A subsampling approach to estimating the distribution of diverging statistics with applications to assessing financial market risks[*]

Patrice Bertail[†]   Christian Haefke[‡]   Dimitris N. Politis[§]

Halbert White[¶]

April 2002

Abstract

In this paper we propose a subsampling estimator for the distribution of statistics diverging at either known or unknown rates when the underlying time series is strictly stationary and strong mixing. Based on our results we provide a detailed discussion of how to estimate extreme order statistics with dependent data and present two applications to assessing financial market risk. Our method performs well in estimating Value at Risk and provides a superior alternative to Hill's estimator in operationalizing Safety First portfolio selection.

JEL Classification: C14, C49, G11
Keywords: Resampling Methods, Extreme Value Statistics, Value at Risk, Portfolio Selection.

[†]Laboratoire de Statistique, CREST, 3, Ave Pierre Larousse, 94205 Malakoff. Email: Patrice.Bertail@ensae.fr

[‡]Department of Economics and Business, Universitat Pompeu Fabra, Ramon Trias Fargas 25–27, 08005 Barcelona, Spain. Email: christian.haefke@econ.upf.es

[§]Department of Mathematics, University of California at San Diego, La Jolla, CA 92093 USA. Email: politis@math.ucsd.edu

[¶]Department of Economics, University of California at San Diego, La Jolla, CA 92093 USA. Email: hwhite@weber.ucsd.edu

1 Introduction

Politis and Romano (1994) introduced the methodology of "subsampling" and showed that it leads to valid large-sample statistical inferences in general estimation situations and data structures (including i.i.d. and stationary data), provided that the employed estimator converges to the unknown parameter at a *known* rate τ_n, and possesses (when normalized by the rate τ_n) a nondegenerate asymptotic distribution. Later, Bertail *et al.* (1999) dispensed with the assumption that the convergence rate τ_n is known, by using a preliminary round of subsampling to consistently estimate the rate τ_n; see also Politis *et al.* (1999) for a review.

In the present paper, we show how subsampling can be used to approximate the sampling distributions of *diverging* (as opposed to converging) statistics that are particularly useful in the context of financial time series and in particular in assessing financial risk. We consider several such statistics in detail as examples. In addition, we deal with the issues of unknown divergence rate and/or unknown rate of "escaping means".

To set up the context for our exposition, let $X^n = (X_1, ..., X_n)$ be an observable stretch of a (strictly) stationary time series $\{X_t, t \in Z\}$ which – for simplicity – will be assumed real-valued. The random variables $\{X_t, t \in Z\}$ are defined on a probability space denoted by $(P, \Omega, \mathcal{A})$, and the first marginal distribution of the $\{X_t, t \in Z\}$ sequence is denoted by F; that is, $P(X_1 \leq x) = F(x)$.

We assume that the time series satisfies a weak dependence condition. In particular, we impose the *strong mixing* condition, namely that $\alpha_X(k) \to 0$ as $k \to \infty$; the Rosenblatt strong mixing coefficients are defined as usual by

$$\alpha_X(k) = \sup_{A,B} |P(A \cap B) - P(A)P(B)|,$$

where A and B are events in the σ-algebras generated by $\{X_t, t < 0\}$ and $\{X_t, t \geq k\}$ respectively; the case where $X_1, ..., X_n$ are independent, identically distributed (i.i.d.) is an important special case in which $\alpha_X(k) = 0$ for all $k > 0$.

A statistic $T_n = T_n(X^n)$ is computed from the data. If T_n estimates an unknown parameter $\theta = \theta(P)$, and if T_n is consistent for θ, then the subsampling methodology of Politis and Romano (1994) immediately applies to provide a consistent approximation to the sampling distribution of T_n under minimal assumptions. In addition, if the rate of convergence of T_n to θ is *a priori* unknown, then it can be estimated via a preliminary round of subsampling and used in the sampling distribution approximation; see Bertail *et al.* (1999) or Politis *et al.* (1999, ch. 8).

In contrast, we are concerned here with *diverging* (as opposed to converging) statistics, i.e., statistics for which[1] $T_n \overset{P}{\longrightarrow} \infty$ as $n \to \infty$. In particular, we will work under the assumption that there exists a (nondegenerate) distribution $K(x, P)$, continuous in x, such that

$$K_n(x, P) \equiv P\{\tau_n T_n \leq x\} \to K(x, P) \tag{1}$$

as $n \to \infty$ for any real number x, where τ_n is a *decreasing* function of n. For example, it may be the case that $\tau_n = n^{-\gamma}$ where γ is a positive constant. Examples of such diverging statistics with well-defined (after normalization) asymptotic distributions are

---

[1] The case where $T_n \overset{P}{\longrightarrow} -\infty$ can be similarly treated by noting that in that case $-T_n \overset{P}{\longrightarrow} \infty$. Note that $\overset{P}{\longrightarrow}$ indicates convergence in probability; the order notations 'in probability', e.g. $O_P(.)$ and $o_P(.)$ will also be used in the sequel – see Brockwell and Davis (1987) for definitions.

given by the extreme order statistics, e.g. the maximum or the kth largest value, from heavy-tailed observations; see Leadbetter *et al.* (1983) where an excellent treatment of extreme order statistics from stationary observations is given. For the converging case, τ_n is an increasing function of n.

Our interest in such statistics is directly motivated by certain quantities relevant for the measurement of exposure to risk in asset markets. To describe these, let $\{P_t\}$ be a sequence of asset prices, such as the daily closing price for the S&P 500 Index, and let $X_t \equiv \ln(P_t/P_{t-1})$ be the one-period return.

Example 1a. The return on the asset from time period 0 to time period n can be measured by

$$\ln(P_n/P_0) = \sum_{t=1}^{n} X_t.$$

A simple form of the efficient markets theory implies that $X_t - \mu$ is a martingale difference sequence with respect to publicly available information, where μ is some constant; note that this is compatible with the assumption that $\{X_t\}$ is a stationary mixing process. In fact, many continuous-time stochastic models of asset price evolution used in the finance literature directly imply that $\{X_t\}$ is a stationary, strong mixing process – examples are the simple geometric Brownian motion *a la* Black-Scholes (1973) and the stochastic volatility models of Hull and White (1987). Taking $T_n = \ln(P_n/P_0) - n\mu$ then gives a simple example of a statistic for which (1) holds: standard central limit results ensure that $\tau_n = 1/\sqrt{n}$. Note, however, that our interest is directed toward the distribution of T_n, i.e., $K_n(x, P)$, rather than the limiting distribution $K(x, P)$. Significantly, this simple case plays a central role in financial risk assessment, through the notion of "value at risk" (VaR) (see for example Hendricks (1996) or Jorion (1997)). One leading measure of VaR is defined as the quantity q such that the probability of the asset value falling below q at the end of a specified time period (of duration n) is a given value p. In other words, the value at risk is simply the quantile

$$q = K_n^{-1}(p, P).$$

The Basle Committee on Banking Supervision has recently proposed the use of value at risk for assessing capital adequacy, while the Securities and Exchange Commission (SEC) has recently required that all corporate treasurers report estimates of value at risk for assets held in corporate portfolios. Standard methods based on assumptions of normality may deliver rather poor estimates of the desired quantiles, due to the well-known excess kurtosis typical of financial market returns. The methods discussed here provide a convenient approach potentially more accurate than current methods (see e.g. Hendricks (1996)).

Example 1b. Another measure of value at risk that is computable without any knowledge of the centering factor $n\mu$ is based on the quantiles of $U_n = \ln(P_n/P_0)$. A simple rescaling generally will not ensure that equation (1) holds with U_n instead of T_n. We therefore have to take into account a centering factor which in general may take the form of a sequence μ_n; in other words, starting with a general statistic U_n, we define $T_n = U_n - \mu_n$, where μ_n is an appropriately chosen sequence such that (1) holds as stated. This is actually a most general set-up, and our statistic U_n will diverge to

infinity if the sequence τ_n is decreasing (the "inflating variance case"), *and/or* if the product $\tau_n \mu_n$ is increasing (the "escaping mean case"). Our section 3 is devoted to this general set-up.[2]

Example 2. Next, consider the worst cumulative return that an asset yields at any point during a fixed time horizon of duration n,

$$T_n = \min_{1 \le t \le n} R_t, \quad R_t = \sum_{\tau=1}^{t} X_\tau.$$

This quantity determines whether "stop" orders are hit or bankruptcy occurs during the horizon and also determines whether "knock-out" or "knock-in" provisions of certain exotic but increasingly popular options are activated. The distribution of T_n gives the probability of occurrence of such events. Further, the quantiles of T_n provide another useful measure of risk exposure, which we shall call "extreme value at risk", (XVaR) as these give the quantities q such that a loss will exceed q at any time during the specified period with specified probability. Analytic analysis of the distribution of T_n is not simple. Nevertheless, our methods provide a simple and direct way to consistently estimate the distribution or quantiles for this choice of T_n.

Example 3. Our third example comes from the literature of "safety first" portfolio selection (Roy, 1952; Arzac and Bawa, 1977; Jansen, et al., 1998). In this theory, it is of interest to know the distribution of the worst possible one-period return over a given time horizon,

$$T_n = \max(-X_1, \ldots, -X_n).$$

This quantity is also the focus of "worst case scenario" analysis, an alternative to value at risk discussed by Boudoukh, Richardson and Whitelaw (1995).

A standard regularity condition on the tail of the distribution F of the X_t's is that, for some $\alpha > 0$ measuring tail thickness and a finite constant Γ, we have

$$x^\alpha (1 - F(x)) \to \Gamma \quad \text{as } x \to \infty.$$

Under an assumption of strong mixing, it can be shown in this case that τ_n is proportional to $n^{-1/\alpha}$ (Leadbetter *et al.*, 1983). Our methods will permit us to estimate not only the distribution K_n and quantiles K_n^{-1}, but also the tail parameters α and Γ.

Without limiting ourselves to a particular example, we work out in Section 2 consistent approximations to the decreasing sequence τ_n and to $K_n(x, P)$ and $K(x, P)$ via subsampling in the case where the centering sequence μ_n is negligible, i.e. where $\mu_n = 0$ or at least $\tau_n \mu_n \to 0$. We further show that the quantiles of $K_n(x, P)$ and $K(x, P)$ can also be estimated consistently; as a result, *prediction intervals* for T_n can be formulated. We also establish that, if the objective is estimation of the sampling distribution of T_b (with b fixed), then a simple subsampling estimator may be applied,

---

[2] Note, however, that the effect of the centering μ_n will be negligible if $\tau_n \mu_n \to 0$, e.g., when the standard deviation of the general statistic U_n increases faster than its mean EU_n; this is actually what happens in the case of the maximum of heavy-tailed observations of our Example 3 – see also section 3.1 in what follows.

and a reference to the large-sample distribution $K(x, P)$ may be unnecessary *even if the form of $K(x, P)$ is explicitly known.* In section 3 , we consider the general case when the centering sequence μ_n is unknown and nonnegligible, and we give convergent estimators of τ_n and μ_n. Our results are illustrated by the case of the maximum when the domain of attraction of the underlying distribution is totally unknown. In Section 4 we focus attention on T_n being the maximum of heavy-tailed stationary observations. In that case, it is well-known that $K(x, P)$ has a specific form (the type II extreme distribution − see Leadbetter *et al.* (1983)) which depends on P only through F, and in particular only on the tail of F. We show how subsampling can be successfully used to give a valid estimate of the 'thickness' parameter α of the tail of F; consequently, the subsampling estimate of the distribution of T_n and/or the limit distribution $K(x, P)$ (with estimated 'thickness' parameter) can be inverted for use in the problem of safety-first portfolio selection. In Section 5 we illustrate applications of our subsampling methodology by obtaining estimates of various quantities relevant for measuring exposure to financial risk, including value at risk and safety first portfolio selection. All technical proofs are placed in Appendix A.

2 Subsampling for diverging statistics: the inflating variance case

2.1 Some definitions and a first result

Define Y_i to be the subsequence $(X_i, X_{i+1}, ..., X_{i+b_n-1})$, for $i = 1, ..., q$, and $q = n - b_n + 1$; note that Y_i consists of b_n consecutive observations from the $X_1, ..., X_n$ sequence, and the order of the observations is preserved. Now let $T_{b_n,i}$ be the value of the statistic T_{b_n} applied to the subsample Y_i. The subsampling distribution of $\tau_n T_n$, based on a subsample size b_n, is defined by

$$K_{b_n}(x \mid X^n, \tau.) \equiv q^{-1} \sum_{i=1}^{q} 1\{\tau_{b_n} T_{b_n,i} \leq x\}, \tag{2}$$

where $1\{A\}$ is the indicator of set A; note that the subsample size b_n is allowed to depend on the actual sample size n.

The following theorem shows that $K_{b_n}(x \mid X^n, \tau.)$ is a consistent distribution estimator; it complements the results of Politis and Romano (1994) by dealing with the case of possibly diverging statistics, i.e., where τ_n may be decreasing.

Theorem 1 *Let the X-sequence be stationary and strong mixing. Let τ_n be a given known function of n.*

(a) If $b_n = b$ (a constant), then, as $n \to \infty$,

$$K_{b_n}(x \mid X^n, \tau.) = K_b(x, P) + o_P(1),$$

for all x.

(b) Suppose the convergence (1) is true for τ_n a known function of n and that

4

$b_n \to \infty$ as $n \to \infty$, but with $b_n/n \to 0$; then

$$\sup_x |K_{b_n}(x \mid X^n, \tau_.) - K(x, P)| = o_P(1), \qquad (3)$$

and

$$\sup_x |K_{b_n}(x \mid X^n, \tau_.) - K_{d_n}(x, P)| = o_P(1), \qquad (4)$$

where d_n is any sequence such that $d_n \to \infty$.

Part (a) of the Theorem indicates that, if our objective is to estimate $K_b(x, P)$ for some fixed b, then this can be achieved in a consistent manner using $K_{b_n}(x \mid X^n, \tau_.)$ with $b_n = b$. If our objective is to estimate $K_{d_n}(x, P)$ for some d_n such that $d_n \to \infty$ (e.g. $d_n = n$), then this can be accomplished using $K_{b_n}(x \mid X^n, \tau_.)$ with $b_n \to \infty$ (but always such that $b_n/n \to 0$).

Remark 1.1 Note that assumption (1) of part (b) in the Theorem concerns the *existence* of a (continuous) limit distribution; the shape of this limit distribution may well be unknown, and can be estimated by (3) above.

Remark 1.2 Looking at the proof of part (a) of the Theorem, note first that for the asymptotic approximation to be useful in a finite-sample situation, b not only has to be fixed, but it has to be small as compared to n (so that b/n is small). Also note that for part (a), existence of a limit distribution (as in (1)) is *not* required; thus, part (a) is true under very few assumptions. Furthermore, it turns out that we do not even need to consider a divergence rate τ_n in order to use a result similar to part (a); this is addressed by the following corollary for which the following definitions are required. Let

$$\bar{K}_{b_n}(x \mid X^n) \equiv K_{b_n}(x \mid X^n, 1),$$

and

$$\bar{K}_b(x, P) \equiv P\{T_b \leq x\}.$$

Corollary 1 *Let the X-sequence be stationary and strong mixing. If $b_n = b$ (a constant), then, as $n \to \infty$,*

$$\bar{K}_b(x \mid X^n) = \bar{K}_b(x, P) + o_P(1),$$

for all x. In addition, if $\bar{K}_b(x, P)$ is continuous in x, we have

$$\sup_x |\bar{K}_b(x \mid X^n) - \bar{K}_b(x, P)| = o_P(1).$$

Another immediate corollary (in the fixed b case) pertains to estimation of the quantiles of $\bar{K}_b(x, P)$ relevant for setting prediction intervals for 'future' observations of T_b. Given a distribution G on the real line and a number $t \in (0, 1)$, we let $G^{-1}(t)$ denote the quantile transformation, i.e., $G^{-1}(t) = \inf \{x : G(x) \geq t\}$, which reduces to the regular inverse of the function G if G happens to be invertible.

Corollary 2 *Let the X-sequence be stationary and strong mixing. If $b_n = b$ (a constant), and if $\bar{K}_b(x, P)$ is continuous in x, then, as $n \to \infty$,*

$$\bar{K}_{b_n}^{-1}(t \mid X^n) = \bar{K}_b^{-1}(t, P) + o_P(1), \qquad (5)$$

for any $t \in (0, 1)$.

2.2 Case of unknown divergence rate τ_n

In Theorem 1 we treated the case of known divergence rate τ_n. Nevertheless, in many cases τ_n is unknown, its form depending on the (unknown) probability measure P. It is often the case that the assumption $\tau_n = n^{-\gamma}$ is reasonable, but then typically the positive constant γ is unknown as it may depend crucially on P. In the case of converging statistics (i.e. $\gamma < 0$), estimation of the rate was treated in Bertail *et al.* (1999) and Politis *et al.* (1999, ch. 8). We will now show how similar ideas are applicable also in the case of diverging statistics ($\gamma > 0$).

The following Lemma relates the quantiles of $\bar{K}_{b_n}(x \mid X^n)$ to those of $K(x, P)$ through the (unknown) rate τ_n; it will be the first step towards constructing an estimator of the rate as in Bertail *et al.* (1999)—see also Politis *et al.* (1999, ch. 8).

Lemma 1 *Let $k_0 = \sup\{x : K(x, P) = 0\}$ and $k_1 = \inf\{x : K(x, P) = 1\}$, and assume that $K(x, P)$ is continuous and strictly increasing on (k_0, k_1) as a function of x. If (3) is true as n tends to infinity, then*

$$\tau_{b_n} \bar{K}_{b_n}^{-1}(t \mid X^n) = K^{-1}(t, P) + o_P(1), \tag{6}$$

for any $t \in (0, 1)$.

For example, if $t = 1/2$, then (6) simply says that $b_n^{-\gamma} \text{median}\{T_{b_n, i}, i = 1, ..., q\} \xrightarrow{P}$ median of $K(x, P)$, where $\text{median}\{T_{b_n, i}, i = 1, ..., q\}$ is the sample median of the collection $\{T_{b_n, i}, i = 1, ..., q\}$, and γ is suitably chosen.

Therefore, if we look at the sample median of the $\{T_{b_n, i}, i = 1, ..., q\}$ and/or other sample quantiles (i.e. different choices of t in (6)) for different choices of b_n (with all choices satisfying $b_n \to \infty$ and $b_n/n \to 0$), then we can estimate the constant γ by a simple regression as discussed next. Having estimated γ, estimates of the limit quantiles $K^{-1}(t, P)$ are also immediately available through (6).

To see this, note that by taking logarithms in (6) and assuming that $\tau_n = n^{-\gamma}$ we get

$$\log\left(\left|\bar{K}_{b_n}^{-1}(t \mid X^n)\right|\right) = \log\left(\left|K^{-1}(t, P)\right|\right) + \gamma \log b_n + o_P(1), \tag{7}$$

so that γ is simply the 'slope' in regressing $\log\left(\left|\bar{K}_{b_n}^{-1}(t \mid X^n)\right|\right)$ on $\log b_n$. So if we consider different subsample sizes $b_{i,n}$, $i = 1, ..., I > 1$, we can use the least squares estimator of slope, namely

$$\gamma_I \equiv \frac{\sum_{i=1}^{I}(y_i - \overline{y})(\log(b_{i,n}) - \overline{\log})}{\sum_{i=1}^{I}(\log(b_{i,n}) - \overline{\log})^2} \tag{8}$$

where for given $t \in (0, 1)$ $y_i = \log\left(\left|\bar{K}_{b_{i,n}}^{-1}(t \mid X^n)\right|\right)$, $\overline{y} = I^{-1}\sum_{i=1}^{I} y_i$, and $\overline{\log} = I^{-1}\sum_{i=1}^{I} \log(b_{i,n})$. Lemma 2 below establishes the consistency of γ_I.

Lemma 2 *Let the X-sequence be stationary and strong mixing. Suppose (1) is true, and $\tau_n = n^{-\gamma}$ where γ is a positive unknown constant. Let $K(x, P)$ be continuous and strictly increasing on (k_0, k_1) as a function of x, where $k_0 = \sup\{x : K(x, P) = 0\}$ and $k_1 = \inf\{x : K(x, P) = 1\}$. Let $b_{i,n} = n^{\beta_i}$, $1 > \beta_1 > \cdots > \beta_I > 0$, and let t be a point in $(0, 1)$. Then $\gamma_I = \gamma + o_P((\log n)^{-1})$.*

We remark that instead of looking at individual quantiles we could use differences of quantiles, e.g. the inter-quartile range, etc., in order to estimate γ; this is because (6) implies that, if $t_1 < t_2$,

$$\tau_{b_n}\left(\bar{K}_{b_n}^{-1}(t_2 \mid X^n) - \bar{K}_{b_n}^{-1}(t_1 \mid X^n)\right) = K^{-1}(t_2, P) - K^{-1}(t_1, P) + o_P(1),$$

and thus

$$\log\left(\bar{K}_{b_n}^{-1}(t_2 \mid X^n) - \bar{K}_{b_n}^{-1}(t_1 \mid X^n)\right) = \log\left(K^{-1}(t_2, P) - K^{-1}(t_1, P)\right) + \gamma \log b_n + o_P(1)$$

as well. Bertail *et al.* (1999) originally developed this idea of looking at ranges, and we will take it up again in section 3.

Lemma 3 below is an immediate corollary of Lemmas 1 and 2; its essence is that the quantiles $K^{-1}(x, P)$ (that represent the 'constant' term in the regression (7)) can be consistently estimated as well, using our estimator of γ in place of the true one, or –for that matter– using any other (at least as accurate) estimator of γ.

Lemma 3 *Let $\hat{\gamma}$ be a statistic calculated from the data X^n such that $\hat{\gamma} = \gamma + o_P((\log n)^{-1})$. Under the assumptions of Lemma 1, and assuming that $\tau_n = n^{-\gamma}$ where γ is a positive unknown constant, we have that*

$$\hat{\tau}_{b_n}\bar{K}_{b_n}^{-1}(t \mid X^n) = K^{-1}(t, P) + o_P(1)$$

where $\hat{\tau}_{b_n} = b_n^{-\hat{\gamma}}$.

Lemma 2 can be generalized to take into account many different quantiles (i.e., t-points) at the same time. So let some points $t_j \in (0, 1)$ be given, where $j = 1, ..., J \geq 1$, together with subsample sizes $b_{i,n}$, $i = 1, .., I > 1$. Equation (7) yields

$$y_{i,j} \equiv \log\left(\left|\bar{K}_{b_{i,n}}^{-1}(t_j \mid X^n)\right|\right) = a_j + \gamma \log(b_{i,n}) + u_{i,j} \tag{9}$$

where $a_j \equiv \log\left(\left|K^{-1}(t_j, P)\right|\right)$ and $u_{i,j} = o_P(1)$, $i = 1, ..., I$ and $j = 1, ..., J$.

Following Bertail *et al.* (1999) we suggest the following ANOVA-type estimator of γ :

$$\gamma_{I,J} \equiv \frac{\sum_{i=1}^{I}(y_{i,.} - \overline{y})(\log(b_{i,n}) - \overline{\log})}{\sum_{i=1}^{I}(\log(b_{i,n}) - \overline{\log})^2} \tag{10}$$

where $y_{i,.} = J^{-1}\sum_{j=1}^{J} y_{i,j} = J^{-1}\sum_{j=1}^{J} \log\left(\left|\bar{K}_{b_{i,n}}^{-1}(t_j \mid X^n)\right|\right)$,
$\overline{y} = (IJ)^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J} y_{i,j}$ and $\overline{\log} = I^{-1}\sum_{i=1}^{I} \log(b_{i,n})$. Theorem 2 now offers a generalization of Lemma 2.

Theorem 2 *Let the X-sequence be stationary and strong mixing. Suppose (1) is true, and $\tau_n = n^{-\gamma}$ where γ is a positive unknown constant. Let $K(x, P)$ be continuous and strictly increasing on (k_0, k_1) as a function of x, where $k_0 = \sup\{x : K(x, P) = 0\}$ and $k_1 = \inf\{x : K(x, P) = 1\}$. Let $b_{i,n} = n^{\beta_i}$, $1 > \beta_1 > \cdots > \beta_I > 0$, and let points $t_j \in (0, 1)$, $j = 1, ..., J \geq 1$ be given. Then $\gamma_{I,J} = \gamma + o_P((\log n)^{-1})$.*

What is perhaps more important than rate estimation *per se*, is that the estimated divergence rate can be used in turn in order to achieve consistent estimation of $K(x, P)$ and its quantiles. In other words, the convergence $\gamma_{I,J} = \gamma + o_P((\log n)^{-1})$ is just fast enough to enable us to plug the estimated rate into the subsampling distribution

7

without spoiling its consistency; this is the subject of Theorem 3 below. Note that this $o_P((\log n)^{-1})$ term is just an upper bound; given additional regularity conditions, the convergence rate of $\gamma_{I,J}$ can be much faster than logarithmic—see e.g. the numerical examples in Politis *et al.* (1999, ch. 8). However, *any* estimator $\hat{\gamma}$ such that $\hat{\gamma} = \gamma + o_P((\log n)^{-1})$ is accurate enough to be used in the subsampling estimator of $K(x, P)$.

Theorem 3 *Let the X-sequence be stationary and strong mixing. Suppose (1) is true, and $\tau_n = n^{-\gamma}$ where γ is a positive unknown constant. Let $K(x, P)$ be continuous and strictly increasing on (k_0, k_1) as a function of x, where $k_0 = \sup\{x : K(x, P) = 0\}$ and $k_1 = \inf\{x : K(x, P) = 1\}$. Let $\hat{\gamma}$ be a statistic calculated from the data X^n and such that $\hat{\gamma} = \gamma + o_P((\log n)^{-1})$; also let $\hat{\tau}_n = n^{-\hat{\gamma}}$. Let b_n be such that $b_n \to \infty$ as $n \to \infty$, but $b_n/n \to 0$. Then*

$$\sup_x |K_{b_n}(x \mid X^n, \hat{\tau}_{\cdot}) - K(x, P)| = o_P(1). \tag{11}$$

Let $t \in (0, 1)$, and let $c_n(t) = K_{b_n}^{-1}(t \mid X^n, \hat{\tau}_{\cdot})$ be the tth quantile of the subsampling distribution $K_{b_n}(x \mid X^n, \hat{\tau}_{\cdot})$. Then

$$P\{\hat{\tau}_{d_n} T_{d_n} \le c_n(t)\} \longrightarrow t, \tag{12}$$

where d_n is any sequence such that $d_n \to \infty$, i.e., $c_n(t)$ consistently estimates the t-quantile of $\hat{\tau}_{d_n} T_{d_n}$.

Remark 3.1. Theorem 3 is a generalization of part (b) of Theorem 1 in the case where the divergence rate τ_n is unknown and must be estimated. Note in particular that equation (12) may be used for the construction of *prediction intervals* for future observations of T_{d_n}.

3   Subsampling for diverging statistics: general case

As mentioned in Example 1b of the Introduction, the divergence of a general statistic U_n may be due either to an inflating variance and/or an escaping mean; this general case is addressed in the present section. We assume that U_n is a computable statistic such that, as $n \to \infty$ for any real number x,

$$K_n(x, P) \equiv P\{\tau_n(U_n - \mu_n) \le x\} \longrightarrow K(x, P); \tag{13}$$

note that equation (13) is identical to (1) with the definition $T_n = U_n - \mu_n$.

Estimating simultaneously the divergence rate τ_n and the centering factor μ_n (or at least its dominant part) is the subject of Section 3.2 below; we also achieve the construction of consistent subsampling distribution estimators. Note that stated in full generality, the problem of estimating μ_n and τ_n is not identifiable: indeed if $\tau_n \mu_n$ converges to some constant $C \ne 0$, then $P(\tau_n U_n \le x) \longrightarrow K'(x, P) = K(x - C, P)$, i.e., we also have an acceptable limiting distribution with the choice $\mu'_n = 0$. To make the problem identifiable, we may –without loss of generality– assume that $\tau_n \mu_n$ is either a strictly monotone function or identically zero (that is $\mu_n = 0$); otherwise, we may assume that the median of $K(x, P)$ is known up to some constant. In the same way the rate of convergence can only be identified up to a multiplicative constant. Note however that these constants have no importance in the construction of confidence intervals because of the invariance properties of subsampling distributions.

3.1 Centering and normalization of the maximum

Before proceeding with our theoretical results we now elaborate on our Example 1 of the Introduction for further motivation; the elaboration consists in allowing for arbitrary distributions generating the data.

To fix ideas we now take $U_n = \max\{X_1, ..., X_n\}$, with X_i being i.i.d. from distribution F; the assumption of independence is not crucial and can be replaced with an assumption of stationarity and weak dependence (mixing) –see Leadbetter *et al.* (1983). Then there are only three possible shapes for the limiting distribution $K(x, P)$ of the appropriately centered and standardized extreme value statistic U_n; the three cases are:

$$Type \ \ I: \ \ K(x, P) = \Lambda(x) \equiv \exp(-\exp(-x)).$$

$$Type \ \ II: \ \ K(x, P) = \Phi_\alpha(x) \equiv \begin{cases} 0 & \text{if } x \leq 0 \\ \exp(-x^{-\alpha}) & \text{if } x > 0 \end{cases}$$

$$Type \ \ III: \ \ K(x, P) = \Psi_\alpha(x) \equiv \begin{cases} \exp(-(-x)^\alpha) & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases}$$

where α is some positive constant. Which of the three cases obtains depends on the generating distribution F. We say that F "falls in the domain of attraction" of distribution G (denoted by $F \in D(G)$) if the limiting distribution $K(x, P)$ of the appropriately centered and standardized extreme value statistic U_n turns out to be G.

Characterizing the domain of attraction of F according to its tail as well as the norming and centering constants τ_n and μ_n has received a lot of attention; let us recall the main facts. Denote by R_α the set of regularly varying functions at ∞ with index α, that is, the set of measurable positive real functions $L(\cdot)$ such that

$$\lim_{x \to \infty} L(\lambda x)/L(x) = \lambda^\alpha,$$

and define the "quantile" inverse of function $H(\cdot)$ by $H^{-1}(x) = \inf\{y : H(y) \geqslant x\}$. We have the following equivalence (see e.g. Bingham et al. (1987, chap. 8)):

(i) $F \in D(\Lambda) \Leftrightarrow H^{-1}(x + y) - H^{-1}(x) \sim yl(e^x)$ when $x \to \infty$, where $l \in R_0$ and we put $H(x) \equiv -\log(1 - F(x))$. In that case, one may choose $\mu_n = H^{-1}(\log n)$ and $\tau_n = \frac{1}{H^{-1}(\log n + 1) - H^{-1}(\log n)}$. It is clear in this case that both the centering factor μ_n and the rate τ_n contribute to the divergence of U_n, and they both depend on the underlying distribution F.

(ii) $F \in D(\Phi_\alpha) \Leftrightarrow 1 - F(x) \in R_{-\alpha}$, ; in this case one may choose $\mu_n = 0$ and $\tau_n = \frac{1}{(1-F)^{-1}(\frac{1}{n})}$. This is the heavy-tailed case of Example 1 of the Introduction, which will be revisited in section 4.2 as well.

(iii) $F \in D(\Psi_\alpha) \Leftrightarrow$ The upper endpoint (say x_0) of F is finite and $1 - F(x_0 - \frac{1}{x}) \in R_{-\alpha}$; in this case one may choose $\mu_n = x_0$ and $\tau_n = \frac{1}{F_1^{-1}(\frac{1}{n})}$, where $F_1(\cdot) = 1 - F(x_0 - \cdot)$. Note that here T_n is actually a converging statistic and that μ_n is not escaping. Thus the results of Bertail et al. (1999) may be applied directly in this case. Recentering the subsampling distribution at T_n would yield a convergent estimator of the true distribution of the maximum. Nevertheless, we show below how this case may be treated simultaneously with the others in a single unifying framework.

Typically, practitioners make some assumptions on the behavior of the tail of F so that the domain of attraction is known and the problem mainly reduces to the estimation of the index α in case (i) and (ii) and to the characterization of the norming and centering constants in case (iii). A general theory that allows the estimation of the distribution of the maximum and the estimation of the norming and centering sequences without this kind of assumption is of interest from a robustness point of view. Moreover, even in case (iii), estimating τ_n and μ_n simultaneously is a difficult problem. In the following we show how the subsampling ideas presented in section 2 may be successfully applied in this general case as well.

3.2 Estimation of the centering and norming sequences

It is clear that by putting $T_n = U_n - \mu_n$, results obtained in Theorems 1 and 2 and the associated lemmas also hold for U_n. In particular we have the following straightforward result relating the quantile of the subsampling distribution of U_n to the norming and centering factors; its proof is similar to the proof of Lemma 1.

Lemma 4 *Assume that the conditions of Lemma 1 hold for $T_n = U_n - \mu_n$ and let*

$$\overline{K}_{U,b_n}(x|X^n) = q^{-1} \sum_{i=1}^{q} 1\{U_{b_n,i} \le x\}$$

denote the unnormalized subsampling distribution of U_n, where $U_{b_n,i}$ is the value of the statistic U_{b_n} as computed from subsample Y_i. Then we have

$$\overline{K}_{U,b_n}^{-1}(t|X^n) = \mu_{b_n} + \tau_{b_n}^{-1} K^{-1}(t,P) + o_P(\tau_{b_n}^{-1}). \tag{14}$$

It follows from this lemma that for any point $0 < t_2 < t_1 < 1$ we have

$$|\bar{K}_{U,b_n}^{-1}(t_2 \mid X^n) - \bar{K}_{U,b_n}^{-1}(t_1 \mid X^n)|/|K^{-1}(t_2,P) - K^{-1}(t_1,P)| = \tau_{b_n}^{-1}(1 + o_P(1)). \tag{15}$$

Suppose first that we are just interested in estimating the shape of τ_n *without* assuming that $\tau_n = n^{-\gamma}$ as in the previous section. Then, since we can not hope to estimate τ_n but up to a multiplicative constant, we can simply choose $\widehat{\tau}_{b_n} = |\bar{K}_{U,b_n}^{-1}(t_2 \mid X^n) - \bar{K}_{U,b_n}^{-1}(t_1 \mid X^n)|^{-1}$ as our rate estimator. In other words, *any range of the subsampling distribution may be used to normalize the subsampling distribution* which is a remarkable fact. In that case it simply means that the limiting distribution is $\widetilde{K}(.,P) = K(.,P)/|K^{-1}(t_2,P) - K^{-1}(t_1,P)|$; that is, the limiting distribution is normalized by imposing $|\widetilde{K}^{-1}(t_1,P) - \widetilde{K}^{-1}(t_2,P)| = 1$. A useful practical choice may be to use the inter-quartile range, i.e., to let $t_1 = 3/4$ and $t_2 = 1/4$.

Because μ_n is a centering factor, information on the median of the asymptotic distribution may be of some interest. For instance, for the type I of section 3.1, the limiting distribution is parameter free with a median equal to $-\log\log 2$. Evaluating equation (14) at the median yields the equation

$$\tau_{b_n} \overline{K}_{U,b_n}^{-1}(1/2| X^n) = \tau_{b_n} \mu_{b_n} - \log\log 2 + o_P(1).$$

Since an obvious estimator of τ_{b_n} is available, it is not difficult to see that a two-step procedure (i.e. plugging the previous estimator $\widehat{\tau}_{b_n}$ in the above equation) yields

a convergent estimator of the function μ_n. Without loss of generality, we might have imposed the assumption that the median of the limiting distribution is zero. This would simply have shifted both the centering factor and the limiting distribution. With this identifying condition, the median of the subsampling distribution is actually the adequate recentering factor. The above considerations suggest the following general result.

Theorem 4 *Assume that there exist sequences τ_n and μ_n such that*

$$K_n(x, P) \equiv P\{\tau_n(U_n - \mu_n) \leq x\} \longrightarrow K(x, P),$$

where the asymptotic distribution $K(x, P)$ may be of unknown shape but is –without loss of generality– assumed to be normalized such that $K^{-1}(1/2, P) = 0$ and $|K^{-1}(t_2, P) - K^{-1}(t_1, P)| = 1$ at some given points of continuity of $K(., P)$, say $0 < t_1 < t_2 < 1$. Let b_n be a subsampling size satisfying the assumptions of Theorem 1(b), and consider the median of the undersampling distribution

$$\widehat{\mu}_{b_n} = \overline{K}_{U,b_n}^{-1}(1/2|\, X^n).$$

Then we have $\widehat{\mu}_{b_n} = \mu_{b_n} + o_P(\tau_{b_n}^{-1})$ as $n \to \infty$.
   Moreover

$$\widehat{\tau}_{b_n} = |\bar{K}_{U,b_n}^{-1}(t_2 \mid X^n) - \bar{K}_{U,b_n}^{-1}(t_1 \mid X^n)|^{-1} = \tau_{b_n}(1 + o_P(1)),$$

and we have

$$q^{-1} \sum_{i=1}^{q} 1\{\widehat{\tau}_{b_n}(U_{b_n,i} - \widehat{\mu}_{b_n}) \leq x\} \to K(x, P).$$

The proof of the theorem follows directly from (14) and (15) and is omitted. Of course, other identifiability conditions may be used to obtain the form of the centering and norming constant. In some cases (for instance in case (iii) of section 3.1), the asymptotic distribution is already "normalized" and entirely known. In that case equation (14) gives straightforward estimators of the normalizing and standardizing constants.

Theorem 5 *Assume that $\tau_n \to 0$ as $n \to \infty$, that b_n is a subsampling size satisfying the assumptions of Theorem 1(b), and that the asymptotic distribution $K(t, P) = K(t)$ is known and independent of P. For any points $(t_{1,i}, t_{2,i})$ $i = 1, ..., I$ such that $|K^{-1}(t_{2,i}, P) - K^{-1}(t_{1,i}, P)| \neq 0$, define*

$$\widehat{\tau}_{b_n} = \left(I^{-1} \sum_{j=1}^{I} |\bar{K}_{U,b_n}^{-1}(t_{2,i} \mid X^n) - \bar{K}_{U,b_n}^{-1}(t_{1,i} \mid X^n)|/|K^{-1}(t_{2,i}, P) - K^{-1}(t_{1,i}, P)|\right)^{-1}.$$

Then $\widehat{\tau}_{b_n} = \tau_{b_n}(1 + o_P(1))$. Define

$$\widehat{\mu}_{b_n, I} = I^{-1} \sum_{i=1}^{I} (K_{U,b_n}^{-1}(t_i|\, X^n) - \widehat{\tau}_{b_n}^{-1} K^{-1}(t_i)).$$

   Then

$$\widehat{\mu}_{b_n, I} = \mu_{b_n} + o_P(\tau_{b_n}^{-1}).$$

Moreover,

$$q^{-1} \sum_{i=1}^{q} 1\{\widehat{\tau}_{b_n} (U_{b_n,i} - \widehat{\mu}_{b_n}) \leq x\} \to K(x).$$

Theorems 4 and 5 are useful in standardizing the subsampling distribution to obtain –for instance– predictive confidence intervals. Theorem 4 can be used to obtain the shape of the normalizing sequence and of the asymptotic distribution when everything is practically unknown. We emphasize however, that the rate τ_n (and the centering product $\tau_n \mu_n$) are only estimable up to a multiplicative constant because of the identifiability issues we discussed previously.

We now follow up on the "difference of quantiles" idea that originated in Bertail et al. (1999)—see also Politis *et al.* (1999, ch. 8). When the functional form of the rate τ_n is known up to some finite dimensional parameter, it is quite easy to see how the ideas of Section 2 may be implemented to obtain a convergent estimator of the norming sequences. Using a difference of quantiles for some fixed $t_1 > t_2$ we define

$$
\begin{aligned}
y(t_1, t_2, b_n) &\equiv \log\left(\left|\bar{K}_{U,b_n}^{-1}(t_2 \mid X^n) - \bar{K}_{U,b_n}^{-1}(t_1 \mid X^n)\right|\right) \qquad (16) \\
&= y(t_1, t_2, \infty) - \log \tau_{b_n}(1 + o_P(1)) \\
\text{with } y(t_1, t_2, \infty) &= \log\left(\left|K^{-1}(t_2, P) - K^{-1}(t_1, P)\right|\right).
\end{aligned}
$$

If we ensure that τ_n is of the form $n^{-\gamma}$ or $h(n)^{-\gamma}$ for some $\gamma \neq 0$ and some (known) real function h then a straightforward regression gives the value of the index γ as before. This method is referred as the "range" method in Bertail et al. (1999) and is illustrated by the following result.

Theorem 6 *Assume that $\tau_n = n^{-\gamma}$ for some $\gamma \neq 0$, , let $(t_{1,j}, t_{2,j})$ $j = 1, ..., J$ be some points in $(0,1)^2$, and $b_{i,n} = n^{\beta_i}$, $1 < i < I$, be some subsampling sizes satisfying the assumptions of Theorem 1(b) ; then the least-squares estimator of γ (say $\widehat{\gamma}_{I,J,n}$) in the model*

$$y(t_{1,j}, t_{2,j}, b_{i,n}) = y(t_{1,j}, t_{2,j}, \infty) + \gamma \log b_{n,i} + u_{j,i}$$

is such that

$$\widehat{\gamma}_{I,J,n} = \gamma + o_P((\log n)^{-1})$$

The proof of this result is similar to the proof of Theorem 2 and is skipped. Note that γ may be positive or negative.

3.3   General rates and slowly varying functions

In many interesting situations, τ_n is a regular varying function of index γ, say $\tau_n = n^{-\gamma} L(n)$ where L is a slowly varying function that is such that, for any $\lambda > 0$, $\lim_{x \to \infty} \frac{L(\lambda x)}{L(x)} = 1$ (see Bingham, Goldie and Teugels (1987)). Since the behavior of L is only interesting for x tending to ∞, the problem of estimating L amounts in fact to finding an "estimated" equivalent of L for large x , up to a multiplicative constant term. The Karamata representation theorem states that there exists a measurable bounded function $\varepsilon(.)$, with $\varepsilon(u) \longrightarrow 0$ as $u \to \infty$ such that $L(n) = \exp(\int_1^n u^{-1} \varepsilon(u) du) = \exp(h_1(\log n))$, where h_1 is a C^∞ function with all derivatives vanishing at ∞ (see

Bingham et al. (1987), Th. 1.3.1 and Th 1.3.3) related to ε by $\varepsilon(t) = h_1^{(1)}(\log t)$, where $h_1^{(1)}$ is the first derivative of h_1. Therefore (16) may be rewritten

$$
\begin{aligned}
y(t_1, t_2, b_n) &= y(t_1, t_2, \infty) + \gamma \log(b_n) + \log L(b_n) + o_P(1) \qquad (17)\\
&= y(t_1, t_2, \infty) + \gamma \log(b_n) + h_1(\log b_n) + o_P(1).
\end{aligned}
$$

With $L(b_n) \neq 1$, the preceeding method of regressing $y(t_1, t_2, b_n)$ on $\log b_n$ at some different points $(t_{1,j}, t_{2,j})$ and different $b_{n,i}$ still yields a convergent estimator of γ but at a rate slower than $\log(n)^{-1}$ (the convergence is due to the fact that $h_1(\log b_n) = o(\log b_n)$). Try for instance $h_1(x) = \log x$ (in that case $L(n) = \log n$) and $b_{n,i} = n^{\beta_i}$ as in Theorem 5; then $\widehat{\gamma}_{I,J,n} = \gamma + O(\frac{\log \log n}{\log n})$. Unfortunately this convergence rate is not sufficient to ensure that $n^{\widehat{\gamma}_{I,J,n}} = n^\gamma(1 + o(1))$ and thus can not be used as a normalizing factor. In that sense, the results in the preceding sections are not robust to the presence of a slowly varying function. It is clear from this situation that γ and h_1 should be estimated simultaneously. As noticed in Bertail et al. (1999), model (17) is known as a partial spline regression model in the econometric literature (see Engle et al. (1986)) and may be estimated using semi-parametric methods. As a particular case, the same method could be applied in the three cases considered in section 3.1 *regardless of the domain of attraction of the distribution.* However it seems to us that this is quite overoptimistic because this method would require the computation of a very large number of subsampling distributions and in any case would require huge sample sizes.

For this reason, we now present a more attractive construction under a mild restriction on the slowly varying function. A quite simple and reasonable hypothesis on L will indeed allow us to give more straightforward estimators of γ and L. The restriction amounts to assuming that

$$
\lim_{x \to \infty} L(x^2)/L(x) = \exp(-C_0), \qquad (18)
$$

where C_0 is some constant. Recall that by the definition of a slowly varying function we have that $L(\lambda x)/L(x)$ converges to 1 for any $\lambda > 0$; that is, in (18) we allow λ to be of the same order as x. This hypothesis clearly holds for the "usual" slowly varying functions of the form $L(x) = \Pi_i(\log^{c_i} x)^{a_i}$ where a_i are some constants and $\log^{c_i}$ denotes the c_i-iterated logarithm.

Then for $\lambda_n = b_n$ such that $b_n^2/n \to 0$ we have

$$
y(t_1, t_2, b_n) - y(t_1, t_2, \lambda_n b_n) = -\gamma \log(b_n) + C_0 + o_P(1) \qquad (19)
$$

It follows from the same arguments as in Section 2 that, using different subsampling sizes, $b_{n,j}$, $j = 1, ..., J \geq 2$, and points $(t_{1,i}, t_{2,i})$ $i = 1, ..., I$, the least squares estimator of γ in (19) gives a convergent estimator $\widehat{\widehat{\gamma}}_{I,J}$ of γ up to $o_P(\log(n)^{-1})$. Now we can plug this estimator in (17) to get (for two fixed points $t_{1,0}$ and $t_{2,0}$)

$$
y(t_{1,0}, t_{2,0}, b_n) - \widehat{\widehat{\gamma}}_{I,J} \log(b_n) = y(t_{1,0}, t_{2,0}, \infty) + h_1(\log b_n) + o_P(1).
$$

Thus the form $h_1(\log b_n)$ can be estimated up to the constant term $y(t_{1,0}, t_{2,0}, \infty)$. Since this constant is of no importance in the construction of the subsampling distribution, we can choose

$$
\widehat{h}_1(\log(b_n)) = y(t_{1,0}, t_{2,0}, b_n) - \widehat{\widehat{\gamma}}_{I,J} \log(b_n)
$$

as a convergent estimator of $h_1(\log b_n)$. This amounts to standardizing the asymptotic distribution by imposing $y(t_{1,0}, t_{2,0}, \infty) = 1$.

In the following we put

$$\widehat{L}(b_n) = \exp(\widehat{h}_1(\log(b_n)))$$

and

$$s(t) = \frac{p(t)}{\int_0^1 p(u)du}$$

with

$$p(t) = \exp(-\frac{1}{x} - \frac{1}{1-x}).$$

We can use the extrapolation scheme of Adamovic (see for instance Bingham et al. (1987), Th. 1.3.4) to construct an asymptotic equivalent of $L(x)$ when $x \to \infty$. Let $b_{n,0}^* = 0 < b_{n,1}^* < b_{n,2}^* < \ldots\ldots < b_{n,K_n}^*$, be a sequence of subsampling sizes. Define for any $x > A > 0$, $x \in [b_{n,i}^*, b_{n,i+1}^*]$, the extrapolation

$$\widehat{L}(x) = \widehat{L}(b_{n,i}^*) + \left(\widehat{L}(b_{n,i+1}^*) - \widehat{L}(b_{n,i}^*)\right) \int_0^{(x-b_{n,i}^*)/(b_{n,i+1}^*-b_{n,i}^*)} s(t)dt. \qquad (20)$$

Now it is apparent that $\widehat{L}$ is a C^∞ slowly varying function that coincides with $\widehat{L}(b_{n,i}^*) = \exp(\widehat{h}_1(\log(b_{n,i}^*)))$ at each $b_{n,i}^*$. The following theorem states that under some regularity assumptions on the subsampling sizes, the suggested construction gives consistent estimators of γ and L; note also that the theorem addresses the general case where γ may be any real number (positive or negative).

Theorem 7 *Assume that we have normalized the asymptotic distribution such that $|K^{-1}(t_{2,0}, P) - K^{-1}(t_{1,0}, P)| = 1$ at some fixed points $t_{1,0}$ and $t_{2,0}$. Assume that $\tau_n = n^{-\gamma}L(n)$ and that $L(x)$ is a normalized slowly varying function satisfying (18); let $(t_{1,i}, t_{2,i})$ $i = 1, ..., I$ be some real numbers such that $t_{1,i} \neq t_{2,i}$ for all $i = 1, ..., I$. Choose $b_{n,j}$, $j = 1, ..., J \geq 2$, such that $b_{n,j}^2$ satisfy the hypotheses of Theorem 1(b). Then the least-squares estimators of γ and C_0 in (19) satisfy*

$$\widehat{\widehat{\gamma}}_{I,J} = \gamma + o_P((\log n)^{-1})$$

and

$$\widehat{C}_o = C_0 + o(1).$$

Moreover, for $b_{n,0}^* = 0 < b_{n,1}^* = n^{1/2} < b_{n,2}^* < \ldots\ldots < b_{n,K_n}^*$ such that $b_{n,;j+1}^*/b_{n,j}^* < e$, $j = 1, ..., M_n$, with $M_n = o(\log n)$ but $M_n \to \infty$ then we have that the estimator in (20) satisfies

$$\frac{L(x)}{\widehat{L}(x)} = 1 + o_P(1), \quad \frac{\widehat{L}(x)}{L(x)} = 1 + o_P(1), \text{ for } x \to \infty.$$

Finally, we have

$$\widehat{\widehat{\tau}}_n \equiv n^{-\widehat{\widehat{\gamma}}_{I,J}}\widehat{L}(n^{1/2})\exp(\widehat{C}_o) = n^{-\widehat{\widehat{\gamma}}_{I,J}}\widehat{L}(b_{n,1})\exp(\widehat{C}_o) = \tau_n(1 + o(1)).$$

14

This theorem states that it is possible both to obtain simple estimates of the index and the slowly varying function (provided that the behavior of $L(x)$ is similar to that of $L(x^2)$). Of course, we need a lot of subsampling distributions (in fact an infinity) to estimate an asymptotic equivalent of L for large x. However –and this makes our method more attractive than the non-parametric method suggested before– we need only *one* subsampling distribution at $b_{n,1}^* = n^{1/2}$ to obtain an asymptotic equivalent of $L(n)$. This is clearly due to the fact that in the smooth case considered here, $L(n)$ is asymptotically equivalent to $L(n^{1/2})$. Because we have imposed $|K^{-1}(t_{2,0}, P) - K^{-1}(t_{1,0}, P)| = 1$ at some points $t_{2,0}$ and $t_{1,0}$, the convergence rate is identifiable: this explains why we need to take into account $\widehat{C}_0$ in the definition of $\widehat{\widehat{\tau}}_n$ although constants do not play a real role in our analysis. Of course, the choice of $t_{2,0}$ and $t_{1,0}$ is completely arbitrary: we suggest using $t_{2,0} = 1 - t_{1,0} = 0.6915$ so as to have some elements of comparison with the reduced normal distribution.

Finally, note that estimation of the centering factor μ_n may be treated in the same fashion under similar assumptions on its shape.

4 The extreme value statistic revisited

4.1 The Type I and Type III Cases

To illustrate the results of sections 3.2 and 3.3 we again consider the extreme value statistic of section 3.1 $T_n = \max(X_1, \ldots, X_n)$. If we know that F belongs to $D(\Lambda)$ then (16) may be used to obtain a parametric or nonparametric estimator of τ_n. Since in that case the asymptotic distribution is entirely known and equal to $K = \Lambda$ there is no reason to standardize, and an estimator of the rate of convergence is given under the hypotheses of Theorem 7 by $\widehat{\widehat{\tau}}_n |\Lambda^{-1}(t_{2,0}) - \Lambda^{-1}(t_{1,0})|$. Other estimators may be proposed in that case since one may also take advantage of the special form of τ_n and μ_n as a function of H^{-1}. In the domain of attraction of type III it is common that τ_n and μ_n are respectively of the form $\log n^{-a}$ and $\log n^b$ with $b > a$. Standard linear regressions may then be used to obtain convergent estimators of a and b.

If one does not know anything about the tail of the distribution, then non-parametric estimators or our Theorem 7 will first give convergent estimators of the rate. Notice that we can get a lot of information directly from the estimated rate. If the estimated rate is increasing with n then this means that $F \in D(\Psi_\gamma)$. In that case we may indifferently use T_n as a recentering factor; alternatively, we may simply use the median of the subsampling distribution. If the estimated rate appears to be decreasing, then the median at several b_n gives the shape of the recentering factor. In any case, it is possible to construct an approximation of the sampling distribution of the maximum *without any information on the tail of the distribution*. Since type II is of particular importance in financial applications we will review that case in the next section.

4.2 The Type II Case

Assume a regularity condition on the tail of F, namely that

$$x^\alpha (1 - F(x)) \to \Gamma \tag{21}$$

as $x \to \infty$, for some $\alpha > 0$ that measures the 'thickness' of the tail of F; here $\Gamma > 0$ is a scaling constant. Under the strong mixing condition and (21) it is well-known that,

if $K_n(x, P)$ has a large-sample limit, then it must be of a specific form (the type II extreme distribution of our Section 3.1); see Leadbetter *et al.* (1983) where sufficient conditions for the existence of a limit for $K_n(x, P)$ are also found by means of the notion of 'distributional mixing'. Furthermore, the divergence rate τ_n is also specified in this case: τ_n is proportional to $n^{-1/\alpha}$.

To fix ideas, we now set $\tau_n = n^{-1/\alpha}$; it now follows that under some mixing conditions and (21) we have [3]

$$K_n(x, P) \equiv P\{n^{-1/\alpha} T_n \le x\} \to K(x, P) = \begin{cases} \exp(-x^{-\alpha}/\Gamma) & \text{if } x > 0 \\ 0 & x \le 0 \end{cases} \qquad (22)$$

as $n \to \infty$.

Therefore, in this specific case, not only do we have existence of a limit distribution $K(x, P)$, but we know the form of $K(x, P)$ as well (up to the two unknown parameters Γ and α). Since $\alpha = 1/\gamma$ (where γ is as in Theorem 6), an estimator of α could be obtained via our Lemma 2 or Theorem 2; these however do *not* take our explicit knowledge of $K(x, P)$ into account.

Nevertheless, we can modify the subsampling methodology of Section 2 to take knowledge of $K(x, P)$ into account. Note that since $K(x, P)$ is continuous, (22) and Polya's theorem imply that

$$\sup_{x > 0} |P\{n^{-1/\alpha} T_n \le x\} - K(x, P)| \to 0,$$

which in turn implies that

$$\sup_{y > 0} |P\{T_n \le y\} - \exp(-\Gamma^{-1} n y^{-\alpha})| \to 0.$$

As in Section 2, we may approximate $P\{T_n \le y\}$ by $\bar{K}_{b_n}(y \mid X^n)$, which suggests that

$$\log \left(-\log \left| \bar{K}_{b_n}(y \mid X^n) \right| \right) = -\log \Gamma + \log b_n - \alpha \log y + o_P(1);$$

treating the above equation as a regression of $\log \left(-\log \left| \bar{K}_{b_n}(y \mid X^n) \right| \right)$ on $\log y$ we can use different values of y to get least squares estimates of both Γ and α.

Alternatively, we can look at quantiles again as in Section 2. Note that due to (22) we have an explicit expression for $K^{-1}(t, P)$ as well, namely

$$K^{-1}(t, P) = (-\Gamma \log t)^{-1/\alpha}.$$

The above equation affords us the opportunity to use (6) to simultaneously estimate Γ and α by least squares; alternatively, we could use an estimate $\hat{\alpha}$ of α (say one obtained by Lemma 2 or Theorem 2) to obtain an estimate of Γ. To elaborate on the latter, note that by Lemma 3 we have

$$b_n^{-1/\hat{\alpha}} \bar{K}_{b_n}^{-1}(t \mid X^n) = (-\Gamma \log t)^{-1/\hat{\alpha}} + o_P(1)$$

---

[3] Note that there is no real discrepancy between eq. (22) and the type II extreme distribution of our Section 3.1; the only difference is that now we are directly interested in the scaling constant Γ as well, and thus we cannot allow the presence of an arbitrary 'floating' proportionality constant in the rate τ_n. In other words, whereas in eq. (22) (and throughout Section 4) we explicitly take $\tau_n = n^{-1/\alpha}$, the rate τ_n that figures in the type II extreme distribution of our Section 3.1 is proportional (but not equal) to $n^{-1/\alpha}$; as a matter of fact, under eq. (21), the rate τ_n that figures in the type II extreme distribution of Section 3.1 satisfies $\tau_n = (\Gamma n)^{-1/\alpha}$ instead.

which can be solved for Γ. To improve upon this estimate, we can look at different t-points and different b_n's and take an average of the corresponding Γ-solutions.

Another possibility for the case of interest here, i.e. where $T_n = \max\{X_1, ..., X_n\}$, is to use the simple and popular Hill's estimator of α instead of our proposals; see de Haan (1994), de Haan *et al.* (1994), and Jansen *et al.* (1998) for a description of Hill's estimator. We note a growing disenchantment in recent literature regarding Hill's estimator; for example, Kearns and Pagan (1997) note that Hill's estimator can be misleading if there is some dependence in the data, whereas Resnick (1997) points out that Hill's estimator can exhibit some occasional strange behavior, and that it certainly cannot account for the possible presence of a slowly varying function in the rate. Extensions of the Hill estimator are given in Beirlant *et al.* (1996a,b), Hols and de Vries (1991), and Huisman *et al.* (2001); in particular, Beirlant *et al.* (1996a,b) consider a regression on the log-transformed quantile plot which gives a generalized version of Hill's estimator.

Recall that to implement Hill's estimator the number k of extreme order statistics that are to be taken into account must be chosen by the practitioner. In this case, k is like a 'smoothing parameter', and its proper choice is crucial for good results; see e.g. Hall (1990) for a discussion. Similarly, to implement our estimators of α using Lemma 2 or Theorem 2, a 'smoothing parameter' –namely the subsample sizes used– must also be chosen. Although the procedures were shown to be valid for $b_{i,n}$ proportional to n^{β_i}, the exact optimal choice of the β_i's (and of the proportionality constants) is an open problem at the moment. Note that a number of papers have dealt with a subsampling and/or bootstrap cross-validation approach having the objective of optimally picking the number k in Hill's estimator; see Danielsson and de Vries (1997), Danielsson *et al.* (2001), de Haan *et al.* (1999), Drees and Kaufman (1998). It is possible to use a similar approach for a data-dependent way of pinpointing the optimal subsample sizes in our method; in fact, this idea has been successfully tried out in Hall (1990) and Hall *et al.* (1995) albeit in more standard settings.

5 Econometric applications

5.1 Data

For our empirical applications we study time series of daily asset returns. For the Value at Risk example, we examine daily returns of the S&P 500 Index, $X_t = \ln(P_t/P_{t-1})$ where P_t is the daily closing S&P 500 index. Our observations start January 3, 1985 and end December 31, 1996, a total of $n = 3129$ daily observations. For our safety first portfolio allocation example, we consider portfolios formed from the S&P 500 (an index of "large cap" stocks), the Russel 2000 (an index of "small cap" stocks) and the 30-day Treasury Bill ("TBill"). Our observations for the Russel 2000 and the TBill cover the same period as for the S&P 500. Summary statistics for the daily returns of our three assets are reported in Table 1.

5.2 Estimating Value-at-Risk

In computing our various VaR estimates we simulate the manner in which these estimates would be deployed in a "live" application. Specifically, we start with a sample of $n_1 = 2348$ observations (1/3/1994) and compute the 1-Day, 10-Day, or 23-Day VaR estimate. The next day we have a sample of $n_2 = 2349$ observations and we repeat the

exercise, continuing in this way until we accumulate an "evaluation" sample of 3 years or 781 observations on estimated VaR. We evaluate the performance of the various estimation methods by computing the observed exceedance rate for the evaluation period. An unbiased method should have exceedance equal to the VaR level. For example, the 5% VaR estimates should have a 5% exceedance for unbiasedness.

We assume $\mu = 0$ and use the subsampling methodology to estimate 1% and 5% VaR for 1, 10, and 23 trading days. The subsampling estimates of the n-day VaR were arrived at by using Lemma 1 and the assumption that the divergence rate τ is proportional to the block length:

$$\tau_n \bar{K}_n^{-1}(p, X^i) = K^{-1}(p, P) + o_p(1)$$
$$\tau_\kappa \bar{K}_\kappa^{-1}(p, X^i) = K^{-1}(p, P) + o_p(1).$$

Dividing the two equations yields:

$$\bar{K}_n^{-1}(p, X^i) = \frac{\tau_\kappa}{\tau_n} \bar{K}_\kappa^{-1}(p, X^i) + o_p(1)$$
$$= \left(\frac{\kappa}{n}\right)^{-\gamma} \bar{K}_\kappa^{-1}(p, X^i) + o_p(1). \tag{23}$$

The choice of the largest block length is based on the following procedure: At every point t in time, split the available data into two subsets. For an initial value of the largest block length estimate the parameter γ based on data up to period $t - 200$ and compute predicted and actual exceedances for the interval $(t - 200, t)$. Then adapt the largest block length if the actual and desired exceedance do not match: reestimate and reevaluate until a choice for the largest block length is found that minimizes the deviation from the actual exceedance over the hold-out period of the past 200 days. Use this optimal block length to determine the VaR for the next consecutive five days. Then again reoptimize over the past 200 observations. The optimal block lengths for the first day of our out-of-sample evaluation, January 3, 1994, are reported in Table 2. Figures 1 and 2 report the optimal block lengths over the complete out-of-sample period and Figures 3 and 4 show the corresponding hold-out sample exceedances. For strictly stationary data it should not be necessary to reoptimize the block size; however, we reoptimize the block size every five trading days in order to take possible nonstationarities in the data into account. This reoptimization leads to time-varying estimates of γ, the evolution of which is shown in Figures 5 and 6. Results are reported in Table 3. The benchmark is computed by simply taking the 1st or 5th percentile of the distribution of historical returns which is also a valid subsampling estimator. Figures $7 - 9$ show the 1 day, 10 day and 23 day out-of-sample Value-at-Risk estimates, respectively.

5.3   Safety first portfolio selection

The safety first method of portfolio selection introduced by Roy (1952) provides an interesting alternative to the traditional mean-variance approach of Markowitz (1952). Recently, Jansen *et al.* (1998) showed how extreme value theory can be used in a multi-period setting to make operational the version of safety first portfolio selection formulated by Arzac and Bawa (1977). Here we show how our subsampling approach can be used in implementing multi-period safety first portfolio selection.

We consider a portfolio of k risky assets, whose one period net random returns are given by r_j, $j = 1, \ldots, k$. When a proportion w_j of the initial portfolio value is held

in asset j, the one period random return on the portfolio is

$$r(w) = \sum_{j=1}^{k} w_j r_j,$$

where w is the $k \times 1$ vector of weights, which by convention sum to unity ($i'w = 1$, where i is the $k \times 1$ vector of 1's). Let the dth quantile of $r(w)$ be given by $q_d(w)$, so that

$$P[r(w) \leq q_d(w)] = d.$$

If the investor can borrow or lend risklessly at the fixed interest rate r over the period, the safety first separation theorem of Arzac and Bawa (1977) specifies that the investor can obtain optimal portfolio weights w^* by solving the problem

$$\max_w [E(r(w)) - r]/[r - q_d(w)].$$

Once optimal weights w^* are given, the amount to be invested is determined as $W + b^*$, where W is the investor's initial wealth and b^* is optimal borrowing (lending if $b^* < 0$), determined as

$$\frac{b^*}{W} = [s - q_d(w^*)]/[q_d(w^*) - r].$$

The parameters s and d summarize the investor's preferences in that, as Arzac and Bawa (1977) prove, the solution just described ensures that the investor maximizes expected return subject to the constraint that wealth at the end of the period (= $(1 + r(w^*))W + b^*(r(w^*) - r)$) falls below a critical level $(1 + s)W$ (typically $s < 0$) with probability d.

Jansen *et al.* (1998) note that investors typically invest for prolonged periods of time, not just for a single period. Nevertheless, a single period of unusually poor performance can be disastrous, especially for fund managers whose clientele tend to be quite mobile. Consequently, Jansen *et al.* (1998) replace the single period return quantile $q_d(w)$ in the safety first optimization above with the quantile for the worst ever single period return $\tilde{q}_d(w)$. They show how extreme value theory permits a useful approximation to these quantiles, making use of Hill's (1975) estimator of the tail index α.

Alternatively, we can use our subsampling estimator (based on Theorem 2) to approximate the desired quantiles. We apply this decision rule to determine portfolios composed of the S&P 500, the Russel 2000, and leverage using TBills. Before we discuss the results we briefly discuss estimation - related issues.

5.3.1 Hill's estimator for the tail index

Let $\{\xi_i\}_{i=1}^n$ be a sequence of random numbers obeying some probability law F which satisfies the appropriate regularity conditions. Let $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ denote the associated order statistics. Furthermore, define γ to be $1/\alpha$. The currently most prominent estimator for the tail index was proposed by Hill (1975) and was derived in a maximum likelihood framework. The estimator is given by

$$\hat{\gamma} = k^{-1} \sum_{i=1}^{k} \left(\ln X_{(n+1-i)} - \ln X_{(n-k)} \right). \tag{24}$$

Asymptotic normality of the Hill estimator has been shown by e.g. Goldie and Smith (1987) who proved that

$$k^{1/2} \left(\hat{\gamma} - \gamma \right) \overset{d}{\to} N \left(0, \gamma^2 \right) \tag{25}$$

if k increases suitably rapidly as $n \to \infty$.

5.3.2 A Resampled Moment Ratio Estimator

Dekkers *et al.* (1989) and Danielsson *et al.* (1996) suggest estimators that are not only based on first moments like the Hill estimator but on higher moments. The estimator suggested by Dekkers *et al.* (1989) is given by

$$\hat{\gamma} \quad = \quad m^{(1)} + 1 - \frac{1}{2} \left[1 - \frac{\left(m^{(1)} \right)^2}{m^{(2)}} \right]^{-1} \tag{26}$$

$$m^{(j)} \quad = \quad k^{-1} \sum_{i+1} k \left(\ln X_{(n-i)} - \ln X_{(n-k)} \right)^j . \tag{27}$$

Dekkers *et al.* (1989) show that their estimator is strongly consistent and asymptotically normal.

The moment ratio estimator suggested by Danielsson *et al.* (1996) is given by

$$\hat{\gamma}^{(j)} \quad = \quad \frac{1}{1+j} \frac{m^{(j+1)}}{m^{(j)}} \tag{28}$$

The $m^{(j)}$ are as defined in (27), so that for $j = 0$ we get the Hill estimator. In a Monte Carlo study Danielsson *et al.* (1996) find that their estimator reduces the potential bias of Hill's estimator.

We implement the double bootstrap strategy suggested by Danielsson *et al.* (2001) and Draisma *et al.* (1999) for a moment ratio estimator given by (28) above, with $j = 1$.

5.3.3 Nonparametric Subsampling

When we perform the nonparametric subsampling approach we assume a divergence rate $\tau_n = n^{-\gamma}$ and estimate γ using a nonparametric estimator obtained from the regressions of sections 2 and 3. The recentering factor μ_n is assumed to be zero, which is the case for distributions in the domain of Type II extreme value distributions. For the problem at hand we can exploit Geluk and De Haan's (1987) proposition 1.7 which states that the density with the fatter tails always dominates in a mixture of two densities. Hence we estimated γ for portfolios with different compositions and kept increasing the block length until the estimates for γ were roughly the same. Based on this method we arrive at estimates for γ of about .42 for the Russel 2000 and .45 for the S&P 500 based on block lengths from 3 to 400.

5.3.4 Out-of-Sample Results

We simulate the decision process for critical losses s of -0.1% and -1% taking $d = 0.01$. The results are reported in Table 4. Figures 12 and 13 graph the estimated tail index γ and the corresponding quantiles for the three estimators. The Hill estimator consistently yields estimates for γ above the resampled moment ratio (MMR) estimator.

Similarly, the estimates for γ based on the subsampling estimator are almost always greater than the ones based on Hill. Therefore it is not surprising that the estimated quantiles based on the subsampling estimator are almost all the time higher in absolute value – thus implying a higher degree of tail fatness – than those based on the Hill estimator. Estimated quantiles based on the MMR approach vary significantly more than those from the other methods.

Even for a conservative strategy of $s = -0.1\%$ all three strategies outperform a simple buy-and-hold strategy in Tbills. In terms of our safety-first criterion MMR strictly dominates Hill by yielding a higher return and fewer exceedances of the wealth threshold. It is interesting to note that MMR does not dominate Hill in the classic mean-variance sense, though. No verdict can be made by comparing the "subsampling" portfolio and the MMR portfolio because both return and exceedance are lower for the former. From Figure 10 we do note, however, that the subsampling portfolio behaves much more smoothly over time.

As s goes from a conservative -0.1% to an aggressive -1% all three portfolios leave the TBill far behind over the out-of-sample period. The intuition why the optimal portfolio weights do not change is the same as in the standard mean-variance framework. As in a standard CAPM framework we assume unrestricted riskless borrowing and lending at the same rate so that it becomes efficient for everybody to hold the same portfolio and then determine leverage according to the risk profile. Now the subsampled portfolio dominates the MMR portfolio both in the safety-first as well as the mean-variance sense. Upon inspection of Figure 11 it seems that the relatively poor performance of the MMR portfolio stems from several substantial one-day losses. We find that the optimal leverage of the MMR strategy by far exceeds those of the other strategies, which is consistent with the lower estimates of γ.

The results just discussed implement the MMR exactly as described in Danielsson *et al.* (1997). Our rationale for this is their statement that "for most economic purposes the usefulness of our procedure resides in the estimation of out-of-sample (P,Q) combinations, rather than in the precise value of the tail index $1/\alpha$". As an alternative we plug the MMR-estimated γ into a type II extreme value distribution and compute the quantile analogously to our procedure for the Hill approach.

Here, for $s = -.1\%$ this alternative MMR is dominated by subsampling, whereas Hill yields higher return and exceedances. For $s = -1\%$ there is no domination, as subsampling yields the lowest returns but with the lowest risk. The disappointing performance of the original MMR portfolio does not occur here because leverage is far more in line with the other methods.

6   Conclusions

In this paper we extend the results of Bertail et al. (1999) and show how subsampling can be used to approximate the sampling distributions of diverging–as opposed to converging–statistics. We provide consistent estimators both in the case of known and unknown divergence rates and illustrate how to estimate measures of tail fatness for dependent data.

Our method performs well in estimating value at risk and our application to safety first portfolio selection indicates that for time series our subsampling estimator constitutes a clear improvement over the Hill and resampled moment ratio estimator which have been devised for iid data.

A   Technical Proofs

Proof of Theorem 1 Fix an x and first note that $EK_{b_n}(x \mid X^n, \tau.) = K_{b_n}(x, P)$ because of stationarity of the X-sequence. Now by an argument similar to the proof of Theorem 3.1 in Politis and Romano (1994) we have that

$$Var\left(K_{b_n}(x \mid X^n, \tau.)\right) = O(b_n/n) + O(n^{-1} \sum_{i=1}^{n} \alpha_X(i));$$

note that both terms on the right-hand side above tend to zero under the assumed conditions – the second term being the Cesaro sum of a null sequence, and the first term due to $b_n/n \to 0$.

Therefore, $K_{b_n}(x \mid X^n, \tau.) = K_{b_n}(x, P) + o_P(1)$. Letting $b_n = b$ (a constant), part (a) is proved. Now letting $b_n \to \infty$, note that $K_{b_n}(x, P) \to K(x, P)$ by assumption (1). Since $K(x, P)$ is assumed continuous, Polya's theorem yields the uniform consistency in (3).

Finally, if $d_n \to \infty$, then note that $\sup_x |K_{d_n}(x, P) - K(x, P)| \to 0$ as well by (1). Thus (4) is a direct consequence of (3). $\square$

Proof of Lemma 1 First note that using arguments similar to those in the proof of Lemma 1 in Bertail *et al.* (1999) or Lemma 8.2.1 in Politis *et al.* (1999), it can be shown here too that

$$K_{b_n}^{-1}(t \mid X^n, \tau.) = K^{-1}(t, P) + o_P(1) \tag{29}$$

for any $t \in (0, 1)$. Also note that by definition of the quantile transform we have

$$\bar{K}_{b_n}\left(x \, \tau_{b_n}^{-1} \mid X^n\right) = K_{b_n}(x \mid X^n, \tau.) \tag{30}$$

and thus, for any $t \in (0, 1)$,

$$K_{b_n}^{-1}(t \mid X^n, \tau.) = \tau_{b_n} \, \bar{K}_{b_n}^{-1}(t \mid X^n) \tag{31}$$

as well. As a consequence of (29) and (31) we immediately have that (6) is true. $\square$

Proof of Lemma 2 The proof follows from the more general Theorem 2 below by letting $J = 1$. $\square$

Proof of Lemma 3 Note that since $\hat{\gamma} = \gamma + o_P((\log n)^{-1})$ it follows that $b_n^{-\hat{\gamma}}/b_n^{-\gamma} \to 1$ in probability, and $b_n^{-\hat{\gamma}} = b_n^{-\gamma}(1 + o_P(1))$. So the LHS of (6) equals $b_n^{-\gamma}\bar{K}_{b_n}^{-1}(t \mid X^n) = b_n^{-\hat{\gamma}}(1 + o_P(1))\bar{K}_{b_n}^{-1}(t \mid X^n) = b_n^{-\hat{\gamma}}\bar{K}_{b_n}^{-1}(t \mid X^n) + o_P(b_n^{-\hat{\gamma}}\bar{K}_{b_n}^{-1}(t \mid X^n)) = b_n^{-\hat{\gamma}}\bar{K}_{b_n}^{-1}(t \mid X^n) + o_P(1)$, since $b_n^{\gamma}\bar{K}_{b_n}^{-1}(t \mid X^n) = O_P(1)$ by (6). $\square$

22

Proof of Theorem 2 Note that since $b_{i,n} = n^{\beta_i}$, $1 > \beta_1 > \cdots > \beta_I > 0$, it follows that $b_{i,n} \to \infty$ and $b_{i,n}/n \to 0$ for all i. Therefore (3) is true, and Lemma 1 validates the ANOVA equation (9), and in particular the fact that $u_{i,j} = o_P(1)$. Thus (10) yields

$$\gamma_{I,J} = \gamma + \frac{\sum_{i=1}^I (u_{i,\cdot} - \overline{u})(\log(b_{i,n}) - \overline{\log})}{\sum_{i=1}^I (\log(b_{i,n}) - \overline{\log})^2},$$

where $\overline{u} = (IJ)^{-1} \sum_{i=1}^I \sum_{j=1}^J u_{i,j}$, and $u_{i,\cdot} = J^{-1} \sum_{j=1}^J u_{i,j}$. As in Theorem 8.2.1 in Politis *et al.* (1999), note that $\overline{\log} = A \log n$, and $\sum_{i=1}^I (\log(b_{i,n}) - \overline{\log})^2 = B(\log n)^2$, for some constants A, B; thus, for fixed I and J, we have $\gamma_{I,J} = \gamma + o_P((\log n)^{-1})$. $\square$

Proof of Theorem 3 Fix an x and note that

$$K_{b_n}(x \mid X^n, \widehat{\tau}_{\cdot}) \equiv q^{-1} \sum_{i=1}^q 1\{b_n^{-\hat{\gamma}} T_{b_n,i} \le x\}.$$

As before, since $\hat{\gamma} = \gamma + o_P((\log n)^{-1})$ it follows that $b_n^{-\hat{\gamma}}/b_n^{-\gamma} \to 1$ in probability, and $b_n^{-\hat{\gamma}} = b_n^{-\gamma}(1 + o_P(1))$. Therefore, $b_n^{-\hat{\gamma}} T_{b_n,i} = b_n^{-\gamma} T_{b_n,i} + o_P(b_n^{-\gamma} T_{b_n,i}) = b_n^{-\gamma} T_{b_n,i} + o_P(1)$, since $b_n^{-\gamma} T_{b_n,i} = O_P(1)$ by (1).

So from (3) of Theorem 1 equation (11) follows. Equation (12) follows by similar arguments as in the proof of Theorem 8.3.1 in Politis *et al.* (1999). $\square$

Proof of Theorem 5 Equation (15) directly yields that $\widehat{\tau}_{b_n} = \tau_{b_n}(1 + o_P(1))$. From (14) we get that at any t_i

$$\overline{K}^{-1}_{U,b_n}(t_i \mid X^n) = \mu_{b_n} + \widehat{\tau}^{-1}_{b_n} K^{-1}(t_i, P)(1 + o_P(1)) + o_P(\tau^{-1}_{b_n}). \tag{32}$$

It follows that

$$\widehat{\mu}_{b_n,I} = I^{-1} \sum_{i=1}^I (K^{-1}_{U,b_n}(t_i \mid X^n)) - \widehat{\tau}^{-1}_{b_n} K^{-1}(t_i) = \mu_{b_n} + I^{-1} \sum_{i=1}^I K^{-1}(t_i) o_P(\tau^{-1}_{b_n}).$$

Since $K(t_i)$ is continuous at each t_i and I fixed, $I^{-1} \sum_{i=1}^I K^{-1}(t_i)$ is bounded and the result follows. $\square$

Proof of Theorem 7 The proof for $\widehat{\overline{\gamma}}_{I,J} = \gamma + o_P(\log n^{-1})$ and $\widehat{C}_0 = C_0 + o(1)$ is similar to the proof of theorem 2. It essentially remains to prove the statement on $\widehat{L}$. By construction of the $b^*_{n,i}$, we have $n^{1/2} \le b^*_{n,j} < e^j n^{1/2}$, $j = 1, ..., M_n$. Since $M_n = o(\log n)$, $b^*_{n,M_n}/n \to 0$ and all the $b^*_{n,j}$ satisfy the hypothesis of Theorem 1b and are thus admissible for the construction of the subsampling distributions. Since $\widehat{\overline{\gamma}}_{I,J} = \gamma + o_P(\log n^{-1})$ holds and since we have fixed $y(t_{1,0}, t_{2,0}, \infty) = 0$, we have directly from (17)

$$\widehat{h}_1(\log(b^*_n)) = h_1(\log b^*_n) + o_P(1).$$

It follows that for any $x \in [b_{n,i}^*, b_{n,i+1}^*]$,

$$\widehat{L}(x) = \{L(b_{n,i}^*)(1 + o_p(1)) + (L(b_{n,i+1}^*)(1 + o_p(1))$$

$$-L(b_{n,i}^*)(1 + o_p(1))) \int_0^{(x-b_{n,i}^*)/(b_{n,i+1}^* - b_{n,i}^*)} s(t)dt + o_P(1)\}.$$

But according to Bingham et al. (1987, p.15 –see the proof of their theorems 1.3.3 and 1.3.4), uniformly in $x \in [b_{n,i}^*, b_{n,i+1}^*]$,

$$L_1(x) = L(b_{n,i}^*) + (L(b_{n,i+1}^*) - L(b_{n,i}^*)) \int_0^{(x-b_{n,i}^*)/(b_{n,i+1}^* - b_{n,i}^*)} s(t)dt$$

is asymptotically equivalent to $L(x)$. Since $s(t)$ is uniformly bounded on each interval, it follows that

$$\widehat{L}(x) = (L(x) + o_P(1) + r_n)$$

with

$$r_n = (L(b_{n,i}^*)o_p(1) - L(b_{n,i+1}^*)o_p(1)).$$

But by the uniform convergence theorem (see theorem 1.2.1 of Bingham et al. (1987)), $L(b_{n,i}^*)/L(x) \to 1$, $L(b_{n,i+1}^*)/L(x) \to 1$ and $L(b_{n,i+1}^*)/L(b_{n,i}^*) \to 1$ uniformly in $x \in [b_{n,i}^*, b_{n,i+1}^*]$. We deduce from this that $r_n = L(x)o_p(1)$ and thus that

$$\widehat{L}(x) = L(x)(1 + o_P(1)),$$

yielding the result of the theorem.

The last result of the theorem follows from the fact that

$$\widehat{L}(n^{1/2}) = L(n^{1/2})(1 + o(1)) = L(n)\exp(C_0)(1 + o(1)) = L(n)\exp(C_0)(1 + o(1).$$

This combined with $\widehat{\widehat{\gamma}}_{I,J} = \gamma + o_P(\log n^{-1})$ and $\widehat{C}_0 = C_0 + o(1)$ yields the convergence of the estimated rate. $\square$

B Tables

<div align="center">

Daily Returns in Percent

	Russel2000	S&P 500	TBill
Mean	0.0514	0.0563	0.0225
Std.	0.7666	0.9499	0.0063
Min	-12.5200	-20.3350	0.0106
Max	7.7230	9.0994	0.0361
n	3129	3129	3129

</div>

Table 1: Descriptive statistics for the data.

<div align="center">

	1 Day	10 Days	23 Days
1% VaR	114	128	239
5% VaR	114	136	227

</div>

Table 2: January 3, 1994: Optimal Block Lengths for the various VaR specifications based on hold-out samples.

<div align="center">

	Benchmark		Subsampling	
	1% VaR	5% VaR	1% VaR	5% VaR
1 Day	0.38	3.59	1.98	8.17
(-QLL)	(.026)	(.086)	(.025)	(.089)
10 Days	0	0.13	1.71	8.56
	(.054)	(.152)	(.049)	(.190)
23 Days	0	0	1.98	7.25
	(.121)	(.254)	(.092)	(.372)

</div>

Table 3: Results for VaR estimation: Actual exceedance (in percent) of the estimated VaR over a period of 781 trading days, 1/3/94 to 12/31/1996. Negative Quasi Log Likelihood[4] (QLL) in parentheses (smaller values indicate a better goodness of fit).

---

[4]We compute $QLL = -n^{-1} \sum_{i=1}^{n} \left| y_i - \widehat{VaR}_i \right| \cdot \left(p 1_{[y_i \geq \widehat{VaR}_i]} + (1-p) 1_{[y_i < \widehat{VaR}_i]} \right)$. This measure weights the observed deviation from the VaR with the probability with which it is supposed to occur.

	T-Bill	S&P 500	RU 2000	s=-1%				s=-.1%			
				SubS	Hill	MR	MR alt.	SubS	Hill	MR	MR alt.
Wealth after 3 years:	1.172	1.867	1.541	2.109	2.330	1.631	2.211	1.265	1.294	1.321	1.243
avg. ann. returns:	5.245	22.312	14.979	27.228	31.372	17.100	29.172	7.893	8.666	9.398	7.275
% of times below -1.0%	0.000	5.371	5.243	10.230	14.322	14.194	10.997	–	–	–	–
% of times below -0.1%	0.000	36.061	32.225	10.230	14.322	14.194	10.997	10.358	13.811	13.555	11.381
% of times below worst ever return	0.128	1.023	1.023	1.151	1.279	1.151	1.407	1.023	0.767	1.151	1.023
Returns:											
Mean:	0.020	0.082	0.057	0.100	0.115	0.115	0.106	0.030	0.033	0.037	0.028
Standard Deviation:	0.002	0.678	0.604	0.919	1.169	3.190	0.965	0.106	0.140	0.542	0.115
Minimum:	0.014	-3.083	-2.747	-4.988	-4.979	-28.325	-4.448	-0.569	-0.671	-5.894	-0.503
Maximum:	0.023	2.728	2.596	3.975	4.582	23.973	3.441	0.406	0.559	7.354	0.431
Leverage:											
Mean:	–	–	–	0.503	0.878	2.590	0.584	-0.823	-0.779	-0.576	-0.813
Standard Deviation:	–	–	–	0.353	0.096	6.436	0.054	0.041	0.011	0.761	0.006
Minimum:	–	–	–	-0.315	0.676	-0.672	0.476	-0.919	-0.803	-0.961	-0.826
Maximum:	–	–	–	1.632	1.166	46.456	0.812	-0.698	-0.749	4.620	-0.792
Portfolio weight for RU2000:											
Mean:	–	–	–	0.557	0.431	0.376	0.467	0.557	0.431	0.376	0.467
Standard Deviation:	–	–	–	0.292	0.313	0.263	0.304	0.292	0.313	0.263	0.304
Minimum:	–	–	–	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum:	–	–	–	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 4: Results for the Safety First Portfolio Selection.

C Figures
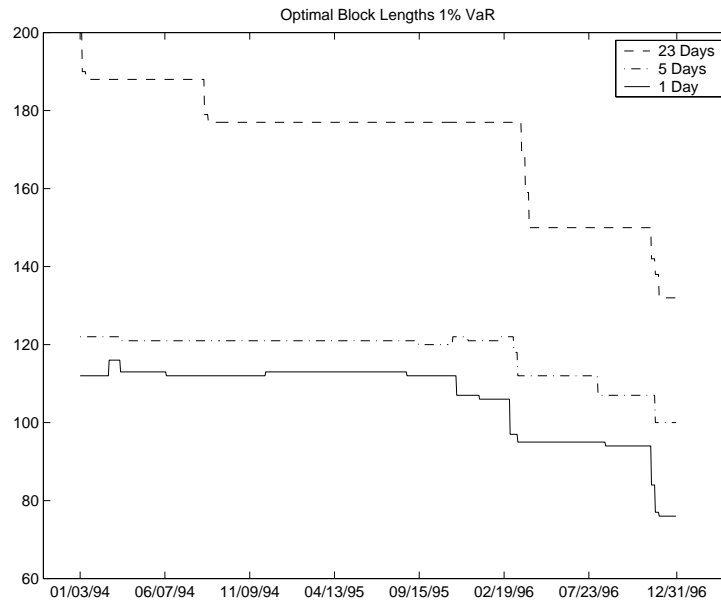
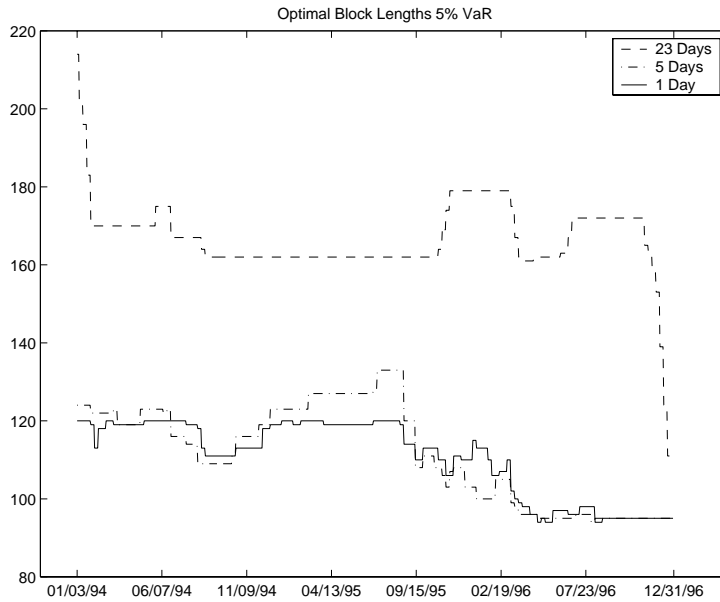Figure 1: Optimal Block Lengths for 1% Value-at-Risk

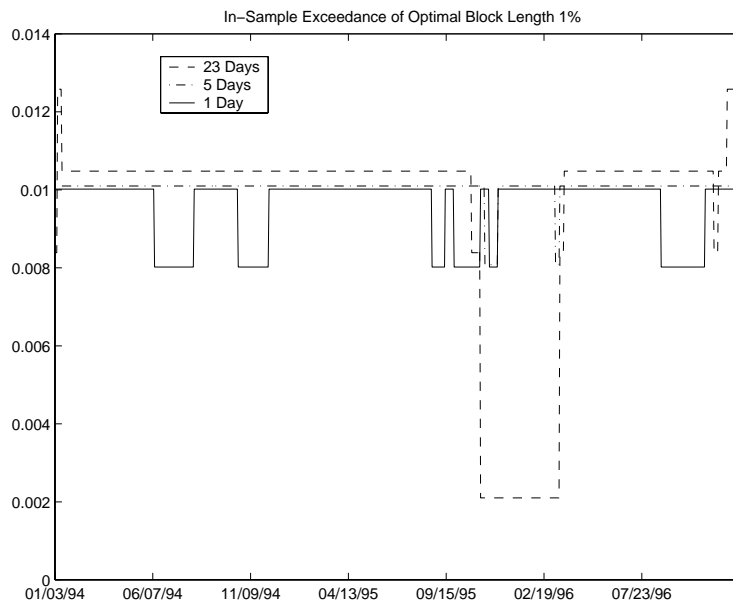Figure 2: Optimal Block Lengths for 5% Value-at-Risk

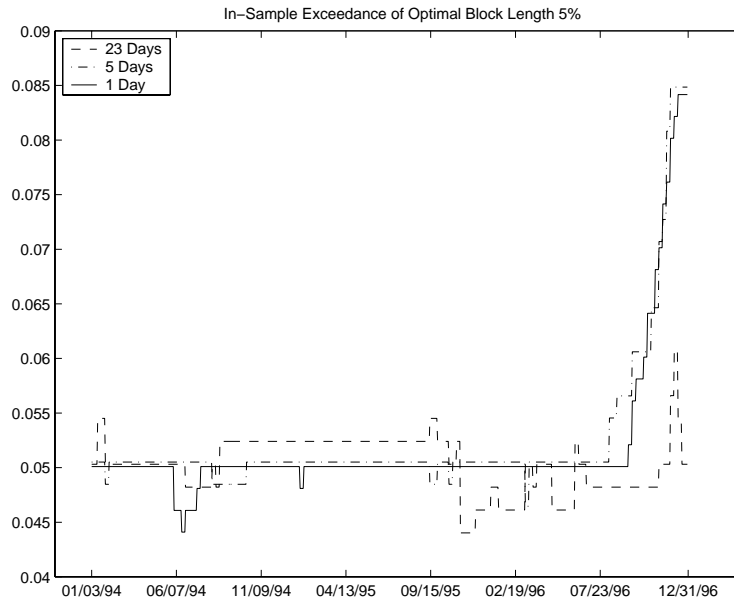Figure 3: In-Sample Exceedances for 1% Value-at-Risk

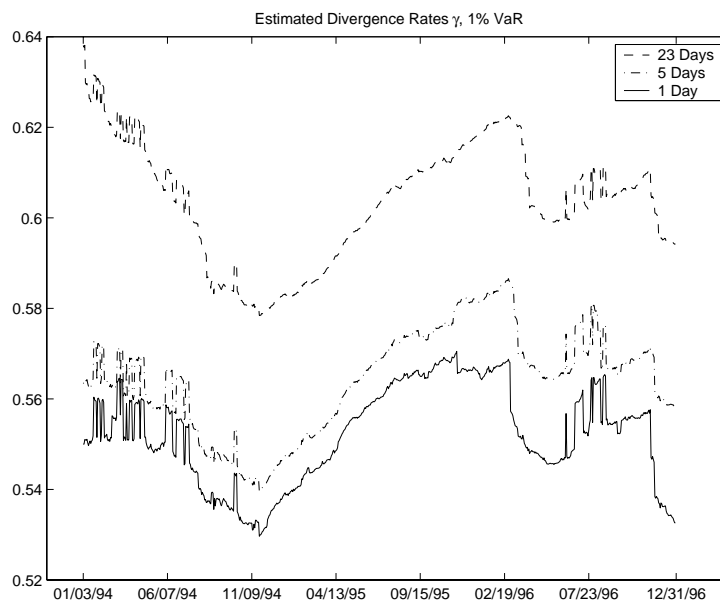Figure 4: In-Sample Exceedances for 5% Value-at-Risk

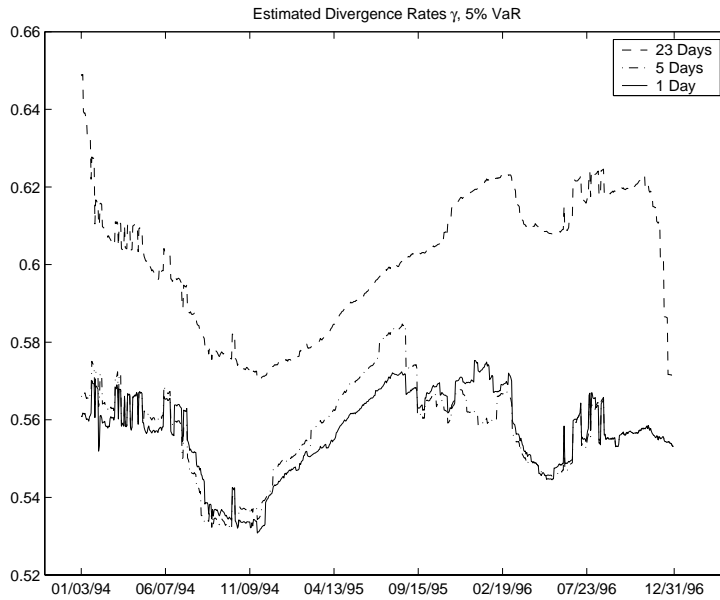Figure 5: Estimated Divergence Rates γ for 1% Value-at-Risk

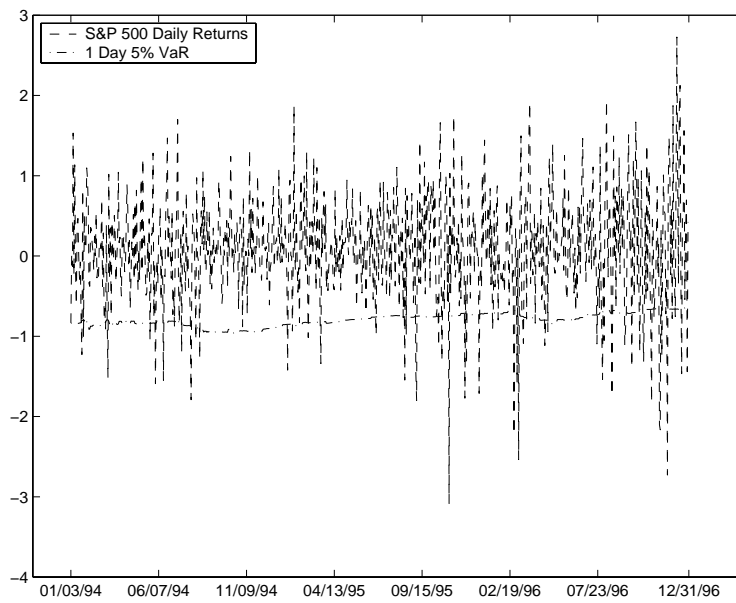Figure 6: Estimated Divergence Rates γ for 5% Value-at-Risk
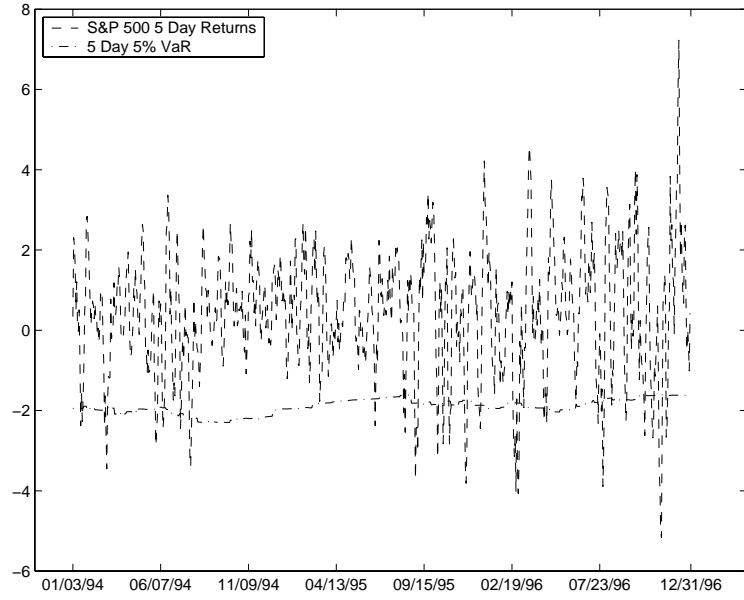

Figure 7: Estimates for 1 Day Value-at-Risk
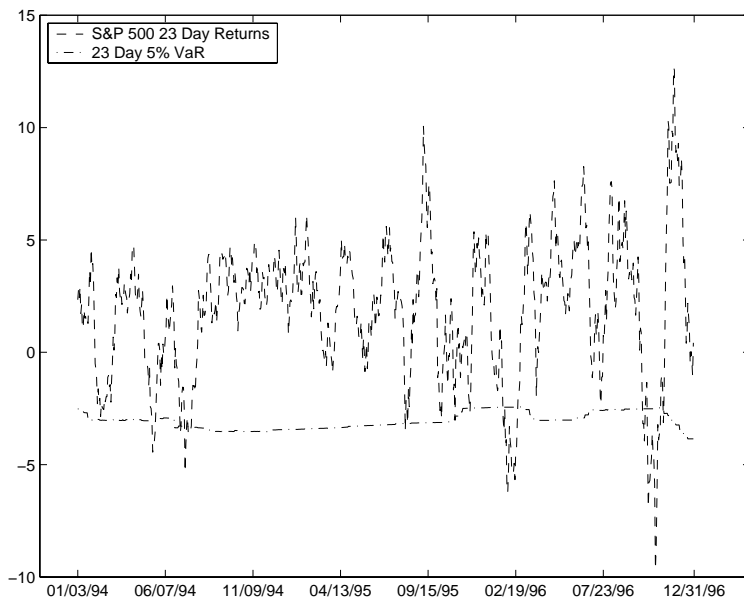
30

Figure 8: Estimates for 10 Day Value-at-Risk

Figure 9: Estimates for 23 Day Value-at-Risk

31
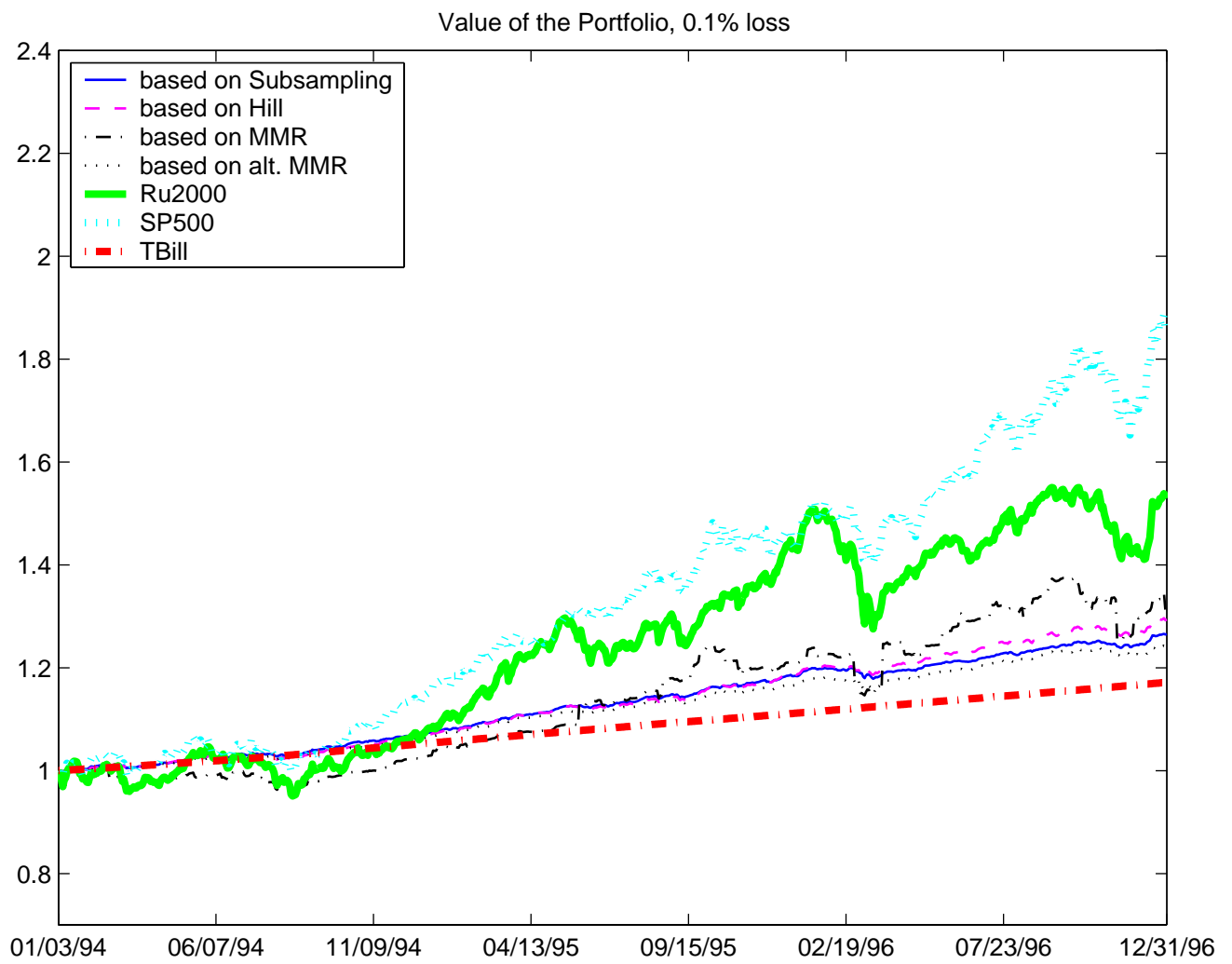
Figure 10: Portfolio Values for $s = -0.1\%$.
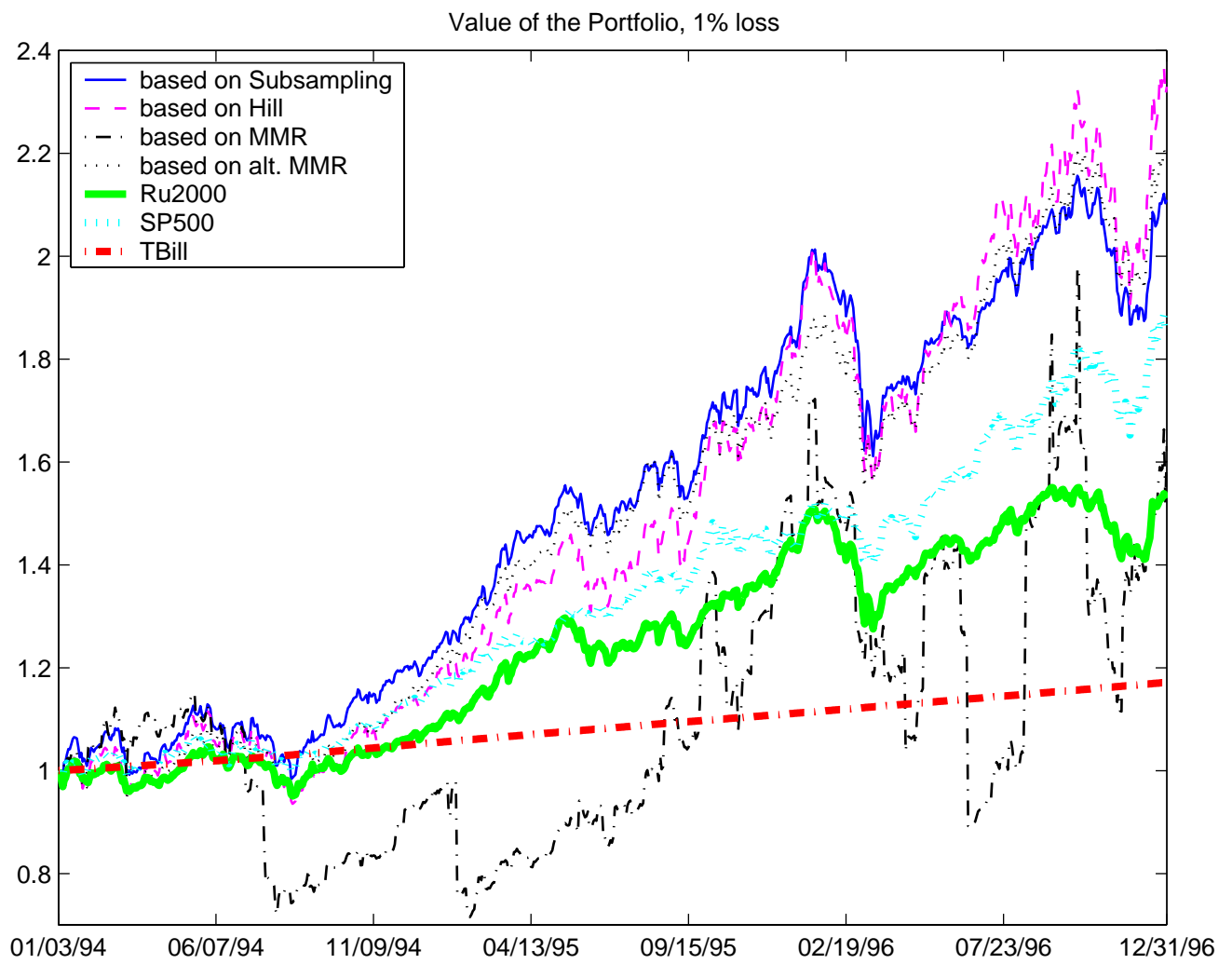
Figure 11: Portfolio Values for $s = -1\%$.

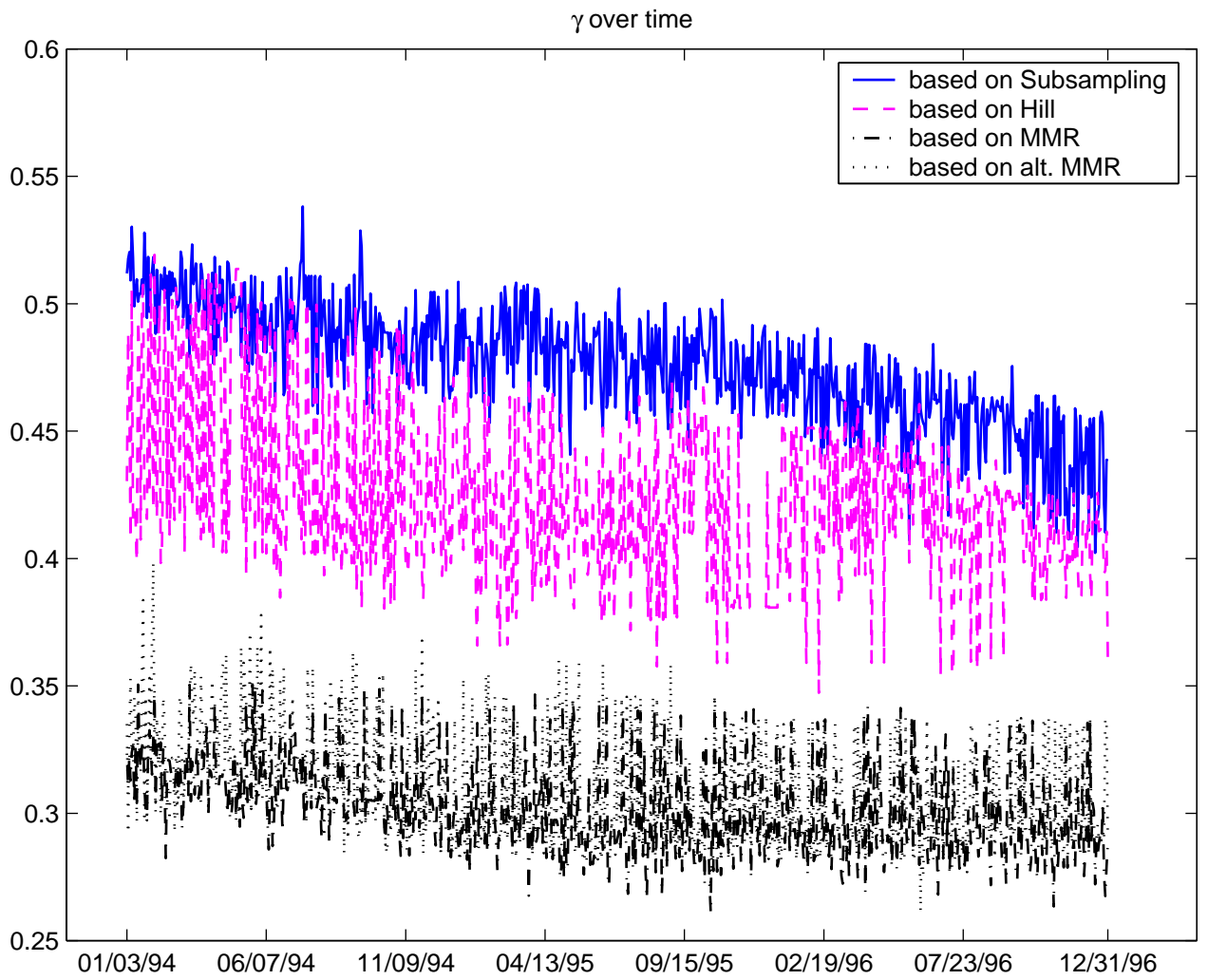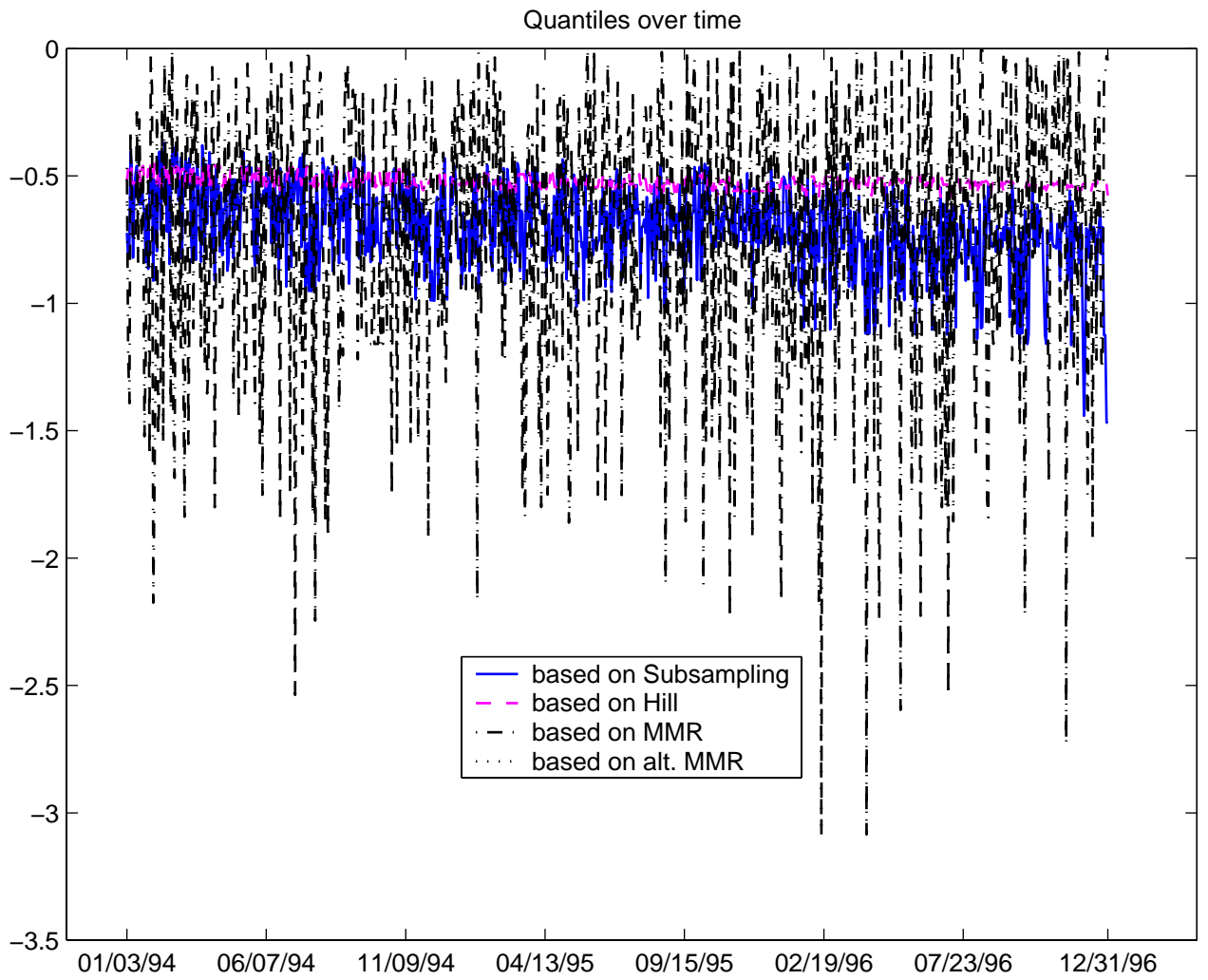Figure 12: Estimated γ.

Figure 13: Estimated Quantiles in %.

References

[1] Arzac, E.R. and Bawa, V.S. (1977), Portfolio Choice and Equilibrium in Capital Markets with Safety - First Investors, *Journal of Financial Economics*, 4, 277 − 288.

[2] Beirlant, J., Vynckier, P., and Teugels, J.L. (1996a), Tail Index Estimation, Pareto Quantile Plots, and Regression Diagnostics, *Journal of the American Statistical Association*, 91, 1659-1667.

[3] Beirlant, J., Vynckier, P., and Teugels, J.L. (1996b), Excess Functions and Estimation of the Extreme-Value Index, *Bernoulli*, 2, 293-318.

[4] Bertail P., Politis, D.N., and Romano, J.P. (1999), On Subsampling Estimators With Unknown Rate of Convergence, *Journal of the American Statistical Association*, 94, 569-579.

[5] Bingham, N.H., Goldie, C.M., and Teugels, J.L. (1987), Regular Variations, *Encyclopedia of Mathematics and its Applications*, Cambridge University Press.

[6] Black, F., and Scholes, M. (1973), The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*, 81, 637 − 659.

[7] Boudoukh, J., Richardson, M.P., and Whitelaw, R.F. (1995), Expect the Worst, *Journal of Risk*, 8, 100 − 101.

[8] Brockwell, P.J., and Davis, R.A. (1987). *Time Series: Theory and Methods*, Springer, New York.

[9] Danielsson, J., Janssen, D.W., and de Vries, C.G. (1996), The Method of Moments Ratio Estimator for the Tail Shape Parameter, *Communications in Statistics*, 25, 711-720.

[10] Danielsson, J. and de Vries, C.G. (1997), Beyond the sample: Extreme Quantile and Probability Estimation, Working Paper, Department of Economics, LSE, London.

[11] Danielsson, J., de Haan, L., Peng, L., and de Vries, C.G. (2001), Using a Bootstrap Method to Choose the Sample Fraction in Tail Estimation, *Journal of Multivariate Analysis*, 76, 226-248.

[12] de Haan, L. (1994), Extreme Value Statistics, in *Extreme Value Theory and Applications*, Galambos, J., Lechner, J., and Simiu, E. (Editors), Kluwer Academic Publ., Dordrecht, 93-122.

[13] de Haan, L., Jansen, D.W., Koedijk, K., and de Vries, C.G. (1994), Safety First Portfolio Selection, Extreme Value Theory and Long Run Asset Risks, in *Extreme Value Theory and Applications*, Galambos, J., Lechner, J., and Simiu, E. (Editors), Kluwer Academic Publ., Dordrecht, 471-488.

[14] Dekkers, A.L.M, Einmahl, H.J., and de Haan, L. (1989), A Moment Estimator for the Index of an Extreme-Value Distribution, *Annals of Statistics*, 76, 226-248.

[15] Draisma, L., de Haan, L., Peng, L., and Pereira, T. (1999), A Bootstrap-Based Method to Achieve Optimality in Estimating the Extreme-Value Index, *Extremes*, 2, 367-404.

[16] Drees, H. and Kaufmann, E. (1998), Selecting the Optimal Sample Fraction in Univariate Extreme Value Estimation, *Stochastic Processes and Applications*, 75, 149-172.

[17] Engle, R.F., Granger, C.W.J., Rice, J., and Weiss, A. (1986), Semiparametric Estimation of the Relation Between Weather and Electricity Sales, *Journal of the American Statistical Association*, 81, 310-320.

[18] Geluk, L. and de Haan, L. (1987). *Regular Variation, Extensions and Tauberian Theorems*, Amsterdam: CWI Tracts 40.

[19] Goldie, C.M. and Smith, R.L. (1987), Slow Variation with Remainder: Theory and Applications, *Quarterly Journal of Mathematics, Oxford 2nd series*, 38, 45–71.

[20] Hall, P. (1990), Using the Bootstrap to Estimate Mean Squared Error and Select Smoothing Parameter in Nonparametric Problems, *Journal of Multivariate Analysis*, 32, 177-203.

[21] Hall, P., Horowitz, J. L., and Jing, B.-Y. (1995), On Blocking Rules for the Bootstrap With Dependent Data, *Biometrika*, 82, 561-574.

[22] Hendricks, D. (1996), The Evaluation of Value at Risk Models Using Historical Data, *Federal Reserve Bank of New York Policy Review*, April, $39 - 69$.

[23] Hill, B.M., (1975), A Simple General Approach to Inference About the Tail of a Distribution, *Annals of Statistics*, 3, 1163–1179.

[24] Hols, M.C.A.B., and de Vries, C.G. (1991), The Limiting Distribution of Extremal Exchange Rate Returns, *Journal of Applied Econometrics*, 6, 287-302.

[25] Huisman, R., Koedijk, K.G., Kool, C.J., and Palm, F. (2001), Tail-Index Estimates in Small Samples, *Journal of Business and Economic Statistics*, 19, 208-216.

[26] Hull, J. and White, A. (1987), The Pricing of Options on Assets with Stochastic Volatility, *Journal of Finance*, 42, $281 - 300$.

[27] Jansen, D.W., Koedijk, K., and de Vries, C.G. (1998), Operationalizing Safety First Portfolio Selection Using Extreme Value Theory, unpublished manuscript.

[28] Jorion, P. (1997). *Value at Risk*, Dow Jones Irvin, New York.

[29] Kearns, P. and Pagan, A. (1997), Estimating the Density Tail Index for Financial Time Series, *Review of Economics and Statistics*, 49, 171-175.

[30] Leadbetter, M.R., Lindgren, G., and Rootzen, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer, New York.

[31] Markowitz, H. (1952), Portfolio Selection, *Journal of Finance*, $77 - 91$.

[32] Politis, D.N. and Romano, J.P.(1994). Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions, *Annals of Statistics*, 22, 2031–2050.

[33] Politis, D.N., Romano, J.P., and Wolf, M. (1999). *Subsampling*, Springer, New York.

[34] Resnick, S.I. (1997), Heavy Tail Modeling and Teletraffic Data, *Annals of Statistics*, 25, 1805–1869.

[35] Roy, A.D., (1952), Safety First and the Holding of Assets, *Econometrica*, 20, 431 − 449.