| $M(0.25)$ | 0.8428±0.1220 | 0.9015±0.0979 | 0.8959±0.0755 | 0.9343±0.0545 | 0.9350±0.0495 |
| $M(0.5)$ | 1.0523±0.1721 | 1.0964±0.1270 | 1.0391±0.0818 | 1.0374±0.0506 | 1.0165±0.0463 |
| | | | $GPD_\gamma$ $(\gamma = 2)(\rho = -2)$ | | |
| $H$ | 2.1099±0.1881 | 2.0214±0.0990 | 2.0849±0.0893 | 2.0389±0.0551 | 2.0606±0.0674 |
| $M$ | 2.0832±0.2040 | 1.9605±0.0895 | 2.0861±0.1023 | 2.0562±0.0661 | 2.0444±0.0558 |
| $\overline{H}_0$ | 2.1310±0.1325 | **2.0030**±0.0954 | 2.0574±0.1051 | 2.0307±0.0610 | 2.0464±0.0540 |
| $\overline{M}_1$ | **1.9728**±0.1796 | 1.9280±0.0875 | **2.0259**±0.0916 | **2.0078**±0.0629 | **2.0042**±0.0473 |
| $H(0)$ | 2.1092±0.1876 | 2.0216±0.0989 | 2.0850±0.0893 | 2.0389±0.0551 | 2.0606±0.0674 |
| $H(0.1)$ | 2.1367±0.1969 | 2.0115±0.1013 | 2.0861±0.0953 | 2.0317±0.0722 | 2.0503±0.0641 |
| $H(0.25)$ | 2.1531±0.2338 | 2.0343±0.1223 | 2.0844±0.1148 | 2.0433±0.0862 | 2.0383±0.0594 |
| $H(0.5)$ | 2.0667±0.2715 | 2.0213±0.0973 | 2.1414±0.1075 | 2.0757±0.0814 | 2.0623±0.0662 |
| $M(0)$ | 2.0828±0.2022 | 1.9605±0.0905 | 2.0863±0.1023 | 2.0563±0.0661 | 2.0444±0.0559 |
| $M(0.1)$ | 2.0689±0.2206 | 1.9666±0.0975 | 2.0844±0.1023 | 2.0537±0.0661 | 2.0440±0.0559 |
| $M(0.25)$ | 2.0524±0.2302 | 1.9592±0.1096 | 2.0949±0.1027 | 2.0573±0.0684 | 2.0453±0.0646 |
| $M(0.5)$ | 2.0348±0.2767 | 1.9249±0.1253 | 2.1156±0.1204 | 2.0813±0.0794 | 2.0322±0.0657 |

Table 3.4
Simulated mean square errors of $H$ (first row) and $REFF$-indicators of $M$, $\overline{H}_0$, $\overline{M}_1$, $H(q)$, and $M(q)$, $q = 0.1, 0.25$, and 0.5, for Fréchet parents with $\gamma = 1$ $(\rho = -1)$

| $n$ | 200 | 500 | 1000 | 2000 | 5000 |
|---|---|---|---|---|---|
| | | | Frechet $(\gamma = 1)(\rho = -1)$ | | |
| $MSE_H$ | 0.0259±0.0004 | 0.0135±0.0002 | 0.0083±0.0001 | 0.0051±0.0000 | 0.0027±0.0000 |
| $REFF_{M\|H}$ | 1.0229±0.0036 | 1.0176±0.0054 | 1.0131±0.0028 | 1.0077±0.0039 | 1.0080±0.0052 |
| $REFF_{\overline{H}_0\|H}$ | **1.3290**±0.0096 | **1.3763**±0.0141 | **1.4731**±0.0071 | **1.5752**±0.0129 | **1.7902**±0.0196 |
| $REFF_{\overline{M}_1\|H}$ | 1.0447±0.0068 | 1.1372±0.0095 | 1.2383±0.0088 | 1.3428±0.0100 | 1.5352±0.0154 |
| $REFF_{H(0)\|H}$ | 0.9065±0.0037 | 0.9172±0.0022 | 0.9238±0.0020 | 0.9279±0.0023 | 0.9370±0.0017 |
| $REFF_{H(0.1)\|H}$ | 0.8144±0.0033 | 0.8131±0.0037 | 0.8132±0.0035 | 0.8101±0.0048 | 0.8120±0.0037 |
| $REFF_{H(0.25)\|H}$ | 0.7416±0.0042 | 0.7421±0.0042 | 0.7405±0.0034 | 0.7382±0.0044 | 0.7413±0.0048 |
| $REFF_{H(0.5)\|H}$ | 0.6251±0.0055 | 0.6284±0.0046 | 0.6307±0.0040 | 0.6295±0.0043 | 0.6316±0.0049 |
| $REFF_{M(0)\|H}$ | 0.9345±0.0033 | 0.9381±0.0048 | 0.9395±0.0027 | 0.9391±0.0042 | 0.9449±0.0049 |
| $REFF_{M(0.1)\|H}$ | 0.8413±0.0038 | 0.8358±0.0051 | 0.8302±0.0027 | 0.8239±0.0043 | 0.8229±0.0044 |
| $REFF_{M(0.25)\|H}$ | 0.7658±0.0045 | 0.7636±0.0049 | 0.7588±0.0026 | 0.7530±0.0042 | 0.7526±0.0042 |
| $REFF_{M(0.5)\|H}$ | 0.6416±0.0048 | 0.6488±0.0050 | 0.6467±0.0029 | 0.6420±0.0039 | 0.6433±0.0039 |

Table 3.5
Simulated mean square errors of $H$ (first row) and $REFF$-indicators of $M$, $\overline{H}_0$, $\overline{M}_1$, $H(q)$, and $M(q)$, $q = 0.1, 0.25$, and $0.5$, for $EV_\gamma$ parents with $\gamma = 0.25$

| $n$ | 200 | 500 | 1000 | 2000 | 5000 |
|---|---|---|---|---|---|
| | | | $EV_\gamma$ $(\gamma = 0.25)(\rho = -0.25)$ | | |
| $MSE_H$ | 0.0402±0.0006 | 0.0246±0.0004 | 0.0176±0.0003 | 0.0127±0.0002 | 0.0085±0.0001 |
| $REFF_{M\|H}$ | 1.0929±0.0122 | 1.2719±0.0109 | 1.3587±0.0137 | 1.4152±0.0071 | 1.5529±0.0143 |
| $REFF_{\overline{H}_0\|H}$ | 1.2339±0.0043 | 1.1713±0.0066 | 1.1332±0.0048 | 1.1023±0.0029 | 1.0711±0.0033 |
| $REFF_{\overline{M}_1\|H}$ | 1.4665±0.0170 | 1.8416±0.0172 | 2.1562±0.0226 | 2.5308±0.0232 | 3.1365±0.0345 |
| $REFF_{H(0)\|H}$ | 1.4959±0.0068 | 1.5169±0.0093 | 1.5336±0.0095 | 1.5407±0.0057 | 1.5597±0.0134 |
| $REFF_{H(0.1)\|H}$ | 1.2293±0.0057 | 1.2136±0.0056 | 1.2072±0.0041 | 1.1994±0.0032 | 1.1902±0.0047 |
| $REFF_{H(0.25)\|H}$ | 1.0880±0.0026 | 1.0810±0.0034 | 1.0779±0.0023 | 1.0751±0.0008 | 1.0721±0.0021 |
| $REFF_{H(0.5)\|H}$ | 0.9095±0.0037 | 0.9147±0.0024 | 0.9183±0.0025 | 0.9202±0.0020 | 0.9256±0.0017 |
| $REFF_{M(0)\|H}$ | **1.5073**±0.0171 | **1.9175**±0.0227 | **2.3055**±0.0208 | **2.7677**±0.0130 | **3.5114**±0.0467 |
| $REFF_{M(0.1)\|H}$ | 1.3995±0.0134 | 1.7507±0.0143 | 2.0598±0.0143 | 2.4442±0.0126 | 3.1323±0.0279 |
| $REFF_{M(0.25)\|H}$ | 1.2298±0.0124 | 1.4341±0.0119 | 1.5345±0.0153 | 1.7109±0.0125 | 2.1631±0.0177 |
| $REFF_{M(0.5)\|H}$ | 0.9544±0.0209 | 1.1164±0.0087 | 1.1985±0.0114 | 1.2547±0.0058 | 1.3204±0.0128 |

Table 3.6
Simulated mean square errors of $H$ (first row) and $REFF$-indicators of $M$, $\overline{H}_0$, $\overline{M}_1$, $H(q)$, and $M(q)$, $q = 0.1, 0.25$, and $0.5$, for $EV_\gamma$ parents with $\gamma = 1$

| $n$ | 200 | 500 | 1000 | 2000 | 5000 |
|---|---|---|---|---|---|
| | | | $EV_\gamma$ $(\gamma = 1)(\rho = -1)$ | | |
| $MSE_H$ | 0.0558±0.0010 | 0.0286±0.0003 | 0.0175±0.0002 | 0.0107±0.0001 | 0.0057±0.0001 |
| $REFF_{M\|H}$ | 1.0262±0.0052 | 1.0307±0.0040 | 1.0242±0.0045 | 1.0194±0.0055 | 1.0165±0.0026 |
| $REFF_{\overline{H}_0\|H}$ | 1.2253±0.0286 | **1.4324**±0.0316 | **1.6822**±0.0327 | **1.9978**±0.0249 | **2.5151**±0.0221 |
| $REFF_{\overline{M}_1\|H}$ | 1.0830±0.0222 | 1.2712±0.0188 | 1.3845±0.0107 | 1.3456±0.0112 | 1.1382±0.0052 |
| $REFF_{H(0)\|H}$ | 1.3168±0.0054 | 1.3331±0.0072 | 1.3390±0.0090 | 1.3476±0.0096 | 1.3565±0.0075 |
| $REFF_{H(0.1)\|H}$ | 1.1832±0.0053 | 1.1813±0.0081 | 1.1786±0.0058 | 1.1764±0.0077 | 1.1755±0.0048 |
| $REFF_{H(0.25)\|H}$ | 1.0773±0.0047 | 1.0781±0.0032 | 1.0731±0.0036 | 1.0719±0.0045 | 1.0731±0.0027 |
| $REFF_{H(0.5)\|H}$ | 0.9096±0.0036 | 0.9131±0.0033 | 0.9141±0.0027 | 0.9141±0.0040 | 0.9143±0.0026 |
| $REFF_{M(0)\|H}$ | **1.3544**±0.0090 | 1.3613±0.0064 | 1.3604±0.0076 | 1.3637±0.0085 | 1.3678±0.0054 |
| $REFF_{M(0.1)\|H}$ | 1.2187±0.0076 | 1.2128±0.0052 | 1.2021±0.0059 | 1.1964±0.0070 | 1.1913±0.0039 |
| $REFF_{M(0.25)\|H}$ | 1.1100±0.0065 | 1.1081±0.0050 | 1.0988±0.0050 | 1.0935±0.0062 | 1.0895±0.0031 |
| $REFF_{M(0.5)\|H}$ | 0.9320±0.0052 | 0.9414±0.0035 | 0.9364±0.0037 | 0.9323±0.0052 | 0.9312±0.0025 |

Table 3.7
Simulated mean square errors of $H$ (first row) and $REFF$-indicators of $M$, $\overline{H}_0$, $\overline{M}_1$, $H(q)$, and $M(q)$, $q = 0.1, 0.25$, and 0.5, for $EV_\gamma$ parents with $\gamma = 2$

| $n$ | 200 | 500 | 1000 | 2000 | 5000 |
|---|---|---|---|---|---|
| | | | $EV_\gamma$ $(\gamma = 2)(\rho = -1)$ | | |
| $MSE_H$ | 0.0129±0.0029 | 0.0612±0.0013 | 0.0372±0.0003 | 0.0226±0.0002 | 0.0118±0.0002 |
| $REFF_{M\|H}$ | 1.0637±0.0050 | 1.0689±0.0042 | 1.0576±0.0039 | 1.0411±0.0048 | 1.0344±0.0028 |
| $REFF_{\overline{H}_0\|H}$ | 0.5819±0.1552 | 0.9426±0.0094 | 1.0074±0.0059 | 1.1089±0.0044 | **1.3571**±0.0028 |
| $REFF_{\overline{M}_1\|H}$ | 0.7405±0.1205 | 1.0356±0.0182 | 1.1190±0.0060 | **1.1616**±0.0091 | 1.1334±0.0051 |
| $REFF_{H(0)\|H}$ | 1.0953±0.0047 | 1.0722±0.0023 | 1.0640±0.0027 | 1.0561±0.0030 | 1.0412±0.0019 |
| $REFF_{H(0.1)\|H}$ | 1.0747±0.0037 | 1.0566±0.0017 | 1.0505±0.0022 | 1.0451±0.0026 | 1.0322±0.0021 |
| $REFF_{H(0.25)\|H}$ | 1.0388±0.0029 | 1.0302±0.0014 | 1.0278±0.0019 | 1.0252±0.0016 | 1.0181±0.0016 |
| $REFF_{H(0.5)\|H}$ | 0.9426±0.0033 | 0.9509±0.0034 | 0.9556±0.0026 | 0.9588±0.0025 | 0.9664±0.0017 |
| $REFF_{M(0)\|H}$ | **1.1839**±0.0096 | **1.1554**±0.0045 | **1.1328**±0.0040 | 1.1052±0.0039 | 1.0836±0.0044 |
| $REFF_{M(0.1)\|H}$ | 1.1563±0.0082 | 1.1365±0.0044 | 1.1162±0.0039 | 1.0918±0.0039 | 1.0733±0.0044 |
| $REFF_{M(0.25)\|H}$ | 1.1131±0.0060 | 1.1056±0.0040 | 1.0896±0.0038 | 1.0683±0.0040 | 1.0561±0.0036 |
| $REFF_{M(0.5)\|H}$ | 0.9958±0.0040 | 1.0134±0.0057 | 1.0060±0.0036 | 0.9959±0.0043 | 0.9947±0.0030 |

Table 3.8
Simulated mean square errors of $H$ (first row) and $REFF$-indicators of $M$, $\overline{H}_0$, $\overline{M}_1$, $H(q)$, and $M(q)$, $q = 0.1, 0.25$, and 0.5, for student parents $t_v$, with $v = 4$ degrees of freedom

| $n$ | 200 | 500 | 1000 | 2000 | 5000 |
|---|---|---|---|---|---|
| | | Student $t_4$ $(\gamma = 0.25)(\rho = -0.5)$ | | | |
| $MSE_H$ | 0.0204±0.0004 | 0.0112±0.0002 | 0.0073±0.0001 | 0.0048±0.0001 | 0.0029±0.0000 |
| $REFF_{M\|H}$ | 0.5277±0.0069 | 0.6198±0.0068 | 0.6696±0.0052 | 0.7078±0.0040 | 0.7481±0.0082 |
| $REFF_{\overline{H}_0\|H}$ | 1.3992±0.0171 | 1.3600±0.0097 | 1.3249±0.0108 | 1.2811±0.0105 | 1.2360±0.0100 |
| $REFF_{\overline{M}_1\|H}$ | 0.5837±0.0079 | 0.6227±0.0066 | 0.6547±0.0042 | 0.6855±0.0053 | 0.7280±0.0083 |
| $REFF_{H(0)\|H}$ | 1.9181±0.0233 | 1.2850±0.0113 | 0.8359±0.0047 | 0.5610±0.0042 | 0.3620±0.0022 |
| $REFF_{H(0.1)\|H}$ | **3.0107**±0.0310 | **3.4637**±0.0290 | **3.9376**±0.0358 | **4.4930**±0.0502 | **5.4485**±0.0625 |
| $REFF_{H(0.25)\|H}$ | 1.7002±0.0156 | 1.7846±0.0127 | 1.8815±0.0126 | 1.9872±0.0144 | 2.1792±0.0143 |
| $REFF_{H(0.5)\|H}$ | 1.0035±0.0011 | 1.0013±0.0009 | 1.0007±0.0004 | 1.0004±0.0000 | 1.0001±0.0004 |
| $REFF_{M(0)\|H}$ | 0.5062±0.0059 | 0.5192±0.0052 | 0.5198±0.0025 | 0.5135±0.0040 | 0.5044±0.0046 |
| $REFF_{M(0.1)\|H}$ | 0.5250±0.0065 | 0.5603±0.0058 | 0.5760±0.0024 | 0.5844±0.0046 | 0.5942±0.0057 |
| $REFF_{M(0.25)\|H}$ | 0.5336±0.0070 | 0.5841±0.0062 | 0.6086±0.0027 | 0.6252±0.0046 | 0.6442±0.0065 |
| $REFF_{M(0.5)\|H}$ | 0.5289±0.0069 | 0.6197±0.0067 | 0.6692±0.0051 | 0.7075±0.0040 | 0.7479±0.0082 |

Table 3.9
Simulated mean square errors of $H$ (first row) and $REFF$-indicators of $M$, $\overline{H}_0$, $\overline{M}_1$, $H(q)$, and $M(q)$, $q = 0.1, 0.25$, and 0.5, for student parents $t_v$, with $v = 2$ degrees of freedom

| $n$ | 200 | 500 | 1000 | 2000 | 5000 |
|---|---|---|---|---|---|
| | | | Student $t_2$ $(\gamma = 0.5)(\rho = -1)$ | | |
| $MSE_H$ | 0.0230±0.0004 | 0.0116±0.0001 | 0.0070±0.0001 | 0.0043±0.0001 | 0.0022±0.0000 |
| $REFF_{M\|H}$ | 0.6813±0.0041 | 0.8123±0.0062 | 0.9086±0.0048 | 1.0057±0.0065 | 1.1488±0.0112 |
| $REFF_{\overline{H}_0\|H}$ | 1.4179±0.0195 | 1.6942±0.0247 | 1.9507±0.0214 | 2.2143±0.0255 | 2.6311±0.0317 |
| $REFF_{\overline{M}_1\|H}$ | 0.6258±0.0058 | 0.7016±0.0084 | 0.7619±0.0066 | 0.8207±0.0107 | 0.8988±0.0076 |
| $REFF_{H(0)\|H}$ | 0.4506±0.0041 | 0.3190±0.0022 | 0.2483±0.0016 | 0.1947±0.0017 | 0.1403±0.0010 |
| $REFF_{H(0.1)\|H}$ | **2.3302**±0.0277 | **2.5868**±0.0176 | **2.8415**±0.0233 | **3.1373**±0.0223 | **3.5726**±0.0243 |
| $REFF_{H(0.25)\|H}$ | 1.9862±0.0166 | 2.2168±0.0163 | 2.4293±0.0166 | 2.6650±0.0189 | 3.0537±0.0364 |
| $REFF_{H(0.5)\|H}$ | 1.0060±0.0018 | 1.0017±0.0008 | 1.0003±0.0007 | 1.0000±0.0004 | 0.9998±0.0003 |
| $REFF_{M(0)\|H}$ | 0.5173±0.0047 | 0.4765±0.0032 | 0.4345±0.0033 | 0.3886±0.0038 | 0.3243±0.0025 |
| $REFF_{M(0.1)\|H}$ | 0.6043±0.0053 | 0.6121±0.0047 | 0.5993±0.0048 | 0.5756±0.0067 | 0.5294±0.0049 |
| $REFF_{M(0.25)\|H}$ | 0.6572±0.0056 | 0.7039±0.0053 | 0.7161±0.0055 | 0.7163±0.0077 | 0.6898±0.0055 |
| $REFF_{M(0.5)\|H}$ | 0.6830±0.0047 | 0.8109±0.0060 | 0.9059±0.0045 | 1.0023±0.0060 | 1.1458±0.0114 |

Table 3.10
Simulated mean square errors of $H$ (first row) and $REFF$-indicators of $M$, $\overline{H}_0$, $\overline{M}_1$, $H(q)$, and $M(q)$, $q = 0.1, 0.25$, and 0.5, for student parents $t_v$, with $v = 1$ degrees of freedom

| $n$ | 200 | 500 | 1000 | 2000 | 5000 |
|---|---|---|---|---|---|
| | | | Student $t_1$ $(\gamma = 1)(\rho = -2)$ | | |
| $MSE_H$ | 0.0370±0.0005 | 0.0166±0.0003 | 0.0095±0.0001 | 0.0053±0.0001 | 0.0025±0.0000 |
| $REFF_{M\|H}$ | 0.8668±0.0074 | 0.9151±0.0068 | 0.9234±0.0052 | 0.9232±0.0044 | 0.9272±0.0044 |
| $REFF_{\overline{H}_1\|H}$ | 0.7966±0.1693 | 1.1591±0.0109 | 1.1584±0.0078 | 1.1610±0.0075 | 1.1636±0.0055 |
| $REFF_{\overline{M}_1\|H}$ | 0.5245±0.0749 | 0.7026±0.0120 | 0.7528±0.0119 | 0.8086±0.0127 | 0.8816±0.0087 |
| $REFF_{H(0)\|H}$ | 0.2529±0.0023 | 0.1712±0.0017 | 0.1303±0.0011 | 0.0976±0.0009 | 0.0671±0.0007 |
| $REFF_{H(0.1)\|H}$ | 0.5569±0.0061 | 0.4928±0.0047 | 0.4602±0.0048 | 0.4290±0.0052 | 0.3909±0.0044 |
| $REFF_{H(0.25)\|H}$ | **1.5400**±0.0146 | **1.6404**±0.0098 | **1.7504**±0.0151 | **1.8564**±0.0185 | **2.0077**±0.0242 |
| $REFF_{H(0.5)\|H}$ | 1.0060±0.0023 | 1.0022±0.0010 | 0.9998±0.0017 | 0.9992±0.0011 | 0.9988±0.0008 |
| $REFF_{M(0)\|H}$ | 0.3431±0.0029 | 0.2423±0.0021 | 0.1879±0.0013 | 0.1435±0.0012 | 0.1004±0.0009 |
| $REFF_{M(0.1)\|H}$ | 0.5361±0.0055 | 0.4734±0.0047 | 0.4455±0.0045 | 0.4177±0.0051 | 0.3826±0.0042 |
| $REFF_{M(0.25)\|H}$ | 0.7588±0.0082 | 0.6971±0.0074 | 0.6559±0.0065 | 0.6139±0.0064 | 0.5614±0.0067 |
| $REFF_{M(0.5)\|H}$ | 0.8698±0.0078 | 0.9179±0.0069 | 0.9234±0.0047 | 0.9225±0.0050 | 0.9264±0.0043 |

Table 3.11
Simulated mean square errors of $H$ (first row) and $REFF$-indicators of $M$, $\overline{H}_0$, $\overline{M}_1$, $H(q)$, and $M(q)$, $q = 0.1, 0.25$, and 0.5, for $GDP$ parents, with $\gamma = 0.5$

| $n$ | 200 | 500 | 1000 | 2000 | 5000 |
|---|---|---|---|---|---|
| | | | $GPD_\gamma$ $(\gamma = 0.5)(\rho = -0.5)$ | | |
| $MSE_H$ | 0.0362±0.0006 | 0.0208±0.0003 | 0.0139±0.0003 | 0.0094±0.0001 | 0.0057±0.0000 |
| $REFF_{M\|H}$ | 1.0687±0.0071 | 1.1037±0.0063 | 1.1145±0.0060 | 1.1192±0.0041 | 1.1317±0.0063 |
| $REFF_{\overline{H}_0\|H}$ | **1.3803**±0.0082 | **1.3390**±0.0069 | **1.3019**±0.0105 | **1.2648**±0.0063 | **1.2336**±0.0061 |
| $REFF_{\overline{M}_1\|H}$ | 1.1892±0.0095 | 1.1682±0.0059 | 1.1557±0.0059 | 1.1486±0.0046 | 1.1493±0.0062 |
| $REFF_{H(0)\|H}$ | 0.9984±0.0000 | 0.9994±0.0004 | 0.9997±0.0003 | 0.9999±0.0028 | 0.9999±0.0004 |
| $REFF_{H(0.1)\|H}$ | 0.9667±0.0013 | 0.9693±0.0019 | 0.9703±0.0009 | 0.9692±0.0015 | 0.9711±0.0014 |
| $REFF_{H(0.25)\|H}$ | 0.9119±0.0031 | 0.9189±0.0040 | 0.9206±0.0022 | 0.9193±0.0024 | 0.9242±0.0020 |
| $REFF_{H(0.5)\|H}$ | 0.7982±0.0048 | 0.8131±0.0063 | 0.8168±0.0045 | 0.8192±0.0023 | 0.8270±0.0018 |
| $REFF_{M(0)\|H}$ | 1.0668±0.0071 | 1.1029±0.0063 | 1.1142±0.0060 | 1.1190±0.0041 | 1.1317±0.0063 |
| $REFF_{M(0.1)\|H}$ | 1.0280±0.0062 | 1.0664±0.0061 | 1.0796±0.0060 | 1.0849±0.0038 | 1.0989±0.0059 |
| $REFF_{M(0.25)\|H}$ | 0.9599±0.0058 | 1.0070±0.0053 | 1.0205±0.0067 | 1.0290±0.0036 | 1.0431±0.0046 |
| $REFF_{M(0.5)\|H}$ | 0.8179±0.0040 | 0.8790±0.0036 | 0.8996±0.0062 | 0.9130±0.0029 | 0.9297±0.0042 |

Table 3.12
Simulated mean square errors of $H$ (first row) and $REFF$-indicators of $M$, $\overline{H}_0$, $\overline{M}_1$, $H(q)$, and $M(q)$, $q = 0.1, 0.25$, and 0.5, for $GDP$ parents, with $\gamma = 2$

| $n$ | 200 | 500 | 1000 | 2000 | 5000 |
|---|---|---|---|---|---|
| | | | $GPD_\gamma$ $(\gamma = 2)(\rho = -2)$ | | |
| $MSE_H$ | 0.0658±0.0012 | 0.0309±0.0006 | 0.0176±0.0003 | 0.0100±0.0001 | 0.0047±0.0001 |
| $REFF_{M\|H}$ | 1.0162±0.0040 | 1.0075±0.0034 | 1.0040±0.0059 | 0.9998±0.0043 | 0.9957±0.0037 |
| $REFF_{\overline{H}_1\|H}$ | **1.1632**±0.0072 | **1.1542**±0.0062 | **1.1490**±0.0060 | **1.1431**±0.0044 | **1.1307**±0.0048 |
| $REFF_{\overline{M}_1\|H}$ | 1.1601±0.0067 | 1.2215±0.0086 | 1.2874±0.0109 | 1.3343±0.0067 | 1.4110±0.0070 |
| $REFF_{H(0)\|H}$ | 0.9979±0.0000 | 0.9991±0.0031 | 0.9996±0.0043 | 0.9998±0.0000 | 0.9999±0.0003 |
| $REFF_{H(0.1)\|H}$ | 0.9563±0.0014 | 0.9553±0.0029 | 0.9563±0.0027 | 0.9583±0.0030 | 0.9579±0.0028 |
| $REFF_{H(0.25)\|H}$ | 0.8860±0.0035 | 0.8856±0.0039 | 0.8887±0.0035 | 0.8900±0.0043 | 0.8903±0.0040 |
| $REFF_{H(0.5)\|H}$ | 0.7469±0.0070 | 0.7472±0.0029 | 0.7511±0.0034 | 0.7533±0.0055 | 0.7549±0.0044 |
| $REFF_{M(0)\|H}$ | 1.0142±0.0039 | 1.0067±0.0033 | 1.0036±0.0059 | 0.9996±0.0043 | 0.9957±0.0037 |
| $REFF_{M(0.1)\|H}$ | 0.9730±0.0031 | 0.9654±0.0027 | 0.9614±0.0054 | 0.9583±0.0040 | 0.9543±0.0014 |
| $REFF_{M(0.25)\|H}$ | 0.9033±0.0040 | 0.8961±0.0026 | 0.8931±0.0056 | 0.8907±0.0039 | 0.8873±0.0039 |
| $REFF_{M(0.5)\|H}$ | 0.7599±0.0062 | 0.7585±0.0019 | 0.7572±0.0067 | 0.7574±0.0047 | 0.7537±0.0042 |

## 3.3 An Application to the Nasdaq Composite Index

As an empirical example, we place ourselves in a context from finance, analyzing the risk for investors holding short positions in the Nasdaq Composite index, i.e., for investors betting on a fall in the index. Since we are interested in the analysis of the risk of holding short positions, we begin with the positive log-returns defined in (1.2.1), assumed to be stationary and weakly dependent. With the purpose of comparison with a case study from Drees (2003), we have used the daily log-returns from 1997-2000, which corresponds to a sample size $n = 1037$. Although there is some increasing trend in the volatility, stationarity is assumed, under the same considerations as in Drees (2003). In Fig. 3.3.1 we display the estimates for the tail index associated to $\hat{\gamma}^H_{n,k}$, $\hat{\gamma}^M_{n,k}$, $\hat{\gamma}^{H(q)}_{n,k}$, and $\hat{\gamma}^{M(q)}_{n,k}$ for some values of $p$. It is clear from the analysis of the $\gamma$-scatterplots that all estimates are positive for $k$ from about 50 up to 450, i.e., there is a strong evidence for a heavy-tailed underlying distribution. However, the patterns exhibited by the different estimators $\hat{\gamma}_{n,k}^{H(q)}$ are significantly different for different values of the tuning parameter $q$. We have been, at a first sight, particularly puzzled with the sample paths of $\hat{\gamma}_{n,k}^{H(0)}$ , and such sample paths immediately suggest a possible non consistency of $\hat{\gamma}_{n,k}^{H(0)}$ due to an infinite left endpoint of the underlying model. We have thus decided to analyze more deeply both tails of the model underlying the sample $r_i$, $1 \le i \le 1036$. For that we have used not only the Hill estimator, but also the MVRB-estimator $\overline{H}_0$ in (3.1.5), which is, for heavy tails, an alternative to the Hill estimator not only at the optimal levels or for large $k$, as happens with the "classical" second-order reduced bias tail index estimators, but for all $k$. It was indeed this estimator that led us to the estimate $\gamma = 0.34$ pictured in Fig. 3.3.1 and consequently to the choice $q = 0.25$ for the class of estimators $H(q)$ in (3.1.7).

**Right Tail Analysis of Nasdaq Data.** In Fig. 3.3.2, and working with the $n_0 = 570$ positive values of the log-returns on NASDAQ data, we picture the sample paths of $\hat{\rho}_0(k)$ and $\hat{\rho}_1(k)$. The algorithm in Sec. 3.2.2 leads us to choose, on the basis of any stability criterion for large values of $k$, the estimate associated to $\tau = 0$. We have considered $\hat{\rho} = \hat{\rho}_0(k_1)$, with $k_1 = n_0^{0.995}$. We have got $\hat{\rho}_0 = \hat{\rho}_0(552) = -0.71$. The use of the $\beta$-estimate suggested in the above mentioned algorithm, led us to the estimate $\hat{\beta}_0 = 1.04$. For the estimation of $\gamma$ through the reduced bias tail index estimators, we have used the heuristic estimate of the level provided in Gomes and Pestana (2007a), i.e., the value $k_{01} \equiv k_{01}(n; \beta, \rho) = (1.96(1-\rho)n_0^{-\rho}/|\beta|)^{2/(1-2\rho)}$. Levels of this type are still levels such that $\sqrt{k}(n/k)^\rho \to \lambda$, finite, and are not yet optimal for the tail index estimation through second-order reduced-bias tail index estimators. However, do not forget that with a tail index estimator like $\overline{H}$, in (3.1.5), we are always safe and able to provide a more reliable estimation than through the Hill estimates. We came to $\hat{k}_{01} = 109$ and to the estimate $\hat{\gamma} = \overline{H}_0(109) = 0.34$. Note that the estimation of the optimal threshold (Hall and Welsh, 1985) for the estimation through the Hill estimator in (3.1.2), leads us to:

$$\hat{k}_0 = \Big(\frac{(1-\hat{\rho})n_0^{-\hat{\rho}}}{\hat{\beta}\sqrt{-2\hat{\rho}}}\Big)^{2/(1-2\hat{\rho})} = 55 \Rightarrow \hat{\gamma}_{n,k}^{H}(\hat{k}_0) = 0.41.$$

**Left Tail Analysis of Nasdaq Data.** Figure 3.3.3 is related to a similar data analysis, carried on the $n_0 = 466$ positive values of the symmetric log returns. We have now obtained $\hat{\rho} = -071$, $\hat{\beta} = 1.05$, $\hat{k}_0 = 48$, $\hat{\gamma}_{n,k}^{H}(\hat{k}_0) = 0.35$, $\hat{k}_{01} = 97$, and $\hat{\gamma} = \overline{H}_0(97) = 0.3$.

This data analysis leads us to the conclusion that the underlying model detains a location median not far from 0. Indeed, when we induce a shift associated to the tuning parameter $q = 0.5$, we get a sample path not a long way from that of the Hill estimator (see Fig.

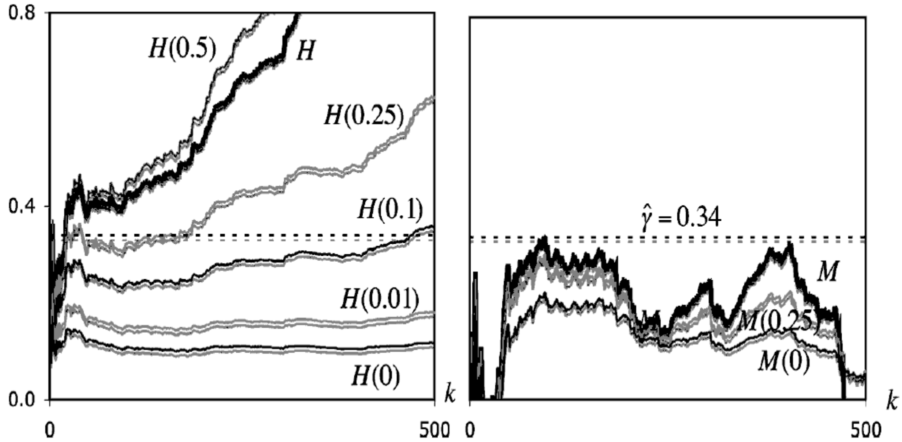



Figure 3.3.1: Tail index estimates based on $\hat{\gamma}^{H}_{n,k}$ and $\hat{\gamma}^{H(q)}_{n,k}$ (left) and on $\hat{\gamma}^{M}_{n,k}$ and $\hat{\gamma}^{M(q)}_{n,k}$ (right), for $q = 0.05, 0.25,\ 0.5$, on NASDAQ data.

3.3.1, left). Moreover, relying on the observed results for the $\gamma$ estimates, it is not sensible to discard the possibility that both tails are heavy (with the right tail underlying the $R_i$ slightly heavier than the left tail ($\hat{\gamma} = 0.34$ for the right tail vs. $\hat{\gamma} = 0.30$ for the left tail). This obviously implies an underlying model with support $(-\infty, +\infty)$. It is then not at all sensible to induce a shift $R_{1:n}$, like it is suggested in Drees (2003). Such a shift is appealing, because it induces for the Hill estimator an almost flat sample path (see again Fig. 3.3.1, left), but as mentioned before, the "flat zone" leads, in this case, to a severe underestimation of the tail index $\gamma$. To support this statement, look again at Figs. 3.2.7 and 3.2.8, with the pattern of mean values ($E$) and mean squared errors ($MSE$) of the PORT-Hill and -moment estimators, respectively, for models from a Student-$t_v$ parent with $v = 4$ degrees of freedom ($\gamma = 0.25$). Although a parametric data analysis of this data is outside the scope of the present article, the similarities between the behaviour of the mean value patterns in Figs. 3.2.7 and 3.2.8 and the sample paths of the Hill and moment PORT-estimators in Fig. 3.3.2, suggest that the cdf underlying these returns is not a long way from a Student-t cdf or its skewed extensions, which are very common models in the area of extremes and finance. For recent references see Jones and Faddy (2003) and McNeil et al. (2005). In this application, and taking into account the previous analysis, it seems sensible to consider as a compromise choice in the PORT-Hill estimator, the shift induced by the first empirical quartile, i.e., to pick the value $q = 0.25$, as we have already seen in Fig. 3.3.1, but the possibility of working simultaneously with other estimators, like the MVRB estimator here considered, should not at all be discarded, because this can help us to better estimate the extreme value index, a parameter of primordial importance in all subsequent extreme value analysis needed.

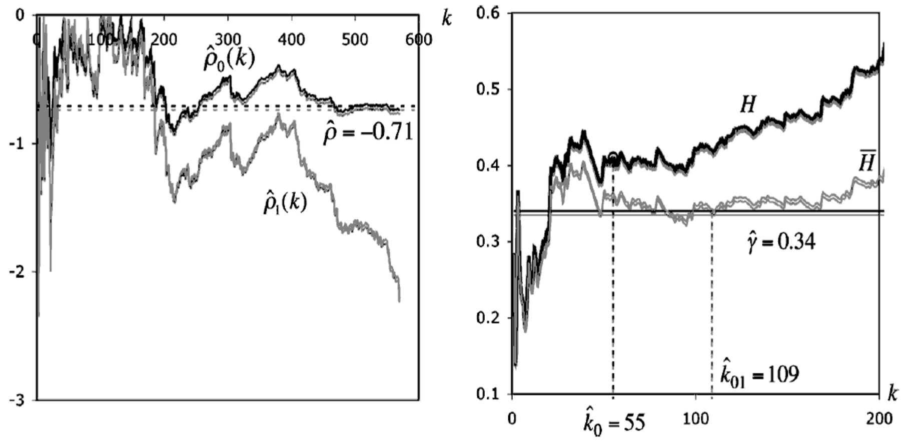



Figure 3.3.2: Estimates of the second-order parameters $\rho$, through $\hat{\rho}_0(k)$ and $\hat{\rho}_1(k)$ (left), and the tail index $\gamma$ (right), for the positive log-returns $X$, on NASDAQ data.

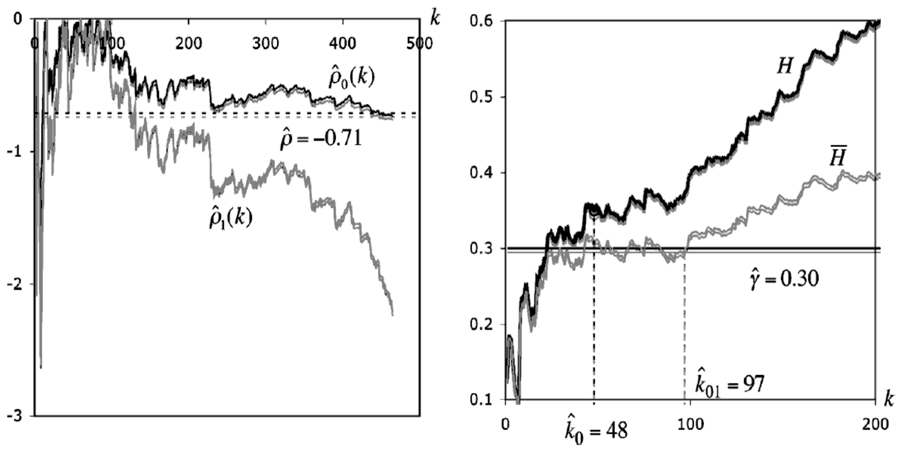



Figure 3.3.3: Estimates of the second-order parameters $\rho$, through $\hat{\rho}_0(k)$ and $\hat{\rho}_1(k)$ (left), and the tail index $\gamma$ (right), for the negative log-returns $L$, on NASDAQ data.

*Paulo Araújo Santos and Isabel Fraga Alves*
*To appear in Computational Statistics and Data Analysis*

# 4

# A New Class of Independence Tests for Interval Forecasts Evaluation

Interval forecasts evaluation can be reduced to examining the unconditional coverage and independence properties of the hit sequence. A new class of exact independence tests for the hit sequence and a definition for tendency to clustering of violations are proposed. The tests are suitable for detecting models with a tendency to generate clusters of violations and are based on an exact distribution that does not depend on an unknown parameter. The asymptotic distribution is also derived. The choice of one test within the class is studied. Moreover, a simulation study provides evidence that, in order to test the independence hypothesis, the suggested tests perform better than other tests presented in the literature. An empirical application is given for a period that includes the 2008 financial crisis.

## 4.1 Introduction

One of the core topics of quantitative financial risk management is the accurate calculation of the Value at Risk (VaR), which amounts to a tail quantile of the forecast profit and loss distribution over a specified time horizon. Owing to the non-iid and non-Gaussian nature of financial asset returns data, the calculation of VaR is not trivial; see, e.g., Kuester et al. (2005) and the references therein for a survey of competing methods. The primary tool for assessing its accuracy is to monitor the binary sequence generated by observing if the return on day t+1 is in the tail region specified by the VaR at time-$t$, or not. This is referred to as the *hit sequence*. In mathematical terms we consider a time series of daily log returns defined in (1.2.1). The corresponding one-day-ahead VaR forecasts made at time-$t$ for time $t+1$, $\text{VaR}_{t+1|t}(p)$, are defined in (1.2.2). Considering that a *violation* occurs when the daily return on the portfolio is lower than the reported VaR, we define the hit function in (1.2.3). Christoffersen (1998) showed that evaluating interval forecasts can be reduced to examining whether the hit sequence, $\{I_t\}_{t=1}^{T}$, satisfies the unconditional coverage (UC) and independence (IND) properties. UC hypothesis means $P[I_{t+1}(p) = 1] = p$, $\forall t$. IND hypothesis means that past information does not hold information about future violations. Clustering of violations is one problematic infraction to the IND hypothesis, which corresponds to several large losses occurring in a short period of time. As noted by Campbell (2007), the IND property represents a more subtle yet equally important property. However, some authors argue that a certain amount of moderate clustering may not be harmful, so that correct UC is somewhat more important than independence (e.g. Jorion, 2002). When both properties are valid we say that forecasts have a correct conditional coverage (CC) and we write

$$P[I_{t+1}(p) = 1|\Omega_t] = p, \forall t. \tag{4.1.1}$$

In Lemma 1 of Christoffersen (1998) it is shown that condition CC (4.1.1) is equivalent to $I_{t+1}(p) \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$. In a recent paper, Berkowitz *et al.* (2009) extended and unified the existing tests by noting that the de-meaned hits $\{I_{t+1} - p\}$ form a martingale difference sequence. Equations (1) and (4.1.1) imply that $E[(I_{t+1} - p)|\Omega_t] = 0$ and then for any variable $Z_t$ in the time-$t$ information set, we must have

$$E[(I_{t+1} - p)Z_t] = 0. \tag{4.1.2}$$

This is the motivation for tests based on the martingale property. The rest of the Chapter is organized as follows. In Section 4.2 we review existent tests for evaluating interval forecasts. In Section 4.3 we present the new class of independence tests and exact and asymptotic distributions are derived for a random variable (rv) related with the test statistic. The choice of one test within the class is also studied. In Section 4.4, and through simulation experiments, we compare the performance with other tests under study. Section 4.5 presents an empirical application. Section 4.6 concludes.

## 4.2 Tests for interval forecasts evaluation

There are several backtesting procedures for evaluating interval forecasts; for a detailed review see Campbell (2007) and Berkowitz *et al.* (2009). The first procedures were mainly concerned with the UC property and the proportion of failures (POF) test proposed by Kupiec (1995) is a well known example. A simple autocorrelation based independence test was proposed by Granger, White and Kamstra (1989). In the last ten years, several tests have been suggested to examine both the IND and the CC properties. The Christoffersen (1998) Markov tests are perhaps the most widely used in the literature. Therein $\pi_{ij}$ is defined as $P[I_t = j|I_{t-1} = i]$, for $i, j \in \{0, 1\}$. In this context, the null hypothesis of the IND test is $H_{0,\text{IND}} : \pi_{01} = \pi_{11}$ and the null hypothesis of the CC test is $H_{0,\text{CC}} : \pi_{01} = \pi_{11} = p$. Denoting by $\pi_1$ the common value of $\pi_{01}$ and $\pi_{11}$ under $H_{0,\text{IND}}$, by $T_0$ the number of zeros in the hit sequence $\underset{\sim}{I}$, $T_1$ the number of ones, $T = T_0 + T_1$ and $T_{ij}$ the number of observations with a $j$ following an $i$, the maximum likelihood estimators are $\hat{\pi}_{01} = T_{01}/T_0$, $\hat{\pi}_{11} = T_{11}/T_1$ and $\hat{\pi}_1 = T_1/T$, the log-likelihood under the alternative hypothesis is

$$\log L(\underset{\sim}{I}, \pi_{01}, \pi_{11}) = (1 - \pi_{01})^{T_0 - T_{01}} \pi_{01}^{T_{01}} (1 - \pi_{11})^{T_1 - T_{11}} \pi_{11}^{T_{11}},$$

the IND test statistic is

$$LR_{\text{IND}} = -2(\ln L(\underset{\sim}{I}, \hat{\pi}_1) - \ln L(\underset{\sim}{I}, \hat{\pi}_{01}, \hat{\pi}_{11})), \tag{4.2.1}$$

and the CC test statistic is

$$LR_{\text{CC}} = -2(\ln L(\underset{\sim}{I}, p) - \ln L(\underset{\sim}{I}, \hat{\pi}_{01}, \hat{\pi}_{11})). \tag{4.2.2}$$

The test statistics (4.2.1) and (4.2.2) are asymptotically distributed as chi-square with one degree of freedom. We use the notation $\text{M}_{\text{IND}}$ for the Markov independence test. If in equation (4.1.2) we set $Z_t$ to be the most recent de-meaned hit we have $E[(I_{t+1} - p)(I_t - p)] = 0$, the only condition explored by the Markov tests. If we set $Z_t = (I_{t-k} - p)$ for any $k \geq 0$, we have $E[(I_{t+1} - p)(I_{t-k} - p)] = 0$. Based on this broader condition Berkowitz *et al.* (2009) suggested the Ljung-Box statistic, for a joint test of whether the first $m$ autocorrelations of $\{I_t\}$ are zero. The testing procedure is based on an asymptotic chi-square distribution with $m$ degrees of freedom.

Considering other data in the information set such as past returns, under CC we have $E[(I_{t+1} - p)g(I_t, I_{t-1}, ..., R_t, R_{t-1}, ...)] = 0$ for any non-anticipating function $g(.)$. In the same line as Engle and Manganelli (2004), Berkowitz *et al.* (2009) consider the autoregression

$$I_t = \alpha + \sum_{k=1}^{n} \beta_{1k} I_{t-k} + \sum_{k=1}^{n} \beta_{2k} g(I_{t-k}, I_{t-k-1}, ..., R_{t-k}, R_{t-k-1}) + \varepsilon_t, \tag{4.2.3}$$

with $n = 1$ and $g(I_{t-k}, I_{t-k-1}, ..., R_{t-k}, R_{t-k-1}) = \text{VaR}_{t-k+1|t-k}(p)$. These authors propose the logit model and test the CC hypothesis with a likelihood ratio test considering for the null $P(I_t = 1) = 1/(1 + e^{-\alpha}) = p$ and the coefficients $\beta_{11}$ and $\beta_{21}$ equal to zero. For the the IND hypothesis the null is $\beta_{11} = \beta_{21} = 0$ and in this case the asymptotic distribution is chi-square with 2 degrees of freedom. We refer to these tests as the CAViaR tests of Engle and Manganelli (CAViaR).

A *duration-based* approach emerged in the literature. There are related works on testing duration dependence (e.g., Kiefer, 1988). As far as we know, the first authors that proposed this approach for interval forecast evaluation were Danielsson and Morimoto (2000), using the chi-square goodness of fit test. In this set-up, let us define the duration between two consecutive violations as

$$D_i := t_i - t_{i-1}, \tag{4.2.4}$$

where $t_i$ denotes the day of violation number $i$ and $t_0 = 0$, which implies that $D_1$ is the time until the first violation. We denote a sequence of $N$ durations by $\{D_i\}_{i=1}^{N}$ . If the CC (4.1.1) hypothesis is valid then $I_{t+1}(p) \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$ and consequently the process $\{D_i\}_{i=1}^{N}$ has a geometric distribution with probability mass function (pmf)

$$f_D(d;\pi) = (1-\pi)^{(d-1)}\pi, \; d \in \mathbb{N}, \tag{4.2.5}$$

with $\pi = p$. We will write the IND hypothesis as

$$D_i \overset{\text{iid}}{\sim} D \sim \text{Geometric}(\pi), \; \text{with } 0 < \pi < 1. \tag{4.2.6}$$

The exponential distribution with probability density function (pdf)

$$f_D(d;\beta) = \beta \exp(-\beta d), \quad d > 0 \text{ and } \beta > 0, \tag{4.2.7}$$

is the continuous analogue of the geometric distribution. Based on the exponential, Christoffersen and Pelletier (2004) suggested tests using the duration based approach, specifying the Weibull, the Gamma and the Exponential Autoregressive Conditional Duration models for the alternative. Haas (2005) showed that tests based on discrete distributions for durations, have higher power.

The Generalised Method of Moments (GMM) test framework suggested by Bontemps and Meddahi (2008) to test for distributional assumptions was extended by Candelon *et al.* (2008) to the case of VaR forecasts accuracy. In the group of duration-based tests it is shown that the proposed GMM tests are the best performers. The orthonormal polynomials associated with the geometric distribution with probability $p$ are defined by the following recursive relationship, $\forall d \in \mathbb{N}$,

$$M_{j+1}(d;p) = \frac{(1-p)(2j+1) + p(j-d+1)}{(j+1)\sqrt{1-p}} M_j(d;p) - \left(\frac{j}{j+1}\right) M_{j-1}(d;p),$$

for any order $j \in \mathbb{N}$, with $M_{-1}(d;p) = 0$ and $M_0(d;p) = 1$. If (4.1.1) is true, then it follows that $E[M_j(D;p)] = 0$, $\forall j \in \mathbb{N}$. The CC property can be expressed as $H_{0,\text{CC}}$ : $E[M_j(D;p)] = 0$ and the IND property can be expressed as $H_{0,\text{IND}}$ : $E[M_j(D;\beta)] = 0$ with $j = \{1, ..., k\}$ and $k > 1$ denoting the number of moment conditions. The parameter

$\beta$, not necessarily equal to $p$, can be either fixed *a priori* or estimated. The GMM test statistic for CC is

$$\mathrm{J}_{\mathrm{CC}}(k) = \Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} M(D_i;p)\Big)'\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} M(D_i;p)\Big), \tag{4.2.8}$$

and for IND is

$$\mathrm{J}_{\mathrm{IND}}(k) = \Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} M(D_i;\hat{\beta})\Big)'\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} M(D_i;\hat{\beta})\Big), \tag{4.2.9}$$

where $M(D_i;p)$ denotes a $(k,1)$ vector whose components are the orthonormal polynomials $M_j(D_i;p)$ in the CC test and $M_j(D_i;\hat{\beta})$ in the IND test, for $j = 1, ..., k$. $\hat{\beta}$ is a consistent estimator of $\beta$. The test statistics (4.2.8) and (4.2.9) follow an asymptotic chi-square distribution with $k$ and $k-1$ degrees of freedom, respectively. For the CC and IND hypothesis, the Markov tests (4.2.1) and (4.2.2) are perhaps the most widely used in the literature and this is why we have chosen the Markov independence test (4.2.1) for the comparative study in Section 4.4. From the available duration-based tests we chose the GMM tests since these have been shown to exhibit the best performance in this group (Berkowitz *et al.*, 2009). We also selected the CAViaR test, the best performer in the simulation study of Berkowitz *et al.* (2009).

## 4.3 A new class of independence tests

### 4.3.1 Motivation

Let $D_{1:N} \leq \cdots \leq D_{N:N}$ be the order statistics (o.s.'s) of durations $D_1, ..., D_N$ defined in (4.2.4). The first motivation behind the class proposed is the following: when violations generated by the hit function occur in clusters, the majority of durations are short (the short durations between violations in the clusters) and some durations are very long (the durations between the last violation of one cluster and the first violation of the following cluster). If the majority of durations are short then the median, $D_{[N/2]:N}$, is short (notation: $[x]$ denotes the integer part of x). If some durations are very long, the maximum, $D_{N:N}$, is very long. Finally, with a short median and a very long maximum, the ratio $D_{N:N}/D_{[N/2]:N}$ is large.

We illustrate this motivation with an example: we chose the returns from the German stock market index (DAX) from January 2, 1997 up until December 30, 2008, and we calculated durations between violations using the popular Historical Simulation (HS) method for VaR(0.05) with a moving window of size 250. The sample size for the hit sequence (T) is 2790 and the sample size for durations (N) is 170. We can calculate the expected values of the o.s.'s, $D_{1:N} \leq \cdots \leq D_{N:N}$, under the independence hypothesis (4.2.6), using the following expression obtained by Margolin and Winokur (1967),

$$E(D_{r:N}) = \sum_{j=N-r-1}^{N} (-1)^{j-N+r-1} \binom{j-1}{N-r} \binom{N}{j} \frac{1}{(1-(1-\pi)^j)}.$$

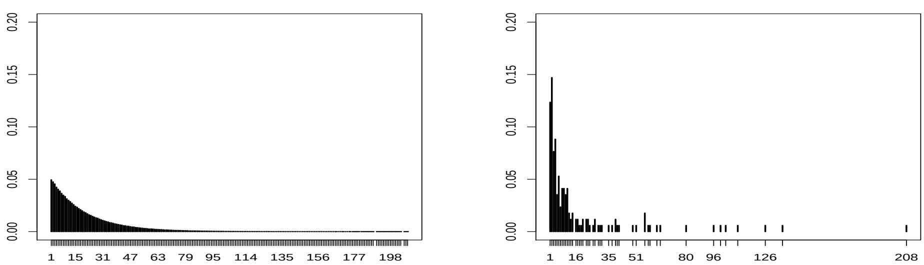



Figure 4.3.1: Geometric($\pi = 0.05$) pmf (left) and frequency of durations (right) between violations for DAX index from January,2 1997 until December 30, 20 2008, based on VaR(0.05), the Historical Simulation technique and on the previous 250 trading days.

Figure 4.3.1 shows the geometric pmf, with $\pi = 0.05$, and the frequency of durations. For short durations, the frequencies in the frequency plot are much higher than the probability masses in the geometric pmf. The majority of durations are short, either equal to or lower than 6 days and the empirical median is 6, contrasting with the expected value of $D_{85:170}$, under IND, which is close to 14. Moreover, for durations above 60 days we note higher frequencies in the frequency plot than the probability masses in the geometric pmf. The maximum duration, $d_{170:170}$, is 208 days, almost double the expected value under IND, which is close to 112. The ratio is 34.66, much higher than the median of $D_{170:170}/D_{85:170}$ under IND, which is 8.03 (see the cdf of Proposition 3.1). In this example, where violations occur in clusters, the majority of durations are short, some durations are very long and a high ratio $D_{N:N}/D_{[N/2]:N}$ gives strong evidence against the IND hypothesis.

A second motivation for the class presented is based on the two-parameter Weibull distribution. The cumulative distribution function (cdf) of the continuous Weibull is

$$F_W(w;\beta,\theta) = 1 - \exp(-(\beta w)^\theta) \quad w > 0,\ \beta > 0,\ \theta > 0. \tag{4.3.1}$$

The Weibull with $\theta < 1$, will generate an excessive number of very short durations and an excessive number of very long durations. The Weibull with $\theta > 1$ will generate the opposite pattern. If we consider $D_w := [W]+1$, we obtain the Nakagawa & Osaki (1975) discrete Weibull, with pmf

$$\begin{aligned} f_{D_w}(d) &= F_W(d) - F_W(d-1) \\ &= \exp(-\beta^\theta)^{(d-1)^\theta} - \exp(-\beta^\theta)^{d^\theta} \\ &= q^{(d-1)^\theta} - q^{d^\theta}, \quad q = 1-p = \exp(-\beta^\theta), \quad d = 1, 2, ... \end{aligned}$$

The method of maximum likelihood (ML) considers the log-likelihood function

$$\log L(q, \theta; d_1, ..., d_n) = \sum_{i=1}^{n} \log \left\{ q^{(d_i-1)^\theta} - q^{d_i^\theta} \right\}.$$

The ML-equations $\partial L/\partial q = 0$ and $\partial L/\partial \theta = 0$, must be solved numerically.

We estimated the parameters $q$ and $\theta$ using the 170 durations from the previous DAX example and with the ML method. The fitted discrete Weibull model with $\hat{q} = 0.82$ and $\hat{\theta} = 0.66$ is presented in Figure 4.3.2. Evidently, the frequency plot pattern is closer to the probability masses of the discrete Weibull with $\theta < 1$ than to the geometric probability masses presented in Figure 4.3.1.

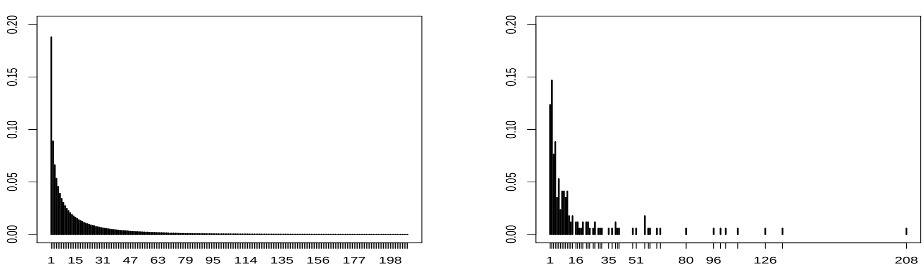



Figure 4.3.2: discrete Weibull($\theta = 0.66, q = 0.82$) pmf (left) and frequency of durations (right) between violations for DAX index from January 2, 1997, up until December 30, 2008, based on the Historical Simulation VaR(0.05) and on the previous 250 trading days.

### 4.3.2 Theoretical results

From now on, we denote

$$a_v = \binom{N-k-1}{v}, \quad b_s = \binom{k-1}{s},$$

$$c_{v,s} = N-k-v+s \quad, \gamma_N = \frac{N!}{(k-1)!(N-k-1)!} \quad \text{and}$$

$$R_{N,k} := \frac{D_{N:N}}{D_{k:N}}, \quad k = 1, ..., N-1. \tag{4.3.2}$$

We will denote $F^W_{R_{N,k}}(r) = P^W[R_{N,k} \leq r]$ the cdf of (4.3.2) under the Weibull model with cdf (4.3.1) and we will denote $F^E_{R_{N,k}}(r) = P^E[R_{N,k} \leq r]$ the cdf of (4.3.2) under the exponential model with pdf (4.2.7).

In Proposition 4.3.1, we present the cdf of $R_{N,k}$ when the underlying model is Weibull, and in Proposition 4.3.2 we prove that the condition $\theta < 1$ is equivalent to state that the median of $R_{N,k}$ is greater than the median under independence.

**Proposition 4.3.1.** *Let $D_1, ..., D_N$, be iid Weibull rv's with common cdf (4.3.1). The cdf of (4.3.2) is*

$$\begin{aligned} F^W_{R_{N,k}}(r) \quad = \quad & 1 - \theta\gamma_N \textstyle\sum_{v=0}^{N-k-1} \sum_{s=0}^{k-1} (-1)^{v+s} a_v b_s \\ & \left([\theta c_{v,s}(v+1)]^{-1} - [\theta c_{v,s}(v+1+c_{v,s}(1/r)^\theta)]^{-1}\right) \end{aligned} \tag{4.3.3}$$

*with $1 \leq r \leq \infty$.*

**Proof:** For the Weibull distribution with cdf (4.3.1), Malik and Trudel (1982) proved that the density of the ratio of the $k$-th and $j$-th o.s.'s with $k < j \leq N$, is

$$\begin{aligned} f^W_{Z_{j,k}}(z;\theta) \quad = \quad & \tfrac{\theta C_j}{(k-1)!(j-k-1)!} \textstyle\sum_{v=0}^{j-k-1} \sum_{s=0}^{k-1} (-1)^{v+s} \binom{j-k-1}{v} \\ & \binom{k-1}{s} z^{\theta-1}[N-j+v+1+(j-k-v+s)z^\theta]^{-2}, \end{aligned}$$

with $0 \leq z \leq 1$ and where $C_j = \prod_{u=1}^{j}(N-u+1)$. We replace $j$ by $N$ to get $Z_{N,k} = D_{k:N}/D_{N:N}$ and the cdf for this ratio is

$$\begin{aligned} F^W_{Z_{N,k}}(z) \quad &= \quad \int_0^z f_{Z_{N,k}}(u)du \\ &= \quad \theta\gamma_N \sum_{v=0}^{N-k-1} \sum_{s=0}^{k-1} (-1)^{v+s} a_v b_s \\ & \qquad \left([\theta c_{v,s}(v+1)]^{-1} - [\theta c_{v,s}(v+1+c_{v,s}z^\theta)]^{-1}\right), \end{aligned}$$

with $0 \leq z \leq 1$. For $R_{N,k} = D_{N:N}/D_{k:N} = 1/Z_{N,k}$ the cdf is

$$F^W_{R_{N,k}}(r) \quad = \quad P^W(D_{k:N}/D_{N:N} > 1/r) = 1 - F^W_{Z_{N,k}}(1/r),$$

and the result follows. □

**Proposition 4.3.2.** *Let $r_{1/2,N,k}$ be the value of $r$ such that $F^E_{R_{N,k}}(r) = 1/2$. If $D_1, ..., D_N$, are iid Weibull rv's with common cdf (4.3.1), then*

$$F^{W\leftarrow}_{R_{N,k}}(1/2) > r_{1/2,N,k} \text{ is equivalent to } \theta < 1,$$

*where $F^{W\leftarrow}_{R_{N,k}}(t) := \inf\{x : F^W_{R_{N,k}}(x) \geq t\}$ denotes the generalized inverse function of $F^W_{R_{N,k}}$.*

**Proof:** $F^{W\leftarrow}_{R_{N,k}}(1/2) > r_{1/2,N,k}$ is equivalent to

$$1/2 > P^W\Big[R_{N,k} \leq r_{1/2,N,k}\Big] = P^W\Big[\Big(\frac{D_{N:N}}{D_{k:N}}\Big)^\theta \leq r^\theta_{1/2,N,k}\Big].$$

Since under the Weibull model $(\beta D)^\theta$ is exponential(1), then

$$1/2 > P^E\Big[\frac{D_{N:N}}{D_{k:N}} \leq r^\theta_{1/2,N,k}\Big] \text{ is equivalent to } \theta < 1.$$

□

**Remark 4.3.1.** *When the underlying model is Weibull, the median of $D_{N:N}/D_{k:N}$ higher than the median under independence is equivalent to $\theta < 1$. However, $F^{W\leftarrow(1/2)}_{R_{N,k}} > r_{1/2,N,k}$ allows for a more general definition of a tendency to clustering, coherent with our first motivation presented in Subsection 4.3.1.*

Now, we can write

$$P^E\Big[R_{N,k} \leq r_{1/2,N,k}\Big] = \frac{1}{2} = P^W\Big[\Big(\frac{D_{N:N}}{D_{k:N}}\Big)^\theta \leq r_{1/2,N,k}\Big]$$

$$= P^W\Big[\theta \leq \log r_{1/2,N,k}\Big(\log\frac{D_{N:N}}{D_{k:N}}\Big)^{-1}\Big],$$

and it follows that

$$\hat{\theta}(k) = \log r_{1/2,N,k}\Big(\log\frac{D_{N:N}}{D_{k:N}}\Big)^{-1}, \tag{4.3.4}$$

is a median unbiased estimator of $\theta$. With $k = 1$ we get the Vogt median unbiased estimator of $\theta$ (Vogt, 1968). Notice that the estimator (4.3.4) is a function of the statistic $D_{N:N}/D_{k:N}$ and for observed values $d_{N:N}/d_{k:N} > r_{1/2,N,k}$ the estimates of $\theta$ are lower than 1. Based on the first motivation presented at the beginning of Subsection 4.3.1 and on Proposition 4.3.2, we propose the following class of statistics

$$S_{N,k} := \frac{D_{N:N} - 1}{D_{k:N}}, \quad k = 1, ..., N-1. \tag{4.3.5}$$

The correction $-1$ made to $D_{N:N}$, allows us to obtain a pivotal test. The Proposition 4.3.3 allows us to do that as well as to present in Proposition 4.3.5 a level $\alpha$ test. In Proposition 4.3.4 we normalize $R_{N,k}$ (4.3.2) to get the asymptotic distribution under the exponential model.

**Proposition 4.3.3.** *Let $D_1, ..., D_N$, be iid rv's whose common pmf is (4.2.5) with $\pi = p \in (0,1)$. If we consider $S_{N,k}$ (4.3.5) and $R_{N,k}$ (4.3.2) under the exponential model with pdf (4.2.7), then*

$$F^{\leftarrow}_{S_{N,k}}(1-\alpha) < F^{E\leftarrow}_{R_{N,k}}(1-\alpha), \text{ for all } 0 < p < 1, \text{ and } 0 < \alpha < 1.$$

**Proof:** Let $Y$ be an exponential rv with pdf (4.2.7) and denote $[Y]$ the integer part of $Y$ and $< Y >$ the fractional part of $Y$. If we define $X = [Y] + 1$, then

$$\begin{aligned} P[X = x] &= F_Y(x) - F_Y(x-1) \\ &= \big(\exp(-\beta)\big)^{(x-1)}\big(1 - \exp(-\beta)\big), \end{aligned}$$

with $x \in \mathbb{N}$. Note that $X$ is distributed as geometric with probability of success $(1 - \exp(-\beta))$. Now, for $p = (1 - \exp(-\beta))$,

$$D_{i:N} \stackrel{d}{=} X_{i:N} = [Y]_{i:N} + 1 \stackrel{d}{=} [Y_{i:N}] + 1,$$

and we have

$$\begin{aligned} S^G_{N,k} &= \frac{D_{N:N} - 1}{D_{k:N}} \stackrel{d}{=} \frac{[Y_{N:N}]}{[Y_{k:N}] + 1} < \frac{[Y_{N:N}] + < Y_{N:N} >}{[Y_{k:N}] + < Y_{k:N} >} \\ &= \frac{Y_{N:N}}{Y_{k:N}}, \end{aligned}$$

which is an rv with cdf $F^E_{R_{N,k}}$. □

**Proposition 4.3.4.** *If we consider $k = [\xi N]$, with $0 < \xi < 1$, and $D_1, ..., D_N$ iid rv's with pdf (4.2.7), then*

$$T^R_{N,k} = -log(1-\xi)R_{N,k} - log\ N \xrightarrow[N\to\infty]{d} G, \tag{4.3.6}$$

*where $G$ stands for a Gumbel rv with cdf*

$$\Lambda_G(g) = exp(-exp(-g)), \quad -\infty < g < +\infty. \tag{4.3.7}$$

**Proof:** It is well known that

$$\sqrt{N}\Big[D_{k:N} - F^{E\leftarrow}(\xi)\Big] \xrightarrow[N\to\infty]{d} \text{Normal}(0, \xi/(1-\xi)),$$

and we can write

$$\begin{aligned} T^R_{N,k} &= \frac{-\log(1-\xi)D_{N:N}}{-\log(1-\xi)+O_p(1/\sqrt{N})} - \log \mathrm{N} \\ &= \frac{D_{N:N} - \log \mathrm{N} - O_p(\log \mathrm{N}/\sqrt{N})}{1+O_p(1/\sqrt{N})} \\ &= \frac{D_{N:N} - \log \mathrm{N}}{1+o_p(1)} + o_p(1). \end{aligned}$$

Since

$$P^E[D_{N:N} - \log \mathrm{N} \le x] = \Big(1 - \frac{e^{-x}}{N}\Big)^N \underset{N\to\infty}{\longrightarrow} \Lambda_G(x),$$

by the Slutsky theorem the result follows. □

**Proposition 4.3.5.** *Let us consider* $\underset{\sim}{D} := \{D_i\}_{i=1}^N$*, the sample of* $N$ *durations (4.2.4) associated with the hit sequence (1). Define the class*

$$\mathcal{T}_k := \Big\{T_{N,k} = -\log(1-\xi)S_{N,k} - \log N,\ \ 0<\xi<1\Big\},$$

*where*

$$k = \begin{cases} [\xi N] & \text{if } [\xi N] \ge 1 \\ 1 & \text{if } [\xi N] < 1. \end{cases}$$

*Denote by* $Med(S_{N,k})$ *the median of* $S_{N,k}$ *and* $r^*_{1/2,N,k}$ *the particular value under geometric distribution (8). At level* $\alpha$*, for testing the IND hypothesis*

$$H_{0,IND} : D_i \overset{iid}{\sim} D \sim Geometric(\pi),\ \ with\ 0<\pi<1\ \ and\ i=1,...,N$$

*against alternatives expressing tendency to clustering patterns*

$$H_1 : Med(S_{N,k}) > r^*_{1/2,N,k},$$

*the rejection region is defined by* $T_{N,k} > t_{\alpha,N,k}$*, where* $t_{\alpha,N,k}$ *denotes a quantile* $1-\alpha$ *of* $T^R_{n,k}$ *(4.3.6) under the exponential model with pdf (4.2.7). For the asymptotic analog of the test use the Gumbel quantiles.*

**Proof:** The proof follows straightforwardly by Proposition 4.3.3, since under the null hypothesis

$$P\Big[S_{N,k} > \frac{t_{\alpha,N,k} + \log \mathrm{N}}{-\log(1-\xi)}\Big] < P^E\Big[R_{N,k} > \frac{t_{\alpha,N,k} + \log \mathrm{N}}{-\log(1-\xi)}\Big] = \alpha.$$

Then use Propositions 4.3.1 and 4.3.4. □

**Remark 4.3.2.** *Propositions 4.3.3 and 4.3.5 show that the critical point $t_{\alpha,N,k}$ implies a conservative approach with $P[\text{type I error}] \leq \alpha$, i.e., we have tests of level $\alpha$ and not of size $\alpha$. Since the distribution of ratios of o.s.'s does not depend on the scale parameter, the tests are pivotal in the sense that they are based on a distribution that does not depend on an unknown parameter. The limitation of the tests is that they can only test IND and not CC.*

**Remark 4.3.3.** *The tests suggested in Proposition 3.5 are based on an exact distribution. The other independence tests, presented in Section 4.2, are based on asymptotic distributions and suffer from small sample bias. To aggravate the problem, the presence of the nuisance parameter $p$ makes it impossible to control the size of the tests using the Monte Carlo testing approach of Dufour (2006) as other authors do for the case of joint testing UC and IND (e.g. Christoffersen and Pelletier (2004), Candelon et al. (2008) and Berkowitz et al. (2009)); see the paper of Dufour (2006) for details.*

### 4.3.3 The choice of k

The class of tests suggested in Proposition 4.3.5 raises one important problem: the choice of $k$. If we assume the continuous analogue of the geometric for the null, the Weibull distribution with $\theta = \theta_1 < 1$ for the alternative, and if we choose the statistics (4.3.2), applying the Proposition 4.3.1, we get the power function

$$\begin{aligned} 1-\beta_{N,k,\theta_1} &= P^W[-\log(1-\xi)R_{N,k} - \log \text{N} > t_{\alpha,N,k}|\theta=\theta_1] \\ &= \theta_1\gamma_N \sum_{v=0}^{N-k-1}\sum_{s=0}^{k-1}(-1)^{v+s}a_v b_s[\theta_1 c_{v,s}(v+1)]^{-1} \\ &\quad -\Big[\theta_1 c_{v,s}\Big(v+1+c_{v,s}\Big(\frac{-\log(1-\xi)}{t_{\alpha,N,k}+\log \text{N}}\Big)^{\theta_1}\Big)\Big]^{-1}. \end{aligned}$$

The $k$ that allows us to obtain the most powerful test is found as the solution of the following discrete maximization problem

$$k^* = \arg\max_k \ (1-\beta_{N,k,\theta_1}), \tag{4.3.8}$$

The optimal choice of $k$ will depend on $N$, $\theta_1$ and on the significance level ($\alpha$). However we are dealing with discrete rv's and we will consider the true process for the null, i.e., the geometric, and the discrete Weibull for the alternative. Taking this into consideration, we replace the power function in (4.3.8) by

$$1-\beta^d_{N,k,\theta_1,q_1} = P[-\log(1-\xi)S_{N,k} - \log \text{N} > t_{\alpha,N,k}|\theta=\theta_1, q=q_1].$$

Now we do not have an equation to compute the power, but we solve the problem by simulation. We have studied extensively the power curves $1-\beta^d_{N,k,\theta_1,q_1}$ with $\alpha = 0.1$,

$0 < \theta_1 < 1$ and $q = 1 - p = 0.95, 0.99$, for fixed values of $N$ and $k$. First, there is no choice of $k$ that leads to a most powerful test against all the alternatives. The test with $k = [0.5N]$ is the most powerful in some cases and in other cases has little less power than the most powerful test. We illustrate this for some cases in Figure 4.3.3.

The discrete Weibull is only one possible process that can lead to clustering of violations. In the following Section we simulate realistic returns processes that generate clustering of violations. We will conclude that for choices of $k = [\xi N]$ with $\xi = 0.3, 0.4, 0.5, 0.6, 0.7$, in a large number of cases we achieve more power with $\xi = 0.5$, and that in other cases, this test is only slightly less powerful. Based on these studies and on Proposition 4.3.2, we suggest the following definition.

**Definition 4.3.1** (Tendency to clustering of violations)**.** *A hit function (1) has a tendency to clustering of violations if the median of $D_{N:N}/D_{[N/2]:N}$ is higher than the median under the independence hypothesis (4.2.6).*

For explicitly testing the IND hypothesis (4.2.6) versus the tendency to clustering of violations, we propose the following test statistic from class $\mathcal{T}_k$, with $\xi = 0.5$

$$T_{N,[N/2]} = \log 2 \frac{D_{N:N} - 1}{D_{[N/2]:N}} - \log \text{N}. \tag{4.3.9}$$

In the Section 4.7 we provide a table with $2 \leq N \leq 100$ and critical values $t_{\alpha,N,[N/2]}$ for $\alpha = 0.1, 0.05, 0.01$. The test is easily implemented for any $N$, computing the upper bound for the p-value by solving $1 - F^W_{R_N}((t_{N,[N/2]} + \log \text{N})/\log 2)$ with $\theta = 1$ or for large $N$ using the asymptotic distribution (4.3.7). The low speed of convergence, $O(\log \text{N}/\sqrt{N})$, increases the importance of the exact distribution.

## 4.4 Comparative Simulation Study

In the context of a Monte Carlo study, we compare the power of the tests we suggest in Proposition 4.3.5 for $\xi = 0.3, 0.4, 0.5, 0.6, 0.7$, with the Markov independence (4.2.1), the CAViaR independence (6.4.1) and the GMM independence (4.2.9) tests. We denote these tests by $T_{N,[\xi N]}$, $\text{M}_{\text{IND}}$, CAViaR and $\text{J}_{\text{IND}}(k)$, respectively. We employ the R language and the fGarch package of Chalabi *et al.* (2008) in order to develop the programs. The R code for implementation of our test and comparisons is available in Araújo Santos (2010). Following other authors (e.g. Christofferson (1998), Christofferson and Pelletier (2004), Haas (2005), Candelon *et al.* (2008) and Berkowitz *et al.* (2009)) we consider a GARCH specification for the returns process. Additionally, we use a APARCH model which nests some of the GARCH models with leverage effect.

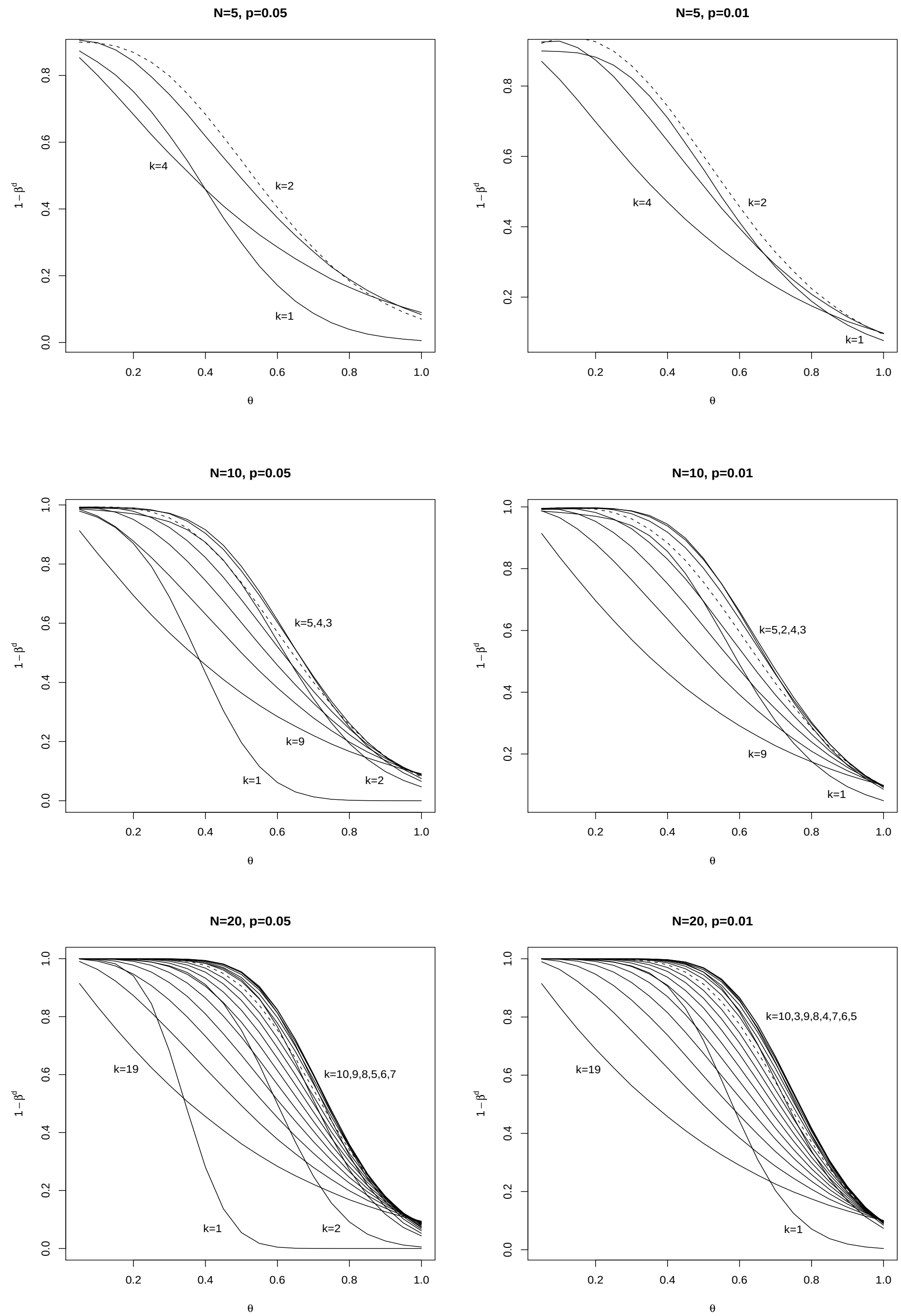



Figure 4.3.3: Power curves with $\alpha = 0.1$, $N = 5, 10, 20$, $p = 0.01, 0.05$ and different $k$ choices. Dashed curves are power curves for $k = [N/2]$.

- Gaussian GARCH(1,1) model (Bollerslev, 1986),

$$R_{t+1} = \sigma_{t+1} Z_{t+1} \quad \text{with} \quad \sigma_{t+1}^2 = w + \alpha R_t^2 + \beta \sigma_t^2, \tag{4.4.1}$$

  where the innovations $Z_{t+1}$´s are drawn independently from a standard normal distribution. As in Christofferson (1998), we chose the parameterisation $w = 0.05$, $\alpha = 0.1$ and $\beta = 0.85$.

- APARCH(1,1) model (Ding *et al.*, 1993),

$$R_{t+1} = \sigma_{t+1} Z_{t+1} \quad \text{with} \quad \sigma_{t+1}^\delta = w + \alpha(|R_t| - \gamma R_t)^\delta + \beta \sigma_t^\delta, \tag{4.4.2}$$

  where the innovations $Z_t$ are drawn independently from a skewed Student´s t($\nu$) distribution with asymmetry coefficient $\varphi$, proposed by Fernandez and Steel (1998). We assume a portfolio that replicates the DAX index and we use daily data from the beginning of 1997 until the end of 2008, for estimation. The parametrization achieved was $w = 0.03$, $\alpha = 0.086$, $\gamma = 0.64$, $\beta = 0.91$, $\delta = 1.15$, $\varphi = 0.88$ and $\nu = 10$.

As in other power studies with the same purpose, we have chosen the Historical Simulation method (HS) which easily generates clusters of violations when applied to heteroskedastic processes. We conducted our power experiment with sample sizes ($T$) equal to 250, 500, 750 and 1000 days. We set the size of the rolling window ($w_s$) equal to 250 and 500 days, and the VaR coverage rate $p$ equal to 0.01 and 0.05. For each $T$, $w_s$ and $p$, we simulated returns using the models (4.4.1) and (4.4.2) and calculated HS VaR´s and the test statistics over 5,000 replications. The empirical power of the tests is obtained by rejection frequencies with 0.1 significance level, excluding the samples with less than 2 violations. We report frequency of excluded samples (FES).

### 4.4.1 Simulation study under the IND hypothesis

For explicitly testing the IND hypothesis, the probability $P[I_{t+1} = 1]$ is unknown, we have a nuisance parameter and it is impossible to simulate finite sample critical values by simulating hit sequences from a binomial distribution with a known $p$, as other authors do for the case of joint testing UC and IND. Therefore, and for all test statistics except (4.3.9), we apply the asymptotic distributions in order to find critical values, conscious of the limitations in the small sample cases. Our tests are based on an exact distribution. Tables 4.1, 4.2, 4.3 and 4.4, present results for the empirical power.

Within our class, the test with $\xi = 0.5$ is most powerful in a large number of cases and in the other cases is only a little less powerful than the best member of our class.

With the exception of very small sample sizes ($p = 0.01$ and $T = 250$) and with the exception of the model (4.4.1) with $w_s = 250$ where our test is a little less powerful than CAViaR, for all other cases, our test is more powerful. The differences in power are, in many cases, quite considerable. Compared with the Markov independence test, the rejection frequency of our test is, for most cases, more than double, and in some cases almost three times the rejection frequency of the Markov test. In comparison with the GMM tests, our test performs better for all cases, with a power that is sometimes twice as large as that of the GMM tests. The GMM tests perform quite well at larger sample sizes ($p = 0.05$ and $T = 1000$) but poorly at small sample sizes. These results contrast with the good results achieved when jointly testing the UC and IND hypotheses. The CAViaR test has more power with very small sample sizes ($p = 0.01$ and $T = 250$) and performs a little better in the case of the model (4.4.1) with $w_s = 250$, but in all other cases our test performs better.

In order to study the empirical type I error rates, we simulate iid Bernoulli samples with $p = 0.01$, $p = 0.05$ and $T = 250, 500, 750, 1000$. Rejection frequencies under the null are calculated over 5,000 replications. In the CAViaR test we generate the VaR regressors with a GARCH model that are independent of the Bernoulli samples, using $w_s = 250$ (CAViaR$_{250}$) and $w_s = 500$ (CAViaR$_{500}$). Table 4.5 presents the results. The Markov and CAViaR tests are undersized for small sample sizes and oversized for large sample sizes. The GMM tests are extremely undersized for small samples. These results confirms that the asymptotic critical values are misleading. For our test, the results confirm the level property, with all empirical type I error rates lower than 0.1.

Table 4.1

Empirical power under IND hypothesis. $\alpha = 0.1$, $w_s = 250$ and Gaussian GARCH(1,1).

| | $p = 0.01$ | | | | $p = 0.05$ | | | |
|---|---|---|---|---|---|---|---|---|
| | T=250 | T=500 | T=750 | T=1000 | T=250 | T=500 | T=750 | T=1000 |
| $T_{N,[0.3N]}$ | 0.181 | 0.41 | 0.571 | 0.673 | 0.288 | 0.526 | 0.667 | 0.733 |
| $T_{N,[0.4N]}$ | 0.24 | 0.479 | 0.596 | 0.673 | 0.354 | 0.576 | 0.693 | 0.758 |
| $T_{N,[0.5N]}$ | 0.28 | 0.497 | 0.564 | 0.621 | 0.377 | 0.579 | 0.694 | 0.75 |
| $T_{N,[0.6N]}$ | 0.278 | 0.466 | 0.516 | 0.547 | 0.377 | 0.561 | 0.661 | 0.713 |
| $T_{N,[0.7N]}$ | 0.287 | 0.427 | 0.442 | 0.445 | 0.362 | 0.521 | 0.608 | 0.648 |
| $\text{M}_{\text{IND}}$ | 0.108 | 0.152 | 0.179 | 0.199 | 0.151 | 0.261 | 0.32 | 0.364 |
| CAViaR | 0.385 | 0.506 | 0.591 | 0.678 | 0.43 | 0.608 | 0.707 | 0.778 |
| $\text{J}_{\text{IND}(3)}$ | 0.083 | 0.202 | 0.284 | 0.365 | 0.213 | 0.465 | 0.641 | 0.749 |
| $\text{J}_{\text{IND}(5)}$ | 0.066 | 0.177 | 0.253 | 0.348 | 0.163 | 0.39 | 0.563 | 0.676 |
| FES | 0.161 | 0.003 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 |

Table 4.2

Empirical power under IND hypothesis. $\alpha = 0.1$, $w_s = 250$ and Skewed t APARCH(1,1).

| | $p = 0.01$ | | | | $p = 0.05$ | | | |
|---|---|---|---|---|---|---|---|---|
| | T=250 | T=500 | T=750 | T=1000 | T=250 | T=500 | T=750 | T=1000 |
| $T_{N,[0.3N]}$ | 0.245 | 0.533 | 0.738 | 0.857 | 0.403 | 0.775 | 0.909 | 0.962 |
| $T_{N,[0.4N]}$ | 0.317 | 0.613 | 0.781 | 0.871 | 0.487 | 0.815 | 0.926 | 0.972 |
| $T_{N,[0.5N]}$ | 0.365 | 0.646 | 0.772 | 0.852 | 0.529 | 0.829 | 0.929 | 0.974 |
| $T_{N,[0.6N]}$ | 0.371 | 0.635 | 0.748 | 0.817 | 0.346 | 0.819 | 0.922 | 0.962 |
| $T_{N,[0.7N]}$ | 0.386 | 0.604 | 0.702 | 0.745 | 0.519 | 0.792 | 0.896 | 0.945 |
| $\text{M}_{\text{IND}}$ | 0.134 | 0.204 | 0.254 | 0.307 | 0.214 | 0.395 | 0.512 | 0.597 |
| CAViaR | 0.462 | 0.562 | 0.636 | 0.712 | 0.526 | 0.715 | 0.809 | 0.876 |
| $\text{J}_{\text{IND}(3)}$ | 0.196 | 0.403 | 0.561 | 0.697 | 0.412 | 0.771 | 0.918 | 0.972 |
| $\text{J}_{\text{IND}(5)}$ | 0.152 | 0.36 | 0.531 | 0.666 | 0.342 | 0.722 | 0.882 | 0.954 |
| FES | 0.223 | 0.011 | 0.000 | 0.000 | 0.012 | 0.000 | 0.000 | 0.000 |

Table 4.3

Empirical power under IND hypothesis. $\alpha = 0.1$, $w_s = 500$ and Gaussian GARCH(1,1).

| | $p = 0.01$ | | | | $p = 0.05$ | | | |
|---|---|---|---|---|---|---|---|---|
| | T=250 | T=500 | T=750 | T=1000 | T=250 | T=500 | T=750 | T=1000 |
| $T_{N,[0.3N]}$ | 0.203 | 0.379 | 0.551 | 0.686 | 0.273 | 0.512 | 0.656 | 0.743 |
| $T_{N,[0.4N]}$ | 0.263 | 0.442 | 0.59 | 0.679 | 0.342 | 0.575 | 0.69 | 0.767 |
| $T_{N,[0.5N]}$ | 0.286 | 0.448 | 0.56 | 0.626 | 0.374 | 0.574 | 0.685 | 0.757 |
| $T_{N,[0.6N]}$ | 0.286 | 0.431 | 0.515 | 0.567 | 0.373 | 0.54 | 0.655 | 0.716 |
| $T_{N,[0.7N]}$ | 0.291 | 0.389 | 0.442 | 0.462 | 0.353 | 0.491 | 0.593 | 0.644 |
| $\mathrm{M_{IND}}$ | 0.111 | 0.175 | 0.202 | 0.213 | 0.151 | 0.254 | 0.315 | 0.367 |
| CAViaR | 0.291 | 0.411 | 0.502 | 0.577 | 0.341 | 0.485 | 0.591 | 0.665 |
| $\mathrm{J_{IND(3)}}$ | 0.094 | 0.185 | 0.281 | 0.366 | 0.204 | 0.447 | 0.629 | 0.754 |
| $\mathrm{J_{IND(5)}}$ | 0.079 | 0.166 | 0.266 | 0.36 | 0.157 | 0.377 | 0.552 | 0.684 |
| FES | 0.291 | 0.047 | 0.003 | 0.000 | 0.002 | 0.000 | 0.000 | 0.000 |

Table 4.4

Empirical power under IND hypothesis. $\alpha = 0.1$, $w_s = 500$ and Skewed t APARCH(1,1).

| | $p = 0.01$ | | | | $p = 0.05$ | | | |
|---|---|---|---|---|---|---|---|---|
| | T=250 | T=500 | T=750 | T=1000 | T=250 | T=500 | T=750 | T=1000 |
| $T_{N,[0.3N]}$ | 0.277 | 0.493 | 0.698 | 0.831 | 0.374 | 0.76 | 0.907 | 0.959 |
| $T_{N,[0.4N]}$ | 0.348 | 0.566 | 0.752 | 0.854 | 0.451 | 0.804 | 0.925 | 0.968 |
| $T_{N,[0.5N]}$ | 0.385 | 0.602 | 0.755 | 0.842 | 0.495 | 0.813 | 0.927 | 0.969 |
| $T_{N,[0.6N]}$ | 0.389 | 0.602 | 0.738 | 0.816 | 0.506 | 0.803 | 0.92 | 0.959 |
| $T_{N,[0.7N]}$ | 0.399 | 0.575 | 0.696 | 0.759 | 0.496 | 0.781 | 0.894 | 0.945 |
| $\mathrm{M_{IND}}$ | 0.152 | 0.214 | 0.289 | 0.342 | 0.208 | 0.383 | 0.526 | 0.63 |
| CAViaR | 0.406 | 0.516 | 0.596 | 0.677 | 0.425 | 0.624 | 0.743 | 0.822 |
| $\mathrm{J_{IND(3)}}$ | 0.217 | 0.393 | 0.556 | 0.68 | 0.367 | 0.747 | 0.912 | 0.968 |
| $\mathrm{J_{IND(5)}}$ | 0.176 | 0.357 | 0.524 | 0.666 | 0.31 | 0.688 | 0.872 | 0.954 |
| FES | 0.375 | 0.091 | 0.012 | 0.001 | 0.045 | 0.001 | 0.000 | 0.000 |

Notes to Tables 4.1,4.2,4.3,4.4: The results are based on 5000 replications. For each sample size ($T$), rolling window size ($w_s$) and coverage rate ($p$) we provide percentage of rejection at a 10% significance level. $T_{N,[\xi N]}$ denotes the test from the class $\tau_k$ with $k = [\xi N]$. $\mathrm{M_{IND}}$ denotes the Markov independence test (4.2.1). CAViaR denotes the independence test (6.4.1) and $\mathrm{J_{IND}}(k)$ denotes the GMM independence test (4.2.9) with k moment conditions. FES denotes frequency of excluded samples.

Table 4.5
Empirical type I error rates with $\alpha = 0.1$.

| | $p = 0.01$ | | | | $p = 0.05$ | | | |
|---|---|---|---|---|---|---|---|---|
| | T=250 | T=500 | T=750 | T=1000 | T=250 | T=500 | T=750 | T=1000 |
| $T_{N,[0.3N]}$ | 0.073 | 0.079 | 0.081 | 0.088 | 0.059 | 0.065 | 0.067 | 0.064 |
| $T_{N,[0.4N]}$ | 0.080 | 0.088 | 0.093 | 0.095 | 0.070 | 0.066 | 0.076 | 0.073 |
| $T_{N,[0.5N]}$ | 0.088 | 0.092 | 0.095 | 0.093 | 0.077 | 0.078 | 0.084 | 0.081 |
| $T_{N,[0.6N]}$ | 0.088 | 0.092 | 0.095 | 0.092 | 0.078 | 0.083 | 0.09 | 0.083 |
| $T_{N,[0.7N]}$ | 0.095 | 0.089 | 0.098 | 0.090 | 0.082 | 0.088 | 0.093 | 0.087 |
| $\mathrm{M_{IND}}$ | 0.023 | 0.029 | 0.039 | 0.037 | 0.054 | 0.111 | 0.158 | 0.134 |
| $\mathrm{CAViaR_{250}}$ | 0.071 | 0.061 | 0.058 | 0.064 | 0.081 | 0.101 | 0.120 | 0.126 |
| $\mathrm{CAViaR_{500}}$ | 0.080 | 0.056 | 0.066 | 0.057 | 0.083 | 0.099 | 0.130 | 0.124 |
| $\mathrm{J_{IND(3)}}$ | 0.003 | 0.006 | 0.011 | 0.015 | 0.018 | 0.032 | 0.045 | 0.045 |
| $\mathrm{J_{IND(5)}}$ | 0.001 | 0.004 | 0.007 | 0.012 | 0.012 | 0.021 | 0.028 | 0.033 |
| FES | 0.292 | 0.038 | 0.004 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

Notes: We simulate iid Bernoulli samples to study the empirical type I error rates. The results are based on 5000 replications. For each sample size ($T$) and coverage rate ($p$) we provide percentage of rejection at a 10% significance level. $T_{N,[\xi N]}$ denotes the test from the class $\tau_k$ with $k = [\xi N]$. $\mathrm{M_{IND}}$ denotes the Markov independence test (4.2.1). CAViaR denotes the independence test (6.4.1) and $\mathrm{J_{IND}}(k)$ denotes the GMM independence test (4.2.9) with k moment conditions. FES denotes frequency of excluded samples.

### 4.4.2 Simulation study under the CC hypothesis

Considering the statistical problem of testing independence of the hit sequence, the main theme of this work, the relevant comparison is the one we presented in Subsection 4.4.1, assuming only IND for the null hypothesis. However, when we apply the tests for backtesting 1% and 5% VaR, we know what $p$ should be, allowing us to use finite sample critical values obtained under the CC hypothesis. Although the independence tests are not designed for testing CC but only IND, it is possible to study their behavior under CC and we did that in this Subsection. The difference between this Subsection and 4.4.1 is that here we apply finite sample critical values assuming $p = 0.01$ and $p = 0.05$. Under this context, our test shows a better performance in many cases but now the differences in the empirical power between our test and the other tests is much smaller (Tables 4.6,4.7,4.8 and 4.9).

Table 4.6
Empirical power under CC hypothesis. $\alpha = 0.1$, $w_s = 250$ and Gaussian GARCH(1,1).

| | $p = 0.01$ | | | | $p = 0.05$ | | | |
|---|---|---|---|---|---|---|---|---|
| | T=250 | T=500 | T=750 | T=1000 | T=250 | T=500 | T=750 | T=1000 |
| $T_{N,[0.3N]}$ | 0.282 | 0.567 | 0.758 | 0.867 | 0.509 | 0.829 | 0.935 | 0.975 |
| $T_{N,[0.4N]}$ | 0.337 | 0.63 | 0.79 | 0.875 | 0.556 | 0.849 | 0.943 | 0.979 |
| $T_{N,[0.5N]}$ | 0.379 | 0.653 | 0.779 | 0.856 | 0.572 | 0.849 | 0.942 | 0.976 |
| $T_{N,[0.6N]}$ | 0.384 | 0.642 | 0.753 | 0.82 | 0.566 | 0.836 | 0.93 | 0.968 |
| $T_{N,[0.7N]}$ | 0.396 | 0.609 | 0.705 | 0.749 | 0.541 | 0.807 | 0.905 | 0.951 |
| $\mathrm{M_{IND}}$ | 0.461 | 0.532 | 0.627 | 0.644 | 0.263 | 0.375 | 0.447 | 0.541 |
| CAViaR | 0.524 | 0.636 | 0.709 | 0.785 | 0.542 | 0.711 | 0.788 | 0.856 |
| $\mathrm{J_{IND(3)}}$ | 0.442 | 0.626 | 0.743 | 0.826 | 0.567 | 0.848 | 0.951 | 0.986 |
| $\mathrm{J_{IND(5)}}$ | 0.43 | 0.635 | 0.765 | 0.859 | 0.557 | 0.842 | 0.943 | 0.985 |
| FES | 0.223 | 0.011 | 0.000 | 0.000 | 0.012 | 0.000 | 0.000 | 0.000 |

Table 4.7
Empirical power under CC hypothesis. $\alpha = 0.1$, $w_s = 250$ and Skewed t APARCH(1,1).

| | $p = 0.01$ | | | | $p = 0.05$ | | | |
|---|---|---|---|---|---|---|---|---|
| | T=250 | T=500 | T=750 | T=1000 | T=250 | T=500 | T=750 | T=1000 |
| $T_{N,[0.3N]}$ | 0.212 | 0.443 | 0.591 | 0.69 | 0.383 | 0.613 | 0.739 | 0.796 |
| $T_{N,[0.4N]}$ | 0.259 | 0.496 | 0.608 | 0.687 | 0.414 | 0.636 | 0.748 | 0.802 |
| $T_{N,[0.5N]}$ | 0.293 | 0.504 | 0.572 | 0.63 | 0.42 | 0.623 | 0.734 | 0.783 |
| $T_{N,[0.6N]}$ | 0.288 | 0.471 | 0.522 | 0.552 | 0.41 | 0.593 | 0.693 | 0.742 |
| $T_{N,[0.7N]}$ | 0.293 | 0.43 | 0.445 | 0.451 | 0.389 | 0.542 | 0.632 | 0.672 |
| $\mathrm{M_{IND}}$ | 0.329 | 0.409 | 0.499 | 0.53 | 0.194 | 0.24 | 0.256 | 0.298 |
| CAViaR | 0.461 | 0.592 | 0.681 | 0.762 | 0.448 | 0.601 | 0.672 | 0.751 |
| $\mathrm{J_{IND(3)}}$ | 0.294 | 0.415 | 0.488 | 0.564 | 0.366 | 0.598 | 0.742 | 0.823 |
| $\mathrm{J_{IND(5)}}$ | 0.295 | 0.441 | 0.536 | 0.618 | 0.375 | 0.594 | 0.734 | 0.815 |
| FES | 0.161 | 0.003 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 |

Table 4.8
Empirical power under CC hypothesis. $\alpha = 0.1$, $w_s = 500$ and Gaussian GARCH(1,1).

| | $p = 0.01$ | | | | $p = 0.05$ | | | |
|---|---|---|---|---|---|---|---|---|
| | T=250 | T=500 | T=750 | T=1000 | T=250 | T=500 | T=750 | T=1000 |
| $T_{N,[0.3N]}$ | 0.238 | 0.412 | 0.576 | 0.703 | 0.295 | 0.529 | 0.678 | 0.756 |
| $T_{N,[0.4N]}$ | 0.282 | 0.459 | 0.603 | 0.687 | 0.355 | 0.586 | 0.703 | 0.774 |
| $T_{N,[0.5N]}$ | 0.302 | 0.456 | 0.569 | 0.632 | 0.382 | 0.584 | 0.694 | 0.763 |
| $T_{N,[0.6N]}$ | 0.301 | 0.437 | 0.523 | 0.573 | 0.378 | 0.545 | 0.661 | 0.723 |
| $T_{N,[0.7N]}$ | 0.299 | 0.395 | 0.447 | 0.47 | 0.358 | 0.496 | 0.596 | 0.648 |
| $\mathrm{M_{IND}}$ | 0.322 | 0.353 | 0.408 | 0.474 | 0.189 | 0.232 | 0.25 | 0.312 |
| CAViaR | 0.361 | 0.477 | 0.576 | 0.662 | 0.362 | 0.479 | 0.558 | 0.625 |
| $\mathrm{J_{IND(3)}}$ | 0.293 | 0.391 | 0.479 | 0.554 | 0.352 | 0.588 | 0.734 | 0.832 |
| $\mathrm{J_{IND(5)}}$ | 0.295 | 0.42 | 0.519 | 0.612 | 0.355 | 0.593 | 0.726 | 0.827 |
| FES | 0.291 | 0.047 | 0.003 | 0.000 | 0.002 | 0.000 | 0.000 | 0.000 |

Table 4.9
Empirical power under CC hypothesis. $\alpha = 0.1$, $w_s = 500$ and Skewed t APARCH(1,1).

| | $p = 0.01$ | | | | $p = 0.05$ | | | |
|---|---|---|---|---|---|---|---|---|
| | T=250 | T=500 | T=750 | T=1000 | T=250 | T=500 | T=750 | T=1000 |
| $T_{N,[0.3N]}$ | 0.311 | 0.524 | 0.726 | 0.847 | 0.393 | 0.771 | 0.913 | 0.962 |
| $T_{N,[0.4N]}$ | 0.371 | 0.584 | 0.766 | 0.86 | 0.466 | 0.812 | 0.928 | 0.971 |
| $T_{N,[0.5N]}$ | 0.398 | 0.612 | 0.762 | 0.846 | 0.504 | 0.818 | 0.931 | 0.97 |
| $T_{N,[0.6N]}$ | 0.4 | 0.611 | 0.743 | 0.818 | 0.512 | 0.808 | 0.922 | 0.961 |
| $T_{N,[0.7N]}$ | 0.406 | 0.582 | 0.699 | 0.762 | 0.502 | 0.785 | 0.897 | 0.946 |
| $\mathrm{M_{IND}}$ | 0.459 | 0.465 | 0.535 | 0.586 | 0.254 | 0.364 | 0.457 | 0.571 |
| CAViaR | 0.468 | 0.571 | 0.657 | 0.744 | 0.442 | 0.621 | 0.723 | 0.799 |
| $\mathrm{J_{IND(3)}}$ | 0.45 | 0.604 | 0.715 | 0.802 | 0.519 | 0.824 | 0.943 | 0.982 |
| $\mathrm{J_{IND(5)}}$ | 0.437 | 0.618 | 0.742 | 0.834 | 0.512 | 0.823 | 0.937 | 0.978 |
| FES | 0.375 | 0.091 | 0.012 | 0.001 | 0.045 | 0.001 | 0.000 | 0.000 |

Notes to Tables 4.6,4.7,4.8,4.9: The results are based on 5000 replications. For each sample size ($T$), rolling window size ($w_s$) and coverage rate ($p$) we provide percentage of rejection at a 10% significance level with finite sample critical values obtained under the CC hypothesis. $T_{N,[\xi N]}$ denotes the test from the class $\tau_k$ with $k = [\xi N]$. $\mathrm{M_{IND}}$ denotes the Markov independence test (4.2.1). CAViaR denotes the independence test (6.4.1) and $\mathrm{J_{IND}}(k)$ denotes the GMM independence test (4.2.9) with k moment conditions. FES denotes frequency of excluded samples.

## 4.5 An Application to the DAX Index

The recent 2008 financial crisis illustrates quite well the importance of explicitly testing the IND property. We apply the popular Historical Simulation (HS) VaR based on the previous 250 trading days to the DAX index. In Figure 4.5.1 we plot the returns and one-day-ahead 1% VaR. In Figure 4.5.2 we plot the hit sequence. We observe a first cluster of 3 violations within 13 days (between January 21 and February 6) and then an impressive cluster of five violations occurring with very short durations in only 13 consecutive trading days (between September 29 and October 15). During this short period of 13 trading days, the index lost almost 20%. With this flagrant pattern of clustering, the backtesting result from the recent regulatory framework, based on a traffic light approach which ignores the IND property, only classify the model as inaccurate with the last violation of the year (see for example Campbell (2007) for details on the traffic light approach).

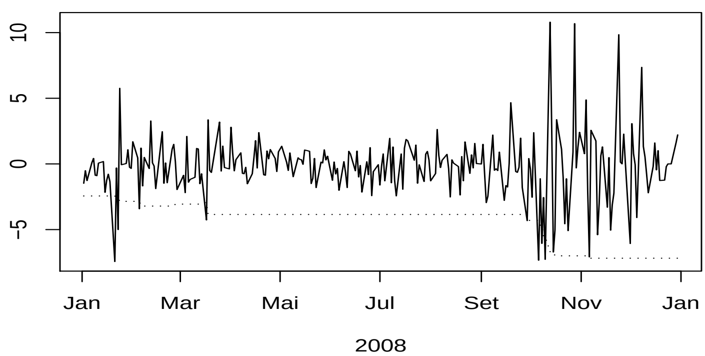



Figure 4.5.1: DAX index returns (solid line) and HS VaR(0.01) forecasts (dotted line)

On the date of the second violation of the five violations cluster, our exact independence test rejects the IND hypothesis with 0.05 significance level. At this date we have a sample of durations of size 6, with $d_{1:6} = 2$, $d_{2:6} = 5$, $d_{3:6} = 9$, $d_{4:6} = 13$, $d_{5:6} = 28$ and $d_{6:6} = 137$ days. The observed value of our test statistic (4.3.9) is $t_{6,3} = \log 2(137-1)/9 - \log 6 = 8.76$. Consulting the Table for $T^E_{N,[N/2]}$ quantiles in the Section 4.7, with $N = 6$ and 0.05 significance level, we get a critical region equal to $[8.00; +\infty[$. The observed value belongs to this region and we reject the IND hypothesis. On this date, the p value obtained with the exact binomial test is 0.108 and the UC hypothesis is not rejected with the usual significance levels. The results show that ignoring

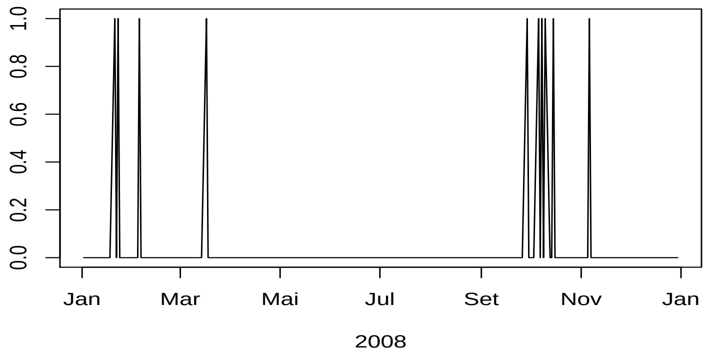



Figure 4.5.2: Hit sequence for the DAX example

the independence property was an important reason for the failure of the traffic light approach during the crisis period. On the date of the third violation, the UC is rejected with the binomial test. With a 0.05 significance level, the IND hypothesis is rejected on the date of the fourth violation, using $J_{IND}(3)$, and on the last day of the five violations cluster, using both $J_{IND}(3)$ and CAViaAR. During the crisis period, our test rejected the independence hypothesis before the other independence tests under study.

## 4.6 Conclusion

In this work we propose a class of independence tests based on an exact distribution that does not depend on an unknown parameter. These tests can be used to test the independence of any hit sequence. In order to test the independence in the context of interval forecasts evaluation, we show that the suggested independence tests perform better than the other tests under study. Although we are usually interested both in correct coverage and independence, specific tests for UC and IND are also important, since they may help to learn about the reasons for the failure of and potential ways to improve actual models used in risk management applications.

## Acknowledgements

We thank Markus Haas, Marc Paolella and other participants at the 3rd International Conference on Computational and Financial Econometrics and 2nd Workshop of the ERCIM Working Group on Computing and Statistics, at Limassol, Cyprus, October

2009, for useful suggestions and comments. Research partially supported by Fundação para a Ciência e Tecnologia (FCT/PROTEC and FCT/POCI 2010 project) and Center of Statistics and Applications of University of Lisbon (CEAUL). Finally, the authors would like to thank the Associate Editor and two referees for their comments and suggestions which lead to substantial improvements of an earlier version of this article.

## 4.7 Table for the $T^E_{N,[N/2]}$ Quantiles

$$P(T^E \geq t_{\epsilon(N)}) = \gamma_N \sum_{v=0}^{N-[N/2]-1} \sum_{s=0}^{[N/2]-1} (-1)^{v+s} a_v b_s [c_{v,s}(v+1)]^{-1} - \Big[c_{v,s}\Big(v+1+c_{v,s}\Big(\frac{-\log(1/2)}{t_{\epsilon(N)}+\log \mathrm{N}}\Big)\Big)\Big]^{-1} = \epsilon$$

| $\epsilon$ | 0.10 | 0.05 | 0.01 | $\epsilon$ | 0.10 | 0.05 | 0.01 |
|---|---|---|---|---|---|---|---|
| $N$ | $t_{0.1(N)}$ | $t_{0.05(N)}$ | $t_{0.01(N)}$ | $N$ | $t_{0.1(N)}$ | $t_{0.05(N)}$ | $t_{0.01(N)}$ |
| 2 | 12.46 | 26.23 | 137.17 | 53 | 3.08 | 4.06 | 6.34 |
| 3 | 28.23 | 59.26 | 308.43 | 54 | 2.87 | 3.81 | 6.01 |
| 4 | 6.72 | 10.90 | 28.50 | 55 | 3.06 | 4.02 | 6.27 |
| 5 | 10.54 | 16.81 | 43.20 | 56 | 2.85 | 3.79 | 5.98 |
| 6 | 5.33 | 8.00 | 17.13 | 57 | 3.04 | 3.99 | 6.25 |
| 7 | 7.40 | 10.90 | 22.82 | 58 | 2.84 | 3.77 | 5.92 |
| 8 | 4.70 | 6.81 | 13.29 | 59 | 3.01 | 3.97 | 6.19 |
| 9 | 6.10 | 8.65 | 16.52 | 60 | 2.82 | 3.75 | 5.89 |
| 10 | 4.30 | 6.11 | 11.30 | 61 | 3.00 | 3.94 | 6.15 |
| 11 | 5.37 | 7.47 | 13.59 | 62 | 2.81 | 3.73 | 5.85 |
| 12 | 4.04 | 5.65 | 10.12 | 63 | 2.98 | 3.92 | 6.10 |
| 13 | 4.90 | 6.74 | 11.81 | 64 | 2.80 | 3.72 | 5.82 |
| 14 | 3.84 | 5.33 | 9.33 | 65 | 2.96 | 3.89 | 6.06 |
| 15 | 4.58 | 6.23 | 10.70 | 66 | 2.79 | 3.69 | 5.79 |
| 16 | 3.70 | 5.09 | 8.74 | 67 | 2.95 | 3.87 | 6.03 |
| 17 | 4.34 | 5.87 | 9.88 | 68 | 2.77 | 3.68 | 5.77 |
| 18 | 3.59 | 4.90 | 8.30 | 69 | 2.93 | 3.85 | 6.00 |
| 19 | 4.14 | 5.58 | 9.30 | 70 | 2.77 | 3.66 | 5.74 |
| 20 | 3.50 | 4.76 | 7.96 | 71 | 2.92 | 3.84 | 5.96 |
| 21 | 3.98 | 5.35 | 8.79 | 72 | 2.76 | 3.65 | 5.72 |
| 22 | 3.41 | 4.62 | 7.66 | 73 | 2.90 | 3.81 | 5.91 |
| 23 | 3.86 | 5.17 | 8.42 | 74 | 2.75 | 3.64 | 5.70 |
| 24 | 3.34 | 4.52 | 7.45 | 75 | 2.88 | 3.79 | 5.88 |
| 25 | 3.75 | 5.00 | 8.11 | 76 | 2.74 | 3.62 | 5.66 |
| 26 | 3.28 | 4.42 | 7.24 | 77 | 2.88 | 3.78 | 5.87 |
| 27 | 3.66 | 4.88 | 7.86 | 78 | 2.73 | 3.61 | 5.66 |
| 28 | 3.23 | 4.35 | 7.06 | 79 | 2.87 | 3.76 | 5.83 |
| 29 | 3.58 | 4.76 | 7.63 | 80 | 2.72 | 3.59 | 5.62 |
| 30 | 3.18 | 4.29 | 6.93 | 81 | 2.85 | 3.74 | 5.79 |
| 31 | 3.51 | 4.66 | 7.45 | 82 | 2.71 | 3.59 | 5.60 |
| 32 | 3.14 | 4.21 | 6.79 | 83 | 2.84 | 3.73 | 5.77 |
| 33 | 3.46 | 4.58 | 7.31 | 84 | 2.71 | 3.58 | 5.59 |
| 34 | 3.11 | 4.16 | 6.70 | 85 | 2.84 | 3.72 | 5.75 |
| 35 | 3.40 | 4.50 | 7.14 | 86 | 2.70 | 3.57 | 5.56 |
| 36 | 3.07 | 4.11 | 6.57 | 87 | 2.82 | 3.70 | 5.73 |
| 37 | 3.35 | 4.43 | 7.02 | 88 | 2.69 | 3.56 | 5.54 |
| 38 | 3.04 | 4.06 | 6.49 | 89 | 2.80 | 3.68 | 5.71 |
| 39 | 3.31 | 4.37 | 6.90 | 90 | 2.68 | 3.55 | 5.52 |
| 40 | 3.02 | 4.03 | 6.43 | 91 | 2.80 | 3.68 | 5.69 |
| 41 | 3.27 | 4.31 | 6.80 | 92 | 2.68 | 3.54 | 5.51 |
| 42 | 2.99 | 3.98 | 6.33 | 93 | 2.79 | 3.66 | 5.67 |
| 43 | 3.23 | 4.25 | 6.71 | 94 | 2.67 | 3.52 | 5.49 |
| 44 | 2.96 | 3.95 | 6.27 | 95 | 2.78 | 3.65 | 5.65 |
| 45 | 3.20 | 4.22 | 6.64 | 96 | 2.67 | 3.51 | 5.47 |
| 46 | 2.94 | 3.91 | 6.22 | 97 | 2.78 | 3.64 | 5.63 |
| 47 | 3.16 | 4.16 | 6.53 | 98 | 2.66 | 3.50 | 5.45 |
| 48 | 2.93 | 3.89 | 6.17 | 99 | 2.77 | 3.63 | 5.61 |
| 49 | 3.14 | 4.13 | 6.46 | 100 | 2.65 | 3.50 | 5.45 |
| 50 | 2.91 | 3.87 | 6.10 | 200 | 2.49 | 3.28 | 5.08 |
| 51 | 3.11 | 4.09 | 6.40 | 1000 | 2.32 | 3.05 | 4.74 |
| 52 | 2.88 | 3.84 | 6.06 | $\infty$ | 2.25 | 2.97 | 4.60 |

*Paulo Araújo Santos and Isabel Fraga Alves*

# 5

# Improved Estimation in the Discrete Weibull Distribution

A new shape parameter estimator for the discrete Weibull distribution, under complete data or under type I censored data, is proposed. This estimator is based on an extension of the Khan, Khalique and Abouammoh (1989) method of proportions. The cumulative distribution function and the moments for the proportions estimator and for a more general class, are derived. Simulations are also carried out to illustrate the substantial improvement achieved in terms of bias and mean square error compared with other estimators. The proposed estimator is applied on a financial dataset dealing with durations between violations in a quantitative risk management environment.

## 5.1 Introduction

In many applications for a wide spectrum of fields, statistical inference models the observed data as a sample from a continuous probability model, implying that the observed data are precisely measured. For example, the time between trades is recorded to the nearest second. As common practice, the actual data available for inference are discrete, either because they are rounded, conditional to the precision of the measuring device, or because the data are themselves discrete; a good example, concerning discrete economic data, is the time periods until the event of interest regarded as countable instead of continuous. In Grimshawa et al. (2005) a study is motivated by the common practice of testing for duration dependence in economic and financial data using the continuous Weibull distribution when the data are really discrete. In many reliability studies, data is measured as discrete random variables such as the number of copies made by a copying machine, number of cycles of a washing machine and so on. Materials, equipment, devices and structures are also frequently monitored only once per period rather than continuously, due to practical restrictions. In these types of reliability studies, the discrete Weibull model plays an important role. For a survey on discrete lifetime distributions see Bracquemond and Gaudoin (2003). Moreover, the discrete Weibull can be applied to other problems, from political renewal analysis (Lin and Guillén 1998) to economic problems involving duration dependence. The "damaged goods" theory implies that the longer the period of unemployment, the more likely the job seeker has some attribute that makes her unemployable, thus less likely to find a job. A discrete Weibull model with a shape parameter lower than one supports the "damaged goods" theory; see Lancaster (1979) for econometric methods for the duration of unemployment. Other examples of duration dependence application are the study of speculative bubbles in stock markets (Harman and Zuehlke 2004) and backtesting Value-at-Risk (Haas 2005). In this work, we propose an improved estimator for the shape parameter of the discrete Weibull version of Nakagawa and Osaki (1975), also known as type I discrete Weibull, with the followings cumulative distribution function (cdf) and probability mass function (pmf):

$$F_D(d) = \begin{cases} 1 - q^{d^\theta}, & d = 1, 2, 3, ...\text{(jump points)} \\ 0, & x < 1 \end{cases} \tag{5.1.1}$$

$$f_D(d) = q^{(d-1)^\theta} - q^{d^\theta}, \quad d = 1, 2, 3, ... \tag{5.1.2}$$

for $1 < q < 0$ and $\theta > 0$. Here $\theta$ is the shape parameter and $q$ is the probability that the *duration* $D$ is greater than one, i.e., $q = P[D > 1]$. Returning to the unemployment example, if unemployment spells have $\theta > 1$, the duration dependence supports the "reservation wage" theory. However, if unemployment spells have $\theta < 1$, the duration dependence supports the "damaged goods" theory; applying lifetime studies terminology, the distribution has increasing failure rate for $\theta > 1$, decreasing failure rate for $0 < \theta < 1$ and reduces to the geometric distribution when $\theta = 1$. If $W$ is a continuous Weibull rv, then a type I discrete Weibull rv can be derived by time discretization $D = [W] + 1$, where $[W]$ denotes the integer part of $W$. Stein and Dattero (1984) introduced a type II discrete Weibull and a type III was proposed by Padgett and Spurrier (1985). Type II has a serious limitation because the support is bounded. The estimation of parameters is difficult in type III. In a detailed study, Bracquemond and Gaudoin (2003), recommended the use of type I discrete Weibull. The rest of the paper is organized as follows. Section 5.2 provides a brief review of estimation methods. In Section 5.3, the cdf and the moments for the proportions estimator and for a more general class, are derived. Based on the study of this class, a new shape parameter estimator is proposed. In Section 5.4, and through simulation experiments, we compare the performance of the new estimator with the method of moments and with the method of proportions. Finally, Section 5.5 presents an empirical application from a quantitative risk management context.

## 5.2 Estimation methods

From the cdf (5.1.1), we have

$$\log[-\log(1 - F_D(d))] = \theta \log d + \log(-\log q).$$

Let $d^*_{1:v} < ... < d^*_{v:v}$, $v \leq n$, be the observed order statistics (o.s.'s) without ties of a sample $d_1, ..., d_n$ from the type I discrete Weibull distribution and $F_n(d) = n^{-1} \sum_{i=1}^{n} I_{\{d_i \leq d\}}$ the empirical cumulative distribution function (ecdf), associated to a random sample $D_1, \cdots, D_n$. If the points

$$\left\{ \log d^*_{i:v}, \log(-\log(1 - F_n(d^*_{i:v})) \right\}_{1 \leq i \leq v}$$

are approximately scattered around a straight line, it can be assumed that the underlying model is (5.1.1) and the parameters estimated by the Probability Plotting method, using these points.

Taking into account (5.1.2), we obtain the first two moments

$$\mu_1 = E[D] = \sum_{d=0}^{\infty} (d+1) q^{d^\theta} - \sum_{d=1}^{\infty} d q^{d^\theta} = 1 + \sum_{d=1}^{\infty} q^{d^\theta}$$

$$\mu_2 = E[D^2] = 2 \sum_{d=1}^{\infty} d q^{d^\theta} + E[D].$$

However, closed forms for these moments are not available, as pointed out in Khan, Khalique and Abouammoh (1989). Consequently, for an observed sample $d_1, d_2, \cdots, d_n$, the Moments estimator, $\hat{\theta}^M_n$, is obtained by a numerical algorithm, which minimizes

$$M(q, \theta) = \left( \left( n^{-1} \sum_{i=1}^{n} d_i \right) - \mu_1 \right)^2 + \left( \left( n^{-1} \sum_{i=1}^{n} d_i^2 \right) - \mu_2 \right)^2. \tag{5.2.1}$$

The method of proportions was proposed by Khan, Khalique and Abouammoh (1989). Since $q = 1 - F_D(1)$, the idea is to use the empirical frequency of observations greater than one

$$\hat{q} = 1 - F_n(1). \tag{5.2.2}$$

In the same way, since $f_D(2) = q - q^{2^\theta}$ and using additionally, the empirical frequency of observations greater than two

$$\hat{\theta}^P_n := \frac{1}{\log 2} \log \frac{\log(1 - F_n(2))}{\log(1 - F_n(1))}. \tag{5.2.3}$$

Denoting $p_d = f_D(d)$, $d = 1, 2, ...$, these authors noted that one may think of using other relation

$$\log q = k^{-\theta} \log(1 - p_1 - ... - p_k), \;\; k = 1, 2, ...$$

In a simulation study it was concluded that the optimum choice of $k$ is 2 and this lead us to (5.2.3).

The method of maximum likelihood considers the log-likelihood function $\log L(q, \theta; d_1, ..., d_n) = \sum_{i=1}^{n} \log \left\{ q^{(d_i-1)^\theta} - q^{d_i^\theta} \right\}$; however, the ML-equations $\partial L/\partial q = 0$ and $\partial L/\partial \theta = 0$, must be solved numerically; in this case, computational problems can occur, despite the good quality of estimates for high values of $q$ (Bracquemond and Gaudoin, 2003). Based on the method of proportions, approximate maximum likelihood estimators were proposed by Kulasekera (1994), both for complete and type I censored data.

## 5.3 Improved shape parameter estimation

The main drawback of Khan, Khalique and Abouammoh (1989) proportions estimator (5.2.3) is that it does not use all the observations but only a few of them, loosing a significant part of the available information. Here, we overcome this limitation.

Using (5.1.1) we have $q^{d^\theta} = 1 - F_D(d)$ and

$$\theta = \frac{1}{\log d} \log \frac{\log(1 - F_D(d))}{\log q} = \frac{1}{\log d} \log \frac{\log(1 - F_D(d))}{\log q}. \tag{5.3.1}$$

Considering equations (5.3.1) for $d = 2, ..., k$, multiplied by constants $c_d$, it is possible to write the following system of equations

$$\left\{ \begin{array}{ccc} c_2\theta & = & \frac{c_2}{\log 2} \log \frac{\log(1-F_D(2))}{\log q} \\ ... & = & ... \\ c_k\theta & = & \frac{c_k}{\log k} \log \frac{\log(1-F_D(k))}{\log q} \end{array} \right. \tag{5.3.2}$$

Now, adding all the equations and solving in order to $\theta$

$$\theta = \sum_{d=2}^{k} \frac{c_d}{\log d} \log \frac{\log(1 - F_D(d))}{\log q} / \sum_{d=2}^{k} c_d. \tag{5.3.3}$$

A class of estimators for $\theta$ is motivated by (5.3.3) through its empirical counterpart. We suggest $c_d = 1$, $\forall_d$, and the estimation of $1 - F_D(d)$ and $q$ by the ecdf, in the same line as Khan, Khalique and Abouammoh, but considering one or more equations (5.3.1) to obtain the following class of estimators defined if $d_{1:n} = 1$ and $d_{n:n} > k$,

$$\hat{\theta}^k := \sum_{d=2}^{k} \frac{1}{\log d} \frac{\log(1 - F_n(d))}{\log(1 - F_n(1))} / (k-1), \tag{5.3.4}$$

where $k \in \{2, 3, ...\}$. The proportions estimator in (5.2.3), $\hat{\theta}_n^P$, is a particular case of estimator $\hat{\theta}^k$ in (5.3.4), for $k = 2$.

### 5.3.1 Theoretical results

In Proposition 5.3.1, for the class (5.3.4), we provide the probability of observing a sample such that the estimator is defined. In Theorem 5.3.1 we provide the cdf and the moments for the class (5.3.4) with $k = 2$. Theorem 5.3.2 generalizes these results for $k \geq 3$. Then, Remark 5.3.1 lead us to propose a new improved estimator for the shape parameter $\theta$. Finally we derive Theorems 5.3.3 and 5.3.4 which alow us to present, in the Corollary 5.3.1, theoretical expressions for the expected value and the variance of this new estimator.

**Proposition 5.3.1.** *Let $D_1, D_2, ..., D_n$ be i.i.d. discrete Weibull rv´s with common cdf (5.1.1). Then, for the class (5.3.4), the probability of observing a sample such that $\hat{\theta}^k$ (5.3.4) is defined, is given by*

$$1 - q^n - \left(1 - q^{k^\theta}\right)^n + \left(q - q^{k^\theta}\right)^n \tag{5.3.5}$$

**Proof:** The conditions for the estimator to be defined are $F_n(1) > 0$ and $F_n(k) < 1$. These conditions are equivalent to

$$D_{1:n} = 1 \quad \text{and} \quad D_{n:n} > k. \tag{5.3.6}$$

Now we compute the probability of observing a sample that satisfies these conditions,

$$\begin{aligned}
&P[D_{1:n} = 1 \wedge D_{n:n} > k] \\
&= 1 - P[D_{1:n} > 1 \vee D_{n:n} \leq k] \\
&= 1 - P[D_{1:n} > 1] - P[D_{n:n} \leq k] + P[D_{1:n} > 1 \wedge D_{n:n} \leq k] \\
&= 1 - P[D_1 > 1 \wedge \cdots \wedge D_n > 1] - P[D_1 \leq 1 \wedge \cdots \wedge D_n \leq k] \\
&\qquad\qquad + P[1 < D_1 \leq k \wedge \cdots \wedge 1 < D_n \leq k] \\
&= 1 - \left(1 - F_D(1)\right)^n - \left(F_D(k)\right)^n + \left(F_D(k) - F_D(1)\right)^n.
\end{aligned}$$

Using the cdf (5.1.1), the result follows.

□

**Theorem 5.3.1 (Cdf and Moments of the Proportions Estimator).** *Let $D_1, D_2, ..., D_n$ be i.i.d. discrete Weibull rv´s with common cdf (5.1.1) and $i_j \in \{1, ..., n\}$ with $j = 1, 2$. Then, conditionally to $D_{1:n} = 1$ and $D_{n:n} > 2$, for the class (5.3.4) with $k = 2$, we have*

*i) The cdf of $\hat{\theta}^k$ (5.3.4) given by*

$$F_{\hat{\theta}^k}(w) = \sum_{\forall i_1, i_2 : \frac{1}{\log 2} \log \frac{\log(1-i_1/n-i_2/n)}{\log(1-i_1/n)} \leq w} \frac{f_{X_1,X_2}(i_1, i_2)}{1 - q^n - (1 - q^{2^\theta})^n + (q - q^{2^\theta})^n}$$

*with*

$$f_{X_1,X_2}(i_1, i_2) = \binom{n}{i_1}\binom{n-i_1}{i_2} p_1^{i_1}(1-p_1)^{n-i_1}\left(\frac{p_2}{1-p_1}\right)^{i_2}\left(1 - \frac{p_2}{1-p_1}\right)^{n-i_1-i_2},$$

$p_1 = 1 - q$ *and* $p_2 = q - q^{2^\theta}$.

*ii) The moments $E\left[(\hat{\theta}^k)^l\right]$, with $l \in \mathbb{N}$, given by*

$$\sum_{\forall i_1, i_2 : 1 < i_1 + i_2 < n \wedge i_1 > 0} \left(\frac{1}{\log 2} \log \frac{\log(1 - i_1/n - i_2/n)}{\log(1 - i_1/n)}\right)^l \frac{f_{X_1,X_2}(i_1, i_2)}{1 - q^n - (1 - q^{2^\theta})^n + (q - q^{2^\theta})^n}.$$

**Proof:** We can write the estimator $\hat{\theta}^k$, with $k = 2$, as

$$\frac{1}{\log 2} \log \frac{\log(1 - X_1/n - X_2/n)}{\log(1 - X_1/n)}$$

where $X_1$ is Binomial($n$, $p_1 = 1-q$) and $X_2$ is Binomial($n$, $p_2 = q - q^{2^\theta}$). The conditional pmf of $X_2$ given that $X_1 = x_1$ is

$$f_{X_2|X_1=i_1}(i_2) = \binom{n-i_1}{i_2}\left(\frac{p_2}{1-p_1}\right)^{i_2}\left(1 - \frac{p_2}{1-p_1}\right)^{n-i_1-i_2},$$

and now we can write the joint pmf of $X_1$ and $X_2$ as

$$\begin{aligned} & f_{X_1,X_2}(i_1, i_2) \\ &= P[X_1 = i_1]P[X_2 = i_2 | X_1 = i_1] \\ &= \binom{n}{i_1} p_1^{i_1}(1-p_1)^{n-i_1}\binom{n-i_1}{i_2}\left(\frac{p_2}{1-p_1}\right)^{i_2}\left(1 - \frac{p_2}{1-p_1}\right)^{n-i_1-i_2}. \end{aligned} \tag{5.3.7}$$

Considering the Proposition 5.3.1, the joint pmf (5.3.7) and the conditional probability definition we get the cdf presented in i). Additionally, taking into account the expected value definition of a function of discrete rv´s we get the moments presented in ii). □

**Theorem 5.3.2.** *Let $D_1, D_2, ..., D_n$ be i.i.d. discrete Weibull rv´s with common cdf (5.1.1) and $i_j \in\in \{1, ..., n\}$ with $j \in \{1, ..., k\}$. Then, conditionally to $D_{1:n} = 1$ and $D_{n:n} > k$, for the class (5.3.4) with $k \geq 3$, we have*

*i) The cdf of $\hat{\theta}^k$ (5.3.4)*

$$F_{\hat{\theta}^k}(w) = \sum_{\forall i_1,\cdots,i_k : \sum_{d=2}^{k} \frac{1}{\log d} \log \frac{\log(1-i_1/n-\cdots-i_k/n)}{\log(1-i_1/n)}/(n-k) \leq w} \frac{f_{X_1,\cdots,X_k}(i_1,\cdots,i_k)}{1-q^n-(1-q^{k^\theta})^n+(q-q^{k^\theta})^n}$$

*with*

$$f_{X_1,\cdots,X_k}(i_1,\cdots,i_k)$$

$$= \binom{n}{i_1} p_1^{i_1}(1-p_1)^{n-i_1}$$

$$\prod_{v=2}^{k} \binom{n-i_1-\cdots-i_{v-1}}{i_v} \Big(\frac{p_v}{1-p_1-\cdots-p_{v-1}}\Big)^{i_v} \Big(1-\frac{p_v}{1-p_1-\cdots-p_{v-1}}\Big)^{n-i_1-\cdots-i_v},$$

*and* $\quad p_d = q^{(d-1)^\theta} - q^{d^\theta}$, *for* $d = 1, 2, \cdots, k$.

*ii) The moments $E\big[(\hat{\theta}^k)^l\big]$, with $l \in \mathbb{N}$,*

$$\sum_{\forall i_1,\cdots,i_k : 1 < i_1+\cdots+i_k < n \wedge i_1 > 0} \Big(\sum_{d=2}^{k} \frac{1}{\log d} \log \frac{\log(1-i_1/n-\cdots-i_k/n)}{\log(1-i_1/n)}/(n-k)\Big)^l$$

$$\times \frac{f_{X_1,\cdots,X_k}(i_1,\cdots,i_k)}{1-q^n-(1-q^{k^\theta})^n+(q-q^{k^\theta})^n}.$$

**Proof:** We can write the estimator $\hat{\theta}^k$ as

$$\sum_{d=2}^{k} \frac{1}{\log d} \log \frac{\log(1-X_1/n...-X_k/n)}{\log(1-X_1/n)}/(n-k)$$

where $X_1, X_2, ..., X_k$ are binomial r.v.´s with parameters $n$ and $p_d = q^{(d-1)^\theta} - q^{d^\theta}$. The conditional pmf of $X_d$ given that $X_1 = i_1 \wedge X_2 = i_2 ... \wedge X_{d-1} = i_{d-1}$ is

$$f_{X_d|X_1=i_1 \wedge X_2=i_2 ... \wedge X_{d-1}=i_{d-1}}(i_d)$$

$$= \binom{n-i_1...-i_{d-1}}{i_d} \Big(\frac{p_d}{1-p_1...-p_{d-1}}\Big)^{i_d} \Big(1-\frac{p_d}{1-p_1...-p_{d-1}}\Big)^{n-i_1...-i_d},$$

and now we can write the joint pmf of $X_1, X_2, ..., X_k$

$$f_{X_1,...,X_k}(i_1, ..., i_k)$$

$$= P[X_1 = i_1]P[X_2 = i_2|X_1 = i_1]...P[X_k = i_k|X_1 = i_1 \wedge X_2 = i_2 ... X_{k-1} = i_{k-1}]$$

$$= \binom{n}{i_1} p_1^{i_1}(1-p_1)^{n-i_1}\binom{n-i_1}{i_2}\Big(\frac{p_2}{1-p_1}\Big)^{i_2}\Big(1-\frac{p_2}{1-p_1}\Big)^{n-i_1-i_2}$$

$$... \times \binom{n-i_1...-i_{k-1}}{i_k}\Big(\frac{p_k}{1-p_1...-p_{k-1}}\Big)^{i_k}\Big(1-\frac{p_k}{1-p_1...-p_{k-1}}\Big)^{n-i_1...-i_k}. \qquad (5.3.8)$$

Considering the Proposition 5.3.1, the joint pmf (5.3.8) and the conditional probability definition we get the cdf presented in i). Additionally we get the moments presented in ii). □

**Remark 5.3.1.** *For a given observed sample $d_1, d_2, ..., d_n$, satisfying the conditions (5.3.6), in order to use the maximum information available, we will choose $k = d_{n:n} - 1$.*

The previous Remark lead us to propose the following estimator defined if $d_{1:n} = 1$ and $d_{n:n} > 2$,

$$\hat{\theta}_n^{IP} := \sum_{d=2}^{K} \frac{1}{\log d} \frac{\log(1-F_n(d))}{\log(1-F_n(1))}/(K-1). \qquad (5.3.9)$$

with $K := D_{n:n} - 1$. Now we use the maximum possible information, choosing $k = d_{n:n} - 1$. Since $F_n(d)$ are consistent estimators of $F_D(d)$, for $d = 2, \cdots, k$, $\hat{\theta}_n^{IP}$ is a consistent estimator of $\theta$. We also notice that $F_n(k)$ is an estimator of $P[D < d_{n:n}]$ with complete data and with type I censored data, allowing us to use (5.3.9) in both cases. The estimator (5.3.9) involves the rv´s $F_n(1)$, $F_n(d)$ and $K = D_{n:n} - 1$. In order to achieve the expected value, variance and mean square error of this estimator, we first derive in Theorem 5.3.3 the moments conditional to $K = k \geq 2$.

**Theorem 5.3.3.** *Let $D_1, D_2, ..., D_n$ be i.i.d. discrete Weibull rv´s with common cdf (5.1.1) and $i_j \in 1, ..., n$ with $j \in \{1, ..., k\}$. Then, conditionally to $D_{1:n} = 1$, for the Improved Proportions estimator (5.3.9), we have*

*i) The cdf of $\hat{\theta}^{IP}$ (5.3.9) conditional to $K = k \geq 2$, given by*

$$F_{\hat{\theta}^{IP}|K=k}(w) = \sum_{\forall i_1, i_2, ..., i_k : \sum_{d=2}^{k} \frac{1}{\log d} \log \frac{\log(1-i_1/n...-i_k/n)}{\log(1-i_1/n)} \leq w} \frac{f_{X_1, X_2, ..., X_k}(i_1, i_2, ..., i_k)}{p^{**}}$$

*with*

$$f_{X_1,X_2,...,X_k}(i_1, i_2, ..., i_k)$$

$$= \binom{n}{i_1}\binom{n-i_1}{i_2}...\binom{n-i_1-...-i_{k-1}}{i_k}\Big(\frac{p_1}{p_1+...+p_{k+1}}\Big)^{i_1}\Big(\frac{p_2+...+p_{k+1}}{p_1+...+p_{k+1}}\Big)^{n-i_1}$$

$$\times\Big(\frac{p_2}{p_2+...+p_{k+1}}\Big)^{i_2}\Big(\frac{p_3+...+p_{k+1}}{p_2+...+p_{k+1}}\Big)^{n-i_1-i_2}$$

$$\times ... \times\Big(\frac{p_k}{p_k+p_{k+1}}\Big)^{i_k}\Big(\frac{p_{k+1}}{p_k+p_{k+1}}\Big)^{n-i_1-...-i_k},$$

$p_1 = 1-q$, $p_d = q - q^{d^\theta}$, *for* $d = 1, ..., k+1$ , *and*

$$p^{**} = \sum_{\forall i_1,i_2,...,i_k:1<i_1+...+i_k<n\wedge i_1>0} f_{X_1,X_2,...,X_k}(i_1,i_2,...,i_k).$$

*ii) The conditional moment* $E\big[(\hat{\theta}^{IP})^l|_{K=k}\big]$, *with* $l \in \mathbb{N}$, *given by*

$$\sum_{\forall i_1,i_2,...,i_k:1<i_1+...+i_k<n\wedge i_1>0}\Big(\sum_{d=2}^{k}\frac{1}{\log d}\log\frac{\log(1-i_1/n...-i_d/n)}{\log(1-i_1/n)}\Big)^l$$
$$\times\frac{f_{X_1,X_2,...,X_k}(i_1,i_2,...,i_k)}{p^{**}}. \tag{5.3.10}$$

**Proof:** We can write the estimator $\hat{\theta}^{IP}$ conditional to $K = k$, as

$$\sum_{d=2}^{k}\frac{1}{\log d}\log\frac{\log(1-X_1/n...-X_d/n),}{\log(1-X_1/n)}$$

such that

$$f_{X_1|_{D=k}}(i_1) = \binom{n}{i_1}\Big(\frac{p_1}{p_1+...+p_{k+1}}\Big)^{i_2}\Big(\frac{p_2+...+p_{k+1}}{p_1+...+p_{k+1}}\Big)^{n-i_1},$$

and for $d = 2, ..., k$,

$$f_{X_d|_{X_1=i_1...\wedge X_{d-1}=i_{d-1}\wedge D=k}}(i_d)$$
$$= \binom{n-i_1...-i_{d-1}}{i_d}\Big(\frac{p_d}{p_d+...+p_{k+1}}\Big)^{i_d}\Big(\frac{p_{d+1}+...+p_{k+1}}{p_d+...+p_{k+1}}\Big)^{n-i_1...-i_d}.$$

Now we get the joint pmf of $X_1, X_2, ..., X_k$, conditional to $D = k$,

$$f_{X_1,X_2,...,X_k}(i_1,i_2,...,i_k)|_{D=k}$$
$$= P[X_1 = i_1|D = k]P[X_2 = i_2|X_1 = i_1 \wedge D = k]$$
$$... \times P[X_k = i_k|X_1 = i_1 \wedge X_{k-1} = i_{k-1} \wedge D = k]$$
$$= \binom{n}{i_1}\Big(\frac{p_1}{p_1+...+p_{k+1}}\Big)^{i_1}\Big(\frac{p_2+...+p_{k+1}}{p_1+...+p_{k+1}}\Big)^{n-i_1}.$$
$$\binom{n-i_1}{i_2}\Big(\frac{p_2}{p_2+...+p_{k+1}}\Big)^{i_2}\Big(\frac{p_3+...+p_{k+1}}{p_2++p_{k+1}}\Big)^{n-i_1-i_2}$$

$$... \times \binom{n - i_1 - ... - i_{k-1}}{i_k} \Big(\frac{p_k}{p_k + p_{k+1}}\Big)^{i_k} \Big(\frac{p_{k+1}}{p_k + p_{k+1}}\Big)^{n-i_1-...-i_k} \tag{5.3.11}$$

With the joint pmf (5.3.11) and the conditional probability definition we get the cdf presented in i). Additionally we get the conditional moments presented in ii). □

**Theorem 5.3.4 (Moments of the Improved Proportions estimator).** *Let $D_1, D_2, ..., D_n$ be i.i.d. discrete Weibull rv´s with common cdf (5.1.1). Denote by $E\big[(\hat{\theta}^{IP})^l|_{K=k}\big]$ the conditional moment (5.3.10), then, conditionally to $D_{1:n} = 1$ and $D_{n:n} > 2$, the moments of de Improved Proportions estimator (5.3.9), with $l \in \mathbb{N}$, are given by*

$$E\big[(\hat{\theta}^{IP})^l\big] = \sum_{k=2}^{\infty} \frac{\Big(1 - q^{n-1}\Big)\Big((1 - q^{(k+1)^\theta})^n - (1 - q^{k^\theta})^n\Big)}{1 - q^n - \Big(1 - q^{k^\theta}\Big)^n + \Big(q - q^{k^\theta}\Big)^n} E\big[(\hat{\theta}^{IP})^l|_{K=k}\big] \tag{5.3.12}$$

**Proof:** For $k \geq 2$,

$$f_{K|_{D_{1:n}=1 \wedge D_{n:n}>2}}(k)$$

$$= \frac{P[D_{1:n} = 1 \wedge K = k]}{P[D_{1:n} = 1 \wedge D_{n:n} > 2]}$$

$$= \frac{P[D_{1:n} = 1|K = k]P[K = k]}{P[D_{1:n} = 1 \wedge D_{n:n} > 2]}.$$

From Proposition 3.1 we get

$$= \frac{P[D_{1:n} = 1|K = k]P[K = k]}{1 - q^n - \Big(1 - q^{k^\theta}\Big)^n + \Big(q - q^{k^\theta}\Big)^n}$$

$$= \frac{\Big(1 - (1 - F_D(1))^{n-1}\Big)\Big(F_K(k) - F_K(k - 1)\Big)}{1 - q^n - \Big(1 - q^{k^\theta}\Big)^n + \Big(q - q^{k^\theta}\Big)^n},$$

and finally

$$f_{K|_{D_{1:n}=1 \wedge D_{n:n}>2}}(k) = \frac{\Big(1 - q^{n-1}\Big)\Big((1 - q^{(k+1)^\theta})^n - (1 - q^{k^\theta})^n\Big)}{1 - q^n - \Big(1 - q^{k^\theta}\Big)^n + \Big(q - q^{k^\theta}\Big)^n}. \tag{5.3.13}$$

If we consider the pmf $f_{K|_{D_{1:n}=1 \wedge D_{n:n}>2}}(k)$ (5.3.13), for a given $l \in \mathbb{N}$, the conditional moment $E\big[(\hat{\theta}^{IP})^l|_K\big]$ can be written as

$$E\big[(\hat{\theta}^{IP})^l|_K\big]$$

$$= \left\{ \begin{array}{ccc} E\big[(\hat{\theta}^{IP})^l|_{K=2}\big] & E\big[(\hat{\theta}^{IP})^l|_{K=3}\big] & ... \\ \frac{\Big(1-q^{n-1}\Big)\Big((1-q^{3^\theta})^n-(1-q^{2^\theta})^n\Big)}{1-q^n-\Big(1-q^{2^\theta}\Big)^n+\Big(q-q^{2^\theta}\Big)^n} & \frac{\Big(1-q^{n-1}\Big)\Big((1-q^{4^\theta})^n-(1-q^{3^\theta})^n\Big)}{1-q^n-\Big(1-q^{2^\theta}\Big)^n+\Big(q-q^{2^\theta}\Big)^n} & ... \end{array} \right. \tag{5.3.14}$$

Computing the mean value of the conditional expectation of $(\hat{\theta}^{IP})^l$ given $K$,

$$E[(\hat{\theta}^{IP})^l] = E\Big[E[(\hat{\theta}^{IP})^l|K]\Big],$$

the theorem follows easily.

□

**Corollary 5.3.1 (Expected Value and Variance of the Improved Proportions estimator).** *Let $D_1, D_2, ..., D_n$ be i.i.d. discrete Weibull rv´s with common cdf (5.1.1). Denote by $E\big[(\hat{\theta}^{IP})^l|_{K=k}\big]$ the conditional moment (5.3.10), then, conditionally to $D_{1:n} = 1$ and $D_{n:n} > 2$, the expected value and variance of the Improved Proportions estimator (5.3.9), with $l \in \mathbb{N}$, are given by*

$$E\big[\hat{\theta}^{IP}\big] = \sum_{k=2}^{\infty} \frac{\Big(1-q^{n-1}\Big)\Big((1-q^{(k+1)^\theta})^n - (1-q^{k^\theta})^n\Big)}{1-q^n-\big(1-q^{k^\theta}\big)^n+\big(q-q^{k^\theta}\big)^n} E\big[\hat{\theta}^{IP}|_{K=k}\big] \tag{5.3.15}$$

$$VAR\big[\hat{\theta}^{IP}\big] = \sum_{k=2}^{\infty} \frac{\Big(1-q^{n-1}\Big)\Big((1-q^{(k+1)^\theta})^n - (1-q^{k^\theta})^n\Big)}{1-q^n-\big(1-q^{k^\theta}\big)^n+\big(q-q^{k^\theta}\big)^n} E\big[(\hat{\theta}^{IP})^2|_{K=k}\big] - \Big(E\big[\hat{\theta}^{IP}\big]\Big)^2 \tag{5.3.16}$$

**Proof:** The proof follows straightforwardly by Theorem 3.4 and the definitions of expected value and variance. □

**Remark 5.3.2.** *The expressions for the expected value and variance of (5.3.9) in the Corollary 5.3.1 involves non closed forms, however we can compute approximations by taking the sum of 2 to $m \in \{3, 4, ...\}$, choosing $m$ such that $F_K(m) \backsimeq 1$ and then dividing this sum by $F_K(m)$.*

**Example 5.3.1.** Consider the Improved Proportions estimator $\hat{\theta}^{IP}$ (5.3.9) and the random sample $D_1, D_2, ..., D_n$, from the discrete Weibull distribution with common cdf (5.1.1) and parameters $\theta = 1.5$ and $q = 0.5$. To illustrate the calculation of approximate values for the expected value, variance and root of mean square error (RMSE), we present in Table 5.1 the conditional moments $E\big[\hat{\theta}^{IP}|_{K=k}\big]$, $E\big[(\hat{\theta}^{IP}|_{K=k})^2\big]$, the conditional probabilities $f_{K|_{D_{1:n}=1 \wedge D_{n:n}>2}}(k)$ and the conditional cdf $F_{K|_{D_{1:n}=1 \wedge D_{n:n}>2}}(k)$. From the last column we observe that $F_K(6) \backsimeq 1$. Using the Theorem 5.3.4 and according to what is refered in Remark 5.3.2 we compute the approximations. In Table 5.1, for each $n \in \{10, 20, 30\}$, adding the products of the values of the third column by the respective values of the fifth column and dividing by $F_K(6) \backsimeq 1$, we get an approximation for the expected values $E\big[\hat{\theta}^{IP}\big]$ presented in Table 5.2. Similarly, using the fourth column instead of the third column, we get an approximation for the second moments $E\big[(\hat{\theta}^{IP})^2\big]$. In the Table 5.2 we also present the approximations for the variance and RMSE.

The expected value, variance and RMSE, for each $n \in \{10, 20, 30\}$, were calculated for the Proportions estimator (5.2.3) using the Theorem 5.3.1 and the results are presented in the Table 5.2. We observe some improvement in terms of bias and RMSE when the Improved Proportions estimator (5.3.9) is used instead of the Proportions estimator (5.2.3) but we will show in the simulation study of Section 5.4 that much more substantial improvements are obtained in the cases $\theta \leq 1$. In these cases we only achieve $F_K(k) \simeq 1$ with a high value of $k$ and therefore the application of Theorem 5.3.4 involves a heavy computational effort.

Table 5.1

Conditional moments of $\hat{\theta}^{IP}$ given that $K = k$ with $\theta = 1.5$, $q = 0.5$ and $l = 1, 2$.

| $n$ | $k$ | $E\big[\hat{\theta}^{IP}\vert_{K=k}\big]$ | $E\big[(\hat{\theta}^{IP}\vert_{K=k})^2\big]$ | $f_{K\vert D_{1:n}=1 \wedge D_{n:n}>2}$ | $F_{K\vert D_{1:n}=1 \wedge D_{n:n}>2}$ |
|---|---|---|---|---|---|
| 10 | 2 | 1.5307 | 2.8199 | 0.69000 | 0.69000 |
| | 3 | 1.3246 | 2.0600 | 0.26012 | 0.95012 |
| | 4 | 1.2046 | 1.6845 | 0.04362 | 0.99375 |
| | 5 | 1.1183 | 1.4423 | 0.00502 | 0.99877 |
| | 6 | 1.0503 | 1.2672 | 0.00045 | 0.99922 |
| 20 | 2 | 1.6107 | 2.8375 | 0.55369 | 0.55369 |
| | 3 | 1.4134 | 2.1378 | 0.36722 | 0.92090 |
| | 4 | 1.2940 | 1.7795 | 0.07008 | 0.99098 |
| | 5 | 1.2088 | 1.5467 | 0.00823 | 0.99921 |
| | 6 | 1.1405 | 1.3735 | 0.00073 | 0.99994 |
| 30 | 2 | 1.6230 | 2.8042 | 0.43018 | 0.43018 |
| | 3 | 1.4561 | 2.2123 | 0.45786 | 0.88803 |
| | 4 | 1.3398 | 1.8632 | 0.09898 | 0.98702 |
| | 5 | 1.2568 | 1.6345 | 0.01184 | 0.99886 |
| | 6 | 1.1893 | 1.4610 | 0.00106 | 0.99992 |

Table 5.2

Expected value, variance and RMSE of $\hat{\theta}^{IP}$ and $\hat{\theta}^{P}$ for $n = 10, 20, 30$.

| Estimator | $n$ | $E\big[\hat{\theta}^{IP}\big]$ | $E\big[(\hat{\theta}^{IP})^2\big]$ | VAR$[\hat{\theta}^{IP}]$ | RMSE$[\hat{\theta}^{IP}]$ |
|---|---|---|---|---|---|
| $\hat{\theta}^{IP}$ | 10 | 1.4605 | 2.5649 | 0.4318 | 0.6583 |
| | 20 | 1.5124 | 2.4947 | 0.2074 | 0.4555 |
| | 30 | 1.5137 | 2.4247 | 0.1333 | 0.3654 |
| $\hat{\theta}^{P}$ | 10 | 1.5001 | 2.7319 | 0.4815 | 0.6939 |
| | 20 | 1.5528 | 2.6545 | 0.2434 | 0.4962 |
| | 30 | 1.5512 | 2.5724 | 0.1661 | 0.4107 |

### 5.3.2 Probability of observing a sample such that $\hat{\theta}^P$ and $\hat{\theta}^{IP}$ are defined

As we show in the previous example and as we will observe in Section 5.4 using simulations, with the Improved Proportions estimator (5.3.9) we can achieve a substantial improvement compared with the Proportions estimator (5.2.3), but this estimator suffers from the same limitation in terms of conditions to be defined. For both estimators we have the conditions (5.3.6) with $k = 2$. Here, we study in which cases this limitation can be relevant. In Figure 5.3.1 we represent the probability (5.3.5), from Proposition 5.3.1, for different values of $\theta$, $q$ and $n$. This probability decreases when $q$ approaches 0 or when $q$ approaches 1. As far as we know, the second case can be relevant in some applications. When the estimators (5.2.3) and (5.3.9) are not defined because $d_{1:n} > 1$, we suggest the following generalization of the Improved Proportions estimator, defined for $d_{n:n} > 2$,

$$\hat{\theta}_n^{IP*} := \sum_{d=2}^{K} \frac{1}{\log d} \frac{\log(1 - F_n(d))}{\log(\hat{q})} / (K - 1),$$

with $K := D_{n:n} - 1$ and $\hat{q}$ an estimator of $q$. If $d_{1:n} = 1$, we can choose $\hat{q} = 1 - F_n(1)$, if $d_{1:n} > 1$ we need to choose other estimator such that the estimate of $q$ is lower than 1.

## 5.4 Simulation Study

Here, we compare the moments estimator $\hat{\theta}_n^M$ (5.2.1), the proportions estimator $\hat{\theta}_n^P$ (5.2.3) and the proposed estimator $\hat{\theta}_n^{IP}$ (5.3.9). For the simulation study we have used the R language. In Figs 5.4.1-5.4.6 we present the simulated mean values and RMSE for $\theta = 0.5, 1, 1.5$ and $q = 05, 0.8$, using 5000 simulations in each sample size ($n = 10, \cdots, 100$). It is possible to calculate estimates both with (5.2.3) and (5.3.9), only if $d_{1:n} = 1$ and $d_{n:n} > 2$. Based on these conditions, some samples were excluded (see Table 5.3). In terms of bias, for all cases, the estimator $\hat{\theta}_n^{IP}$, performs much better than the others estimators under study. In terms of RMSE, for $\theta < 1$ (decreasing failure rate), the estimator $\hat{\theta}_n^{IP}$ performs much better. For $\theta \geq 1$ (increasing failure rate or geometric distribution), the performance in terms of RMSE of the moments estimator (5.2.1) and $\hat{\theta}_n^{IP}$ is almost the same or in some cases ($q = 0.8$), (5.2.1) performs slightly better.

Table 5.3

Frequency of excluded samples

| Sample | $\theta = 0.5$ | | $\theta = 1$ | | $\theta = 1.5$ | |
|---|---|---|---|---|---|---|
| size $n$ | $q = 0.5$ | $q = 0.8$ | $q = 0.5$ | $q = 0.8$ | $q = 0.5$ | $q = 0.8$ |
| 10 | 0.0088 | 0.1020 | 0.0538 | 0.1094 | 0.2284 | 0.1072 |
| 20 | 0.0000 | 0.0116 | 0.0028 | 0.0114 | 0.0482 | 0.0116 |
| 30 | 0.0000 | 0.0022 | 0.0000 | 0.0018 | 0.0114 | 0.0002 |
| 40 | 0.0000 | 0.0000 | 0.0000 | 0.0004 | 0.0022 | 0.0000 |

## 5.5 An Application to the Volkswagen stock returns

In the previous Section we show that for the case $\theta < 1$, the proposed Improved Proportions estimator (5.3.9) performs much better than the other estimators under study. Here, as an empirical example where the case $\theta < 1$ is relevant, we place ourselves in a context from quantitative risk management. We consider the Volkswagen share price from January 3, 2003 to January 29, 2010, and the daily log returns defined in (1.2.1). The data come from Web site http://chart.yahoo.com/ with ticker symbol **vow.de**. The corresponding one-day-ahead VaR forecasts made at time $t$ for time $t+1$, $VaR_{t+1|t}(p)$, are defined in (1.2.2). Considering a *violation* the event that a return is lower than the reported VaR, we define the hit function in (1.2.3) and the duration between two consecutive violations as $D_i := t_i - t_{i-1}$, where $t_i$ denotes the day of violation number $i$. Christoffersen (1998) showed that evaluating interval forecasts can be reduced to examining whether the hit sequence satisfies the unconditional coverage (UC) and independence (IND) properties. It is possible to write the IND property as

$$D_i \overset{\text{iid}}{\sim} D \sim \text{discrete Weibull}(\theta = 1).$$

A problematic non verification of IND is the one that leads to clustering of violations, which corresponds to several large losses occurring in a short period. With clustering, we have an excessive number of very short durations and an excessive number of very long durations. The discrete Weibull with $\theta < 1$ will generate this pattern, for this reason, the estimate of the shape parameter can be used to identify a model that violates IND in this way. Using the popular Historical Simulation (HS) method for VaR(0.05), we calculate 95 durations with a moving window of size 250. The obtained estimates were $\hat{q}^M = 0.847$, $\hat{q}^P = 0.832$, $\hat{\theta}^M = 0.712$, $\hat{\theta}^P = 0.794$ and $\hat{\theta}^{IP*} = 0.67$. All estimators gives evidence that the HS VaR method used, leads to clustering of violations, with estimates of $\theta$ lower than one. We consider three models, $F_M$, $F_P$ and $F_{IP}$ fitted with the methods (5.2.1), (5.2.3) and (5.3.9). To asses how well these distributions fits the Volkswagen durations data set, Figure 5.5.1 contains the plot of the ecdf along with $F_M$, $F_P$ and $F_{IP}$ cdf's. We also plot the ecdf along with the cdf of the geometric (0.05) which corresponds to the UC and IND hypothesis. To measure the discrepancy between the ecdf and the cdf's, the Kolmogorov-Smirnov and Chi-Square statistics are given in Table 5.4. Clearly, the moments and the improved proportions methods provide much better fit than the proportions method. These two methods performs well with the real data set under study, but the improved proportions is based on a simple equation while the method of moments involves equations that cannot be solved easily by ordinary techniques.

Table 5.4
Goodness-of-fit statistics for fits of the Geometric(0.05), $F_M$, $F_P$ and $F_{IP}$ distributions to the Volkswagen durations data set

| Statistic | Geometric(0.05) | $F_M$ | $F_P$ | $F_{IP}$ |
|---|---|---|---|---|
| Kolmogorov-Smirnov | 0.221 | 0.052 | 0.095 | 0.0514 |
| Chi-Square | 46.7 | 6.16 | 35.8 | 5.84 |

## Acknowledgements

Research partially supported by Fundação para a Ciência e Tecnologia (FCT/PROTEC and FCT/OE) and Center of Statistics and Applications of University of Lisbon (CEAUL).

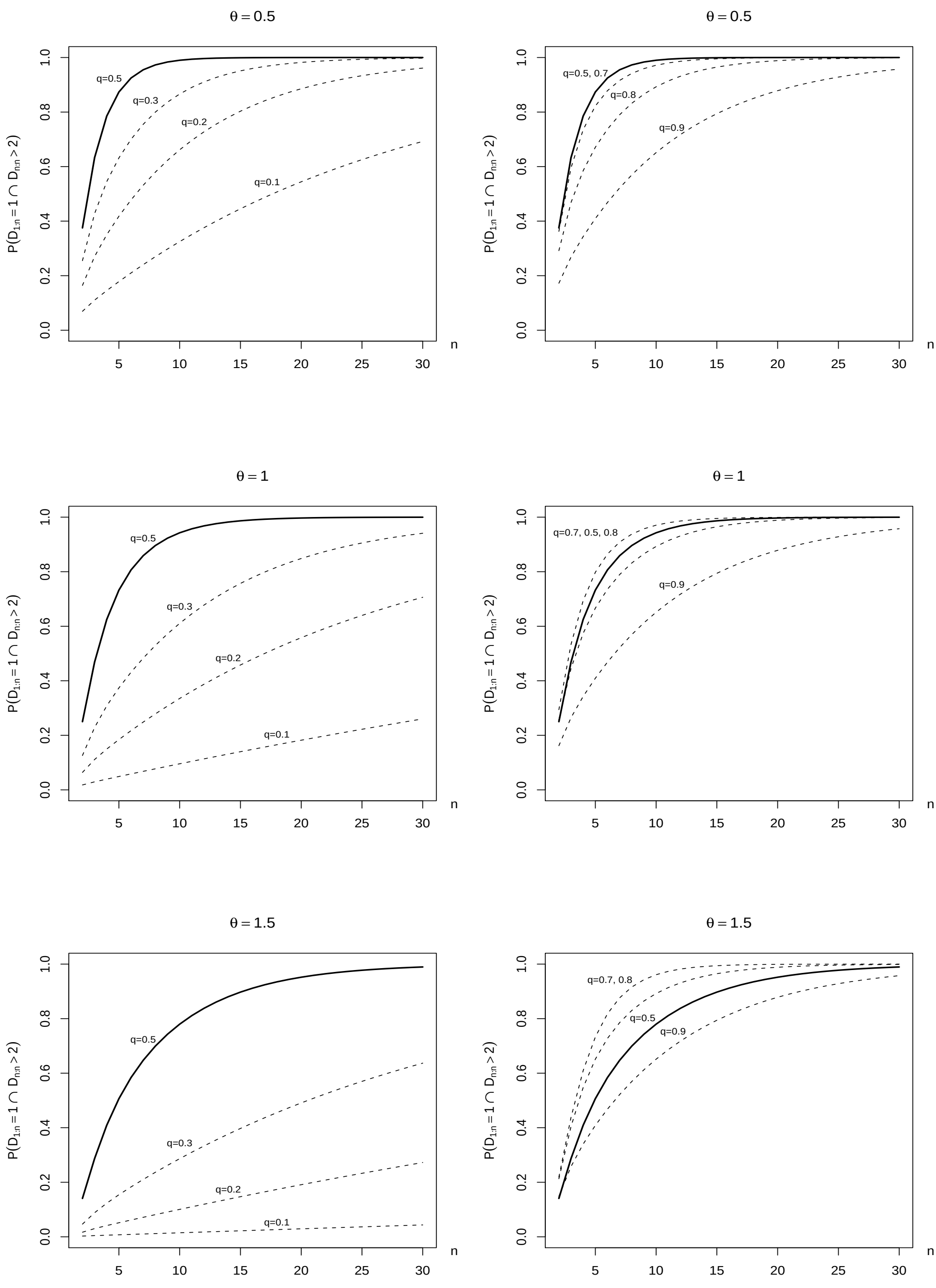



Figure 5.3.1: Probability of observing a sample such that $\hat{\theta}^P$ and $\hat{\theta}^{IP}$ are defined with $\theta \in \{0.5, 1, 1.5\}$, $q \in \{0.1, 0.2, 0.3, 0.5, 0.7, 0.8, 0.9\}$ and sample sizes from 2 to 30.

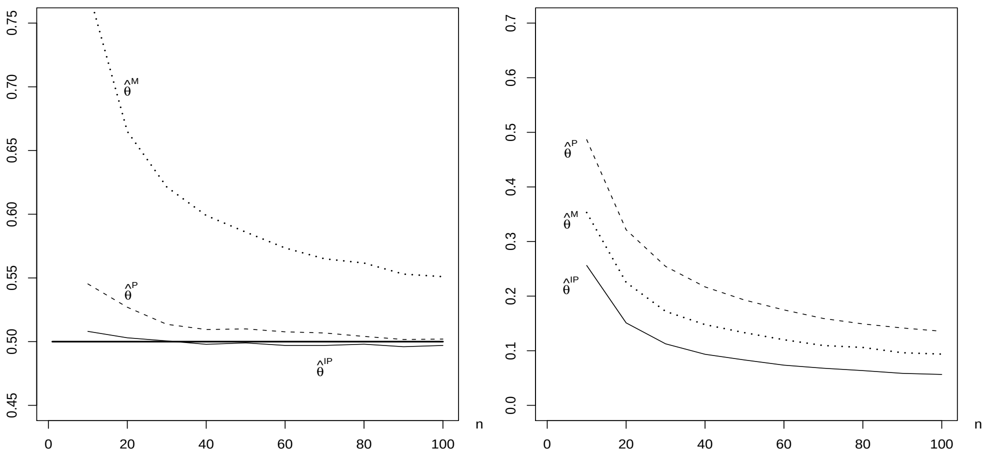



Figure 5.4.1: Simulated Mean values (left) and root mean squared errors (right), of $\hat{\theta}^M$, $\hat{\theta}^P$ and $\hat{\theta}^{IP}$, from a discrete Weibull model with $q = 0.5$ and $\theta = 0.5$ (decreasing failure rate).

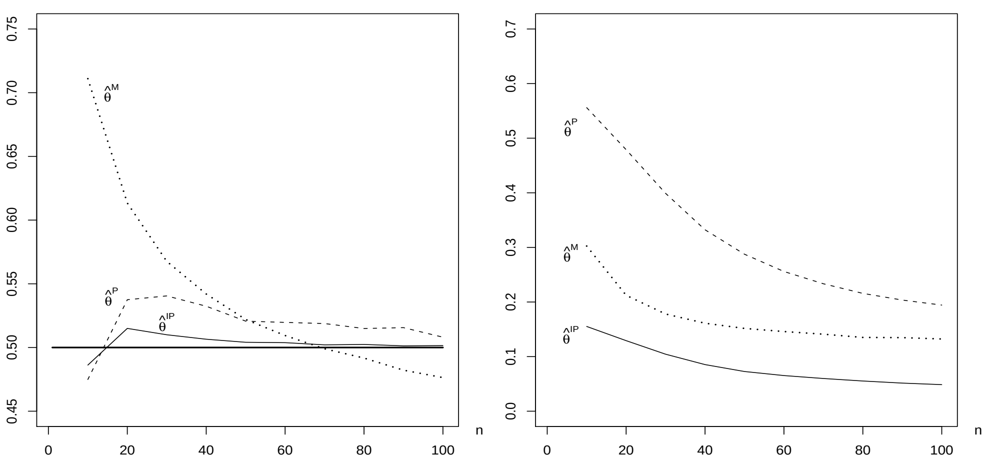



Figure 5.4.2: Simulated Mean values (left) and root mean squared errors (right), of $\hat{\theta}^M$, $\hat{\theta}^P$ and $\hat{\theta}^{IP}$, from a discrete Weibull model with $q = 0.8$ and $\theta = 0.5$ (decreasing failure rate).

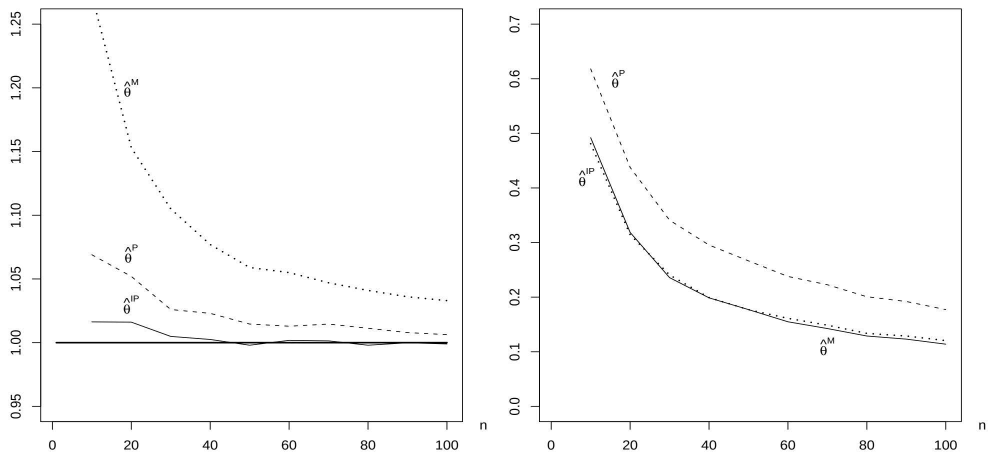



Figure 5.4.3: Simulated Mean values (left) and root mean squared errors (right), of $\hat{\theta}^M$, $\hat{\theta}^P$ and $\hat{\theta}^{IP}$, from a discrete Weibull model with $q = 0.5$ and $\theta = 1$ (geometric distribution).

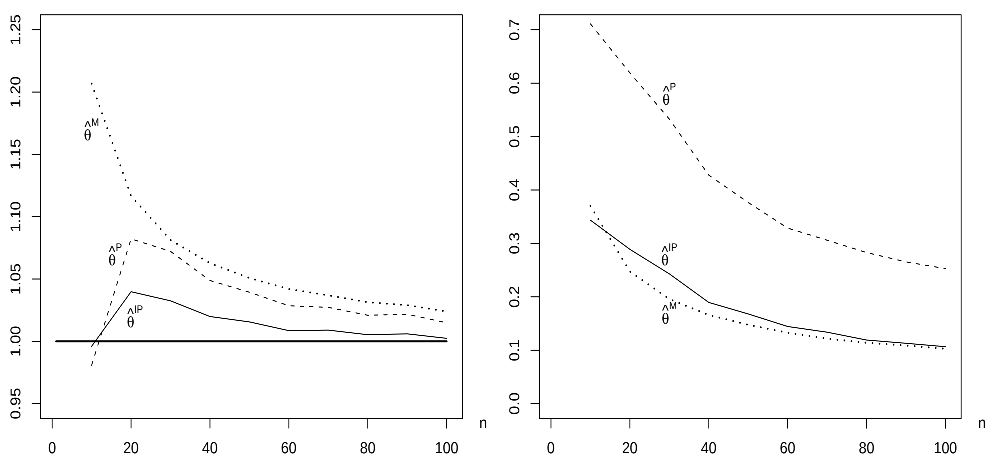



Figure 5.4.4: Simulated Mean values (left) and root mean squared errors (right), of $\hat{\theta}^M$, $\hat{\theta}^P$ and $\hat{\theta}^{IP}$, from a discrete Weibull model with $q = 0.8$ and $\theta = 1$ (geometric distribution).

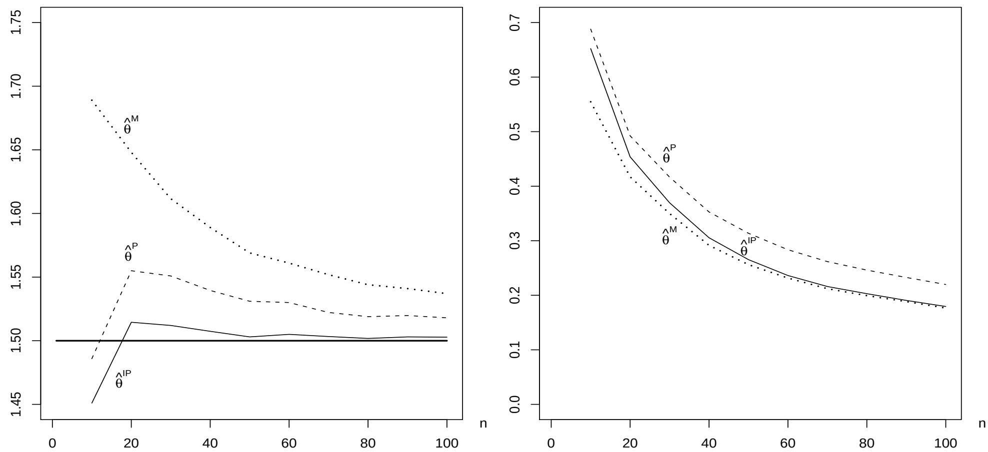



Figure 5.4.5: Simulated Mean values (left) and root mean squared errors (right), of $\hat{\theta}^M$, $\hat{\theta}^P$ and $\hat{\theta}^{IP}$, from a discrete Weibull model with $q = 0.5$ and $\theta = 1.5$ (increasing failure rate).

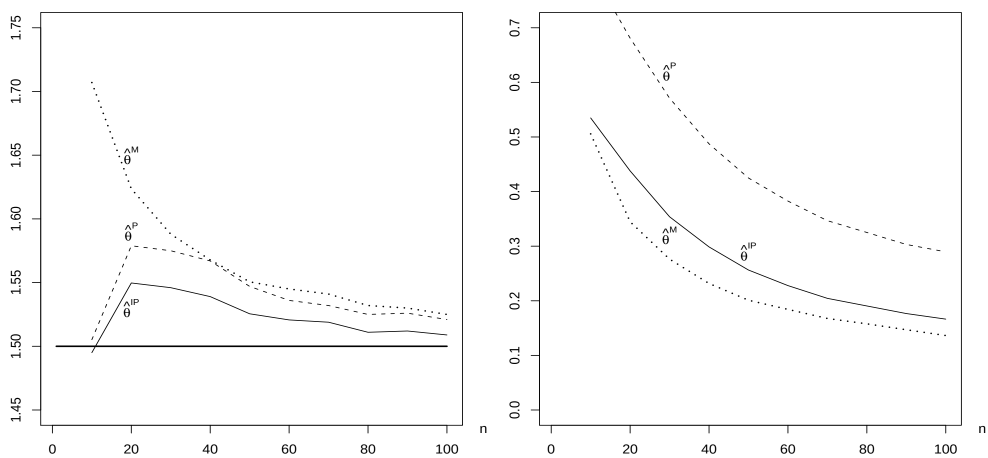
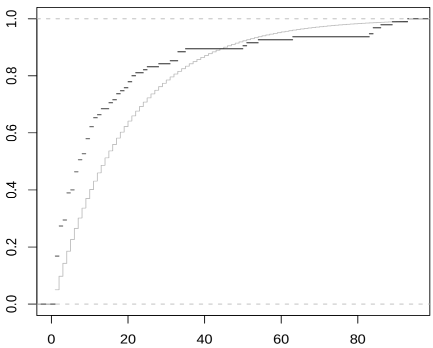
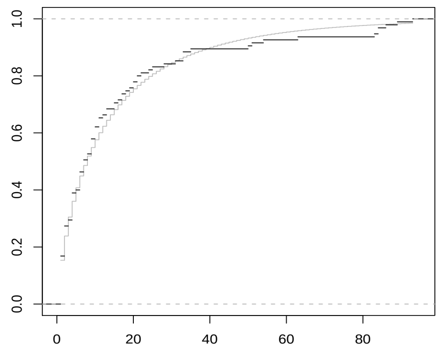


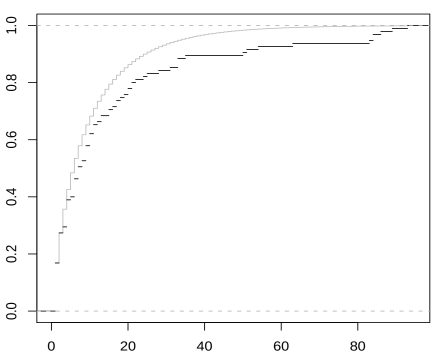

Figure 5.4.6: Simulated Mean values (left) and root mean squared errors (right), of $\hat{\theta}^M$, $\hat{\theta}^P$ and $\hat{\theta}^{IP}$, from a discrete Weibull model with $q = 0.8$ and $\theta = 1.5$ (increasing failure rate).



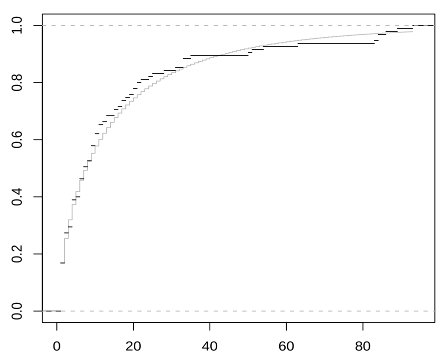

(a) Geometric(0.05) cdf (grey) and ecdf (black)




(b) cdf $F_M$ (grey) and ecdf (black)




(c) cdf $F_P$ (grey) and ecdf (black)




(d) cdf $F_{IP}$ (grey) and ecdf (black)

Figure 5.5.1: Comparison of Geometric(0.05), $F_M$, $F_P$ and $F_{IP}$ cdf's and ecdf for Volkswagen durations data set.

*Paulo Araújo Santos and Isabel Fraga Alves*

# 6 Forecasting Value-at-Risk with a Duration based POT method

Threshold methods, based on fitting a stochastic model to the excesses over a threshold, were developed under the acronym POT (peaks over threshold). In order to eliminate the tendency to clustering of violations, a model based approach within the POT framework, that uses the durations between excesses as covariates, is proposed. Based on this approach, models for forecasting one-day-ahead Value-at-Risk were suggested and applied to real data. Comparative studies provide evidence that they can perform better than state-of-the art risk models and much better than the widely used RiskMetrics model, both in terms of out-of-sample accuracy and under the Basel II Accord.

## 6.1 Introduction

Investors and traders must pay attention not only to the expected return from their activities but also to the risks that they incur. It is widely accepted that risk-adjusted performance measures can guide institutions toward a better risk/return profile and can play a relevant role to achieve a more secure financial system. This justify the interest of developing more accurate risk models. Value-at-Risk (VaR) aggregates several components of risk into a single number and has emerged as the standard measure in quantitative risk management. In terms of regulation, the Basel II Accord requires that banks and other Authorized Deposit-taking Institutions (ADIs) to report their daily VaR forecasts to the monetary authorities (typically, a central bank) at the beginning of each trading day and defines daily capital requirements based on these forecasts (for a detailed discussion of VaR, see Jorion, 2000). We will deal with the excesses over a high threshold and for this reason, in this Chapter, instead of the daily log returns we will consider the symmetric of daily log returns, $R_{t+1} = -\log(P_{t+1}/P_t) \times 100$, where $P_t$ is the value of the portfolio at time $t$. Consequently, in this Chapter instead of (1.2.2), we define the one-day-ahead VaR forecast made at time $t$ for time $t+1$, $VaR_{t+1|t}(p)$, as

$$P[R_{t+1} > VaR_{t+1|t}(p)|\Omega_t] = p,$$

where $\Omega_t$ is the information set up to time-$t$ and $p$ is the *coverage rate*. A *violation* occurs when the symmetric daily return exceeds the reported VaR, i.e., when $R_{t+1} > VaR_{t+1|t}(p)$. The rest of the Chapter is organized as follows. In Section 6.2, we review the peaks over threshold (POT) method with an example that illustrates the problem of *tendency to clustering of violations.* In Section 6.3, in order to solve this problem, we propose risk models based on durations and within the POT framework. Comparisons between the proposed risk models and other models are made in Section 6.4. Finally, conclusions and directions for future research are given in Section 6.5.

## 6.2 The POT method and the tendency to clustering of violations problem

The Generalized Pareto Distribution (GPD) has the form

$$GPD_{\gamma,\sigma}(y) = \begin{cases} 1-(1+\gamma y/\sigma)^{-1/\gamma}, & \gamma \neq 0 \\ 1-\exp\left(-y/\sigma\right), & \gamma = 0, \end{cases} \tag{6.2.1}$$

where $\sigma > 0$, and the support is $y \geq 0$ when $\gamma \geq 0$ and $0 \leq y \leq -\sigma/\gamma$ when $\gamma < 0$. The expected value and variance are given by

$$E[Y] = \frac{\sigma}{1-\gamma} \quad (\gamma < 1), \qquad VAR[Y] = \frac{\sigma^2}{(1-\gamma)^2(1-2\gamma)} \quad (\gamma < 1/2).$$

Generally, with $\gamma > 0$, $E[Y^c]$ does not exist for $\gamma \geq 1/c$. The probability that the random variable (r.v.) $X$ assumes a value that exceeds a threshold $u$ by at most $y$, given that it does exceed the threshold, is given by the *excess distribution*

$$F_u(y) = P[X-u \leq y|X > u] = \frac{F(y+u)-F(u)}{1-F(u)}, \tag{6.2.2}$$

for $0 \leq y < x^F - u$, where $x^F$ is the (finite or infinite) right endpoint of F, defined by $x^F := \sup\{x : F(x) < 1\}$. The EVT, with the following theorem (Balkema and de Haan (1974) and Pickands (1975)), suggests the GPD (6.2.1) as an approximation for the excess distribution (6.2.2), for a sufficiently high threshold $u$.

**Theorem 6.2.1** (Pickands-Balkema-de Haan Theorem)**.** *It is possible to find a function* $\beta(u)$ *such that*

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

*if and only if* $F$ *is in the maximum domain of attraction of an extreme value distribution.*

For a wide class of distributions, the excess distribution (6.2.2) over a high threshold $u$ can be approximated by the GPD (6.2.1) and this result holds for essentially all common continuous distributions; more precisely, Theorem 6.2.1 holds for all distributions in some max-domain of attraction of an extreme value distribution, i.e., distributions for which the sequence of maxima linearly normalized converges to one non degenerate limit law of theorem 1.1.1. To estimate the parameters $\gamma$ and $\sigma$ we fit the GPD to the excesses over the conveniently chosen threshold $u$. For $\gamma > -1/2$, the standard properties of the maximum likelihood (ML) estimators have been proved by Smith (1987) and extended for $\gamma > -1$ by Zhou (2010). Furthermore, it is possible to show, using simulations, that inference is often robust to choice of the threshold $u$, when $u$ is big enough. Smith (1987) proposed a tail estimator based on a GPD approximation to the excess distribution. We denote $n$ the number of excesses above $u$ in a sample $X_1, ..., X_{n_x}$. Using $n/n_x$ as estimator of $\bar{F}(u)$ the relation $\bar{F}_u(x-u) = \bar{F}(x)/\bar{F}(u)$ and $\bar{F}_u(x-u)$ estimated by a GPD approximation, we obtain the tail estimator

$$\hat{\bar{F}}(x) = \frac{n}{n_x}\Big(1 + \hat{\gamma}\frac{x-u}{\hat{\sigma}}\Big)^{-1/\hat{\gamma}}, \quad \text{valid for } x > u. \tag{6.2.3}$$

For $p < \bar{F}(u)$ and inverting the tail estimator formula (6.2.3), we get the VaR POT estimator

$$\hat{\text{VaR}}^{POT}_{t+1|t}(p) = u + \frac{\hat{\sigma}}{\hat{\gamma}}\Big(\Big(\frac{n}{n_x p}\Big)^{\hat{\gamma}} - 1\Big). \tag{6.2.4}$$

Now, turning theory into practice, one example is presented to illustrate the problem of tendency to clustering of violations which occurs when we apply the VaR POT estimator (6.2.4) to financial time series. The data consist of 15190 daily returns of Standard & Poor's Index (S&P 500), from January 4, 1950 through May 18, 2010. We choose the threshold, $u = x_{13671:15190} = 0.9897$, such that 10% of the values are larger than the threshold; see McNeil and Frey (2000) for a simulation study that support a similar choice. In Figure 6.2.1 we present the returns with the threshold (grey line) and a histogram where we can observe how the GPD, with the parameters estimated by ML estimation, adjust very well to the excesses. In this example, we obtain a VaR(0.05) equal to 1.42 and a VaR(0.01) equal to 2.67. In Figure 6.2.2(a), instead of considering 15190 daily returns to obtain one VaR(0.01) estimate, we present one-day-ahead VaR forecasts with a rolling window of size 1000 ($n_w = 1000$). The percentage of days where the symmetric returns exceeds the correspondent VaR forecast, i.e., the percentage of violations, equals 1.367% of the 14190 days used for the out-of-sample forecasts, when the expected is 1%.

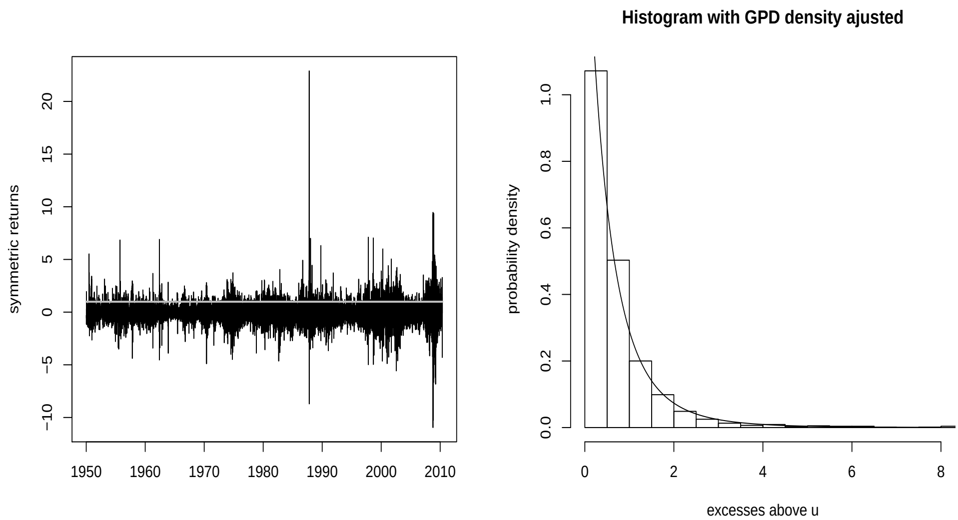
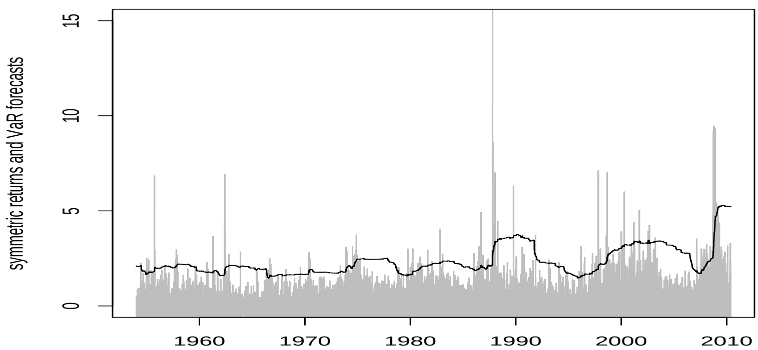
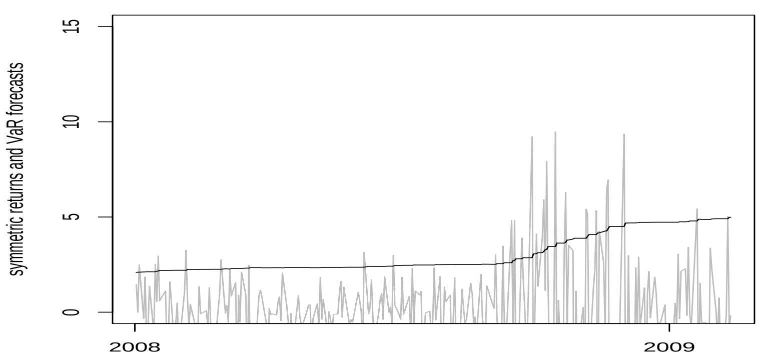



Figure 6.2.1: Symmetric returns (left) and Histogram of 1519 excesses above the threshold $u = x_{13671:15190} = 0.9897$ (right) for the S&P 500 Index from January 4, 1950 through May 18, 2010.

However, the serious problem of POT method and other unconditional models, is tendency to clustering of violations associated with the volatility clustering phenomenon. Figure 6.2.2(b) illustrates this problem during the 2008 financial crisis period. Between January 2, 2008 and February 12, 2009, we have a large number of violations in a short period of time. Over this period, the number of violations was 29, representing 10.28% of the 282 trading days, when the expected value for the percentage of violations is 1%.




(a) January 4, 1950 through May 18, 2010 (14190 trading days)




(b) January 2, 2008 through February 12, 2009 (282 trading days)

Figure 6.2.2: Symmetric returns of S&P 500 Index (grey line) and one-day-ahead VaR(0.01) forecasts with POT method (black line) and a rolling window of size 1000.

## 6.3 A duration based POT method (DPOT)

Our main goal is to eliminate the tendency to clustering of violations that occurs with the POT method. To achieve this goal, within the POT framework we propose the presence of durations between excesses as covariates. Smith (1990), developed ML and Least Squares estimation procedures under the POT framework with the shape and scale parameters dependent on covariates. For a general overview of EVT and its application to VaR, including the use of explanatory variables, see, for instance, Tsay (2010). For details about the mathematical theory of EVT and its applications to risk management, see Embrechts et al. (2001).

Let $y_1, ..., y_n$ be the excesses above a high threshold $u$, $d_1$ the duration until the first excess and $d_2, ..., d_n$, defined by

$$d_i = t_i - t_{i-1}, \tag{6.3.1}$$

where $t_i$ denotes the day of excess $i$. We propose to use from the information set up to time t ($\Omega_t$), the last $v$ durations between excesses, $d_n, d_{n-1}, ..., d_{n-v+1}$ and the duration since the excess $n$ which we define by $d^t$. With the durations $d_i, ..., d_{i-v+1}$, it is possible to consider at the time of excess number $i$, the duration since the preceding $v$ excesses, defined by

$$d_{i,v} = d_i + ... + d_{i-v+1} = t_i - t_{i-v}. \tag{6.3.2}$$

At day $t$, after the excess $n$, we define $d_{t,1} = d^t$, $d_{t,2} = d^t + d_n$ and for $v = 3, 4, \cdots$,

$$d_{t,v} = d^t + d_{n,v-1} = d^t + d_n + \cdots + d_{n-v+2},$$

which represents the duration until $t$ since the preceding $v$ excesses.

### 6.3.1 Empirical Motivation

The motivation for the presence of durations between excesses as covariates has mainly been based on the relation between the amount of the excess and durations which we observe in various financial time series. Figure 6.3.1 (left) presents for the S&P 500 Index example of Section 6.2, the scatterplot of excesses ($y_i$) and durations since the preceding excess ($d_i$). Clearly, large excesses tend to be associated with short durations and small excesses tend to be associated with long durations. In Figure 6.3.1 (right) we observe a similar pattern for excesses and durations between the 2 preceding excesses ($d_{i-1}$). Table 6.1 gives Pearson correlations between excesses, durations and the inverse of durations. The linear association between excesses and durations is weak, but increases when we take the inverse of durations, as expected. Adding durations we get the duration since the preceding $v$ excesses defined in (6.3.2) and the correlation increases a little more when we compute the correlation between excesses and the inverse of these durations. In short, the empirical results show some nonlinear association between excesses and durations. We also observe that the excesses have higher mean and higher variance with short durations, and lower mean and lower variance with long durations. Based on these empirical results, we propose to define the expected value and variance of the excesses dependent on the durations.

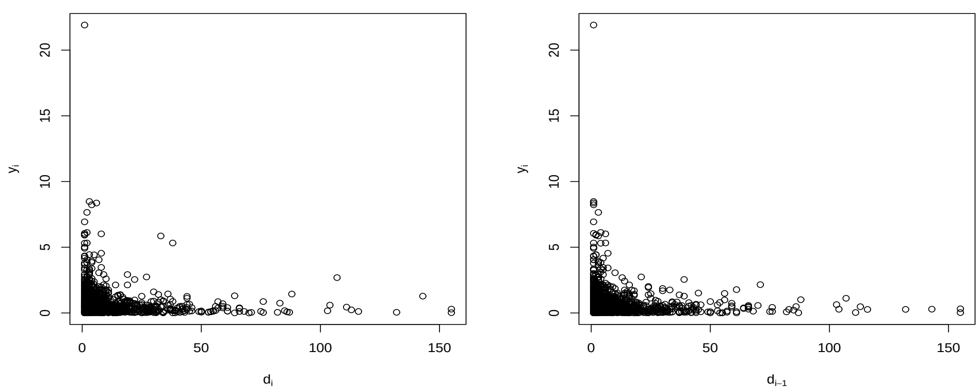



Figure 6.3.1: S&P 500 Index from January 4, 1950 through May 18,2010. Scatter plot of excesses above a high threshold ($u = 0.9897$) and durations since the preceding excess (left) and scatter plot of excesses and durations between the 2 preceding excesses (right).

Table 6.1

S&P 500 Index. Pearson correlation between $y_i$, $d_{i-j}$, $\frac{1}{d_{i-j}}$ and $\frac{1}{d_{i,v}}$.

| $j$ | $\text{Corr}(y_i, d_{i-j})$ | $\text{Corr}(y_i, \frac{1}{d_{i-j}})$ | $v$ | $\text{Corr}(y_i, \frac{1}{d_{i,v}})$ |
|---|---|---|---|---|
| 0 | -0.123 | 0.193 | 2 | 0.284 |
| 1 | -0.127 | 0.174 | 3 | 0.325 |
| 2 | -0.096 | 0.149 | 4 | 0.335 |
| 3 | -0.126 | 0.148 | 5 | 0.346 |

### 6.3.2 DPOT Model

With the durations (6.3.1) and the duration since the excess $n$, $d^t$, we assume the GPD for the excesses $Y_i$ above $u$, such that

$$Y_t \sim GPD\Big(\gamma, \sigma_t = g(\alpha_1, ..., \alpha_k, ..., d^t, d_n, d_{n-1}, ..., d_{n-v+2})\Big),$$

where $\gamma, \alpha_1, ..., \alpha_k$, are parameters to be estimated. And we propose the following class of estimators

$$\widehat{\text{VaR}}_{t+1|t}^{DPOT}(p) = u + \frac{\hat{\sigma_t}}{\hat{\gamma}}\Big(\Big(\frac{n}{n_x p}\Big)^{\hat{\gamma}} - 1\Big), \tag{6.3.3}$$

with $\hat{\sigma}_t = g(\hat{\alpha_1}, ..., \hat{\alpha_k}, ..., d^t, d_n, d_{n-1}, ..., d_{n-v+2})$.

The proposed *DPOT* method implies, for $\gamma < 1$, a conditional expected value for excesses, and for $\gamma < 1/2$, a conditional variance, both dependent on $d^t$ and the last $v$ durations between excesses,

$$E[Y_t|\Omega_t] = \frac{\sigma_t}{1-\gamma} \quad (\gamma < 1), \quad VAR[Y_t|\Omega_t] = \frac{(\sigma_t)^2}{(1-2\gamma)} \quad (\gamma < 1/2).$$

The empirical results of Section 6.3.1 suggest a inverse relation between excesses and the durations since the preceding $v$ excesses, with $1/(d_{i,v})^c$, $c > 0$, which leads to the specification $\sigma_t = \alpha\frac{1}{(d_{t,v})^c}$ and the VaR estimator

$$\widehat{\text{VaR}}_{t+1|t}^{DPOT(v,c)}(p) = u + \frac{\hat{\alpha}}{\hat{\gamma}(d_{t,v})^c}\Big(\Big(\frac{n}{n_x p}\Big)^{\hat{\gamma}} - 1\Big), \tag{6.3.4}$$

where $\hat{\gamma}$ and $\hat{\alpha}$ are estimators of the parameters $\gamma$ and $\alpha$. Applying the maximum likelihood theory to estimate the parameters, the log likelihood obtained is

$$\begin{aligned}
\log \text{L}(\gamma, \alpha) &= \log \prod_{i=v}^{n} f_{Y_i}(y_i) \\
&= \log \prod_{i=v}^{n} \Big(\frac{\alpha}{(d_{i,v})^c}\Big)^{-1}\Big(1 + \frac{\gamma}{\alpha} y_i (d_{i,v})^c\Big)^{-(1/\gamma+1)} \\
&= -\sum_{i=v}^{n} \log\Big(\frac{\alpha}{(d_{i,v})^c}\Big) - \Big(\frac{1}{\gamma}+1\Big)\sum_{i=v}^{n} \log\Big(1 + \frac{\gamma}{\alpha} y_i (d_{i,v})^c\Big).
\end{aligned} \tag{6.3.5}$$

We present results for $v = 3$, $c \in \{0.8, 0.75, 0.7\}$ and apply an implementation of Nelder and Mead algorithm, using the stats package of R (R Development Core Team, 2008), to maximize (6.3.5).

Using the proposed models with the S&P 500 Index returns presented in the Section 6.2 example, we obtain for 14190 one-day-ahead VaR forecasts, 138 (0.9725%), 134 and 134 (0.9443%) violations, respectively with $c = 0.8$, $c = 0.75$ and $c = 0.7$. These percentages are much closer to the expected 1% than the 1.367% obtained with the unconditional POT model. In Figure 6.3.2, the grey line corresponds to the S&P 500 returns, the dotted, the longslash, the solid and the longslash grey lines, correspond to one-day-ahead VaR forecasts calculated respectively with the DPOT($c = 0.8$), DPOT($c = 0.75$), DPOT($c = 0.7$) and the POT models. For the 2008 global financial crises period, Figure 6.3.2, shows how the DPOT models solve the problem of tendency to clustering of violations, producing much better risk forecasts that adjust quickly to the high volatility in the returns during September and October. Within this period of 282 days, the number of violations with DPOT($c = 0.8$) was 8, with DPOT($c = 0.75$) was 8 and with DPOT($c = 0.7$) was 11, much less than the 29 violations obtained with the unconditional POT method. Moreover, notice that with some exceptions, in the majority of the days the difference between DPOT($c = 0.8$), DPOT($c = 0.75$) and DPOT($c = 0.7$) forecasts, is very small, suggesting that the method is robust for different values of $c$ in the interval between 0.7 and 0.8. Empirical findings in Section 6.4 will suggest that a choice of $c = 0.75$ is preferable. We also study the model with $c$ estimated, but we achieve poor results.

## 6.4 Comparative studies

Using the returns from S&P 500 Index, German stock market Index (DAX) and Financial Times London Stock Exchange Index (FTSE), we compare the proposed DPOT method with a two-stage hybrid method which combines a time-varying volatility model with the EVT approach, known as Conditional EVT, and with two conditional parametric models. We employ the R language in order to develop the programs. The web site http://finance.yahoo.com was the source of the data. In Section 6.4.1 we briefly review the Conditional EVT method, the Asymmetric Power Autoregressive Conditional Heteroscedasticity (APARCH) model and the widely used RiskMetrics model. In Section 6.4.2 we evaluate the accuracy of out-of-sample interval forecasts produced with the risk models and in Section 6.4.3 we compare the performance under the Basel II Accord.

### 6.4.1 Conditional EVT, APARCH and RiskMetrics

The EVT procedure described in Section 6.2 is unconditional, however, to solve or reduce the problem of clustering, we can apply EVT to returns adjusted by some dynamic structure. It is usual to assume for the returns, $R_t = \mu_t + \varepsilon_t$, where $\varepsilon_t$ is the unpredictable component and $\mu_t$ the conditional mean expressed as a $s$th order autoregressive process,

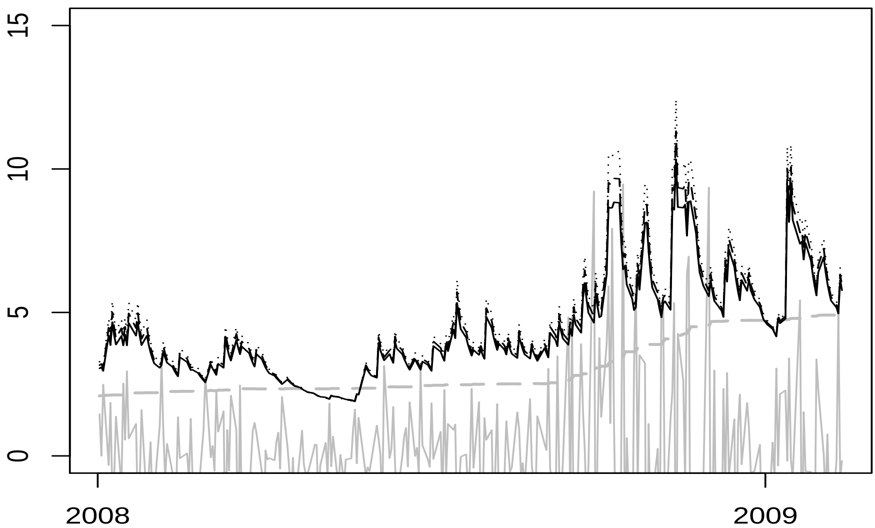



Figure 6.3.2: Symmetric returns of S&P 500 Index from January 2, 2008 through February 12,2009 (solid grey), and one-day-ahead VaR(0.01) forecasts with DPOT($c = 0.8$) (dotted), DPOT($c = 0.75$) (longdash), DPOT($c = 0.7$) (solid), POT method (longdash grey) and a rolling window of size 1000.

AR(s),

$$\mu_t = \phi_0 + \sum_{i=1}^{s} \phi_i R_{t-i}.$$

The unpredictable component can be expressed by $\varepsilon_t = Z_t \sigma_t$, where the innovations, $Z_t$, are a sequence of independently and identically distributed random variables with zero mean and unit variance, and the conditional variance is

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2,$$

where $\alpha_i > 0$ and $\beta_j > 0$, for $i = 0, 1, ..., p$ and $j = 1, 2, ..., q$. This time-varying volatility model for the unpredictable component, is a Generalised Autoregressive Conditional Heteroscedasticity (GARCH) process, proposed by Bollerslev (1986). The GARCH model with $p = 1$ and $q = 1$, usually captures with success several stylized facts of financial time series. Diebold et al. (1998) proposed in a first step the standardization of the returns through the conditional means and variances estimated with a time-varying volatility model, and in a second step, estimation of a $p$ quantile using EVT and the standardized returns. McNeil and Frey (2000) combine a AR(1)-GARCH(1,1) process assuming normal innovations with the POT method from EVT. We will denote this model as CEVT-n. The filter with normal innovations, while capable of removing the majority of clustering, will frequently be a misspecified model for returns. For accommodate this misspecification, Kuester et al. (2006) suggested a filter with the skewed-$t$ distribution. We will denote this model as CEVT-sst. Applying Conditional EVT, the VaR estimator is

$$VaR_{t+1|t}^{CEVT}(p) = \hat{\mu}_{t+1|t} + \hat{\sigma}_{t+1|t}\hat{z}_p,$$

where $\hat{\mu}_{t+1|t}$ and $\hat{\sigma}_{t+1|t}$ are the estimated conditional mean and conditional standard deviation for $t+1$, obtained with a AR(1)-GARCH(1,1) process. Moreover, $\hat{z}_p$ is a quantile $p$ estimate, obtained with the POT method and the standardized residuals calculated as

$$(z_{t-n+1}, ..., z_t) = \left(\frac{r_{t-n+1} - \hat{\mu}_{t-n+1}}{\hat{\sigma}_{t-n+1}}, ..., \frac{r_t - \hat{\mu}_t}{\hat{\sigma}_t}\right).$$

Several studies conclude that conditional EVT is the method with better out-of-sample performance to forecast one-day-ahead VaR (e.g. McNeil and Frey (2000), Byström (2004), Bekiros and Georgoutsos (2005), Kuester et al. (2006), Ghorbel and Trabelsi (2008), Ozun et al. (2010)), and this is the reason wy we choose CEVT-n and CEVT-sst models for the comparative studies.

Empirical evidence shows that the increase in volatility is larger when the returns are negative than when they are positive. This asymmetric evolution of the conditional variance is known as leverage effect (Black, 1976). We also choose for the comparative study one asymmetric GARCH-type model, the APARCH model introduced by Ding, Granger and Engle (1993). The conditional variance of the APARCH$(p, q)$ model can be written as

$$\sigma_t^\delta = w + \sum_{i=1}^{p} \alpha_i(|\varepsilon_{t-i}| - \gamma_i\varepsilon_{t-i})^\delta + \sum_{j=1}^{q} \beta_j\sigma_{t-j}^\delta,$$

where $\delta > 0$ and $-1 < \gamma < 1$. The asymmetric coefficient $\gamma$, takes the leverage effect into account. We consider this model as it is a very general GARCH-type model, including as special cases several GARCH-type models and asymmetric GARCH-type models: ARCH Model of Engle ($\delta = 2$, $\gamma_i = 0$ and $\beta_j = 0$), GARCH Model of Bollerslev ($\delta = 2$, $\gamma_i = 0$), TS-GARCH Model of Taylor and Schwert ($\delta = 2$, $\gamma_i = 0$), GJR-GARCH Model of Glosten, Jagannathan and Runkle ($\delta = 2$), T-ARCH Model of Zakoian ($\delta = 1$), N-ARCH Model of Higgens and Bera ($\gamma_i = 0$ and $\beta_j = 0$) and log-ARCH Model of Geweke and Pentula ($\delta \to 0$). The model chosen for the comparative studies was the AR-APARCH(1,1) with skewed-$t$ innovations, which we denote by APARCH-sst.

$$\widehat{\mathrm{VaR}}_{t+1|t}^{APARCH}(p) = \hat{\phi}_0 + \hat{\phi}_1 r_t + s_p \times \hat{\sigma}_{t+1|t},$$

with $s_p$ a quantile $p$ of the skewed-$t$ distribution with parameters estimated using the data. In a comparative study for the Asian markets, Tu, Wong and Chang (2008) found that the APARCH model with the skewed-$t$ distribution performs better than with the normal or with the student distribution. GARCH-type models with skewed-$t$ innovations have been frequently found to provide excellent forecast results; see, for example, Mittnik and Paolella (2000), Giot and Laurent (2004).

Finally, for the comparative study, we also choose the widely used RiskMetrics model developed by J.P. Morgan (J.P. Morgan´s Riskmetrics Technical Document, 1996). This model assumes that the return follows a conditional normal distribution $\mathrm{N}(0, \sigma_t^2)$, with the dynamic of volatility modeled using a exponential weighted moving average (EWMA) method

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1-\lambda)\varepsilon_{t-1}^2.$$

RiskMetrics (1996) suggests $\lambda = 0.94$ for daily data. The recursion can be initialized by the sample variance ($\sigma_1^2 = \hat{\sigma}^2$) or the square of the first return ($\sigma_1^2 = r_1^2$).

$$\widehat{\mathrm{VaR}}_{t+1|t}^{RM}(p) = z_p \times \hat{\sigma}_{t+1|t},$$

with $z_p$ a quantile $p$ of the standard normal distribution. The empirical results of the following Section will clearly suggest that with the normality assumption we obtain underestimated VaR forecasts and more violations than the expected.

### 6.4.2 Out-of-Sample studies with SP 500, DAX and FTSE indexes

In this Section we compare the CEVT-sst, CEVT-n, APARCH-sst, RiskMetrics and DPOT models with $v = 3$, $c \in \{0.8, 0.75, 0.7\}$, denoted respectively by DPOT(0.8), DPOT(0.75) and DPOT(0.7). We examine the one-day-ahead VaR(0.01) forecasts performance with the S&P 500 Index, DAX Index and FTSE Index, considering returns produced by all the historical data until May 18, 2010. Using a rolling window of size 1000 we obtain 14190, 3917 and 5599 one-day-ahead VaR(0.01) forecasts for each model, respectively with the S&P 500, DAX and FTSE. As usual, the threshold $u$ was chosen such that 10% of the values are larger than the threshold. The primary tool for assessing the accuracy of the interval forecasts is to monitor the binary sequence generated by observing if the return on day $t+1$ is in the tail region specified by the VaR at time-$t$, or not. This is referred to as the hit sequence (1.2.3). To test the UC hypothesis we apply the Kupiec test (Kupiec, 1995). To test the IND hypothesis we apply two tests. In the same line as Engle and Manganelli (2004), Berkowitz *et al.* (2009) consider the autoregression

$$I_t = \alpha + \beta_1 I_{t-1} + \beta_2 \mathrm{VaR}_{t|t-1}(p) + \varepsilon_t, \tag{6.4.1}$$

and propose the logit model. We can test the IND hypothesis with a likelihood ratio test considering for the null $\beta_1 = \beta_2 = 0$ and in this case the asymptotic distribution is chi-square with 2 degrees of freedom. We refer to this test as the CAViaR independence test of Engle and Manganelli (CAViaR). The other independence test applied was recently introduced in the literature (Araújo Santos and Fraga Alves, 2010) and is based on the ratio $(D_{N:N} - 1)/D_{[N/2]:N}$, where $D_{N:N}$ and $D_{[N/2]:N}$, are the maximum and the median of durations between consecutive violations and until the first violation. This new test is suitable for detect models with a tendency to generate clusters of violations, is based on an exact distribution, is pivotal in the sense that is based on a distribution that does not depend on an unknown parameter and outperforms, in terms of power, existing procedures in realistic settings. We refer to this test as MM ratio test.

The empirical findings, with the p values of the tests, are presented in Tables 6.2, 6.3 and 6.4. Table 6.5, summarize the results in terms of number of times that the hypotheses are rejected. As the unconditional POT model do not account for volatility clustering, is unable to produce iid violations and both independence tests reject the IND hypothesis with very small p values. With a violation frequency equal to 0.01367, the UC hypothesis is also clearly reject in the case of the POT model. We only considered the POT method in the Table 6.2. The performance of RiskMetrics is very poor. The violation frequency is even much worse than with the unconditional POT model. With RiskMetrics the violation frequencies equals 0.018675, 0.016845 and 0.018222, respectively for the S&P 500, DAX and FTSE indexes, much higher than the expected 0.01. For this model and with all indexes the UC hypothesis is rejected with very small $p$-values. Both DPOT, CEVT and APARCH-sst models performs very well in terms of the UC hypothesis, taking into account that in no case the hypothesis is rejected since all $p$-values are very high. It is interesting to note the impressive performance of CEVT models in terms of UC in Table 6.2, with 142 violations in 14190 out-of-sample forecasts it was impossible to obtain a better result (the violation frequency is equal to 0.01000705). The same impressive performance occurs with the DPOT(0.7) in Table 6.4, with 56 violations in 5599 out-of-sample forecasts was impossible to obtain a better result (the violation frequency is equal to 0.01000179). In terms of IND hypothesis, the DPOT models performs clearly better than the CEVT models and than APARCH-sst. Considering the eighteen cases with three DPOT models, three indexes and two independence tests, with DPOT models the IND hypothesis is rejected only in one case. For the CEVT-n, CEVT-sst and APARCH-sst models, the IND hypothesis is rejected, respectively, 3, 3 and 4 times. Table 6.5, summarize these results. This empirical evidence shows that the DPOT models can be successful in removing the tendency to clustering of violations, which was our main objective, can perform better than state of the art risk models and much better than the widely used RiskMetrics model.

Table 6.2
Out-of-sample accuracy for VaR(0.01) applied to S&P 500 Index returns from January 4, 1950 until May 18, 2010, with a rolling window of size 1000. Unconditional coverage and independence tests.

| Model | Violation frequencies | Kupiec $p$-value | CAViaR $p$-value | MM Ratio $p$-value |
|---|---|---|---|---|
| POT | 0.013672 | 0.0000 | 0.0000 | 0.0000 |
| DPOT(0.8) | 0.009725 | 0.7410 | 0.0189 | 0.7902 |
| DPOT(0.75) | 0.009443 | 0.5011 | 0.1018 | 0.1048 |
| DPOT(0.7) | 0.009443 | 0.5011 | 0.8659 | 0.0566 |
| CEVT-n | 0.010007 | 0.9933 | 0.0145 | 0.0166 |
| CEVT-sst | 0.010007 | 0.9933 | 0.0236 | 0.0314 |
| APARCH-sst | 0.009015 | 0.2305 | 0.0064 | 0.0717 |
| RiskMetrics | 0.018393 | 0.0000 | 0.0000 | 0.3401 |

Table 6.3
Out-of-sample accuracy for VaR(0.01) applied to DAX Index returns from November 27, 1990 until May 18, 2010, with a rolling window of size 1000. Unconditional coverage and independence tests.

| Model | Violation frequencies | Kupiec $p$-value | CAViaR $p$-value | MM Ratio $p$-value |
|---|---|---|---|---|
| DPOT(0.8) | 0.008425 | 0.3085 | 0.6918 | 0.8821 |
| DPOT(0.75) | 0.008935 | 0.4953 | 0.7175 | 0.8597 |
| DPOT(0.7) | 0.010722 | 0.6533 | 0.2786 | 0.1886 |
| CEVT-n | 0.010467 | 0.7706 | 0.0156 | 0.6180 |
| CEVT-sst | 0.009446 | 0.7250 | 0.0030 | 0.7227 |
| APARCH-sst | 0.009191 | 0.6058 | 0.0079 | 0.0358 |
| RiskMetrics | 0.016083 | 0.0004 | 0.0936 | 0.5245 |

Table 6.4
Out-of-sample accuracy for VaR(0.01) applied to FTSE Index returns from April 3, 1984 until May 18, 2010, with a rolling window of size 1000. Unconditional coverage and independence tests.

| Model | Violation frequencies | Kupiec $p$-value | CAViaR $p$-value | MM Ratio $p$-value |
|---|---|---|---|---|
| DPOT(0.8) | 0.009109 | 0.4962 | 0.1033 | 0.6359 |
| DPOT(0.75) | 0.009109 | 0.4962 | 0.3405 | 0.6373 |
| DPOT(0.7) | 0.010002 | 0.9989 | 0.8646 | 0.5687 |
| CEVT-n | 0.011073 | 0.4275 | 0.4037 | 0.7410 |
| CEVT-sst | 0.011073 | 0.4275 | 0.4143 | 0.7423 |
| APARCH-sst | 0.008573 | 0.2143 | 0.0047 | 0.2338 |
| RiskMetrics | 0.018575 | 0.0000 | 0.2704 | 0.5607 |

Table 6.5
Number of rejections of the UC and IND hypotheses with significance level equal to 0.05.

| | Number of rejections | |
|---|---|---|
| | UC hypothesis | IND hypothesis |
| DPOT(0.8) | 0 | 1 |
| DPOT(0.75) | 0 | 0 |
| DPOT(0.7) | 0 | 0 |
| CEVT-n | 0 | 3 |
| CEVT-sst | 0 | 3 |
| APARCH-sst | 0 | 4 |
| RiskMetrics | 3 | 1 |

### 6.4.3 Minimization of capital requirements under the Basel II Accord

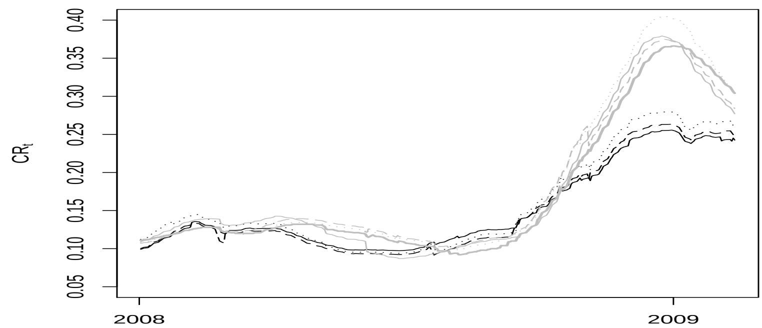
Under the Basel II Accord, ADIs have to communicate their daily risk forecasts to the monetary authority (typically a central bank) at the beginning of the trading day, using a VaR model. Too high forecasts will lead to large capital requirements. On the other hand, too low forecasts will lead to excessive violations and consequently to a penalty that increases capital requirements. The penalty can be an increase in a multiplicative factor to calculate capital requirements or the imposition of a standard model when the number of violations exceeds 10. Let us consider an ADI that invest at day $t+1$ an amount $\mathrm{A}_{t+1}$ in a portfolio of risky assets. The portfolio is financed by deposits ($\mathrm{D}_{t+1}$) and equity ($\mathrm{E}_{t+1}$). At day $t+1$ the ADI must satisfy capital requirements for market risk ($\mathrm{CR}_{t+1}$) such that $\mathrm{E}_{t+1} \geq \mathrm{CR}_{t+1}\mathrm{A}_{t+1}$. Note that for a given $CR_{t+1}$, to satisfy this inequality the ADI can increase the equity or reduce the amount invested. Of course, even without this rule, risk averse investors will reduce this amount during periods of high risk. The Basel II Accord stipulates $\mathrm{CR}_{t+1}$ as

$$\mathrm{CR}_{t+1} = \sup\Big\{(3+k)\overline{\mathrm{VaR}}_{60}, \mathrm{VaR}_t\Big\}, \tag{6.4.2}$$

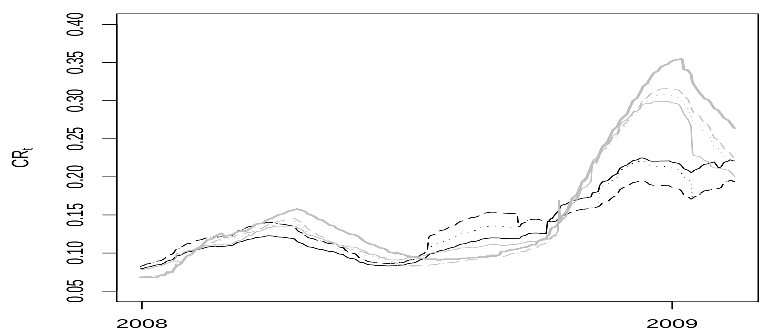
where $\overline{\mathrm{VaR}}_{60}$ is the average VaR over the previous 60 trading day´s and $k$ is a multiplicative factor that depends on the number of violations in the previous 250 trading days ($N_v$), according to the following function,

$$k = \begin{cases} 0 & \text{if } N_v \leq 4 \\ 0.3 + 0.1(N_v - 4) & \text{if } 5 \leq N_v \leq 6 \\ 0.65 & \text{if } N_v = 7 \\ 0.65 + 0.1(N_v - 7) & \text{if } 8 \leq N_v \leq 9 \\ 1 & \text{if } N_v = 10. \end{cases}$$

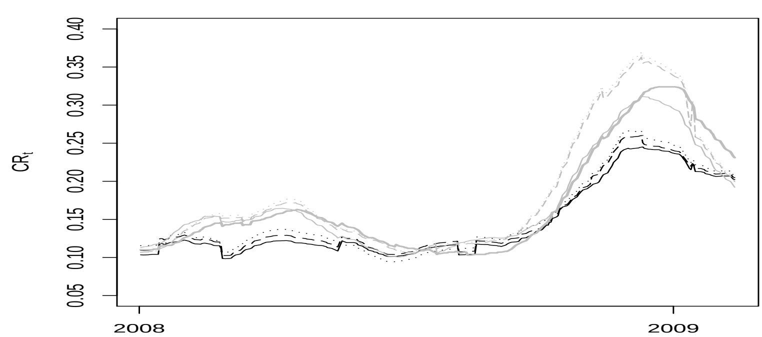
In the same way as McAleer et al. (2009), we can write the ADI profit for day $t+1$ as $\Pi_{t+1} = r_{At+1}A_{t+1} - r_{Dt+1}D_{t+1} - r_{Et+1}E_{t+1}$, where $r_{At+1}$ denotes the return on the ADI portfolio on day $t+1$, $r_{Dt+1}$ the rate for deposits on day $t+1$ and $r_{Et+1}$ the cost of holding equity. An increase in $\mathrm{E}_{t+1}$ will reduce expected profits and for that reason an ADI is interested in the minimization of $\mathrm{CR}_{t+1}$. In a recent work, McAleer et al. (2009c) compare, in terms of minimization of capital requirements, well known and widely used time-varying volatility models applied in one-day-ahead VaR forecast. These authors advanced the idea and conclude that optimal risk management within the Basel II Accord requires to use combinations of models. In this Section we choose the S&P 500 index, DAX index and FTSE index returns for the period, January 2, 2008, to February 12, 2009, which includes the global financial crisis, taking into account the comparability with this previous study. Using equation (6.4.2) and the DPOT, CEVT, APARCH-sst and RiskMetrics models, we calculated $\mathrm{CR}_t$ for each model over this period and for each index. In Figure 6.4.1, the dotted, longdash, solid, dotted grey, longdash grey, solid grey and bold grey lines, correspond to daily capital requirements obtained respectively with the DPOT(0.8), DPOT(0.75), DPOT(0.7), CEVT-n, CEVT-sst, APARCH-sst and RiskMetrics models. From Figure 6.4.1, it is evident that, for these indexes and this period, the DPOT models perform much better than the other models under study. Only in few days and with very small differences, the DPOT models produced higher capital requirements and this is mainly before the high volatile period, suggesting that DPOT models anticipate better these periods than the other models. Tables 6.6, 6.7 and 6.8 gives the maximum number of violations in the previous 250 trading days and the average capital requirements. In terms of number of violations, the DPOT models with $c = 0.8$ and $c = 0.75$ perform better than with $c = 0.7$ with which we had 11 violations in the previous 250 trading days. Although this occurs during a very severe crisis, exceeds 10 violations and falls in the red zone defined by the Basel II Accord. In terms of capital requirements, Tables 6.6, 6.7 and 6.8 show that, in the period under study, the DPOT models lead to substantially lower average capital requirements than the other models under study. The differences are in the majority of cases higher than 200 basis points and in some cases higher than 300 basis points.




(a) S&P 500 index




(b) DAX index




(c) FTSE index

Figure 6.4.1: Daily capital requirements ($CR_t$)between 2 January 2008 and 12 February 2009, under the Basel II Accord, applying the DPOT(0.8) (dotted), DPOT(0.75) (longdash), DPOT(0.7) (solid), CEVT-n (dotted grey), CEVT-sst (longdash grey), APARCH-sst (solid grey) and RiskMetrics (bold grey) models.

Table 6.6
Maximum $N_v$ and average $CR_t$ for S&P 500 index from January 2, 2008 until February 12, 2009.

| Model | Maximum $N_v$ | Average capital requirements ($CR_t$) |
|---|---|---|
| DPOT(0.8) | 8 | 0.1583 |
| DPOT(0.75) | 8 | 0.1495 |
| DPOT(0.7) | 9 | 0.1505 |
| CEVT-n | 10 | 0.1825 |
| CEVT-sst | 8 | 0.1781 |
| APARCH-sst | 7 | 0.1739 |
| RiskMetrics | 11 | 0.1715 |

Table 6.7
Maximum $N_v$ and average $CR_t$ for DAX index from January 2, 2008 through February 12, 2009.

| Model | Maximum $N_v$ | Average capital requirements ($CR_t$) |
|---|---|---|
| DPOT(0.8) | 5 | 0.1385 |
| DPOT(0.75) | 5 | 0.1373 |
| DPOT(0.7) | 11 | 0.1351 |
| CEVT-n | 4 | 0.1457 |
| CEVT-sst | 4 | 0.1484 |
| APARCH-sst | 5 | 0.1474 |
| RiskMetrics | 11 | 0.1601 |

Table 6.8
Maximum $N_v$ and average $CR_t$ for FTSE index from January 2, 2008 until February 12, 2009.

| Model | Maximum $N_v$ | Average capital requirements ($CR_t$) |
|---|---|---|
| DPOT(0.8) | 9 | 0.1522 |
| DPOT(0.75) | 10 | 0.1500 |
| DPOT(0.7) | 11 | 0.1446 |
| CEVT-n | 9 | 0.1850 |
| CEVT-sst | 9 | 0.1821 |
| APARCH-sst | 7 | 0.1688 |
| RiskMetrics | 11 | 0.1705 |

## 6.5 Conclusions

In this work we propose a POT method that uses the durations between excesses as covariates. Based on this method, three DPOT models for forecasting one-day-ahead VaR were compared with other models. Empirical findings presented in Section 6.4.2 show that they perform very well in terms of unconditional coverage and better than state-of-the art models in terms of removing the tendency to clustering of violations. In terms of out-of-sample accuracy, DPOT models perform much better than the widely used RiskMetrics model. Moreover, the empirical findings presented in Section 6.4.3, suggest that the DPOT models can have an important role in the minimization of capital requirements under the Basel II Accord. In the period under study, the DPOT models lead to substantially lower average capital requirements. It is possible that we can achieve lower average capital requirements by integrating DPOT in a combination of models strategy or, for example, in a dynamic learning strategy such as the one proposed by McAleer et al. (2009). The study of these issues remains for future research. Finally, we notice that in order to deal with the volatility clustering, the proposed models do not assume a parametric distribution for the entire distribution of the returns, as the CEVT or GARCH-type models, but assumes a parametric model only on the tail and based on solid asymptotic theory.

## Acknowledgements

A preliminary version of this paper was presented at the 30th International Symposium on Forecasting, at San Diego CA, USA, June 2010, and at the 4rd International Conference on Computational and Financial Econometrics and 3rd Workshop of the ERCIM Working Group on Computing and Statistics, at University of London, UK, December 2010. We thank Juan-Ángel Jiménez-Martin, Pilar Abad, Sonia Benito and other participants, for useful suggestions and comments. Research partially supported by Fundação para a Ciência e Tecnologia (FCT/PROTEC and FCT/OE) and Center of Statistics and Applications of the University of Lisbon (CEAUL).

*Paulo Araújo Santos and Isabel Fraga Alves*

# 7

# Extremal Quantiles Estimation with quasi-PORT and DPOT Methodologies - an application to Value-at-Risk

Under the context of high quantiles, Value-at-Risk (VaR) models based on the PORT Hill estimator, VaR models based on the DPOT method and other unconditional and conditional models are compared through a out-of-sample accuracy study. To obtain a reasonable number of violatios for backtesting, the log returns from the Down Jones Industrial Average index, which constitute a financial time series with a very large data size, were used.

## 7.1 Introduction

In this Chapter we are concerned with extraordinary events in financial markets - the so known as “Black Swans” events - such as the Black Thursday (stock market crash on 24 October, 1929), the Black Monday (stock market crash on 19 October, 1987), the turmoil in the bond market in February 1994 and the recent 2008 financial crisis. These crisis are characterized by extreme price changes and a major concern for regulators and owners of financial institutions is the adequacy of capital to ensure that they can still be in business after such extreme price changes. VaR defined in (1.2.2), emerged as the primary tool for financial risk assessment. Here, we are dealing with rare events and thus with much lower probabilities than the usual $p = 0.01$ used for daily capital requirements calculations under the Basel II Accord. In this Chapter it will be considered the probability of an adverse extreme price movement that is expected to occur approximately once every four years (p=0.001) or once every eight years (p=0.0005); therefore, we fall in the context of high quantiles defined in (1.1.2). This context may have interest in the development of stress tests (e.g., Longin, 2001; Tsay, 2010), which are directly related to the occurrence of extremes in financial markets. Some authors (e.g., Danielsson and Vries, 1997) argued that when small probabilities come into play, an unconditional approach is better suited for VaR estimation, because extreme price changes do not appear to be related to a particular level of volatility, nor exhibit time dependence. In fact, it is demonstrated by de Haan, Resnick, Rootzén and de Vries (1989), that for certain dependent processes, such as ARCH, volatility clustering vanishes at the level of extremes. Moreover, Resnick and Starica (1996) have shown the consistency of the Hill estimator under certain types of dependence, such as GARCH.

In this Chapter, both unconditional and conditional VaR models are compared. We have chosen two unconditional VaR models based on the PORT Hill estimator (2.2.3) proposed in Chapter 2 and two conditional VaR models based on the DPOT methodology proposed in Chapter 6. Additionally, other unconditional and conditional models are also used in the comparisons. In Section 7.2, a recent approach in EVT, involving the reduction of bias, is briefly reviewed and the VaR methods used in the comparative study are summarized. In Section 7.3, the results of the comparative out-of-sample study are presented and conclusions.

## 7.2 VaR models

For the out-of-sample study, the following models were considered.

### 7.2.1 Quasi-PORT

The Hill estimator for the tail index (Hill, 1975), presented in (2.1.6) and denoted by $\hat{\gamma}^H_{n,k}$, may exhibit a high asymptotic bias, i.e., as $n \to \infty$, $\sqrt{k}(\hat{\gamma}^H_{n,k} - \gamma)$ is asymptotically normal with variance $\gamma^2$ and a non-null mean value, equal to $\lambda_A/(1-\rho)$, whenever $\sqrt{k}A(n/k) \to \lambda_A \neq 0$, finite, with $A(.)$ the function in (2.1.3). This non-null asymptotic bias, together with a rate of convergence of the order of $1/\sqrt{k}$, leads to sample paths with a high bias for large $k$ and high variance for small $k$. Recent developments in EVT, involve the reduction of bias (see Peng (1998), Beirlant, Dierckx, Goegebeur and Matthys (1999), Feuerverger and Hall (1999), Gomes, Martins and Neves (2000, 2002b), Gomes and Martins (2001), Caeiro and Gomes (2002), Gomes, Figueiredo and Mendonça (2004), among others). They achieved $\gamma$ estimators with asymptotic variance equal or higher than $(\gamma(1-\rho)/\rho)^2 > \gamma^2$. More recently, Caeiro, Gomes and Pestana (2005), Gomes and Pestana (2007a), Gomes, Martins e Neves (2007b) and Gomes, de Haan and Henriques-Rodrigues (2008b), have proposed minimum variance reduced bias (MVRB) estimators for $\gamma$. They reduce bias without increasing the asymptotic variance, which is kept at the value $\gamma^2$. A simple class of MVRB-estimators is the one introduced in Caeiro, Gomes and Pestana (2005), studied in Chapter 3 and already presented in (3.1.5), with the functional form that we recall here,

$$\hat{\gamma}^{\overline{H}}_{n,k,\hat{\beta},\hat{\rho}} := \hat{\gamma}^{H}_{n}\Big(1 - \frac{\hat{\beta}}{1-\hat{\rho}}\Big(\frac{n}{k}\Big)^{\hat{\rho}}\Big),$$

where $H(k)$ is the Hill estimator and $\hat{\rho}$ and $\hat{\beta}$ are consistent estimators of the second order parameters $\rho$ and $\beta$. See Fraga Alves, Gomes and de Hann (2003) for $\rho$ estimation and Gomes, de Haan and Henriques-Rodrigues (2008b) for $\beta$ estimation.

The MVRB tail index estimators in (3.1.5) are not location invariant, but they are much less sensitive to changes in location than the classical Hill estimator, thus, they are "approximately" location invariant. Gomes, Figueiredo, Henriques-Rodrigues and Miranda (2010) have proposed to use the PORT Hill estimator (2.2.3) instead of the Hill estimator (2.1.6) in the MVRB estimator (3.1.5). This estimator was named "quasi-PORT" tail index estimator and its functional form is

$$\hat{\gamma}^{\overline{H}(q)}_{n,k,\hat{\beta},\hat{\rho}} := \hat{\gamma}^{H(q)}_{n,k}\left\{1 - \frac{\hat{\beta}}{1-\hat{\rho}}\left(\frac{n}{k}\right)^{\hat{\rho}}\right\}, \qquad 0 \le q < 1, \tag{7.2.1}$$

where $H^{(q)}(k)$ is the PORT-Hill estimator (2.2.3), and $\hat{\rho}$ and $\hat{\beta}$ are consistent estimators of the second order parameters $\rho$ and $\beta$. For the case of high quantiles, Gomes, Figueiredo, Henriques-Rodrigues and Miranda (2010) proposed to use the "quasi-PORT" tail index estimator (7.2.1) instead of the PORT Hill estimator (3.1.5) in the PORT-Weissman-Hill high quantile estimator (2.3.1). This estimator was named "quasi-PORT" $\text{VaR}_p$ estimator and its functional form is

$$\widehat{\chi}^{\overline{H}(q)}_{p_n} := (X_{n-k_n:n} - X_{n_q:n})\left(\frac{k_n}{np_n}\right)^{\hat{\gamma}^{\overline{H}(q)}_{n,k,\hat{\beta},\hat{\rho}}} + X_{n_q:n}, \quad 0 \le q < 1. \tag{7.2.2}$$

With $q = 0.25$ and $q = 0.5$, two unconditional VaR models based on the estimator (7.2.2) were chosen. The estimates of $\rho$ and $\beta$ were obtained using the algorithm suggested in Gomes and Pestana (2007).

### 7.2.2 DPOT

In Section 6.3, a duration based POT model was proposed and the out-of-sample performance was compared with other models, for forecasting one-day-ahead VaR(0.01) denoted by $\text{VaR}_{t+1|t}(0.01)$. This is the VaR used by financial institutions to compute daily capital requirements under the Basel II Accord. Here, for very small values of $p$, the DPOT model was used in the comparative study of Section 7.3 with the most simple specification ($v = 1, c = 1$) and with the specification with better out-of-sample results in the comparative studies of Chapter 6 ($v = 3, c = 0.75$).

### 7.2.3 Other models

We have chosen more three models from EVT, the unconditional POT model presented in (6.2.4), the conditional EVT models presented in (6.4.1) and denoted respectively by CEVT-n and CEVT-sst. Finally, three parametric conditional models were used in the study, the RiskMetrics model (6.4.1), the AR-APARCH model (6.4.1) with normal innovations (APARCH-n) and the AR-APARCH model (6.4.1) with skewed $t$ innovations (APARCH-sst).

## 7.3 Out-of-Sample study with the DJIA index

Under the context of high quantiles, we set $p = 0.001$ and $p = 0.0005$. To achieve a reasonable number of violations for backtesting, it is important to have a very large data size and this lead us to use the log returns of the Down Jones Industrial Average index, one of the oldest stock indexes. From October 2, 1928, until March 25, 2011, we compute 20713 returns and with a moving windows of size $n_w = 1000$ days, we obtain 19713 one-day-ahead VaR forecasts for each model. As in previous studies, for the EVT methods, we choose the number of top order statistics $k = 100$; see McNeil and Frey (2000) for a simulation study that supports a similar choice. To test the UC hypothesis we apply the Kupiec test (Kupiec, 1995) and to test the IND hypothesis we apply the maximum to median ratio test (Araújo Santos and Fraga Alves, 2010) denoted by MM independence test. The programs were written in the R language and with the fGarch (Chalabi,Wuertz e Miklovic, 2008) and POT (Ribatet, 2009) packages.

Tables 7.1 and 7.2, summarize the results respectively for $p = 0.001$ and $p = 0.0005$. The APARCH-sst based on Skwed-t errors perform well in terms of UC under $p = 0.001$. Empirical findings show that the Skewed-t is clearly preferable than the normal for the distribution of the errors. The performance of conditional parametric models based on the normal distribution (RiskMetrics and APARCH-n) is disastrous with the number of violations exceeding more than five times the expected under UC.

Table 7.1
Out-of-sample accuracy for VaR(0.001) applied to Down Jones Industrial Average index returns from October 2, 1928 until March 25, 2011, with a rolling window of size 1000.

| Model | Number of violations | Violation frequencies | Kupiec $p$-value | MM Ratio $p$-value |
|---|---|---|---|---|
| Unconditional EVT models: | | | | |
| POT | 36 | 0.001826 | 0.0000 | 0.1254 |
| Quasi-PORT($q = 0.25$) | 31 | 0.001573 | 0.0190 | 0.1048 |
| Quasi-PORT($q = 0.5$) | **20** | **0.001015** | **0.9486** | **0.1849** |
| Conditional EVT models: | | | | |
| DPOT($v = 1$) | **22** | **0.001116** | **0.6130** | **0.1966** |
| DPOT($v = 3$) | 32 | 0.001623 | 0.0112 | 0.5849 |
| CEVT-n | 31 | 0.001573 | 0.0190 | 0.8919 |
| CEVT-sst | 31 | 0.001573 | 0.0190 | 0.9631 |
| Conditional parametric models: | | | | |
| RiskMetrics | 128 | 0.006493 | 0.0000 | 0.0015 |
| APARCH- n | 101 | 0.005124 | 0.0000 | 0.0141 |
| APARCH- sst | **22** | **0.001116** | **0.6130** | 0.0564 |

Note to Table 7.1: For each model, the number of one-day-ahead VaR(**0.001**) forecasts is 19713 and the expected number of violations under the UC hypothesis is **19.713**.

Table 7.2
Out-of-sample accuracy for VaR(0.0005) applied to Down Jones Industrial Average index returns from October 2, 1928 until March 25, 2011, with a rolling window of size 1000.

| Model | Number of violations | Violation frequencies | Kupiec $p$-value | MM Ratio $p$-value |
|---|---|---|---|---|
| Unconditional EVT models: | | | | |
| POT | 27 | 0.001370 | 0.0000 | 0.2102 |
| Quasi-PORT($q = 0.25$) | 14 | 0.000710 | 0.2146 | 0.0981 |
| Quasi-PORT($q = 0.5$) | **11** | **0.000558** | **0.7206** | **0.1562** |
| Conditional EVT models: | | | | |
| DPOT($v = 1$) | **10** | **0.000507** | **0.9636** | **0.2035** |
| DPOT($v = 3$) | 24 | 0.001217 | 0.0001 | 0.1048 |
| CEVT-n | 24 | 0.001217 | 0.0001 | 0.7834 |
| CEVT-sst | 25 | 0.001268 | 0.0001 | 0.9922 |
| Conditional parametric models: | | | | |
| RiskMetrics | 101 | 0.005124 | 0.0000 | 0.0288 |
| APARCH- n | 74 | 0.003754 | 0.0000 | 0.0105 |
| APARCH- sst | 16 | 0.000812 | 0.0729 | 0.0799 |

Note to Table 7.2: For each model, the number of one-day-ahead VaR(**0.0005**) forecasts is 19713 and the expected number of violations under the UC hypothesis is **9.8565**.

Further to the tail, more disastrous are the results. With the smaller probability level $p = 0.0005$, RiskMetrics produced 101 violations which represents more than ten times the expected value equal to 9.8565, under UC. The APARCH-n model produced 74 violations, more than seven times the expected. These results confirm what is well known in the literature (see for instance Danielsson and Vries, 1997). On the other hand, the accuracy of the best performers Quasi-PORT($q = 0.5$) and DPOT($v = 1$) is very good, with the number of violations very close to the expected under UC. These two models have also good results in terms of independence. In the group of EVT models they perform clearly better than the classic POT model and than the CEVT hybrid model. Finally, it is interesting to note that one of the best performers, Quasi-PORT($q = 0.5$), is based on the iid assumption and this provides evidence that the iid assumption can work well when we are dealing with very small probability levels.

As future research, we plan to extend the out-of-sample study presented in this Chapter, to other indexes and other types of large financial time series, such as individual stocks and foreign currencies.

## Acknowledgements

Research partially supported by Fundação para a Ciência e Tecnologia (FCT/PROTEC and FCT/OE) and Center of Statistics and Applications of University of Lisbon (CEAUL).

# A

# List of programs

## A.1 Interval forecasts evaluation: R programs for the new independence test proposed in Chapter 4 and comparisons

Subsection A.1.1, presents a R program for the implementation of the independence test proposed in Chapter 4, with any hit sequence. Subsection A.1.2, presents a R program for a comparative simulation study with Markov, CAViaR and GMM independence tests. In Subsection A.1.3, the table needed for the program of the Subsection 2.2, is given.

### A.1.1 R program for the new independence test

```
## Example of a hit sequence:
hit<-c(0,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0)
## Or use the following code to read the hit sequence from a text
file: ## hit <- read.table("hit.txt")[,1]

tt <- length(hit) #### Durations
no_hit_duration <- 0
j<-1
zeros <- 0
for(i in 1:tt) {
if (hit[i]<1){
zeros <- zeros+1
 }
 else {
 no_hit_duration[j]<- zeros+1
 zeros <- 0
 j <- j+1
  }
 }
no <- no_hit_duration
n <- length(no)

#### T[0.5] Independence Test
no <- sort(no)
observed_T <-(no[n]-1)/no[floor(0.5*n)]
observed_T

#### Simulation of the p value upper bound
```

```
v <- 0
replicas <- 250000
v <- rep(0, times=replicas)
print('wait for p value upper bound simulation')
for(i in 1:replicas) {
if((i/10000-floor(i/10000))==0){print(replicas-i)}
 u <- runif(n)
 y <- -log(1-u)
 no_simul <- sort(y)
 v[i] <- log(2)*(no_simul[n]/no_simul[floor(n/2)])-log(n)
 }

simulated_p_value_upper_bound <- length(v[v>=observed_T])/replicas
observed_T
simulated_p_value_upper_bound
```

### A.1.2 R program for the comparative simulation study

```
library(fGarch)
table <- read.table("table_T50.txt")[,2]

replicas <- 5000
tt <- 500 ## size of the hit sequence
ws <- 500 ##window size
coverage <- 0.01
v1 <- 0
v2 <- 0
v3 <- 0
v4 <- 0
var <- 0
reject_freq <- 0
failures <-0

for(t in 1:replicas) {
print (t)
##############################  MODEL1: Gaussian GARCH(1,1) #######################
##   model = garchSpec(model = list(omega = 0.05, alpha = 0.1, beta = 0.85))
##   a <- garchSim(model, n = tt+ws)
##################################################################################

#############################  MODEL2: Skewed t APARCH(1,1) #######################
     model = garchSpec(model = list(mu = 0, omega = 0.03,
     alpha = c(0.086), gamma = c(0.64), beta = 0.91, delta = 1.15,
     shape = 10, skew=0.88), cond.dist = "sstd")
     a <- garchSim(model, n = tt+ws)
##################################################################################

#### Hit function
 hit <-runif(tt)
 for(i in 1:tt) {
 iws <- i+ws
 m_iws <- iws-1
 b <- a[i:m_iws]
 th <- quantile(b, probs=coverage)
 var[i] <- th
 if(a[i+ws]<th){hit[i]=1}
 else {hit[i]=0}
 }

#### Durations
 no_hit_duration <- 0
 j<-1
 zeros <- 0
 for(i in 1:tt) {
 if (hit[i]<1){ zeros <- zeros+1 }
 else {
 no_hit_duration[j]<- zeros+1
```

```
 zeros <- 0
 j <- j+1
  }
 }
 no <- no_hit_duration
 n <-length(no)

#### Exclude samples with size less than 2
 if (n<2){
 v1[t] <- -1
 v2[t] <- -1
 v3[t] <- -1
 v4[t] <- -1
 failures <- failures+1 }
 else{

#### T[0.5] Independence Test
 no <- sort(no)
 p50 <- no[floor(0.5*n)]
 max <- no[n]
 observed_T  <- (max-1)/p50
 if (observed_T  > table[n]){reject_freq <- reject_freq+1}

#### Markov Independence Test
 zz <- 0
 umz <- 0
 zum <- 0
 umum <- 0
 m_tt <- tt-1

 for(k in 1:m_tt) {
 i<-k+1
 if (hit[k]==0 & hit[i]==0){
 zz <- zz +1
 }
 else if (hit[k]==0 & hit[i]==1){
 zum <- zum +1
 }
 else if (hit[k]==1 & hit[i]==1){
 umum <- umum +1
 }
 else{
 umz <- umz +1
  }
 }

p00 <- zz/(zz+zum)
p01 <- zum/(zz+zum)
p10 <- umz/(umz+umum)
p11 <- umum/(umz+umum)
llp <- (zum+umum)/(zz+umz+zum+umum)
ll2 <- ((1-llp)^(zz+umz))*(llp^(zum+umum))
ll1 <- (p00^zz)*(p01^zum)*(p10^umz)*(p11^umum)
v1[t] <- -2*log(ll2/ll1)

#### Caviar Independence Test
hit1 <- hit[1:m_tt]
hit2 <- hit[2:tt]
var2 <- var[2:tt]

mylogit <- glm(hit2~hit1+var2,
family=binomial(link="logit"),na.action=na.pass) logLik(mylogit)

alpha <- -log(length(hit)/sum(hit)-1)
loglik1 <- -sum(1-hit2)*alpha-(tt-1)*log(1+exp(-alpha))

emv <- mylogit$coefficients
emv1 <- emv[1]
emv2 <- emv[2]
emv3 <- emv[3]
```

```
loglik2 <- -sum((1-hit2)*(emv1+emv2*hit1+emv3*var2))-sum(log(1+exp(-emv1-emv2*hit1-emv3*var2)))
v2[t]<- -2*(loglik1-loglik2)

##### GMM Independence Tests
p <- n/tt
m1 <- (1-pp*no)/sqrt(1-pp)
m2 <- (3-pp-pp*no)*(1-pp*no)/(2-2*pp)-0.5
m3 <-(5-2*pp-pp*no)/(3*sqrt(1-pp))*m2-(2/3)*m1
m4 <-(7-3*pp-pp*no)/(4*sqrt(1-pp))*m3-(3/4)*m2
m5 <-(9-4*pp-pp*no)/(5*sqrt(1-pp))*m4-(4/5)*m3
mm1 <- sum(m1)/sqrt(n)
mm2 <- sum(m2)/sqrt(n)
mm3 <- sum(m3)/sqrt(n)
mm4 <- sum(m4)/sqrt(n)
mm5 <- sum(m5)/sqrt(n)
v3[t] <- (mm1^2)+(mm2^2)+(mm3^2)
v4[t] <-(mm1^2)+(mm2^2)+(mm3^2)+(mm4^2)+(mm5^2)
 }
}

#### Empirical Power of Tests and Frequency of Excluded Samples
T_test <- (reject_freq)/(replicas-failures)

M_ind <- length(v1[v1>2.706])/(replicas-failures) ### Asymptotic critical values
CAViaR <- length(v2[v2>4.605])/(replicas-failures)
J_ind3 <- length(v3[v3>4.605])/(replicas-failures)
J_ind5 <- length(v4[v4>7.779]) /(replicas-failures)
FSE <- failures/replicas

T_test
M_ind
CAViaR
J_ind3
J_ind5
FSE
```

### A.1.3 Table for the *table_50.txt* file

| | |
|---|---|
| 1 | -1 |
| 2 | 18.97 |
| 3 | 42.31 |
| 4 | 11.69 |
| 5 | 17.55 |
| 6 | 10.26 |
| 7 | 13.51 |
| 8 | 9.76 |
| 9 | 11.96 |
| 10 | 9.52 |
| 11 | 11.19 |
| 12 | 9.43 |
| 13 | 10.77 |
| 14 | 9.36 |
| 15 | 10.50 |
| 16 | 9.35 |
| 17 | 10.35 |
| 18 | 9.34 |
| 19 | 10.22 |
| 20 | 9.36 |
| 21 | 10.13 |
| 22 | 9.40 |
| 23 | 10.0 |
| 24 | 9.41 |
| 25 | 10.06 |
| 26 | 9.44 |
| 27 | 10.04 |
| 28 | 9.47 |
| 29 | 10.03 |
| 30 | 9.51 |
| 31 | 10.01 |
| 32 | 9.54 |
| 33 | 10.02 |
| 34 | 9.57 |
| 35 | 10.02 |
| 36 | 9.60 |
| 37 | 10.04 |
| 38 | 9.64 |

```
39    10.05
40     9.68
41    10.06
42     9.70
43    10.09
44     9.74
45    10.09
46     9.78
47    10.12
48     9.81
49    10.13
50     9.85
51    10.15
52     9.86
53    10.17
54     9.89
55    10.20
56     9.92
57    10.22
58     9.96
59    10.23
60     9.98
61    10.25
62    10.02
63    10.27
64    10.05
65    10.30
66    10.07
67    10.30
68    10.10
69    10.34
70    10.12
71    10.37
72    10.14
73    10.37
74    10.18
75    10.39
76    10.20
77    10.42
78    10.22
79    10.43
80    10.25
81    10.45
82    10.27
83    10.48
84    10.29
85    10.50
86    10.33
87    10.51
88    10.34
89    10.53
90    10.36
91    10.55
92    10.37
93    10.57
94    10.41
95    10.59
96    10.44
97    10.60
98    10.46
99    10.62
100   10.47
101   10.64
102   10.49
103   10.65
104   10.52
105   10.67
106   10.52
107   10.69
108   10.56
109   10.71
110   10.56
111   10.72
112   10.59
113   10.74
114   10.61
115   10.76
116   10.64
117   10.77
118   10.65
119   10.79
120   10.66
121   10.80
122   10.68
123   10.82
124   10.70
125   10.83
126   10.71
127   10.85
128   10.74
129   10.87
130   10.74
131   10.88
```

```
132   10.77
133   10.89
134   10.79
135   10.92
136   10.80
137   10.92
138   10.81
139   10.95
140   10.83
141   10.96
142   10.84
143   10.98
144   10.86
145   10.98
146   10.88
147   10.99
148   10.89
149   11.01
150   10.91
151   11.02
152   10.92
153   11.04
154   10.94
155   11.05
156   10.96
157   11.07
158   10.97
159   11.08
160   10.98
161   11.09
162   10.99
163   11.11
164   11.01
165   11.12
166   11.02
167   11.13
168   11.02
169   11.14
170   11.04
171   11.16
172   11.07
173   11.17
174   11.08
175   11.17
176   11.08
177   11.19
178   11.10
179   11.21
180   11.10
181   11.21
182   11.12
183   11.23
184   11.15
185   11.24
186   11.16
187   11.24
188   11.16
189   11.27
190   11.18
191   11.28
192   11.18
193   11.28
194   11.20
195   11.30
196   11.22
197   11.30
198   11.23
199   11.32
200   11.24
```

## A.2 R programs for the improved shape parameter estimator of Chapter 5

Subsection A.2.1, presents a program to simulate the moments of the improved shape parameter estimator, conditional to $K = k$. The results of the simulation can be compared with the results of the R program presented in Subsection A.2.2 that calculates these moments based on the Theorem 5.3.3. The second program constitute part of the R code used to implement the example 5.3.1. The other calculations needed to implement the example are easily carried out by adapting the program given in Subsection A.2.2.

### A.2.1 R program to simulate the moments of the improved shape parameter estimator, conditional to $K = k$

```
shape <- 1.5
q <- 0.5
scale <- 1/((-log(q))^(1/shape))

sum=0
sum_est=0
sum_est_quad=0
replicas=1000000

for (i in 1:replicas){

a <- rweibull(10, shape, scale)
a <- floor(a)+1
if(min(a)==1 & max(a)==5){
Fn1=length(a[a<=1])/n
Fn2=length(a[a<=2])/n
Fn3=length(a[a<=3])/n
Fn4=length(a[a<=4])/n

sum_est=sum_est+((1/log(2))*log(log(1-Fn2)/log(1-Fn1))
+(1/log(3))*log(log(1-Fn3)/log(1-Fn1))+(1/log(4))*log(log(1-Fn4)/log(1-Fn1)))/3

sum_est_quad=sum_est_quad+(((1/log(2))*log(log(1-Fn2)/log(1-Fn1))
+(1/log(3))*log(log(1-Fn3)/log(1-Fn1))+(1/log(4))*log(log(1-Fn4)/log(1-Fn1)))/3)^2

sum=sum+1
 }
}

sum
sum/replicas
mom1=sum_est/sum
mom2=sum_est_quad/sum
mom1
mom2
```

### A.2.2 R program to calculate, based on the Theorem 5.3.3, the moments of the improved shape parameter estimator, conditional to $K = k$

```
comb <- function(a,b){
co <- factorial(a)/(factorial(a-b)*factorial(b))
return(co) }

n=10
theta=1.5
q=0.5

sum=0
p_ast=0
mom1=0
mom2=0
gx1=0
gx2=0
nn=n-1

p1=1-q
p2=q-q^(2^theta)
p3=q^(2^theta)-q^(3^theta)
p4=q^(3^theta)-q^(4^theta)
p5=q^(4^theta)-q^(5^theta)
```

```
for(i1 in 1:nn){
for(i2 in 0:nn){
for(i3 in 0:nn){
for(i4 in 0:nn){

if(i1+i2+i3+i4<n){
fmp_a=comb(n,i1)*((p1/(p1+p2+p3+p4+p5))^i1)*(((p2+p3+p4+p5)/(p1+p2+p3+p4+p5))^(n-i1))
fmp_b=comb(n-i1,i2)*((p2/(p2+p3+p4+p5))^i2)*(((p3+p4+p5)/(p2+p3+p4+p5))^(n-i1-i2))
fmp_c=comb(n-i1-i2,i3)*((p3/(p3+p4+p5))^i3)*(((p4+p5)/(p3+p4+p5))^(n-i1-i2-i3))
fmp=fmp_a*fmp_b*fmp_c*comb(n-i1-i2-i3,i4)*((p4/(p4+p5))^i4)*((p5/(p4+p5))^(n-i1-i2-i3-i4))
p_ast=p_ast+fmp
}}}}

for(i1 in 1:nn){
for(i2 in 0:nn){
for(i3 in 0:nn){
for(i4 in 0:nn){

if(i1+i2+i3+i4<n){
fmp_a=comb(n,i1)*((p1/(p1+p2+p3+p4+p5))^i1)*(((p2+p3+p4+p5)/(p1+p2+p3+p4+p5))^(n-i1))
fmp_b=comb(n-i1,i2)*((p2/(p2+p3+p4+p5))^i2)*(((p3+p4+p5)/(p2+p3+p4+p5))^(n-i1-i2))
fmp_c=comb(n-i1-i2,i3)*((p3/(p3+p4+p5))^i3)*(((p4+p5)/(p3+p4+p5))^(n-i1-i2-i3))
fmp=fmp_a*fmp_b*fmp_c*comb(n-i1-i2-i3,i4)*((p4/(p4+p5))^i4)*((p5/(p4+p5))^(n-i1-i2-i3-i4))

gx1=((1/log(2))*log(log(1-i1/n-i2/n)/log(1-i1/n))
+(1/log(3))*log(log(1-i1/n-i2/n-i3/n)/log(1-i1/n))
+(1/log(4))*log(log(1-i1/n-i2/n-i3/n-i4/n)/log(1-i1/n)))/3

gx2=(((1/log(2))*log(log(1-i1/n-i2/n)/log(1-i1/n))
+(1/log(3))*log(log(1-i1/n-i2/n-i3/n)/log(1-i1/n))
+(1/log(4))*log(log(1-i1/n-i2/n-i3/n-i4/n)/log(1-i1/n)))/3)^2
mom1=mom1+gx1*fmp/p_ast mom2=mom2+gx2*fmp/p_ast sum=sum+fmp/p_ast }
}}}}

mse=mom2-(mom1)^2+(mom1-theta)^2
rmse=sqrt(mse)

1-p_ast
rmse
mom1
mom2
```

## A.3 R program to implement the DPOT model proposed in Chapter 6

```
#### For running this example it is necessary to download the daily
#### prices of SP 500 index until May 28, 2010, or at least the
#### first 2002 days, compute the returns and save them in the file
#### with the name SP_1950_Maio2010.txt
#### Choose c <- 0.75 to implement the DPOT(c=0.75)
c <- 0.75

#### log-likelihood function which takes tree arguments: theta is the vector of
#### parameters, y the excesses and x the durations
gpdlik <- function(theta,y,x){
alpha1 <- theta[1]
gamma <- theta[2]
n<-length(y)

logl<- -sum(log(alpha1*(1/x)^c))-(1/gamma+1)*sum(log(1+gamma*y/(alpha1*(1/x)^c)))
return(-logl)
}

#### Here we read the log returns from the text file SP_1950_Maio2010.txt.
```

```
#### Then we compute the symmetric of log returns
#### and choose the returns from day 1001 until day 2000, to illustrate the
#### calculation of one forecast for day 2001
xx <- read.table("SP_1950_Maio2010.txt")
a <- xx*-1
a <- a[,1]
b <- a[1001:2000]

#### Calculation of excesses and durations since the preceding 3 excesses
b_sort<-sort(b)
u<-b_sort[900]
bb <- b[b>u]
bb <- bb-u
duration <- 1
j <- 1
xexc <- rep(0,times=length(bb))

for(ii in 1:1000){
if (b[ii]>u){
xexc[j] <- duration
duration <- 1
j <- j+1
}
else {
duration <-duration+1
 }
}

lag1_xexc <-rep(0,times=length(bb))
d2 <-rep(0,times=length(xexc))
limite <- length(xexc)-1
xxxx <- xexc[1:limite]
lag1_xexc <- c(0, xxxx)
limite2 <- length(xexc)-2
xxxx <- xexc[1:limite2]
lag2_xexc <- c(0, 0, xxxx)
limite3 <- length(bb)
bb <- bb[3:limite3]
xexc <- xexc[3:limite3]
lag1_xexc <- lag1_xexc[3:limite3]
lag2_xexc <- lag2_xexc[3:limite3]
d3 <- xexc+lag1_xexc+lag2_xexc
#### v=3, durations since the preceding 3 excesses

#### Here we use the optim with Nelder and Mead algorithm to
#### maximize the log likelihood
modelo <- optim(c(0.5,0.5), gpdlik, y=bb, x=d3)
mle1 <- modelo$par[1] mle2 <- modelo$par[2]

## Finally with the VaR DPOT estimator we compute the forecast
delta <- mle1*(1/(duration+xexc[length(xexc)]+xexc[length(xexc)-1]))^c
var_forecast <- u + ((0.1/coverage)^mle2-1)*(delta/mle2)

## One-day-ahead VaR forecast:
var_forecast


################# Now we explain the following message
## Warning message:
## In log(1 + gamma * y/(alpha1 * (1/x)^c)) :NaNs produced
###########################################################

#### If we write
modelo
#### we obtain the following message indicating successful convergence
# $convergence
# [1] 0
#

#### The Nelder and Mead Algorithm to implement DPOT always converges
```

```
#### with all the data we use until now.
#### The optimizer chooses values based on a deterministic search
#### algorithm and the warnings messages occur when the values do not obey to:
#### 0 <= y_i <= -sigma_t/gamma  when  gamma < 0
#### (support when gamma is negative)
#### When this occur, we have log of a negative number and then the
#### message NaN, but this do not create a problem because the
#### optimizer continue to other interactions choosing other values
#### until reach convergence.
#### For example: If we stop the interaction at interaction 32,
#### we don´t have any message because until interaction 32 the
#### values always obey to the support condition

modelo <- optim(c(0.5,0.5), hgplik, y=bb, x=d3, control=list(maxit=32))

#### If we try more one interaction we have a Warning message

modelo <- optim(c(0.5,0.5), hgplik, y=bb, x=d3, control=list(maxit=33))

#### With the following code we can observe what happened:
as.list(body(hgplik))
trace("hgplik", quote(if(any(is.nan(logl))) {browser() }), at=6, print=F)
modelo <- optim(c(0.5,0.5), hgplik, y=bb, x=d3)
where
# log(1 +-0.0590332031249936 * bb/(5.128759765625 * (1/d3)^c))

#### With a estimate of gamma = -0.05903320312499 < 0,
#### considering the estimate of alpha, the execess and the
#### durations, we have one NaN. However, the algorithm continues until
#### convergence without problems
```

## A.4 R programs for the out-of-sample studies of Chapters 6 and 7

### A.4.1 POT model

```
library(POT)
##############
x <- read.table("DJI_1928_Março2011.txt")
a <- x*-1
a <- a[,1]
tt <- 19713
##############

#### coverage rate or probability level:
coverage <- 0.001

#### rolling window size:
ws <- 1000

hit <-runif(tt)
varforecast <- runif(tt)

for(i in 1:tt) {
print(i)
iws <- i+ws
iws_m <- iws-1
b <- a[i:iws_m]

#### VaR POT
b <- sort(b)
u <- b[900]
y <- b[b>u]
mle <- fitgpd(y, u, "mle")$param
qpot <- u + mle[1]/mle[2]*((0.1/coverage)^(mle[2])-1)
```

```
varforecast[i] <- qpot

if (a[iws]<varforecast[i]){hit[i]=0} else {hit[i]=1}
 }

#### Save the forecasts in a excel file:
write.table(varforecast, file="VaR_pot_dji.xls")
```

### A.4.2 Quasi-PORT model

```
##############
x <- read.table("DJI_1928_Março2011.txt")
a <- x*-1
a <- a[,1] tt <- 19713
##############

#### coverage rate or probability level:
coverage <- 0.001

#### rolling window size:
ws <- 1000

hit <-runif(tt)
varhill  <- runif(tt)

for(i in 1:tt) {
print(i)
iws <- i+ws
iws_m <- iws-1
b <- a[i:iws_m]

#### VaR Quasi-PORT
b <- sort(b)
b_posi <- b[b>0]
n_0 <- length(b_posi)
k_0 <- floor(n_0^0.995)
k_2 <- floor(n_0^0.999)
ciclo <- k_2-k_0+1

rho_1 <- rep(0, times=ciclo)
rho_0 <- rep(0, times=ciclo)
yy <- rep(0, times=k_0)
ind <- k_0:1

##### Algorithm for estimate rho and beta
for(ii in 1:ciclo) {
k <- k_0+ii-1
pos <- 1000-k
u <- b[pos]
y <- b[b>u]
y <- sort(y)

M_1 <- (sum(log(y/u)))/k
M_2 <- (sum(log(y/u)^2))/k
M_3 <- (sum(log(y/u)^3))/k

T_1 <- (M_1 - (M_2/2)^0.5)/((M_2/2)^0.5-(M_3/6)^(1/3))
T_0 <- (log(M_1)-0.5*log(M_2/2))/(0.5*log(M_2/2)-(1/3)*log(M_3/6))

rho_1[ii] <- -abs(3*(T_1-1)/(T_1-3))
rho_0[ii] <- -abs(3*(T_0-1)/(T_0-3))
 }

u <- b[1000-k_0]
pos_y <- 1000-k_0+1
y <- b[pos_y:1000]
y <- sort(y)
```

```
s_1 <- sum((rho_1-median(rho_1))^2)
s_0 <- sum((rho_0-median(rho_0))^2)
if (s_0 <= s_1) tau <- 0 else tau <- 1
if (s_0 <= s_1) rho <- rho_0[1] else rho <- rho_1[1]

#### beta estimation
yy[1] <- u
for (iii in 2:k_0){
yy[iii] <- y[iii-1]
 }

d_n <- sum((ind/k_0)^(-rho))/k_0
d_p <- sum((ind/k_0)^(rho))/k_0

D_0 <- sum(ind*log(y/yy))/k_0
D_n1 <- sum((ind/k_0)^(-rho)*ind*log(y/yy))/k_0
D_n2 <- sum((ind/k_0)^(-2*rho)*ind*log(y/yy))/k_0
beta <- ((k_0/1000)^(rho))*(d_n*D_0-D_n1)/(d_n*D_n1-D_n2)

k <- 100
pos <- 1000-k

#### Sample of Excesses
# b[500] for q=0.5 and b[250] for q=0.25
q_emp <- b[500]
b <- b - q_emp
u <- b[pos]
y <- b[b>u]

#### Hill MVRB
hill <- (sum(log(y/u)))/k
hill_mvrb <- hill*(1-beta/(1-rho)*(ws/k)^rho)
c_n <- k/(ws*coverage)

#### Quasi-PORT VaR
quantil_quasiPORT <- (u*(c_n^hill_mvrb))+q_emp
varhill[i] <- quantil_quasiPORT

if (a[iws]<varhill[i]){hit[i]=0} else {hit[i]=1}
}

#### Save the forecasts in a excel file:
write.table(varhill, file="VaR_pot_dji.xls")
```

### A.4.3 DPOT(v=1) model

```
c <- 0.75
hgplik <- function(theta,y,x){
alpha1 <- theta[1]
gamma <- theta[2]
n <- length(y)
logl <- -sum(log(alpha1*(1/x)^(c)))-(1/gamma+1)*sum(log(1+gamma*y/(alpha1*(1/x)^(c))))
return(-logl)
 }

##############
x <- read.table("DJI_1928_Março2011.txt")
a <- x*-1
a <- a[,1]
tt <- 19713
##############

#### coverage rate or probability level:
coverage <- 0.001

#### rolling window size:
ws <- 1000
```

```
hit <-runif(tt)
varforecast <- runif(tt)

for(i in 1:tt) {
print(i)
iws <- i+ws
iws_m <- iws-1
b <- a[i:iws_m]

###### VaR DPOT(v=1)
b_sort <- sort(b)
u <- b_sort[900]
bb <- b[b>u]

bb <- bb-u
duration <- 1
j<-1
xexc <-rep(0,times=length(bb))

for(ii in 1:1000){
if (b[ii]>u){
xexc[j] <- duration
duration <- 1
j <- j+1
 }
else {
duration <-duration+1
  }
 }

d <- xexc

##
modelo <- optim(c(0.5,0.5), hgplik, y=bb, x=d)
mle1 <- modelo$par[1]
mle2 <- modelo$par[2]

delta <- mle1*(1/duration)^(c)
varforecast[i] <- u + ((0.1/coverage)^mle2-1)*(delta/mle2)

if (a[iws]<varforecast[i]){hit[i]=0} else {hit[i]=1}
}

#### Save the forecasts in a excel file:
write.table(varforecast, file="VaR_dpot_v1_dji.xls")
```

### A.4.4 DPOT(v=3) model

```
c <- 0.75
hgplik <- function(theta,y,x){
alpha1 <- theta[1]
gamma <- theta[2]
n <- length(y)
logl <- -sum(log(alpha1*(1/x)^(c)))-(1/gamma+1)*sum(log(1+gamma*y/(alpha1*(1/x)^(c))))
return(-logl)
 }

##############
x <- read.table("DJI_1928_Março2011.txt")
a <- x*-1
a <- a[,1]
tt <- 19713
##############

#### coverage rate or probability level:
coverage <- 0.001
```

```
#### rolling window size:
ws <- 1000

hit <-runif(tt)
varforecast <- runif(tt)

for(i in 1:tt) {
print(i)
iws <- i+ws
iws_m <- iws-1
b <- a[i:iws_m]

###### VaR DPOT(v=3)
b_sort<-sort(b)
u<-b_sort[900]

bb <- b[b>u]
bb <- bb-u
duration<-1
j<-1
xexc <-rep(0,times=length(bb))

for(ii in 1:1000){
if (b[ii]>u){
xexc[j] <- duration
duration <- 1
j <- j+1
}

else {
duration <-duration+1
 }
}

lag1_xexc <-rep(0,times=length(bb))
d2 <-rep(0,times=length(xexc))
limite <- length(xexc)-1
xxxx <- xexc[1:limite]
lag1_xexc <- c(0, xxxx)
limite2 <- length(xexc)-2
xxxx <- xexc[1:limite2]
lag2_xexc <- c(0, 0, xxxx)
limite3 <- length(bb)
bb <- bb[3:limite3]
xexc <- xexc[3:limite3]
lag1_xexc <- lag1_xexc[3:limite3]
lag2_xexc <- lag2_xexc[3:limite3]
d3 <- xexc+lag1_xexc+lag2_xexc
#### v=3, durations since the preceding 3 excesses

##
modelo <- optim(c(0.5,0.5), hgplik, y=bb, x=d3)
mle1 <- modelo$par[1]
mle2 <- modelo$par[2]

delta <- mle1*(1/(duration+xexc[length(xexc)]+xexc[length(xexc)-1]))^(c)
varforecast[i] <- u + ((0.1/coverage)^mle2-1)*(delta/mle2)

if (a[iws]<varforecast[i]){hit[i]=0} else {hit[i]=1}
}

#### Save the forecasts in a excel file:
write.table(varforecast, file="VaR_dpot_v3_dji.xls")
```

### A.4.5 Conditional EVT model

```
library(fGarch)
library(POT)
```

```
##############
x <- read.table("DJI_1928_Março2011.txt")
a <- x*-1
a <- a[,1]
tt <- 19713
##############

#### coverage rate or probability level:
coverage <- 0.001

#### rolling window size:
ws <- 1000

hit <- runif(tt)
varforecast <- runif(tt)

for(i in 1:tt) {
print(i)
iws <- i+ws
iws_m <- iws-1
b <- a[i:iws_m]

#### VaR Conditional EVT

#### For normal innovations:
#argarch <- garchFit(~arma(1,0)+garch(1,1), data = b, cond.dist =
#"norm", include.mean=TRUE, trace = FALSE)

#### For skewed-t innovations:
argarch <- garchFit(~arma(1,0)+garch(1,1), data = b, cond.dist =
"sstd", include.mean=TRUE, trace = FALSE)

coef <- argarch@fit$params$params
sigma <- argarch@sigma.t
fitted <- argarch@fitted
resid <- (b-fitted)/sigma

b <- resid
b <- sort(b)
u <- b[900]
y <- b[b>u]
mle <- fitgpd(y, u, "mle")$param

qpot <- u + mle[1]/mle[2]*((0.1/coverage)^(mle[2])-1)
sig_sq <- coef[3]+coef[4]*(a[iws_m]-fitted[1000])^2+ coef[6]*(sigma[1000])^2
varforecast[i] <- qpot*sqrt(sig_sq)+coef[1]+coef[2]*a[iws_m]

if (a[iws]<varforecast[i]){hit[i]=0} else {hit[i]=1}
 }

#### Save the forecasts in a excel file:
write.table(varforecast, file="VaR_cevt_dji.xls")
```

### A.4.6 RiskMetrics model

```
##############
x <- read.table("DJI_1928_Março2011.txt")
a <- x*-1
a <- a[,1]
tt <- 19713
##############

#### coverage rate or probability level:
coverage <- 0.001

#### rolling window size:
ws <- 1000
```

```
hit <- runif(tt)
varforecast <- runif(tt)

for(i in 1:tt) {
print(i)
iws <- i+ws
iws_m <- iws-1
b <- a[i:iws_m]

#### VaR RiskMetrics
sig_sq <- runif(1000)
##sig_sq[1] <- (b[1])^2
## Or we can start the recursion by:
sig_sq[1] <- (sd(b))^2

for(vv in 2:1001) {
sig_sq[vv] <- 0.94*sig_sq[vv-1]+0.06*(b[vv-1])^2
}

#### p=0.01
# varforecast[i] <- 2.326348*sqrt(sig_sq[1001])

#### p=0.001
varforecast[i] <- 3.090232*sqrt(sig_sq[1001])

#### p=0.0005  #### qnorm(0.9995)=3.290527
# varforecast[i] <- 3.290527*sqrt(sig_sq[1001])

if (a[iws]<varforecast[i]){hit[i]=0} else {hit[i]=1}
 }

#### Save the forecasts in a excel file:
write.table(varforecast,file="VaR_rm_dji.xls")
```

### A.4.7 APARCH model

```
library(fGarch)
##############
x <- read.table("DJI_1928_Março2011.txt")
a <- x*-1
a <- a[,1] tt <- 19713
##############

#### coverage rate or probability level:
coverage <- 0.001

#### rolling window size:
ws <- 1000

hit <- runif(tt)
varforecast <- runif(tt)

for(i in 1:tt) {
print(i)
iws <- i+ws
iws_m <- iws-1
b <- a[i:iws_m]

#### VaR APARCH
p_quantile <- 1 - coverage

#### For normal innovations:
# aparch <- garchFit(~arma(1,0)+aparch(1,1), data = b,
# cond.dist="norm", include.mean=TRUE, trace = FALSE)

#### For skwed-t innovations:
```

```
aparch <- garchFit(~arma(1,0)+aparch(1,1), data = b,
cond.dist="sstd", include.mean=TRUE, trace = FALSE)

coef <- aparch@fit$params$params
sigma <- aparch@sigma.t
res <- residuals(aparch) media <- coef[1]+coef[2]*a[iws_m]

#### For normal innovations:
#sig_sq <- coef[3]+coef[4]*res[1000]^2+coef[6]*sig[1000]^2+indicator*coef[5]*res[1000]^2
#varforecast[i] <- qnorm(p_quantile, mean = media, sd = sqrt(sig_sq))

#### For skwed-t innovations:
sig<-(coef[3]+coef[4]*(abs(res[1000]-coef[5]*res[1000]))^coef[7]
+coef[6]*sigma[1000]^coef[7])^(1/coef[7])
varforecast[i] <- qsstd(p_quantile, mean = media, sd = sig,
nu = coef[9], xi = coef[8])

if (a[iws]<varforecast[i]){hit[i]=0} else {hit[i]=1}
 }
```

### A.4.8 UC and IND tests

```
##### Durations for the MM independence test
no_hit_duration <- 0
j <- 1
zeros <- 0

for(i in 1:tt) {
 if (hit[i]<1){
  zeros <- zeros+1
  }
else {
no_hit_duration[j]<- zeros+1
zeros <- 0
j <- j+1
  }
 }

no <- no_hit_duration
n <- length(no)

#### MM test
no <- sort(no)
PMR <- (no[n]-1)/no[floor(0.5*n)]
PMR
observed_T <- log(2)*(no[n]-1)/no[floor(0.5*n)]-log(n)

#### Simulation of the upper bound for the p value
v <- 0
replicas <- 100000
v <- runif(replicas)

print('wait for p value upper bound simulation')
for(i in 1:replicas) {
if((i/10000-floor(i/10000))==0){print(replicas-i)}
u <- runif(n)
y <- -log(1-u)
no_simul <- sort(y)
v[i] <- log(2)*(no_simul[n]/no_simul[floor(n/2)])-log(n)
  }
simulated_p_value_upper_bound <- length(v[v>=observed_T])/replicas

#####Caviar Test
lim_s <- tt-1
hit1 <- hit[1:lim_s]
hit2 <- hit[2:tt]
var2 <- varforecast[2:tt]
```

```
mylogit<- glm(hit2~hit1+var2, family=binomial(link="logit"),
na.action=na.pass) logLik(mylogit)

alpha <- -log(length(hit)/sum(hit)-1)
loglik1 <- -sum(1-hit2)*alpha-(tt-1)*log(1+exp(-alpha))

emv <- mylogit$coefficients
emv1 <- emv[1]
emv2 <- emv[2]
emv3 <- emv[3]

loglik2 <- -sum((1-hit2)*(emv1+emv2*hit1+emv3*var2))
-sum(log(1+exp(-emv1-emv2*hit1-emv3*var2)))
caviar <- -2*(loglik1-loglik2)

#### percentage of violations
viol <- sum(hit)
pv <- viol/tt
viol
pv

#### Kupiec test p value #To change the coverage, change the following line:
#coverage <- 0.01
num <- (1-pv)^(tt-viol)*pv^viol
denom <- (1-coverage)^(tt-viol)*coverage^viol
log_R <- 2*log(num/denom)
1-pchisq(log_R,df=1)

#### MM independence test p value
simulated_p_value_upper_bound

#### Caviar independence test p value
1-pchisq(caviar,df=2) ####
```

# Bibliography


[1] Araújo Santos, P. and Fraga Alves, M.I. (2011). Forecasting Value-at-Risk with a Duration based POT method. Notas e Comunicações CEAUL 17/2011.

[2] Araújo Santos, P. and Fraga Alves, M.I. (2010). A new class of independence tests for interval forecasts evaluation. Computational Statistics and Data Analysis. In press. doi:10.1016/j.csda2010.10.002.

[3] Araújo Santos, P., (2010). Interval Forecasts Evaluation: R programs for a new independence test. Notas e Comunicações CEAUL 17/2010.

[4] Araújo Santos, P. and Fraga Alves, M.I. (2010). VaR Prediction with a Duration based POT method. Proceedings of the ISF2010, 30th International Symposium on Forecasting, San Diego, CA, USA.

[5] Araújo Santos, P. and Fraga Alves, M.I. (2009). Evaluating Forecast Models with an Exact Independence Test. CEAUL. Technical Report 10/2009.

[6] Araújo Santos, P., Fraga Alves, M. I., Gomes, M. I. (2007). Comparação do desempenho de diferentes níveis aleatórios na metodologia PORT. In Ferrão, M.E., Nunes, C. and Braumann, C.A. (eds.), Estatística Ciência Interdisciplinar, pp. 717-728.

[7] Araújo Santos, P., Fraga Alves, M. I., Gomes, M. I. (2006). Peaks over random threshold methodology for tail index and quantile estimation. Revstat 4(3), 227-247.

[8] Araújo Santos, P., (2004). Quantis Extremos em Caudas Pesadas. Dissertação de Mestrado, D.E.I.O., Faculdade de Ciências da Universidade de Lisboa.

[9] Balkema, A.A. and de Haan, L. (1974). Residual Life Time at Great Age. Ann. Probab., 2,792-804.

[10] Basel Commitee on Banking Supervision, (1995), An Internal Model-Based Approach to Market Risk Capital Requirements, BIS, Basel, Switzerland.

[11] Basel Commitee on Banking Supervision, (1996), Amendment to the Capital Accord to incorporate market risks, BIS, Basel, Switzerland.

[12] Basel Commitee on Banking Supervision, (1996), Supervisory Framework for the Use of "Backtesting" in Conjunction with the Internal Model-Based Approach to Market Risk Capital Requirements, BIS, Basel, Switzerland.

[13] Basel Commitee on Banking Supervision, (2006), International Convergence of Capital Measurement and Capital Standards, a Revised Framework Comprehensive Version, BIS, Basel, Switzerland.

[14] Bekiros, S.D. and Georgoutsos, D.A. (2005). Estimation of Value-at-Risk by extreme value and conventional methods: a comparative evaluation of their predictive performance. Journal of International Financial Markets, Institutions and Money, 15, Issue 3, 2009-228.

[15] Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., (2004). Statistics of Extremes: Theory and Applications. Wiley, England.

[16] Beirlant,J., Dierckx, G., Goegebeur, Y. and Matthys, G. (1999). Tail index estimation and exponential regression model. Extremes, 2, 177-200.

[17] Berkowitz, J., Christoffersen P. and Pelletier D., (2009). Evaluating Value-at-Risk models with desk-level data, Management Science, Published online in Articles in Advance.

[18] Bingham, N. H., Goldie, C. M., Teugels, J. L. (1987). Regular Variation. Cambridge: Cambridge University Press.

[19] Black R., (1976). Studies in stock price volatility changes. Proceedings of the 1976 Business Meeting of the Business and Economics Statistics Section. American Statistical Association, 177-181.

[20] Bollerslev, T., (1986). Generalized Autoregressive Conditional Heteroskedasticity, Journal of Econometrics, 31, pp. 307-327.

[21] Bontemps, C. and Meddahi, N., (2008). Testing distributional assumptions: A GMM approach, Working Paper.

[22] Bracquemond C. and Gaudoin O., (2003). A survey on discrete lifetime distributions. International Journal of Reliability, Quality and Safety Engineering, vol 10, 69-98.

[23] Bystrom, H. (2004). Managing Extreme Risks in Tranquil and Volatile Markets Using Conditional Extreme Value Theory. International Review of Financial Analysis, 13, no.2, 133-152.

[24] Caeiro, F., Gomes, M. I. (2008). Minimum-varaince reduced-bias tail index and high quantile estimation. Revstat 6(1), 1-20.

[25] Caeiro, F., Gomes, M. I. and Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator. RevStat 3 (2), 111-136.

[26] Caeiro, F., Gomes, M. I. (2002). A class of asymptotically unbiased semi-parametric estimators of the tail index. Test, 11:2, 345-364.

[27] Campbell, S.D., (2007). A review of backtesting and backtesting procedures, Journal of Risk, 9(2), 1-18.

[28] Campbell, J.Y., Lo, A.W. and Mackinlay, A.C. (1997). The Econometrics of Financial Markets. Princeton University Press, New Jersey.

[29] Candelon, B., Colletaz, G., Hurlin, C., and Tokpavi, S., (2008). Backtesting value-at-Risk: A GMM Duration-Based Test, HAL, Working Paper.

[30] Castillo, E., Hadi A.S. Balakrishnan, N. and Sarabia, J.M., (2005). Extreme Value and Related Models with Applications in Engineering and Science, John Wiley & Sons, Hoboken, New Jersey.

[31] Christoffersen, P. and Pelletier, D., (2004). Backtesting Value-at-Risk: A Duration-Based Approach, Journal of Financial Econometrics, 2,1,84-108.

[32] Christoffersen, P. (2003). Elements of Financial Risk Management. Amsterdam: Academic Press.

[33] Christoffersen, P., (1998). Evaluating Intervals Forecasts, International Economic Review, 39, 841-862.

[34] Coles, s. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer.

[35] Danielsson, J. and Morimoto, Y., (2000). Forecasting Extreme Financial Risk: A Critical Analysis of Practical Methods for the Japanese Market, Monetary and Economic Studies, 18(2), 25-48.

[36] de Haan, L., Peng, L. (1998). Comparison of extreme value index estimators. Statistica Neerlandica 52, 60-70.

[37] Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. Ann. Statist. 17, 1833-1855.

[38] Diebold, F.X., Schuermann, T. and Stroughair, J.D. (1998). Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management. Working Paper, 98-10, Wharton School, University of Pennsylvania.

[39] Ding, Z., Engle, R.F and Granger, C.W.J., (1993). A long memory property of stock market return and a new model, Journal of Empirical Finance, 1, 83-106.

[40] Dufour, J.M., (2006). Monte Carlo tests with nuisance parameters: a general approach to finite sample inference and nonstandard asymptotics, Journal of Econometrics, 127(2),443-477.

[41] Dowd, K., (2002). Measuring Market Risk. Chichester: John Wiley & Sons.

[42] Drees, H. (2003). Extreme quantile estimation for dependent data, with applications to finance. Bernoulli 9(4), 617-657.

[43] Drees, H. (2003). Extreme quantile estimation for dependent data, with applications to finance. Bernoulli 9(4), 617-657.

[44] Embrechts, P., Klüppelberg, C. and Mikosch, T. (2001). Modeling Extremal Events for Insurance and Finance (Berlin: Springer, 3rd ed. ).

[45] Engle, R.F. and Manganelli, S., (2004). CAViaR: Conditional Autoregressive Value-at-Risk by Regression Quantiles, Journal of Business and Economics Statistics, 22, 367-381.

[46] Fernández, C. and Steel, M.F.J., (1998). On Bayesian modelling of fat tails and skewness, Journal of the American Statistical Association, 93, 359-371.

[47] Granger, C.W.J., H. White and M. Kamstra, (1989). Interval forecasting: An analysis based upon. ARCH-quantile estimators. Journal of Econometrics, 40, 87-96.

[48] Falk, M., Husler, J. and Reiss, R.D. (2010) Laws of Small Numbers: Extremes and Rare Events, 3rd ed., Springer-Basel.

[49] Ferreira, A. de Haan, L. and Peng, L. (2003).On optimizing the estimation of high quantiles of a probabilitydistribution. Statistics 37, 401-434.

[50] Fisher, R.A. and Tippett, L.H.C., (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proc. Cambridge Philos. Soc. 24, 180-190.

[51] Fraga Alves, M.I. and Araújo Santos, P. (2004). Extreme Quantiles Estimation with Shifted Data From Heavy Tails. Notas e Comunicações CEAUL 11/2004.

[52] Fraga Alves, M.I., Gomes, M.I. e de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter. Portugaliae Mathematica 60:1, 193-213.

[53] Geluk, J. and de Haan, L. (1987). Regular Variation, Extensions and Tauberian Theorems. CWI Tract 40, Center of Mathematics and Computer Science, Amsterdam, Netherlands.

[54] Ghorbel, A. and Trabelsi, A. (2008). Predictive performance of conditional Extreme Value Theory in Value-at-Risk estimation. Int. J. Monetary Economics and Finance, 1, No.2, 121-147.

[55] Glosten, L., Jagannathan and R. and Runkle, D. (1992). On the relation between the expected value and volatility of nominal excess return on stocks, Journal of Finance, 46, 1779-1801.

[56] Genedenko, B.V., (1943). Sur la distribution limite du terme du maximum d'une série aléatoire. Ann. Math. 44, 423-453.

[57] Gomes, M.I., Figueiredo, F., Henriques-Rodrigues, L. and Miranda, M.C. (2010). A quasi-PORT methodology for VaR based on second-order reduced-bias estimation. Notas e Comunicações CEAUL 05/2010.

[58] Gomes, M.I., Fraga Alves, M.I. and Araújo Santos, P. (2008a). PORT Hill and Moment Estimators for Heavy-Tailed Models. Commun. Statist. Simul. and Comput. 37, 1281-1306

[59] Gomes,M.I., de Haan, L. e Henriques Rodrigues, L. (2008b). Tail Index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. J. Royal Statistical Society, B70, Issue 1, 31-52.

[60] Gomes, M. I., Henriques Rodrigues, L., Vandewalle, B., Viseu, C. (2008c). A heuristic adaptive choice of the threshold for bias-corrected Hill estimators. To appear in Journal of Statistical Computation and Simulation 78(2):133-150.

[61] Gomes, M.I. e Pestana D. (2007a). A sturdy reduced bias extreme quantile (VaR) estimator. J. Amer. Statist. Assoc., 102, No. 477, 280-292.

[62] Gomes, M.I., Martins, M.J. e Neves, M.M. (2007b). Improving second order reduced-bias tail index estimator. Revstat, 5(2), 177-207.

[63] Gomes, M. I., Pestana, D. (2007c). A simple second order reduced bias extreme value index estimator. Journal of Statistical Computation and Simulation 77(6):487-504.

[64] Gomes, M. I., Martins, M. J., Neves, M. (2007d). Revisiting the second order reduced bias "maximum likelihood" tail index estimators. Revstat 5(2):177-207.

[65] Gomes, M.I. and Figueiredo, F. (2006) Bias Reduction in risk modelling: semi-parametric quantile estimation. Test, 15(2): 375-396.

[66] Gomes, M.I., Caeiro, F. e Figueiredo, F. (2004). Bias Reduction of a Tail Index Estimator Trough an External Estimation of the Second Order Parameter. Statistics, 38, 497-510.

[67] Gomes, M. I. and Oliveira, O. (2003). How can non-invariant statistics work in our benefit in the semi-parametric estimation of parameters of rare events. Commun. Stat., Simulation Comput. 32(4), 1005-1028.

[68] Gomes, M.I., Martins, M.J. (2002b). "Asymptotically unbiased" estimators of the tail index based on theexternal estimation of the second order parameter Extremes, 5:1, 5-31.

[69] Gomes, M.I., Martins, M.J. e Neves M.M. (2002c). Generalized Jackknife semi-parametric estimators of the tail index. Portugaliae Mathematics, 59:4, 393-408.

[70] Gomes, M.I. and Martins, M.J. (2001). Generalizations of the Hill estimator - sysmptotic versus finite sample behaviour. J. Statist. Planning and Inference, 93, 161-180.

[71] Gomes, M.I., Martins, M.J. e Neves M.M. (2000). Alternatives to a semi-parametric estimator of parameters of rare events - the Jackknife methodology. Extremes, 3:3, 207-229.

[72] Gumbel, E.J. (2004, 1st ed 1958). Statistics of Extremes, Columbia University Press, New York.

[73] Grimshaw S.D., McDonaldb J., McQueenc G.R., Thorleyc S., (2005). Estimating Hazard Functions for Discrete Lifetimes. Communications in Statistics-Simulation and Computation Volume 34, Issue 2, 2005, 451-463.

[74] Haas, M., (2005). Improved duration-based backtesting of Value-at-Risk, Journal of Risk, 8(2), 17-36.

[75] Hall, P., Welsh, A. H. (1985). Adaptative estimates of parameters of regular variation. Annals of Statistics 13:331-341.

[76] Hall, P. (1982). On some Simple Estimates of an Exponent of Regular Variation. J. R. Statist. Soc. 44, no. 1, 37-42.

[77] Harman Y.S. and Zuehlke T.W., (2004). Duration Dependence Testing for Speculative Bubbles Journal of Economics and Finance, vol.28, no.4, 17-36.

[78] Hill, B.M. (1975). A Simple General Approach to Inference about the Tail of a Distribution. Ann. Statist. 3, no. 5, 1163-1174.

[79] Holton, G.A. (2003). Value-at-Risk Theory and Practice. Academic Press.

[80] Jones, M. C., Faddy, M. J. (2003). A skew extension of the t-distribution, with applications. Journal of the Royal Statistical Society B 65(1):159-174.

[81] Jorion, P., (2002). Fallacies about the Effects of Market Risk Management Systems. Journal of Risk, 5, 75-96.

[82] Jorian, P. (2000). Value at Risk: The New Benchmark for Managing Financial Risk, McGraw-Hill, New York.

[83] Kiefer, N., (1988). Economic Duration Data and Hazard Functions, Journal of Economic Literature, 26, 646-679.

[84] Khan M.S.A., Khalique A. and Abouammoh A.M., (1989). On Estimating Parameters in a Discrete Weibull Distribution. IEEE Transactions Reliability, vol. 38, Aug., 348-350.

[85] Kotz, S., Nadarajah, S. (2000). Extreme Value Distributions: Theory and Applications. London: Imperial College Press.

[86] Kuester, K., Mittik, S. and Paolella, M.S., (2006). Value-at-Risk Prediction: A Comparison of Alternative Strategies. Journal of Financial Econometrics, 4(1), 53-89.

[87] Kulasekera K.B., (1994). Approximate MLE's of the Parameters of a discrete Weibull Distribution with Type I Censored Data. Microelectronics Reliability, vol. 34, no. 7, 1185-1188.

[88] Kupiec, P., (1995). Techniques for verifying the accuracy of risk measurement models, Journal of Derivatives, 3, 73-84.

[89] Lancaster T., (1979). Econometric Methods for the Duration of Unemployment. Econometrica, 47, 939-956.

[90] Lin T. and Guillén M., (1998). The rising Hazards of Party Incumbency. A Discrete Renewal Analysis. Political Analysis. An Annual Pulication of the Methodology Section of the American Political Science Association. vol. 7, 31-57.

[91] Longin, F. (2001). From VaR to stress testing: the extreme value approach, Journal of Banking and Finance, 24, 1097-1130.

[92] Mandelbroat, Benoit B. (1963). The variation of certain speculative prices, *The Journal of Business*, 36, 394-419.

[93] Malik, R.J. and Trudel, R., (1982). Probability density function of quotient of order statistics from the pareto, power and weibull distributions, Communications in Statistics-Theory and Methods, 11(7), 801-814.

[94] Margolin, B. H. and Winokur, H.S., Jr., (1967). Exact moments of the order statistics of the geometric distribution and their relation to inverse sampling and reliability of redundant systems, Journal of the American Statistical Association, 62, 915-925.

[95] Martins, M.J.(2000). Estimacão de Caudas Pesadas - Variantes ao Estimador de Hill. Tese de Doutoramento, D.E.I.O., Faculdade de Ciências da Universidade de Lisboa.

[96] Mittnik, S. and Paolella, M.S. (2000). Conditional Density and Value-at-Risk Prediction of Asian Currency Exchange Rates. Journal of Forecasting, 19, 313-333.

[97] McAleer, M., Jiménez-Martin, J.A. and Pérez-Amaral, T. (2009). A decision rule to minimize daily capital charges in forecasting value-at-risk. To appear in Journal of Forecasting.

[98] McAleer, M., Jiménez-Martin, J.A. and Pérez-Amaral, T. (2009b). Optimal Risk Management Before, During and After the 2008-09 Financial Crisis. Medium for Econometric Applications, 17(3), 20-27.

[99] McAleer, M., Jiménez-Martin, J.A. and Pérez-Amaral, T. (2009c). Has the Basel II Accord Encouraged Risk Management During the 2008-09 Financial Crisis? Tinbergen Institute Discussion Paper. TI 2009-039/4.

[100] McNeil, A. J., Frey, R., Embrechts, P. (2005). Quantitative Risk Management: Concepts, Techniques and Tools. Princeton: Princeton University Press.

[101] McNeil, A.J. and Frey, R. (2000). Estimation of Tail-related Risk Measures for Heteroscedastic Financial Time Series: An extreme value approach. Journal of Empirical Finance, 7, 271-300.

[102] Nakagawa, T. and Osaki, S., (1975). The discrete Weibull distribution, IEEE Transactions Reliability, vol 24, Dec., 300-301.

[103] Nelson, D.B. (1991). Conditional heteroscedasticity in asset returns: a new approach, Econometrica, 59, 347-370.

[104] Neves, C. and Fraga Alves, M. (2008). The ratio of maximum to the sum for testing super heavy tails. In Arnold, B., Balakrishnan, N., Sarabia, J., and Minguez, R., editors, Advances in Mathematical and Statistical Modeling, 181-194 Birkhauser, Boston.

[105] Ozun, A., Cifter, A. and Yilmazer, S. (2010). Filtered Extreme Value Theory for Value-At-Risk Estimation: Evidence from Turkey. The Journal of Risk Finance Incorporating Balance Sheet, 11, no.2, 164-179.

[106] Padgett W.J. and Spurrier J.D., (1985). Discrete failure models. IEEE Transactions Reliability, vol. 34, no.3, 253-256.

[107] Peng, L. (1998). Asymptotically unbiased estimator for the extreme-value-index. Statistics and Probability Letters, 38(2), 107-115.

[108] Pickands III, J. (1975). Statistical Inference using Extreme value Order Statistics, Ann. Statist., 3, 119-131.

[109] R Development Core Team, (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

[110] Reiss, R.D., and Thomas, M. (2007). Statistical Analysis of Extreme Values, with Application to Insurance, Finance, Hydrology and Other Fields, 3rd edition, Birkhäuser Verlag.

[111] Resnick, S.I. (1997). Heavy tail modeling and teletraffic data. Annals of Statistics 25(5): 1805-1869.

[112] Resnick, S.I. and Starica, C. (1996). Tail Index estimation for Dependent Data. mimeo, School of ORIE, Cornell University, 1996.

[113] Ribatet, M. (2009). POT: Generalized Pareto Distribution and Peaks Over Threshold. R package version 1.0-9. http://people.epf1.ch/mathieu.ribatet,http://r-forge.r-project.org/projects/pot/.

[114] Riskmetrics (1996), J.P. Morgan Technical Document, $4^{th}$ Edition, New York, J.P. Morgan.

[115] Smith, R. (1987). Estimating tails of probability distributions. Ann. Statist., (15)1174-1207.

[116] Smith, R. (1990). Models for Exceedances over High Thresholds. J. R. Statist. Soc. B, (52), no.3, 393-442.

[117] Stein W.E. and Dattero R., (1984). A new discrete Weibull distribution, IEEE Transactions Reliability, vol R-33, Jun., 196-197.

[118] Tsay, R.(2010). Analysis of Financial Time Series. Wiley Series in Probability and Statistics.

[119] Tu, A.H., Wong, W.K. and Chang, M.C. (2008). Value-at-Risk for Long and Short Positions of Asian Stock Markets. International Research Journal of Finance and Economics, 22, 135-143.

[120] Vogt, H., (1968). Zur Parameter- und Prozentpunktschätzung von Lebensdauerverteilungen bei kleinem Stichprobenumfang, Metrika, 14, 117-131.

[121] Weissman, I. (1978). Estimation of Parameters and Large Quantiles Based on the $k$ Largest Observations. J. Amer. Statist. Assoc., 73, 812-815.

[122] Wuertz, D., Chalabi, Y. and Miklovic, M., (2008). fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling, R package version 290.76. http://www.rmetrics.org.

[123] Zhou, C.(2010). The extent of the maximum likelihood estimator for the extreme value index, J. Multivariate Anal. 101-4, 971-983.