Kernel Methods for Small Sample and Asymptotic Tail Inference for Dependent, Heterogeneous Data

[Running Title: Kernel Methods for Tail Inference]

Jonathan B. Hill*
Dept. of Economics
Florida International University, Miami, FL

March 17, 2007

Abstract

This paper considers tail shape inference techniques robust to substantial degrees of serial dependence and heterogeneity. We detail a new kernel estimator of the asymptotic variance and the exact small sample mean-squared-error, and a simple representation of the bias of the B. Hill (1975) tail index estimator for dependent, heterogeneous data.

Under mild assumptions regarding the tail fractile sequence, memory and heterogeneity, choosing the sample fractile by non-parametrically minimizing the mean-squared-error leads to a consistent and asymptotically normal estimator.

A broad simulation study demonstrates the merits of the resulting minimum MSE estimator for autoregressive and GARCH data. We analyze the distribution of a standardized Hill-estimator in order to asses the accuracy of the kernel estimator of the asymptotic variance, and the distribution of the minimum MSE estimator. Finally, we apply the estimators to a small study of the tail shape of equity markets returns.

1. INTRODUCTION The use of extreme value theory has reached into risk management in finance, damage and catastrophe modeling in the engineering, actuarial and meteorological sciences, and the analysis of asset market contagion and hyper-inflation. See, e.g., Mandelbrot (1963), Fama (1965), McCulloch (1996), Embrechts, Klüppelberg, and Mikosch (1997), Finkenstadt

*Dept. of Economics, Florida International University, Miami, FL; www.fiu.edu/~hilljona; jonathan.hill@fiu.edu.

AMS classification: 62G32.

Keywords: Hill estimator; regular variation; extremal near epoch dependence; kernel estimator; mean-square-error.

and Rootzén (2003), Bradley and Taqqu (2003), Rachev (2003), and Beirlant, Goegebeur, Segers, Teugels, and De Waal (2004).

The general framework for analyzing extremes begins by assuming the distribution tail is regularly varying at ∞: there exists some $\alpha > 0$ such that for all t

$$\bar{F}_t(x) := P(X_t > x) = x^{-\alpha} L(x), \quad x > 0, \quad \text{where } L \text{ is slowly varying.} \tag{1}$$

Distributions satisfying (1) include the domain of attraction of the stable laws with $\alpha < 2$, coincide with the maximum domain of attraction of the extreme value distributions $\exp\{-x^{-\alpha}\}$, and characterize the tails of GARCH processes. See de Haan (1970), Leadbetter, Lindgren and Rootzén (1983), Bingham, Goldie and Teugels (1987), Resnick (1987), and Basrak, Davis, and Mikosch (2002).

Denote by $X_{(i)} > 0$ the i^{th} order statistic of the sample path $\{X_1, ..., X_n\}$: $X_{(1)} \geq X_{(2)} \geq ...$, and let $m = \{m_n\}_{n \in \mathbb{N}}$ denote a sequence of integers satisfying $m_n \to \infty$ as $n \to \infty$, and $m_n = o(n)$. B. Hill (1975) proposed the following estimator of α^{-1}:

$$\hat{\alpha}_{m_n}^{-1} := 1/m_n \sum_{i=1}^{m_n} \ln X_{(i)} / X_{(m_n+1)}.$$

The Hill-estimator has been widely used in the economics, finance, and telecommunications literatures, in particular on data for which substantial evidence suggests serial dependence and/or heterogeneity via volatility clustering. See Akgiray and Boothe (1988), Cheng and Rachev (1995), Resnick and Rootzén (2000) and Chan, Deng, Peng and Xia (2005) and Hill (2005b), to name a very few.

The question of selecting the tail fractile m_n for dependent, heterogeneous data, however, remains entirely ignored. Tail fractile selection is a long standing problem in the extreme value theory literature because tail shape and tail dependence estimates may be highly sensitive to the chosen tail region. See Figure 1 for plots of $\hat{\alpha}_{m_n}^{-1}$ and $\hat{\alpha}_{[n^\delta]}^{-1}$ for *iid*, AR(1) and IGARCH data with Stable or Pareto innovations. Positive serial dependence increases the likelihood that neighboring observations are close in value, diminishing observable tail thickness and rendering the Hill-estimator positively biased (e.g. Stable AR(1)); volatility clustering augments the detectable degree of tail thickness, leading to negatively biased estimates (e.g. Pareto IGARCH). Such plots are essentially non-formative for dependent or heterogeneous data.

INSERT FIGURE 1 HERE

Graphical "Hill-plot" methods for tail fractile selection for *iid* data are considered at length in Renick and Stărică (1997) and Drees, de Haan, and Resnick (2000). See, also, Beirlant, Vynckier and Teugels (1996) and Beirlant, Dierckx, and Guillou (2005). Bootstrap and adaptive selection methods for selecting m_n in the *iid* case are considered in Hall and Welsh (1985), Hall (1990), Drees and Kaufman (1997), and Draisma, de Haan, Peng and Periera (1997).

Mean-squared-error (MSE) and bias reduction methods for selecting m_n are developed in the seminal contributions of Hall (1982) and Hall and Welsh (1985), and recently in Beirlant,Vynckier and Teugels (1996), Resnick and Stărică (1997), Danielsson, de Haan, Peng, and de Vries (1998), Hall (1990), and Huisman Koedijk, Kool, and Palm (2001). An *iid* assumption is universally imposed in this literature, and several methods are suggested without theoretical riguer. Consider Beirlant,Vynckier and Teugels (1996) and Huisman, Koedijk, Kool, and Palm (2001).

It is a possible misconception that selecting the sequence $\{m_n\}$ by minimizing the MSE for each n leads to a inconsistent estimator $\hat{\alpha}_{m_n}$. Hall (1982) and Hall and Welsh (1985) use the tail shape

$$P(X_t > x) = Cx^{-\alpha}(1 + O(x^{-\theta})), \alpha, \theta > 0, \tag{2}$$

and focus on sequences $m_n \sim \lambda n^{2\theta/(2\theta+\alpha)}$, $\lambda > 0$. The MSE is easy to characterize for *iid* data, and is parametrically minimized with respect to λ. The resulting estimator $\hat{\alpha}_{m_n}$ is inconsistent *due simply to the chosen class of fractiles* $\lambda n^{2\theta/(2\theta+\alpha)}$. See Hall (1982: Theorem 2) and Hsing (1991: Theorems 2.4). See also Huisman *et al* (2001).

In this paper we analyze non-parametric methods for selecting the tail fractile m_n for the B. Hill (1975) estimator for processes with substantial degrees of dependence and heterogeneity restricted only to extremes. We do not impose any structure on non-extremes. For processes satisfying (1) we discuss a consistent kernel estimator of the exact MSE and asymptotic variance. We then characterize the small sample bias for tails satisfying (2), construct a bias-corrected MSE, and select $\{m_n\}$ by minimizing the MSE.

We only consider a class of sequences $\{m_n\}$ for which the Hill-estimator is known to be consistent and asymptotically normal under general conditions, cf. Hill (2005a). The class (2) is not too restricted, and includes stochastic recurrence equations, including the marginal distributions of Multivariate GARCH processes. See Basrak *et al* (2002).

We focus on the Hill-estimator because a broad asymptotic theory already exists for dependent, heterogeneous data. The minimum MSE estimator is consistent and asymptotically normal for processes $\{X_t\}$ with extremes that are Near-Epoch-Dependent on the mixing extremes of some shock process $\{\epsilon_t\}$. This covers at least nonlinear distributed lags, ARFIMA(p, d, q) and FIGARCH(p, d, q) processes with fractional difference $d \in [0, 1)$, bilinear processes, and random coefficient and threshold autoregressions. Apparently this is the most general theory available for this or any other tail estimator (e.g. Pickands 1975; Smith 1987; Drees, Ferreira, and de Haan 2004), and for this or any other fractile selection method (e.g. Hall and Welsh 1985; Drees, de Haan, and Resnick 2000).

In a large scale simulation study we analyze the merits of the kernel MSE estimator, say $\hat{\sigma}^2_{m_n}$, for sample fractile selection. We also analyze $\hat{\sigma}^2_{m_n}$ as an estimator of the asymptotic variance. We perform Cramer-von Mises tests of standard normality on standardized ratios $z_{m_n} \equiv \sqrt{m_n}(\hat{\alpha}^{-1}_{m_n} - \alpha^{-1})/\hat{\sigma}_{m_n}$ and find the range of over which $z_{m_n} \approx N(0, 1)$ at the 5%-10% levels nearly always includes the optimally selected m_n by minimizing the MSE.

Section 2 lays out extremal dependence definitions and properties of the Hill-estimator. In Sections 3 and 4 we characterize the mean-squared-error and bias of the Hill estimator for general data. Section 5 presents fractile selection methods and key asymptotic theory. Sections 6 and 7 contain a simulation study and an application to daily equity market returns. All figures and tables are placed at the end of the paper.

2. TAIL SHAPE AND TAIL MEMORY We assume X_t has for each t a common marginal distribution tail (1) with support on $[0, \infty)$, and σ-field $\Im_t = \sigma(X_\tau : \tau \leq t)$. In practice this setting applies to an absolute series $X_t = |Y_t|$, or tail preserving transforms $X_t = -Y_t \times I(Y_t < 0)$ and $X_t = Y_t \times I(Y_t > 0)$.

Assume $\bar{F}_t(x)/\bar{F}_t(x-) \to 1$ such that there exists sequences $\{b_{m_n}\}_{n\geq 1}$ and $\{m_n\}_{n\geq 1}$, $b_{m_n} \to \infty$, satisfying

$$(n/m_n)P\left(X_t > b_{m_n}\right) \to 1. \tag{3}$$

See Leadbetter *et al* (1983: Theorem 1.7.13). We must restrict the tail shape in order to expedite asymptotic normality. Cf. Goldie and Smith (1987: property SR1), Hsing (1991) and Hill (2005a).

Assumption A $\{X_t\}$ *satisfies* (1) *for some* $\alpha > 0$. *For some positive measurable* $g : \mathbb{R}_+ \to \mathbb{R}_+$,

$$L(\lambda x)/L(x) - 1 = O(g(x)) \text{ as } x \to \infty. \tag{4}$$

The function g *has bounded increase: there exists* $0 < D, z_0 < \infty$ *and* $\tau < 0$ *such that* $g(\lambda z)/g(z) \leq D\lambda^\tau$ *some for* $\lambda \geq 1$ *and* $z \geq z_0$. *Specifically,* $\{m_n\}_{n\geq 1}$, $\{b_{m_n}\}_{n\geq 1}$, *and* $g(\cdot)$ *satisfy*

$$\sqrt{m_n}\, g(b_{m_n}) \to 0, \quad \text{where } m_n \to \infty,\ m_n = o(n). \tag{5}$$

Remark 1: Tails satisfying Assumption A include $\bar{F}_t(x) = cx^{-\alpha}(1 + O((\ln x)^{-\theta}))$ and $\bar{F}_t(x) = cx^{-\alpha}(1 + O(x^{-\theta}))$. See Haeusler and Teugels (1985). Tails of the latter form have been widely exploited in the applied and theoretical statistics and econometrics literatures, and characterizes the tails of GARCH processes. See Hall (1982), Hall and Welsh (1985), Chan and Tran (1989), Caner (1998), Basrak *et al* (12002) and Hill (2005b) to name a new.

Remark 2: Property (5) restricts the rate at which $m_n \to \infty$, and is key to ensuring consistency and asymptotic normality for $\hat{\alpha}_{m_n}$ for general classes of data. See Hsing (1991) and Hill (2005a). If, for example, the tail satisfies (2) then (5) holds if

$$m_n = o(n^{2\theta/(2\theta+\alpha)}).$$

See Haeusler and Teugels (1985) for this and other examples. Property (5), therefore, does *not* include the sequence $m_n \sim \lambda n^{2\theta/(2\theta+\alpha)}$ exploited in Hall

(1982), Hall and Welsh (1985) and Huisman *et al* (2001). Such sequences render $\hat{\alpha}_{m_n}$ an inconsistent estimator of α. See Hall (1982) and Hsing (1991).

We require extremal versions of mixing and Near-Epoch-Dependence properties developed in Hill (2005a). Consult that source for complete details, and see Hall and Heyde (1980) and Gallant and White (1988) for details on conventional mixing and NED processes.

Denote by $a_n = \{a_{n,t}\}$ a sequence of constant real thresholds, $a_{n,t} \to \infty$ as $n \to \infty$, and denote by $E_{n,t} \equiv E_{n,t}(a_n)$ a measurable extremal functional of ϵ_t for $t \in \{1,...,n\}$, and $E_{n,t} = 0$ for any other $t \notin \{1,...,n\}$. Examples include the extreme event $I(|\epsilon_t| > a_{n,t})$, exceedance $(|\epsilon_t| - a_{n,t})_+$, and value $|\epsilon_t| \times I(\epsilon_t > a_{n,t})$. Define

$$\digamma_{n,s}^{t} := \sigma(E_{n,\tau}(a_n) : s \leq \tau \leq t),$$

and define the coefficients

$$\varepsilon_{n,q_n} \equiv \sup_{A_{n,t} \in \digamma_{n,t}, B_{n,t+q_n} \in \digamma_{n,t+q_n}^{+\infty} : t \in \mathbb{Z}} |P(A_{n,t} \cap B_{n,t+q_n}) - P(A_{n,t})P(B_{n,t+q_n})|$$

$$\varpi_{n,q_n} \equiv \sup_{A_{n,t} \in \digamma_{n,t}, B_{n,t+q_n} \in \digamma_{n,t+q_n}^{+\infty} : t \in \mathbb{Z}} |P(B_{n,t+q_n} | A_{n,t}) - P(B_{n,t+q_n})|.$$

where $\{q_n\}$ is a sequence of positive integers satisfying $1 \leq q_n \leq n$, $q_n \to \infty$ as $n \to \infty$ and $q_n = o(n)$.

E-Mixing *If* $q_n^{\lambda} \varepsilon_{q_n} \to 0$ *as* $n \to \infty$ *we say* $\{\epsilon_t\}$ *is Extremal-Strong Mixing with size* $\lambda > 0$*. If* $q_n^{\lambda} \varpi_{q_n} \to 0$ *as* $n \to \infty$ *we say* $\{\epsilon_t\}$ *is Extremal-uniform mixing with size* $\lambda > 0$.

L_2**-E-NED** $\{X_t\}$ *is* L_p*-Extremal-NED on some array of* σ*-fields* $\{\digamma_{n,1}^{t}\}$ *with size* $\lambda > 0$, $p > 0$, *if for any* $u \in \mathbb{R}$

$$\left\| P(X_t > b_{m_n} e^u | \Im_{t-q_n}^{t+q_n}) - P(X_t > b_{m_n} e^u | \digamma_{n,t-q_n}^{t+q_n}) \right\|_p \leq d_{n,t}(u) \psi_{q_n},$$

where $d_{n,t} : \mathbb{R} \to \mathbb{R}_+$ *is Lebesgue measurable on* $\mathbb{R}_+$, $\sup_t d_{n,t}(u) = O((m_n/n)^{1/r})$ *for each* $u \in \mathbb{R}$, *and* $(n/m_n)^{1/p - 1/r} q_n^{\lambda} \psi_{q_n} \to 0$ *for some* $r \geq p$.

Remark 1: $\{X_t\}$ is E-NED on $\{\epsilon_t\}$ if extremal information induced by extremes of ϵ_t can be used to forecast the extreme event $X_t > b_{m_n} e^u$ with an almost surely zero prediction error as $n \to \infty$. Memory restrictions are not imposed on the non-extremal support: $X_t \leq b_{m_n} e^u \to \infty$.

Remark 2: The E-NED property allows for a large degree of extremal and non-extremal memory and heterogeneity, including ARFIMA(p, d, q) and FIGARCH(p, d, q) processes with $d \in [0, 1)$, bilinear processes, nonlinear distributed lags, random coefficient autoregressions, and extremal threshold autoregressions. See Hill (2005a: Section 5). Thus, random walk and IGARCH processes have not been shown to have near epoch dependent extremes.

Assumption B *$\{X_t\}$ is L_2-E-NED with size $1/2$ on an E-Mixing process $\{\epsilon_t\}$. The base $\{\epsilon_t\}$ is E-Uniform Mixing with size $r/[2(r-1)]$ for some $r \geq 2$, or E-Strong Mixing with size $r/(r-2)$ for some $r > 2$.*

Under Assumptions A and B, Theorem 5 of Hill (2005a) delivers

$$\sqrt{m_n}\left(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1}\right)/\sigma_{m_n} \Rightarrow N(0,1) \quad \text{and} \quad \sqrt{m_n}\left(\hat{\alpha}_{m_n} - \alpha\right)/\tilde{\sigma}_{m_n} \Rightarrow N(0,1),$$

where

$$\sigma_{m_n}^2 := E(\sqrt{m_n}(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1}))^2 \quad \text{and} \quad \tilde{\sigma}_{m_n}^2 = \alpha^4 \sigma_{m_n}^2.$$

3. MEAN-SQUARED-ERROR KERNEL ESTIMATOR

Note $\sigma_{m_n}^2$ is identically the asymptotic variance and exact small sample MSE of $\hat{\alpha}_{m_n}^{-1}$. If the data are *iid* then Hall (1982) shows $\sigma_{m_n}^2 \to \alpha^{-2}$. In general when an analytic expression for $\sigma_{m_n}^2$ is not available Hill (2005a) proposes a kernel estimator

$$\hat{\sigma}_{m_n}^2 = m_n^{-1} \sum_{s=1}^{n} \sum_{t=1}^{n} w((s-t)/\gamma_n) \hat{Z}_s \hat{Z}_t$$

where $\hat{Z}_t := [\left(\ln X_t / X_{(m_n+1)}\right)_+ - (m_n/n)\hat{\alpha}_{m_n}^{-1}]$, and $w((s-t)/\gamma_n)$ denotes a standard kernel function with bandwidth γ_n, $1 \leq \gamma_n < m_n$, $\gamma_n \to \infty$ as $n \to \infty$.

Assumption C *Let $m_n/n^{1/2} \to \infty$, $\gamma_n \to \infty$* as *$n \to \infty$, $\gamma_n/m_n = o(n^{-1/2})$, and $1/m_n \sum_{s,t=1}^{n} |w((s-t)/\gamma_n)| = o(n^{1/2})$. Moreover, $w(\cdot)$ satisfies Assumption 1 of de Jong and Davidson (2000).*

Under Assumptions A-C, Theorem 6 of Hill (2005a) delivers

$$|\hat{\sigma}_{m_n}^2 - \sigma_{m_n}^2| \to 0.$$

This includes Bartlett, Parzen, Quadratic Spectral and Tukey-Hanning kernels.

The estimator $\hat{\sigma}_{m_n}^2$, alone, provides an enormous improvement in theory over existing MSE representations in which an *iid* assumption is universally imposed, leading to a potentially massive underestimate of the true MSE of $\hat{\alpha}_{m_n}^{-1}$ for highly dependent and heterogeneous data. See Section 6 for evidence.

Nevertheless, notice that $\hat{\sigma}_{m_n}^2$ incorporates the possibly biased $\hat{\alpha}_{m_n}^{-1}$ in $\{\hat{Z}_t\}$. In the next section we exploit a new characterization of the small sample bias to improve both $\hat{\alpha}_{m_n}^{-1}$ and the MSE for small samples.

4. SMALL SAMPLE BIAS

The small sample bias is exactly

$$B_{m_n} := E(\sqrt{m_n}(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1})).$$

If $X_t \overset{iid}{\sim} (2)$ such that $P(X_t > x) = cx^{-\alpha}(1 + dx^{-\theta} + o(x^{-\theta}))$, and $m_n \sim \lambda n^{2\theta/(2\theta+\alpha)}$, then Hall (1982: Theorem 2) proves $B_{m_n} \to dc^{-\theta/\alpha}\alpha\theta\lambda^{\theta/\alpha+1/2}/(\alpha + \theta)$. Under Assumptions A and B, however, the Hill-estimator is asymptotically unbiased.

THEOREM 1 *Under Assumptions A and B*

$$B_{m_n} = \sqrt{m_n} \times \alpha^{-1}[(1 + O(g(b_{m_n})))^2 - 1] + o(1) = o(1). \tag{6}$$

Inspecting (6), a simple estimate of the bias can be achieved if the non-parametric term $O(g(b_{m_n}))$ can be expressed analytically. For example, if the tail probability satisfies (2) then Haeusler and Teugels (1985) show

$$g(b_{m_n}) = b_{m_n}^{-\theta},$$

and $\sqrt{m_n}g(b_{m_n}) \to 0$ if $m_n = o(n^{2\theta/(2\theta+\alpha)}) = 0$. The m/n^{th}-quantile b_{m_n} can be easily estimated by the $\sqrt{m_n}$-consistent $X_{(m_n+1)}$, where consistency is established for processes E-NED on an E-Mixing base in Hill (2005a: Theorem 5).

Hall and Welsh (1985) argue that tails (2) typically arise (i) as powers of smooth distributions; (ii) from Type II extreme value distributions $\exp\{-x^{-\alpha}\}$; or (iii) from stable distributions. They argue that case (i) typically involves $\theta = \alpha$ or $\theta = 2\alpha$; case (ii) renders $\theta = \alpha$; and case (iii) implies $\theta = \alpha$ if $\alpha < 1$, and $\alpha/2 < \theta \leq \alpha$ if $1 < \alpha < 2$.

In most cases, therefore, $\theta = \alpha$, $\theta = 2\alpha$, or θ lies in a known range and in most cases, they argue, $\alpha \geq \theta$. Hall (1990) and Huisman *et al* (2001) therefore simply assume $\theta = \alpha$.

COROLLARY 2 *Suppose*

$$\bar{F}_t(x) = cx^{-\alpha}(1 + x^{-\alpha}), \quad \alpha, c > 0, \tag{7}$$

and $m_n = o(n^{2/3})$. *Under Assumptions A and B*

$$B_{m_n} = \sqrt{m_n}\alpha^{-1} \times .5 \times b_{m_n}^{-\alpha} + o(1).$$

Remark 1: Under Assumption B tail shape (2) require $m_n = o(n^{2\theta/(2\theta+\alpha)}) = o(n^{2/3})$ when $\theta = \alpha$.

Remark 2: For small samples and *any* process satisfying (2), or more generally (1), $\hat{B}_{m_n}$ is at best a rough approximation of the true bias. Pareto random variables, for example, have $O(x^{-\theta}) = 0$, hence $\hat{B}_{m_n}$ over-estimates the bias.

The bias estimator $\hat{B}_{m_n}$ and bias-correct Hill-estimator $\hat{\alpha}_{m_n}^{-1}(\hat{B})$ simultaneously solve

$$\begin{aligned}\hat{B}_{m_n} &= \sqrt{m_n}\hat{\alpha}_{m_n}^{-1}(\hat{B}) \times .5 \times X_{(m_n+1)}^{-\hat{\alpha}_{m_n}(\hat{B})} \\ &= \sqrt{m_n}\hat{\alpha}_{m_n}^{-1} \times .5 \times X_{(m_n+1)}^{-1/(\hat{\alpha}_{m_n}^{-1} - \hat{B}_{m_n}/\sqrt{m_n})} \\ &\quad - \hat{B}_{m_n} \times .5 \times X_{(m_n+1)}^{-1/(\hat{\alpha}_{m_n}^{-1} - \hat{B}_{m_n}/\sqrt{m_n})}\end{aligned} \tag{8}$$

$$\hat{\alpha}_{m_n}^{-1}(\hat{B}) = \hat{\alpha}_{m_n}^{-1} - \hat{B}_{m_n}/\sqrt{m_n}$$

The solution $\{\hat{\alpha}^{-1}_{m_n}(\hat{B}), \hat{B}_{m_n}\}$ can be computed numerically for each m_n. However, for large samples the mean-value-theorem implies the following result.

LEMMA 3 *Under the conditions of Corollary 2 and* $m_n \sim n^{\delta}$, $0 < \delta < 2/3$, *if* $\hat{B}_{m_n} = o_p(\sqrt{m_n})$ *then* $\hat{B}_{m_n} \to 0$ *and*

$$\left| \hat{B}_{m_n} - \frac{\sqrt{m_n}\hat{\alpha}^{-1}_{m_n} \times X^{-\hat{\alpha}_{m_n}}_{(m_n+1)}}{2 + \hat{\alpha}_{m_n} \times X^{-\hat{\alpha}_{m_n}}_{(m_n+1)} \times \left(\ln X_{(m+1)}\right) + X^{-\hat{\alpha}_{m\,n}}_{(m_n+1)}} \right| \to 0. \qquad (9)$$

The argument used to prove Lemma 3 shows if $\hat{B}_{m_n}/\sqrt{m_n}$ degenerates to zero, then it must be the case that $\hat{B}_{m_n} \to 0$. In the sequel, therefore, we simply assert the following.

ASSUMPTION D *Any solution* $\{\hat{\alpha}^{-1}_{m_n}(\hat{B}), \hat{B}_{m_n}\}$ *to (8) satisfies* $\hat{B}_{m_n} = o_p(\sqrt{m_n})$ *and* $\hat{\alpha}^{-1}_{m_n}(\hat{B}) = \hat{\alpha}^{-1}_{m_n} + o_p(1)$ *under Assumptions A-B.*

Finally, a small sample bias-corrected MSE estimator $\hat{\sigma}^2_{m_n}(\hat{B})$ uses $\hat{\alpha}^{-1}_{m_n}(\hat{B})$:

$$\hat{\sigma}^2_{m_n}(\hat{B}) := m_n^{-1} \sum\nolimits_{s=1}^{n} \sum\nolimits_{t=1}^{n} w((s-t)/\gamma_n) \times \hat{Z}_s(\hat{B}) \times \hat{Z}_t(\hat{B})$$
$$\hat{Z}_t(\hat{B}) := \left(\ln X_t / X_{(m_n+1)}\right)_+ - (m_n/n)\hat{\alpha}^{-1}_{m_n}(\hat{B}).$$

5. Fractile Selection The simplest criteria for selecting the sample fractile is the minimization of the MSE over some set of appropriate sequences $\{m_n\}$. The trick for ensuring both asymptotic normality *and* consistency is to restrict attention to the set of *proportional* sequences under Assumption A:

$$S_m = \{m_i = \{m_{i,n}\}_{n\in\mathbb{N}} : m_{i,n} \to \infty \text{ as } n \to \infty, \text{(4)-(5) hold}, m_{i,n}/m_{j,n} = 1 + o(1), \forall i,j\}.$$

The next result shows that $\{\hat{\alpha}_{m_n}, \hat{\sigma}^2_{m_n}\}$ is asymptotically equivalent to $\{\hat{\alpha}_{\tilde{m}_n}, \hat{\sigma}^2_{\tilde{m}_n}\}$ for any pair $m, \tilde{m} \in S_m$.

THEOREM 4 *Let* $m, \tilde{m} \in S_m$. *Under Assumptions A and B,* $\hat{\alpha}_{m_n} = \hat{\alpha}_{\tilde{m}_n} + o_p(1/\sqrt{\tilde{m}_n})$. *If additionally Assumption C holds, then* $\hat{\sigma}^2_{m_n} = \hat{\sigma}^2_{\tilde{m}_n} + o_p(1)$.

Now select that fractile m_n^* that minimizes the mean-squared-error $\hat{\sigma}^2_{m_n}$ or the bias-corrected the mean-squared-error $\hat{\sigma}^2_{m_n}(\hat{B})$ for each n over some set of feasible integers $M_n \subseteq \{1, ..., n\}$. Each integer m_n in M_n must belong to a sequence $\{m_n\}$ in S_m.

Under (2) consider

$$S_m = \{m_i = \{m_{i,n}\}, m_{i,n} \sim n^{\delta},\ 0 < \delta < 2\theta/(2\theta + \alpha\}.$$

For each n a candidate set of tail fractile integers $m_{i,n}$ is

$$\begin{aligned} M_n &= \{m_{i,n}\}_{i=1}^{n_M} \\ &= \{[n^{\delta}] - e_1[n^{\delta-\iota}] + i\}_{i=1}^{n_M}, \text{ where } n_M = e_2[n^{\delta-\iota}] + 1, \end{aligned}$$

and each $e_i \in \mathbb{N}$ does not depend on n above a fixed threshold, set to ensure $1 \leq m_{i,n} \leq n$ for the minimal n encountered.

For example, if $n = 5000$ and a rolling-window tail analysis involves windows of minimum width $n = 1000$, the subset

$$M_n = \{[n^{.66}] - 1 \times [n^{.65}] + i\}_{i=1}^{n_M}, \quad n_M = 5 \times [n^{.65}] + 1$$

is valid, where $M_{1000} = \{6, ..., 547\}$ and $M_{5000} = \{22, ..., 1544\}$.

The optimal fractile m_n^* can be written as

$$m_n^* = [n^{\delta}] - [e_1 n^{\delta-\iota}] + i_n^*, \quad i_n^* \in \{1, ..., n_M\}, \quad i_n^* = o(n).$$

The resulting sequence $\{m_n^*\}_{n\geq 1}$ is itself an element of S_m. By construction each

$$m_n^* = [n^{\delta}] - [e_1 n^{\delta-\iota}] + i_n^* \in \{\ [n^{\delta}] - [e_1 n^{\delta-\iota}], ..., [n^{\delta}] + (e_2 - e_1)[n^{\delta-\iota}]\}\ ,$$

hence for any $m_n \in S_m$

$$1 \longleftarrow \frac{[n^{\delta}] - [e_1 n^{\delta-\iota}] + i_n^*}{[n^{\delta}] + (e_2 - e_1)[n^{\delta-\iota}]} \leq \frac{m_n^*}{m_n} \leq \frac{[n^{\delta}] - [e_1 n^{\delta-\iota}] + i_n^*}{[n^{\delta}] - [e_1 n^{\delta-\iota}]} \rightarrow 1.$$

THEOREM 5 *Under Assumptions A-C for any subset* $M_n \subseteq S_m$ *and*

$$m_n^* \ \in \ \left\{\arg\min_{m \in M_n} \hat{\sigma}^2_{m_n}, \ \ \arg\min_{m \in M_n} \hat{\sigma}^2_{m_n}(B)\right\}$$

we have $\sqrt{m_n^*}(\hat{\alpha}^{-1}_{m_n^*} - \alpha^{-1})/\hat{\sigma}_{m_n^*} \Rightarrow N(0,1)$ *and* $\sqrt{m_n^*}(\hat{\alpha}^{-1}_{m_n^*}(\hat{B}) - \alpha^{-1})/\hat{\sigma}_{m_n^*}(\hat{B}) \Rightarrow N(0,1)$.

Remark: Theorem 5 has far more applications than a minimum MSE. *Any criterion* that selects the fractile sequence from a set of proportional sequences that satisfies Assumption B will not affect consistency and asymptotic normality of $\hat{\alpha}_{m_n^*}$, nor consistency of $\hat{\sigma}^2_{m_n^*}$.

6. MONTE CARLO STUDY We draw random samples of *iid* mean-zero innovations $\{\epsilon_t\}$ from either a symmetric Stable distribution with unit scales, or a symmetric Paretian tail

$$\begin{aligned} f(-z) &= f(z) = \alpha z^{-\alpha-1} + 2\alpha z^{-\alpha-1}, \quad z \geq 2 \\ &= .1492704, \quad z \in [0,2]. \end{aligned}$$

By construction $\int_{-\infty}^{\infty} f(z)dx = 1$. In both cases $\alpha = 1.7$. Simulation results for other values of α are qualitatively similar. We choose $\alpha = 1.7$ because it is near

2 and we want to access the accuracy of tests of finite variance. The sample size is $n = 1000$.

6.1 Step Up

We simulate AR(1) and Power-GARCH(1,1) processes using $\{\epsilon_t\}$. In the AR(1) case

$$X_t = \phi X_{t-1} + \epsilon_t, \quad \phi \in \{0, .2, .6, .9\}.$$

In the Power-GARCH(1,1) case

$$X_t = \sigma_{t-1}\epsilon_t, \quad \sigma_t^p = \theta_0 + \theta_1|\epsilon_{t-1}|^p + \theta_2\sigma_{t-1}^p, \ p = 1.5.$$

We randomly select $\theta_i \in [.01, .5]$ from a uniform distribution, retaining only those $\theta_1 + \theta_2 < .99$.

We simulate $3n$ observations and retain the last n. We generate $10,000$ series of each process and compute $\hat{\alpha}_m$ for the absolute series $|X_t|$. All reported values are averages over all simulated series.

6.2 Estimation

We estimate $\hat{\alpha}_{m_n}$ with accompanying asymptotic 95% confidence bands using the kernel estimators $\hat{\hat{\sigma}}_{m_n}^2 = \hat{\alpha}_{m_n}^4 \hat{\sigma}_{m_n}^2$ with a Bartlett kernel and bandwidth $\gamma_n = [m^{1/5}]$. Parzen and Tukey-Hanning kernels produce qualitatively similar results. We compute the bias $\hat{\check{B}}_{m_n}$ according to formula (9), and estimate tail parameters for all $m_n \in M_n$,

$$M_n = \{10, ..., 550\}.$$

See Figure 1, and see Tables 1 and 2 for confidence band for selected $m_n \in M_n$. All reported values are averages over all repetitions.

6.3 Tail Index Estimates

We now return to the issues raised in the introduction. As the degree of positive serial dependence increases observations cluster with greater probability hence the observable tail thickness is large. More observations from the tail are required to obtain a sharp estimate of α for persistent data. Consider Tables 1.1 and 2.1. For Stable or Paretian AR(1) processes with $\phi = .9$ we require at least $m_n = 420$ tail observations in order not to incur a Type II error for one-sided tests of $\alpha \geq 2$ at the 5%-level, and as many as $m_n = 510$ tail observations will render a 95% confidence band that contains $\alpha = 1.7$. Using $m_n = 420$ (not shown) the confidence band for Stable random variables is $1.68 \pm .24$, and using $m_n = 510$ (not shown) the band is $1.51 \pm .19$.

In general, for GARCH(1,1) data we require fewer tail observations to obtain a sharp index estimate that autoregressive data with positive serial dependence. This is fairly intuitive: positive serial dependence increases the likelihood of clustering such that more observations are required to discern the true tail shape (the Hill-estimator is positively biased for small m_n). On the other hand, volatility clustering augments tail thickness. The use of fewer tail observations in this case improves the likelihood that they will not be neighbors and hence not strongly heteroscedastically related (the Hill-estimator is negatively biased for large m_n).

6.4 Kernel Asymptotic Variance Estimation

The asymptotic variance of $\hat{\alpha}_{m_n}$ is α^2 for benchmark *iid* random variables. In this case the kernel variance estimator satisfies $\widehat{\hat{\sigma}}_{m_n} \approx \hat{\alpha}_{m_n}$ for large $m_n \geq 350$ for Stable processes, and even larger $m_n \geq 400$ for Paretian tailed processes. See column one of Tables 1 and 2. As the degree of dependence and heterogeneity increases, however, the kernel variance increases well beyond α^2. For Stable AR(1) processes with *iid* errors, $\phi = .9$, and $m_n = 400$, the kernel estimate is $\widehat{\hat{\sigma}}_{m_n} = 4.08 > 1.7$.

For exact Pareto tails, $P(X_t > x) = x^{-\alpha}$, the performance of the Hill-estimator significantly improves, in particular for *iid* data. This is well known in the literature. In simulations not shown here the estimated kernel variance is between 1.6 and 1.8 for any $m_n \geq 50$.

Of course this may simply point out the inability of $\hat{\sigma}^2_{m_n}$ to approximate the variance of $\hat{\alpha}^2_{m_n}$ for dependent data. We perform a unique simulation study to assess the merits of $\hat{\sigma}^2_{m_n}$ by testing how well the standard normal distribution approximates the true distribution of

$$z_{j,m_n} = \sqrt{m_n}\left(\hat{\alpha}^{-1}_{j,m_n} - \alpha^{-1}\right)/\hat{\sigma}_{j,m_n}.$$

In this case $\hat{\alpha}^{-1}_{j,m_n}$ represents the tail index estimate of the j^{th} series, say $\{X_{t,j}\}_{t=1}^{1000}$, and each series is generated independently of any other series. We perform Cramer-von Mises tests of standard normality on sequence $\{z_{j,m_n}\}_{j=1}^{1000}$ for each simulated process.

In Figure 2 we plot Cramer-von Mises test statistics over $m_n \in M_n$ for *iid*, AR(1) with $\phi = .9$ and GARCH(1,1). In general a large number of tail observations is required to ensure approximate standard normality at the 1% and 5% levels for dependent data, in particular for Stable random variables. A broad spectrum of fractile values can generate $z_{j,m_n} \approx N(0,1)$ at the 10%-level for GARCH(1,1) data.

6.5 Minimum MSE Estimates

We select the optimal fractile m_n^* by minimizing the MSE $\hat{\sigma}^2_{m_n}$ and the bias-corrected MSE $\hat{\sigma}^2_{m_n}(\hat{B})$. In general each criterion generates sharp estimates, in particular for serially dependent data. The optimally selected fractile is nearly always within the fractile range over which the Hill-estimator is approximately normally distributed at the 5-10% level. See the last two rows of Tables 1 and 2.

The only challenging cases involve the minimum bias-corrected MSE estimator. The estimator tends to be positively biased for *iid* and low memory AR data. Nevertheless, the resulting estimator is exceptional for strongly dependent data (AR(1) with $\phi = .9$).

7. EMPIRICAL APPLICATION We now study daily log-returns $\{Y_t\}$ to the NASDAQ and S&P500 composite indices over the period Jan. 1, 2001 to Dec. 31, 2005. Market closures are treated as missing values, and each series is filtered through a standard 5 day dummy regression to remove daily effects. After differencing and removing missing values the sample size is 1422.

See Figure 2 for plots of Y_t, and $\hat{\alpha}_{m_n}$ based on the absolute series $X_t \equiv |Y_t|$. The only distribution characteristics that matter regarding asymptotics for $\hat{\alpha}_{m_n}$ (assuming the tails are regularly varying) is dependence and heterogeneity in the extremes. In order to assess the degree of dependence in the extremes we estimate the first order tail dependence coefficient $r_{m_n}(1)$ defined as

$$r_{m_n}(1) := (n/m)\left(P(X_t > b_{m_n}, X_{t-1} > b_{m_n}) - P(X_t > b_{m_n})P(YX_{t-1} > b_{m_n})\right).$$

A nonparametric estimator is simply

$$\hat{r}_{m_n}(1) = \frac{1}{m_n}\sum\nolimits_{t=1}^{n}\left(I(X_t > X_{(m_n+1)}) - (m_n/n)\right)\left(I(X_t > X_{(m_n+1)}) - (m_n/n)\right) .$$

See Hill (2006) for a literature review, asymptotic theory and a robust kernel variance estimator associated with $\hat{r}_{m_n}(1)$ under Assumptions A-C. Specifically, under Assumptions A-C

$$\sqrt{m_n}(\hat{r}_{m_n}(1) - r_{m_n}(1))/v_{m_n} \Rightarrow N(0,1)$$

where $v^2_{m_n} := E(\sqrt{m_n}(\hat{r}_{m_n}(1) - r_{m_n}(1))^2$, and a kernel estimator, a la $\hat{\sigma}^2_{m_n}$, satisfies $\hat{v}^2_{m_n} - v^2_{m_n} \rightarrow 0$.

Both equity returns series display significant levels of positive first order extremal dependence. Unless we impose a parametric model and explicitly work out a parametric expression for the asymptotic variance of the Hill-estimator (if one exists analytically), the evidence supports the use of the kernel estimators.

Minimum MSE [MMSE] estimates for either series are close to 2. Non-bias corrected MMSE estimates $\hat{\alpha}_{m_n} \pm 1.96\hat{\hat{\sigma}}_{m_n}/\sqrt{m_n}$ are $1.71 \pm .204$ and $2.02 \pm .257$ for the NASDAQ and SP500, respectively. The "bias-corrected" MMSE estimates are identical to the uncorrected MMSE estimates (up to four decimal places) because the estimated bias for each equity market is tiny (between $-.01$ and $-.12$) relative to the magnitude of the MSE itself (between 5 and 80). See Figure 3 for plots of $\hat{\alpha}^{-1}_{m_n}$, $\hat{\alpha}^{-1}_{m_n}(\hat{B}_{m_n})$ and $\hat{B}_{m_n}$.

Appendix 1: Proofs of Main Results

Proof of Theorem 1. Define for any $u \in \mathbb{R}$

$$\{U_{m,t}\} := \left\{(\ln X_t/b_{m_n})_+ - E[(\ln X_t/b_{m_n})_+]\right\} \tag{10}$$
$$\{U^*_{m,t}(u/\sqrt{m_n})\} := \left\{I\left(X_t > b_{m_n}e^{u/\sqrt{m_n}}\right) - E\left[I\left(X_t > b_{m_n}e^{u/\sqrt{m_n}}\right)\right]\right\}.$$

From Lemma A.1 in Appendix 2, Assumptions A and D imply

$$\begin{aligned}\sqrt{m_n}\left(\hat{\alpha}_m^{-1}-\alpha^{-1}\right) &= m_n^{-1/2}\sum_{t=1}^{n}\left(U_{m,t}-\alpha^{-1}U^*_{m,t}(u/\sqrt{m_n})\right)\\ &\quad+\sqrt{m_n}\left(1/m_n\sum_{t=1}^{n}E(\ln X_t/b_{m_n})_+-\alpha^{-1}\right)+e_n,\end{aligned}$$

where $E[m_n^{-1/2}\sum_{t=1}^{n}(U_{m,t} - \alpha^{-1}U^*_{m,t}(u/\sqrt{m_n}))] = 0$ by construction, $e_n = o_p(1)$ and $E[e_n] = o(1)$. Analogous to arguments in Hsing (1991), by properties (1) and (3), dominated convergence and arguments in Smith (1982: eq. 2.2)

$$\begin{aligned}&E\left[\sqrt{m_n}\left(\hat{\alpha}_m^{-1}-\alpha^{-1}\right)\right]\\ &=\sqrt{m_n}\left(1/m_n\sum_{t=1}^{n}E(\ln X_t/b_{m_n})_+-\alpha^{-1}\right)+E[e_n]\\ &=\sqrt{m_n}\left(\frac{n}{m_n}E(\ln X_t/b_{m_n})_+-\alpha^{-1}\right)+o(1)\\ &=\sqrt{m_n}\left(\frac{n}{m_n}\int_0^\infty P\left(X_t>b_{m_n}e^u\right)du-\alpha^{-1}\right)+o(1)\\ &=\sqrt{m_n}\left(\frac{n}{m_n}P\left(X_t>b_{m_n}\right)\int_1^\infty y^{-1}\frac{P\left(X_t>b_{m_n}y\right)}{P\left(X_t>b_{m_n}\right)}dy-\alpha^{-1}\right)+o(1)\\ &=\sqrt{m_n}\left((1+O(g(b_{m_n})))\times\int_1^\infty y^{-1-\alpha}dy\,(1+O(g(b_{m_n})))-\alpha^{-1}\right)+o(1)\\ &=\sqrt{m_n}\alpha^{-1}\left((1+O(g(b_{m_n})))^2-1\right)+o(1).\end{aligned}$$

■

Proof of Corollary 2. Similar to the above argument, suppose $\bar{F}_t(x) = cx^{-\alpha}(1 + x^{-\varpi\times\alpha})$ for some $\varpi > 0$. Then

$$\begin{aligned}&E\left[\sqrt{m_n}\left(\hat{\alpha}_m^{-1}-\alpha^{-1}\right)\right]\\ &=\sqrt{m_n}\left(\frac{n}{m_n}\int_1^\infty y^{-1}P\left(X_t>b_{m_n}y\right)dy-\alpha^{-1}\right)+o(1)\\ &=\sqrt{m_n}\left(\frac{n}{m_n}\int_1^\infty y^{-1}cb_{m_n}^{-\alpha}y^{-\alpha}(1+b_{m_n}^{-\alpha\varpi}y^{-\alpha\varpi})dy-\alpha^{-1}\right)+o(1)\\ &=\sqrt{m_n}\left(c\frac{n}{m_n}b_{m_n}^{-\alpha}\left[\int_1^\infty y^{-1-\alpha}dy+b_{m_n}^{-\varpi\alpha}\int_1^\infty y^{-1-(1+\varpi)\alpha}dy\right]-\alpha^{-1}\right)+o(1)\\ &=\sqrt{m_n}\left(c\frac{n}{m_n}b_{m_n}^{-\alpha}\left[\alpha^{-1}+(1/1+\varpi))\alpha^{-1}b_{m_n}^{-\varpi\alpha}\right]-\alpha^{-1}\right)+o(1)\\ &=\sqrt{m_n}\alpha^{-1}\left((1+o(1/\sqrt{m_n}))\times\left[1+(1+\varpi)^{-1}b_{m_n}^{-\varpi\alpha}\right]-1\right)+o(1)\\ &=\sqrt{m_n}\alpha^{-1}(1+\varpi)^{-1}b_{m_n}^{-\varpi\alpha}+o(1),\end{aligned}$$

where $c = (m_n/n)b_{m_n}^{\alpha}(1 + o(1/\sqrt{m_n}))$ by Lemma A.2. Simply put $\varpi = 1$ to complete the proof. ■

Proof of Lemma 3.

Step 1 ($\hat{B}_{m_n} = o_p(\sqrt{m_n})$): Suppose $\hat{B}_{m_n} = o_p(\sqrt{m_n})$. In order to prove $\hat{B}_{m_n} = o_p(1)$ from (8) we need only show

$$\sqrt{m_n} X_{(m_n+1)}^{-1/(\hat{\alpha}_{m_n}^{-1} - \hat{B}_{m_n}/\sqrt{m_n})} = o_p(1).$$

Define

$$\hat{c}_{m_n} := (m_n/n) X_{(m+1)}^{\hat{\alpha}_{m_n}}$$

and write

$$\ln\left((n/m_n) X_{(m_n+1)}^{-1/(\hat{\alpha}_{m_n}^{-1} - \hat{B}_{m_n}/\sqrt{m_n})}\right) = -\frac{\ln \hat{c}_{m_n}}{1 - \hat{\alpha}_{m_n} \times \hat{B}_{m_n}/\sqrt{m_n}} - \frac{\ln(n/m_n)}{\sqrt{m_n}} \times \left[\frac{\hat{\alpha}_{m_n} \times \hat{B}_{m_n}}{1 - \hat{\alpha}_{m_n} \times \hat{B}_{m_n}/\sqrt{m_n}}\right]. \tag{11}$$

Under Assumptions A and B, Theorem 1 of Hill (2005b) states

$$\hat{c}_{m_n} \to c,$$

and $m_n \sim n^{\delta}$ with $\delta < 2/3$ implies

$$(\ln n/m_n)/\sqrt{m_n} = O\left(m_n^{-1/2} \ln n\right) = o(1).$$

Thus, $\hat{B}_{m_n}/\sqrt{m_n} \to 0$, $\hat{c}_{m_n} \to c$ and (11) imply

$$(n/m_n) X_{(m_n+1)}^{-1/(\hat{\alpha}_{m_n}^{-1} - \hat{B}_{m_n}/\sqrt{m_n})} \to c^{-1}.$$

Use $m_n/n^{2/3} \to 0$ to conclude

$$\sqrt{m_n} X_{(m_n+1)}^{-1/(\hat{\alpha}_{m_n}^{-1} - \hat{B}_{m_n}/\sqrt{m_n})} = O_p\left(\sqrt{m_n}(m_n/n)\right) = o_p(1).$$

Step 2: By the mean-value-theorem there exists for each m_n a random variable $B^*_{m_n} \in [0, \hat{B}_{m_n}]$ that satisfies

$$\begin{aligned}
\hat{B}_{m_n} = {} & \sqrt{m_n}\hat{\alpha}_{m_n}^{-1} \times .5 \times X_{(m_n+1)}^{-\hat{\alpha}_{m_n}} \\
& - .5 \times \sqrt{m_n}\hat{\alpha}_{m_n}^{-1} \times \left(\ln X_{(m_n+1)}\right) \times \frac{X_{(m_n+1)}^{-1/(\hat{\alpha}_{m_n}^{-1} - B^*_{m_n}/\sqrt{m_n})}}{\sqrt{m_n}(\hat{\alpha}_{m_n}^{-1} - B^*_{m_n}/\sqrt{m_n})^2} \times \hat{B}_{m_n} \\
& - .5 \times X_{(m_n+1)}^{-1/(\hat{\alpha}_{m_n}^{-1} - B^*_{m_n}/\sqrt{m_n})} \times \hat{B}_{m_n} \\
& + .5 \times \frac{B^*_{m_n}\left(\ln X_{(m_n+1)}\right) X_{(m_n+1)}^{-1/(\hat{\alpha}_{m_n}^{-1} - B^*_{m_n}/\sqrt{m_n})}}{\sqrt{m_n}(\hat{\alpha}_{m_n}^{-1} - B^*_{m_n}/\sqrt{m_n})^2} \times \hat{B}_{m_n}.
\end{aligned}$$

Let $\hat{B}_{m_n} = o_p(\sqrt{m_n})$ such that $B^*_{m_n} \to 0$ by Step 1: $B^*_{m_n} \in [0, \hat{B}_{m_n}] \to 0$. Notice $(\ln X_{(m_n+1)}) \times X_{(m_n+1)}^{-1/\hat{\alpha}_{m_n}^{-1}} \to 0$. Therefore

$$\left| \hat{B}_{m_n} - \frac{\sqrt{m_n}\hat{\alpha}_{m_n}^{-1} \times X_{(m_n+1)}^{-\hat{\alpha}_{m_n}}}{2 + \hat{\alpha}_{m_n} \times X_{(m_n+1)}^{-\hat{\alpha}_{m_n}} \times \left(\ln X_{(m+1)}\right) + X_{(m_n+1)}^{-\hat{\alpha}_{m_n}}} \right| \to 0.$$

■

Proof of Theorem 4.

Claim 1: Consider $\left\{U_{m_n,t}, U^*_{m_n,t}(u/\sqrt{m_n})\right\}$ defined in (10). Lemma A.1 implies for any $(m,\tilde{m}) \in S_m$

$$m_n^{-1/2}\sum_{t=1}^{n}\left(U_{m_n,t} - \alpha^{-1}U^*_{m_n,t}(u/\sqrt{m_n})\right) \\ - (m_n/\tilde{m}_n)^{1/2} \times \tilde{m}_n^{-1/2}\sum_{t=1}^{n}\left(U_{\tilde{m}_n,t} - \alpha^{-1}U^*_{\tilde{m}_n,t}(u/\sqrt{\tilde{m}_n})\right) = o_p(1),$$

where $(m_n/\tilde{m}_n)^{1/2} = 1 + o(1)$ by assumption, hence

$$m_n^{-1/2}\sum_{t=1}^{n}\left(U_{m_n,t} - \alpha^{-1}U^*_{m_n,t}(u/\sqrt{m_n})\right) \\ -\tilde{m}_n^{-1/2}\sum_{t=1}^{n}\left(U_{\tilde{m}_n,t} - \alpha^{-1}U^*_{\tilde{m}_n,t}(u/\sqrt{\tilde{m}_n})\right) = o_p(1).$$

Lemma A.3 implies

$$\sqrt{m_n}\left(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1}\right) - \sqrt{m_n}\left(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1}\right) \\ = m_n^{-1/2}\sum_{t=1}^{n}\left(U_{m_n,t} - \alpha^{-1}U^*_{m_n,t}(u/\sqrt{m_n})\right) \\ - \tilde{m}_n^{-1/2}\sum_{t=1}^{n}\left(U_{\tilde{m}_n,t} - \alpha^{-1}U^*_{\tilde{m}_n,t}(u/\sqrt{\tilde{m}_n})\right) = o_p(1).$$

Therefore

$$\sqrt{m}\left(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1}\right) - \sqrt{\tilde{m}_n}\left(\hat{\alpha}_{\tilde{m}_n}^{-1} - \alpha^{-1}\right) \\ = \sqrt{m_n}\left(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1}\right)\left(1 - \sqrt{\tilde{m}_n}/\sqrt{m_n}\right) + \sqrt{\tilde{m}_n}\left(\hat{\alpha}_{m_n}^{-1} - \hat{\alpha}_{\tilde{m}_n}^{-1}\right) = o_p(1).$$

By Theorem 5 of Hill (2005a) $\hat{\alpha}_{m_n}^{-1} - \alpha^{-1} = O(1/\sqrt{m_n})$, and by assumption $1 - \sqrt{\tilde{m}_n}/\sqrt{m_n} = o(1/\sqrt{m_n})$. We deduce $\sqrt{m_n}(\hat{\alpha}_{mn}^{-1} - \hat{\alpha}_{mn}^{-1}) = o_p(1)$, hence $\hat{\alpha}_{m_n}^{-1} = \hat{\alpha}_{\tilde{m}_n}^{-1} = o_p(1/\sqrt{\tilde{m}_n})$ as claimed.

Claim 2: Clearly

$$\left|\hat{\sigma}^2_{m_n} - \hat{\sigma}^2_{\tilde{m}_n}\right| \leq \left|\hat{\sigma}^2_{m_n} - \sigma^2_{m_n}\right| + \left|\hat{\sigma}^2_{\tilde{m}_n} - \sigma^2_{\tilde{m}_n}\right| + \left|\sigma^2_{m_n} - \sigma^2_{\tilde{m}_n}\right|.$$

By Theorem 6 of Hill (2005a) $\hat{\sigma}^2_{m_n} - \sigma^2_{m_n} = o(1)$ for all $m \in S_{m_n}$. Hence, it suffices to prove $\sigma^2_{mn} = \sigma^2_{\tilde{m}_n} + o_p(1)$. This follows from Claim 1: $\hat{\alpha}_{mn}^{-1} = \hat{\alpha}_{\tilde{m}_n}^{-1} + o_p(1/\sqrt{\tilde{m}_n})$ and $m_n/\tilde{m}_n = 1 + o(1)$ $\forall\{m_n,\tilde{m}_n\} \in S_n$ imply

$$\sqrt{\tilde{m}_n}\left(\hat{\alpha}_{\tilde{m}_n}^{-1} - \alpha^{-1}\right) - \sqrt{m_n}\left(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1}\right) \Rightarrow 0$$

in distribution, hence (e.g. Theorem 2.1 of Billingsley, 1999)

$$\sigma^2_{m_n} - \sigma^2_{\tilde{m}_n} = E(\sqrt{m_n}(\hat{\alpha}^{-1}_{m_n} - \alpha^{-1}))^2 - E(\sqrt{\tilde{m}_n}(\hat{\alpha}^{-1}_{\tilde{m}_n} - \alpha^{-1}))^2 \to 0.$$

■

Proof of Theorem 5. The claim follows from Theorem 4, asymptotic normality $\sqrt{m_n}(\hat{\alpha}^{-1}_{m_n} - \alpha^{-1})/\sigma_{m_n} \Rightarrow N(0,1)$ and consistency $\hat{\sigma}^2_{m_n} - \sigma^2_{m_n} \to 0$, cf. Theorems 5 and 6 of Hill (2005a). ■

Appendix 2: Supporting Lemmeta

LEMMA A.1 *Under Assumptions A and B,*

$$\begin{aligned}\sqrt{m_n}\left(\hat{\alpha}^{-1}_{m_n} - \alpha^{-1}\right) &= m_n^{-1/2}\sum_{t=1}^{n}\left(U_{m_n,t} - \alpha^{-1}U^*_{m_n,t}(u/\sqrt{m_n})\right) \\ &\quad + \sqrt{m_n}\left[E\left(m_n^{-1}\sum_{t=1}^{n}(\ln X_t/b_{m_n})_+\right) - \alpha^{-1}\right] + e_n \\ &= m_n^{-1/2}\sum_{t=1}^{n}\left(U_{m_n,t} - \alpha^{-1}U^*_{m_n,t}(u/\sqrt{m_n})\right) + o_p(1),\end{aligned}$$

where $e_n = o_p(1)$ and $E[e_n] = o(1)$, and e_n is independent of scale.

LEMMA A.2 *Under the conditions of Corollary 2, $c = (m_n/n)b^{\alpha}_{m_n}(1 + o(1/\sqrt{m_n}))$.*

LEMMA A.3 *Under Assumptions A and B, $\forall\{m_n, m_n\} \in S_{m_n}$*

$$\begin{aligned}&1/\sqrt{m_n}\sum_{t=1}^{n} U_{m_n,t} - 1/\sqrt{\tilde{m}_n}\sum_{t=1}^{n} U_{\tilde{m}_n,t} = o_p(1)\\ &1/\sqrt{m_n}\sum_{t=1}^{n} U^*_{m_n,t}(u/\sqrt{m_n}) - 1/\sqrt{\tilde{m}_n}\sum_{t=1}^{n} U^*_{\tilde{m}_n,t}(u/\sqrt{\tilde{m}_n}) = o_p(1).\end{aligned}$$

Proof of Lemma A.1. We can always write

$$\begin{aligned}&\sqrt{m_n}\left(\hat{\alpha}^{-1}_{m_n} - \alpha^{-1}\right)\\ &\quad = \sqrt{m_n}\left(1/m_n\sum_{i=1}^{m_n}\ln X_{(i)}/X_{(m_n+1)} - \alpha^{-1}\right)\\ &\quad = \sqrt{m_n}\left(1/m_n\sum_{i=1}^{m_n}\ln X_{(i)}/b_{m_n} - \alpha^{-1}\right) - \sqrt{m_n}\ln X_{(m_n+1)}/b_{m_n}\\ &\quad = \sqrt{m_n}\left(1/m_n\sum_{i=1}^{m_n}\ln X_{(i)}/b_{m_n} - E\left(1/m_n\sum_{t=1}^{n}(\ln X_t/b_{m_n})_+\right)\right)\\ &\qquad - \sqrt{m_n}\left(\ln X_{(m_n+1)}/b_{m_n}\right) + \sqrt{m_n}\left[E\left(1/m_n\sum_{t=1}^{n}(\ln X_t/b_{m_n})_+\right) - \alpha^{-1}\right].\end{aligned}$$

Using Assumption A and arguments in Hsing (1991: p. 1554), $\sqrt{m_n}[E(m_n^{-1}\sum_{t=1}^{n}(\ln X_t/b_{m_n})_+) - \alpha^{-1}] = o(1)$.

Moreover, Lemma 4 of Hill (2005a) and a Cramér-Wold device give

$$\sqrt{m_n}\left(1/m_n\sum_{t=1}^{n}U_{m_n,t},\ \alpha^{-1}1/m_n\sum_{t=1}^{n}U^*_{m_n,t}(u/\sqrt{m_n})\right) \Rightarrow (Z_1, Z_2),$$

where each $Z_i \sim N(0, \sigma_i^2)$, $\sigma_i^2 < \infty$.

Furthermore, Lemma 1 of Hill (2005a) states $|\ln X_{([\rho m_n])} - \ln b_{(\rho m)}| \to 0$ for all ρ in an arbitrary neighborhood of 1. Therefore Theorem 2.2 of Hsing (1991: eq. 2.4-2.7) can be used to show

$$\sqrt{m_n}\left(1/m_n \sum_{i=1}^{m_n} \ln X_{(i)}/b_{m_n} - E\left(1/m_n \sum_{t=1}^{n} (\ln X_t/b_{m_n})_+\right)\right) \Rightarrow Z_1$$
$$\sqrt{m_n}\left(\ln X_{(m_n+1)}/b_{m_n}\right) \Rightarrow Z_2,$$

hence

$$\begin{aligned}\sqrt{m_n}\left(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1}\right) &= 1/\sqrt{m_n}\sum_{t=1}^{n}\left(U_{m_n,t} - \alpha^{-1}U_{m_n,t}^{*}(u/\sqrt{m_n})\right) \\ &\quad + \sqrt{m_n}\left[E\left(1/m_n \sum_{t=1}^{n} (\ln X_t/b_{m_n})_+\right) - \alpha^{-1}\right] + e_n \\ &= 1/\sqrt{m_n}\sum_{t=1}^{n}\left(U_{m_n,t} - \alpha^{-1}U_{m_n,t}^{*}(u/\sqrt{m_n})\right) + o_p(1),\end{aligned}$$

where

$$\begin{aligned}e_n &= \left(1/\sqrt{m_n}\sum_{i=1}^{m_n} \ln X_{(i)}/b_{m_n} - E\left(1/m_n \sum_{t=1}^{n} (\ln X_t/b_{m_n})_+\right)\right) - 1/m_n \sum_{t=1}^{n} U_{m_n,t} \\ &\quad + \sqrt{m_n}\left(\ln X_{(m_n+1)}/b_{m_n}\right) - \alpha^{-1} 1/\sqrt{m_n}\sum_{t=1}^{n} U_{m_n,t}^{*}(u/\sqrt{m_n}) \\ &= o_p(1).\end{aligned}$$

Finally $E[e_n] \to 0$ in probability by Fatou's Lemma: $\limsup_{n\geq 1} E[|e_n|] \leq E[\limsup_{n\geq 1} |e_n|] = 0$, hence $|E[e_n]| \leq E[|e_n|] \to 0$. ■

Proof of Lemma A.2. By the construction of the sequence $\{b_{m_n}\}_{n\geq 1}$ and the distribution tail $\bar{F}_t(b_{m_n}) = c b_{mn}^{-\alpha}(1 + b_{mn}^{-\alpha})$ we can write

$$(m_n/n) b_{m_n}^{\alpha} \sim c(1 + b_{m_n}^{-\alpha}) \Rightarrow b_{m_n}^{\alpha} \sim c(n/m_n).$$

Use $m_n = o(n^{2/3})$ to conclude $(m_n/n) b_{m_n}^{\alpha} \sim c(1 + c^{-1}(m_n/n)) = c + (m_n/n)$ $= c + o(m_n^{-1/2})$. ■

Proof of Lemma A.3. For simplicity let $\tilde{m}_n \leq m_n$ $\forall n$. Write

$$\begin{aligned}&1/m_n^{1/2}\sum_{t=1}^{n} U_{m_n,t} - 1/\tilde{m}_n^{1/2}\sum_{t=1}^{n} U_{m_n,t} \\ &\quad = 1/m_n^{1/2}\sum_{t=1}^{n}\left(U_{m_n,t} - (m_n/\tilde{m}_n)^{1/2} U_{\tilde{m}_n,t}\right),\end{aligned} \tag{12}$$

where

$$\begin{aligned}&U_{m_n,t} - (m_n/\tilde{m}_n)^{1/2} U_{\tilde{m}_n,t} \\ &\quad = (\ln X_t/b_{m_n})_+ - (m_n/\tilde{m}_n)^{1/2}(\ln X_t/b_{\tilde{m}_n})_+ \\ &\qquad + (m_n/\tilde{m}_n)^{1/2} E(\ln X_t/b_{\tilde{m}_n})_+ - E(\ln X_t/b_{m_n})_+ .\end{aligned}$$

Using Assumption A and $(\ln X_t/b_{m_n})_+ \geq 0$, the expectations difference is $o(m_n^{1/2}/n)$ because

$$
\begin{aligned}
&(m_n/\tilde{m}_n)^{1/2} E(\ln X_t/b_{\tilde{m}_n})_+ - E(\ln X_t/b_{m_n})_+ \qquad (13)\\
&= (m_n/\tilde{m}_n)^{1/2} \int_0^\infty \bar{F}(b_{\tilde{m}_n} e^u)\, du - \int_0^\infty \bar{F}(b_{m_n} e^u)\, dudu \\
&= (m_n/\tilde{m}_n)^{1/2} \bar{F}(b_{\tilde{m}_n}) \int_0^\infty \frac{\bar{F}(b_{\tilde{m}_n} e^u)}{\bar{F}(b_{\tilde{m}_n})} du - \bar{F}(b_{m_n}) \int_0^\infty \frac{\bar{F}(b_{m_n} e^u)}{\bar{F}(b_{m_n})} du \\
&\sim (m_n/\tilde{m}_n)^{1/2} \bar{F}(b_{\tilde{m}_n}) \int_0^\infty e^{-\alpha u} du - \bar{F}(b_{m_n}) \int_0^\infty e^{-\alpha u} du \\
&= \alpha^{-1} \times \bar{F}(b_{m_n}) \left[\left(\frac{m_n}{\tilde{m}_n} \right)^{1/2} \frac{\bar{F}(b_{\tilde{m}_n})}{\bar{F}(b_{m_n})} - 1 \right] \\
&= \alpha^{-1} \times O(m_n/n) \times o(1/m_n^{1/2}) = o(m_n^{1/2}/n),
\end{aligned}
$$

where the last line follows from $\bar{F}(b_{m_n}) = O(m_n/n)$ by the construction of b_{m_n}, and Lemma A.3.1 below.

It is easy to show $b_{m_n} \leq b_{\tilde{m}_n}$ as $n \to \infty$ follows from $\tilde{m}_n \leq m_n$. Assume n is sufficiently large such that $b_{m_n} \leq b_{\tilde{m}_n}$. Define

$$A_{t,m_n,1} = \{t : b_{\tilde{m}_n} \geq X_t > b_{m_n}\}, \quad A_{t,\tilde{m}_n,2} = \{t : X_t \geq b_{\tilde{m}_n}\}.$$

Then (12) and (13) imply

$$
\begin{aligned}
&\left| m_n^{-1/2} \sum\nolimits_{t=1}^n \left(U_{m_n,t} - (m_n/\tilde{m}_n)^{1/2} U_{\tilde{m}_n,t} \right) \right| \qquad (14)\\
&\leq \left| m_n^{-1/2} \sum\nolimits_{t=1}^n \left((\ln X_t/b_{m_n})_+ - (m_n/\tilde{m}_n)^{1/2} (\ln X_t/b_{\tilde{m}_n})_+ \right) \right| + o(1) \\
&\leq m_n^{-1/2} \sum\nolimits_{t \in A_{t,mn,1}} \ln X_t/b_{m_n} \\
&\qquad + \left| m_n^{-1/2} \sum\nolimits_{t \in A_{t,\tilde{m}_n,2}} \left(\ln X_t/b_{m_n} - (m_n/\tilde{m}_n)^{1/2} \ln X_t/b_{\tilde{m}_n} \right) \right| + o(1) \\
&\leq m_n^{-1/2} \sum\nolimits_{t \in A_{t,1}} \ln X_t/b_{m_n} + (\ln b_{\tilde{m}_n}/b_{m_n}) \times 1/\tilde{m}_n^{1/2} \sum\nolimits_{t=1}^n I(X_t \geq b_{\tilde{m}_n}) + o(1)
\end{aligned}
$$

If we show each term on the right-hand-side of (14) is $o_p(1)$ then the claim is proven by Chebyshev's inequality. First, as $n \to \infty$

$$
\begin{aligned}
&\left\| 1/m_n^{1/2} \sum\nolimits_{t \in A_{t,1}} \ln X_t/b_{m_n} \right\|_1 \\
&\qquad \leq \frac{n}{m_n^{1/2}} P(X_t > b_{m_n}) \times \int_0^{\ln b_{\tilde{m}_n}/b_{m_n}} \frac{P(X_t > b_{m_n} e^u)}{P(X_t > b_{m_n})} du
\end{aligned}
$$

$$= \frac{n}{m_n^{1/2}} \frac{m_n}{n} \left(1 + o(m_n^{-1/2})\right) \times \int_0^{\ln b_{\tilde{m}n}/b_{mn}} e^{-\alpha u} du$$
$$= m_n^{1/2} \left(1 + o(m_n^{-1/2})\right) \times \alpha^{-1} \times \left(1 - b_{\tilde{m}_n}^{\alpha}/b_{m_n}^{\alpha}\right)$$
$$= m_n^{1/2} \left(1 + o(m_n^{-1/2})\right) \times \alpha^{-1} \times o(m_n^{-1/2}) = o(1).$$

The first equality follows from dominated convergence and Assumption A:

$$\begin{aligned} P(X_t &> b_{m_n} e^u) \sim e^{-\alpha u} P(X_t > b_{m_n}) \\ &= (m_n/n) \times (1 + O(g(b_{m_n}))) \\ &= (m_n/n) \times (1 + o(1/\sqrt{m_n})). \end{aligned}$$

The third equality follows from Lemma A.3.1.

For the second term in (14), arguments similar to the above and Lemma A.3.1 give

$$\begin{aligned} &(\ln b_{\tilde{m}_n}/b_{m_n}) \times \left\| \tilde{m}_n^{-1/2} \sum\nolimits_{t=1}^{n} I\left(X_t \geq b_{\tilde{m}_n}\right) \right\|_1 \\ &\leq o(\tilde{m}_n^{-1/2}) \times (n/\tilde{m}_n^{1/2}) \times P\left(X_t \geq b_{m_n}\right) \\ &= o(\tilde{m}_n^{-1/2}) \times (m_n/\tilde{m}_n^{1/2}) \times \left(1 + o(m_n^{-1/2})\right) = o(1). \end{aligned}$$

■

LEMMA A.3.1 *Under the conditions of Lemma A.3,* $\forall m_n, m_n \in S_{m_n}$

$$\left(\frac{m_n}{\tilde{m}_n}\right)^{1/2} \frac{\bar{F}(b_{\tilde{m}_n})}{\bar{F}(b_{m_n})} = 1 + o(m_n^{-1/2}) \quad \text{and} \quad b_{\tilde{m}_n}^{\alpha}/b_{m_n}^{\alpha} = 1 + o(m_n^{-1/2}).$$

Proof. Properties (1)-(5) imply

$$\begin{aligned} (n/m_n)\bar{F}(b_{m_n}) &= (n/m_n) b_{m_n}^{-\alpha} L(b_{m_n}) \\ &= 1 + O(g(b_{m_n})) = 1 + o(m_n^{-1/2}) \end{aligned}$$

hence

$$\left(\frac{m_n}{\tilde{m}_n}\right) \frac{\bar{F}(b_{\tilde{m}_n})}{\bar{F}(b_{m_n})} = \frac{(n/\tilde{m}_n) b_{\tilde{m}_n}^{-\alpha}}{(n/m_n) b_{m_n}^{-\alpha}} \frac{L(b_{\tilde{m}_n})}{L(b_{m_n})} = \frac{1 + o(\tilde{m}_n^{-1/2})}{1 + o(m_n^{-1/2})} = 1 + o(m_n^{-1/2}), \quad (15)$$

given $m_n/\tilde{m}_n = o(1)$. But this implies $b_{m_n}/b_{\tilde{m}_n} \to 1$, hence $L(b_{m_n})/L(b_{\tilde{m}_n}) \to 1$. Moreover, for some N, every $n \geq N$, and any $\lambda > 1$

$$\left| \frac{L(b_{\tilde{m}_n})}{L(b_{m_n})} - 1 \right| \leq \left| \frac{L(b_{m_n}\lambda)}{L(b_{m_n})} - 1 \right| = O(g(b_{m_n})) = o(m_n^{-1/2}). \quad (16)$$

From (15) we now deduce

$$\frac{(n/m_n)b_{m_n}^{-\alpha}}{(n/\tilde{m}_n)b_{\tilde{m}_n}^{-\alpha}}\left(1+o(m_n^{-1/2})\right)=1+o(m_n^{-1/2}),$$

hence

$$b_{\tilde{m}_n}^{\alpha}/b_{m_n}^{\alpha}=1+o(m_n^{-1/2}). \tag{17}$$

Together (15)-(17) imply $\bar{F}(b_{m_n})/\bar{F}(b_{m_n})=1+o(m_n^{-1/2})$, hence

$$\left(\frac{m_n}{\tilde{m}_n}\right)^{1/2}\frac{\bar{F}(b_{\tilde{m}_n})}{\bar{F}(b_{m_n})}=1+o(m_n^{-1/2}).$$

■

References

[1] Akgiray, V. and G. Booth (1988). "The Stable Law Model of Stock Returns," *Journal of Business and Economic Statistics* 6, 51-57.

[2] Basrak, B., R.A. Davis, and T. Mikosch (2002). "Regular Variation of GARCH Processes," *Stochastic Processes and their Applications* 99, 95-115.

[3] Beirlant, J., P. Vynckier and J. Teugels (1996). "Tail Index Estimation, Pareto Quantile Plots, and Regression Diagnostics," *Journal of the American Statistical Association* 91, 1659-67.

[4] Beirlant, J., D. Dierckx, and A. Guillou (2005). "Estimation of the Extreme Value Index and Generalized Plots," *Bernoulli* 11, 949-970.

[5] Beirlant, J. Y. Goegebeur, J. Segers, J. Teugels, D. De Waal (2004). "Statistics of Extremes: Theory and Applications," (John Wiley & Sons: New York).

[6] Bingham, N.H., C.M. Goldie and J. L. Teugels (1987). "Regular Variation," (Cambridge Univ. Press: Great Britain).

[7] Bradley, B.O., and M.S. Taqqu (2003). "Financial Risk and Heavy Tails. In Handbook of Heavy-Tailed Distributions in Finance," S. T. Rachev (ed.), Elsevier: Amsterdam.

[8] Brown, B. and R. W. Katz (1993). "Regional Analysis of Temperature Extremes: Spatial Analog for Climate Change?" *Journal of Climate* 8, 108–119.

[9] Caner, M. (1998). "Tests for Cointegration with Infinite Variance Errors," *Journal of Econometrics* 86, 155-175.

[10] Chan, N.H., and L.T. Tran (1989). "On the First Order Autoregressive Process with Infinite Variance," *Econometric Theory* 5, 354-362.

[11] Chan, N.H., S.D. Deng, L. Peng, and Z. Xia (2005). "Interval Estimation for the Conditional Value-at-Risk Based on GARCH with Heavy Tailed Innovations," *Journal of Econometrics* (forthcoming).

[12] Cheng, B.N. and S.T. Rachev (1995). "Multivariate Stable Futures Prices," *Mathematical Finance* 5, 133-153.

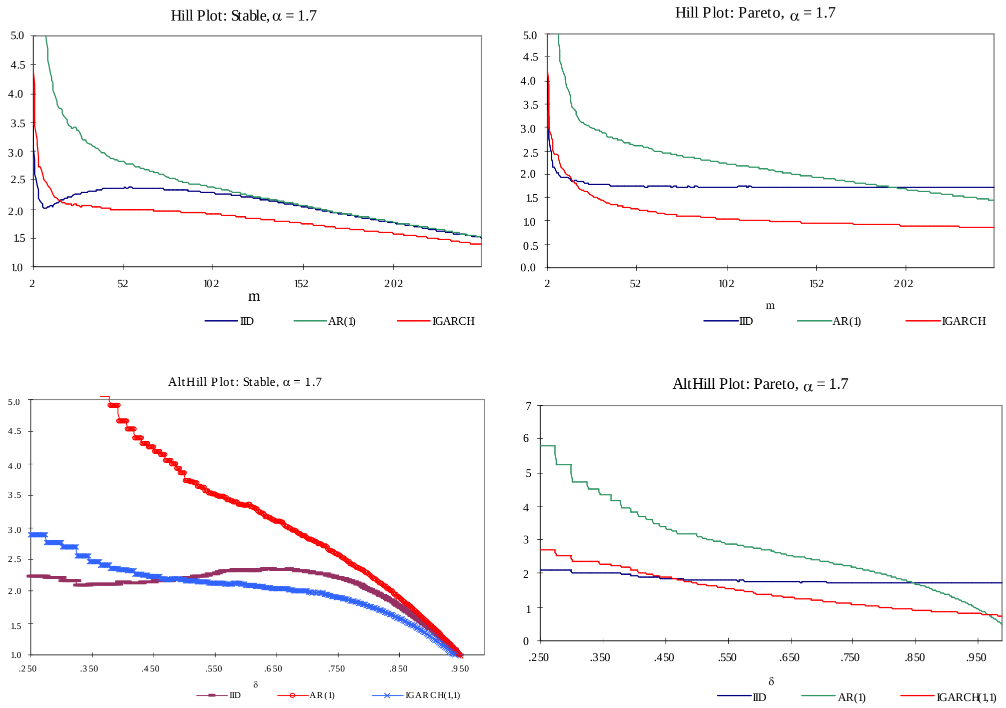
[13] Cline, D. B. H. (1986). "Convolution Tails, Product Tails and Domains of Attraction," *Probability Theory and Related Fields* 72, 525-557.

[14] Danielsson, J., L. de Haan, L. Peng, and C. G. de Vries (1998), "Using a Bootstrap

Method to Choose the Sample Fraction in Tail Index Estimation," mimeo, Eramus University Rotterdam.
[15] de Haan, L. (1970). "On Regular Variation and Its Applications to the Weak Convergence of Sample Extremes," *MC Tract 32, Mathematisch Centrum*, Amsterdam.
[16] de Jong, R.M., and J. Davidson (2000). "Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices," *Econometrica* 68, 407-423.
[17] Draisma, G., L. de Haan, L. Peng, T.T. Periera (1997). "A Bootstrap-based Method to Achieve Optimality in Estimating the Extreme-value Index," Eramus University, Econometric Institute Report EI 2000-18/A.
[18] Drees, H., de Haan, L. and Resnick, S. (2000). "How to Make a Hill Plot," *Annals of Statistics* 28, 254-274.
[19] Drees, H., Ferreira A, and L. de Haan (2004). On Maximum Likelihood Estimation of the Extreme Value Index, *Annals of Applied Probability* 14, 1179-1201.
[20] Drees, H., and E. Kaufman (1997). "Selecting the Optimal Sample Fraction in Univariate Extreme Value Estimation," *Stochastic Processes and their Applications* 75, 149-172.
[21] Dumouchel, W.H. (1983). "Estimating the Stable Index α in order to Measure Tail Thickness," *Annals of Statistics* 11 1019-1036.
[22] Embrechts, P. and C. M. Goldie (1980). "On Closure and Factorization Properties of Subexponential Distributions," *Journal of Aust. Math. Soc. Ser. A* 29, 243-256.
[23] Embrechts, P. and C. M. Goldie (1982). "On Convolution Tails," *Stochastic Processes and their Applications* 13, 263-278.
[24] Embrechts, P., C. Kluppelberg, and T. Mikosch (1997). "Modelling Extremal Events for Insurance and Finance," (Springer: New York).
[25] Fama, E. (1965). "Portfolio Analysis in a Stable Paretian Market," *Management Science* 11, 404-419.
[26] Finkenstadt, B., and H. Rootzén (2003). "Extreme Values in Finance, Telecommunications and the Environment," (Chapman and Hall).
[27] Gallant, A. R. and H. White (1988). A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models. Basil Blackwell: Oxford.
[28] Goldie, C.M., and R.L. Smith (1987). "Slow Variation with Remainder: Theory and Applications," *Quarterly Journal of Mathematics* 38, 45-71.
[29] Haeusler, E., and J. L. Teugels (1985). "On Asymptotic Normality of Hill's Estimator for the Exponent of Regular Variation," *Annals of Statistics* 13, 743-756.
[30] Hall, P. (1982). "On Some Estimates of an Exponent of Regular Variation," *Journal of the Royal Statistical Society* 44, 37-42.
[31] 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.
[32] Hall, P., and C. C. Heyde (1980). Martingale Limit Theory and Its Applications. Academic Press: New York.
[33] Hall, P., and A.H. Welsh (1985). "Adaptive Estimates of Parameters of Regular Variation," *Annals of Statistics* 13, 331-341.
[34] Hill, B.M. (1975). "A Simple General Approach to Inference about the Tail of a Distribution," *Annals of Mathematical Statistics* 3, 1163-1174.

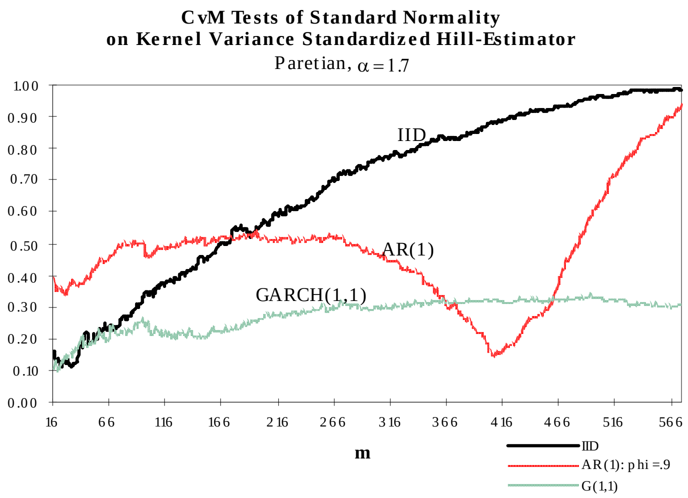
[35] Hill, J.B. (2005a). "On Tail Index Estimation for Dependent, Heterogeneous Data," Dept. of Economics, Florida International University; available at http://www.fiu.edu/∼hill_het_dep.pdf,
[36] Hill, J.B. (2005b). "Gaussian Tests of 'Extremal White Noise' for Dependent, Heterogeneous, Heavy Tailed Stochastic Processes with an Application", Conditionally accepted by *Journal of Econometrics*, available at www.fiu.edu/∼co-relation-test.pdf.
[37] Hill, J.B. (2006). "Robust Tests of Extremal Volatility Spillover with an Application to Extremal Asset Market Contagion", Dept. of Economics, Florida International University.
[38] Hsing, T. (1991). "On Tail Index Estimation Using Dependent Data," *Annals of Statistics* 19, 1547-1569.
[39] Huisman R., K.G. Koedijk, C.J.M. Kool, and F. Palm (2001), "Tail-Index Estimates in Small Samples," *Journal of Business and Economic Statistics* 19, 208-216.
[40] Kearns, P. and A. Pagan (1997). "Estimating the Density Tail Index for Financial Time Series," *The Review of Economics and Statistics* 79, 171-175.
[41] Leadbetter, M.R., G. Lindgren and H. Rootzén (1983). "Extremes and Related Properties of Random Sequences and Processes," (Springer-Verlag: New York).
[42] Mandelbrot, B. (1963). "The Variation of Certain Speculative Prices," *Journal of Business* 36, 394-419.
[43] Manski, C. (1983). "Closest Empirical Distribution Estimation," *Econometrica* 51, 304-319.
[44] McCulloch, J. H. (1996). "Financial Applications of Stable Distributions," in G.S. Maddala and C.R. Rao, eds., *Statistical Methods in Finance* (Handbook of Statistics 14), 393-425, Elsevier Science: Amsterdam.
[45] Mikosch,T., and C. Stărică (2000). "Limit Theory for the Sample Autocorrelations and Extremes of a GARCH(1,1) Process," *Annals of Statistics* 28, 1427-145.
[46] Pickands, J. (1975). Statistical-Inference using Extreme Order Statistics, *Annals of Statistics* 3, 119-131.
[47] Poon, S., M. Rockinger and J.A. Tawn (2002). "Modelling Extreme-Value Dependence in International Stock Markets," working paper, University of Strathclyde.
[48] Rachev, S. T. (2003). "Handbook of Heavy Tailed Distributions in Finance," (Elsevier Science: New York).
[49] Resnick, S. (1987). "Extreme Values, Regular Variation and Point Processes," (Springer-Verlag: New York).
[50] Resnick, S. (1987). "Heavy Tail Modeling and Teletraffic Data," *Annals of Statistics* 25, 1805-1869.
[51] Resnick, S., and H. Rootzén (2000). "Self-Similar Communication Models and Very Heavy Tails," *Annals of Applied Probability* 10, 753–778.
[52] Resnick, S. and C. Stărică (1997). "Smoothing the Hill Estimator," *Advances in Applied Probability* 29, 271-293.
[53] Smith, R.L. (1982). Uniform Rates of Convergence in Extreme-Value Theory, Advances in Applied Probability 14, 600-622.
[54] Smith, R.L. (1987). "Estimating Tails of Probability-Distributions," *Annals of Statistics* 15, 1175-1207.

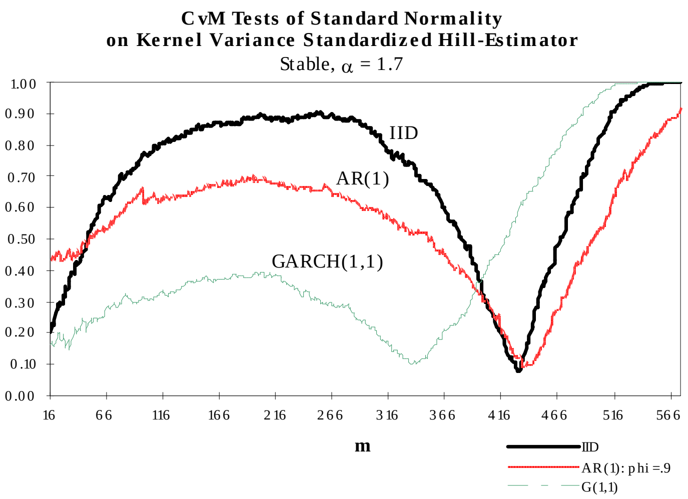
Figure 1
Hill-Plots and Alt-Hill Plots for iid, AR(1) and IGARCH



Notes: The data are based on randomly generated symmetric Stable or Pareto innovations $\{\epsilon_t\}$ with unit scale and index $\alpha = 1.7$. In the AR(1) case $X_t = .9X_{t-1} + \epsilon_t$. In the IGARCH case $X_t = \sigma_{t-1}\epsilon_t$, $\sigma_t^p = \theta_0 + (1-\theta_1)\epsilon_{t-1}^p + \theta_1\sigma_{t-1}^p$, where $\{\theta_0, \theta_1\} \sim U[.1, .9]$.

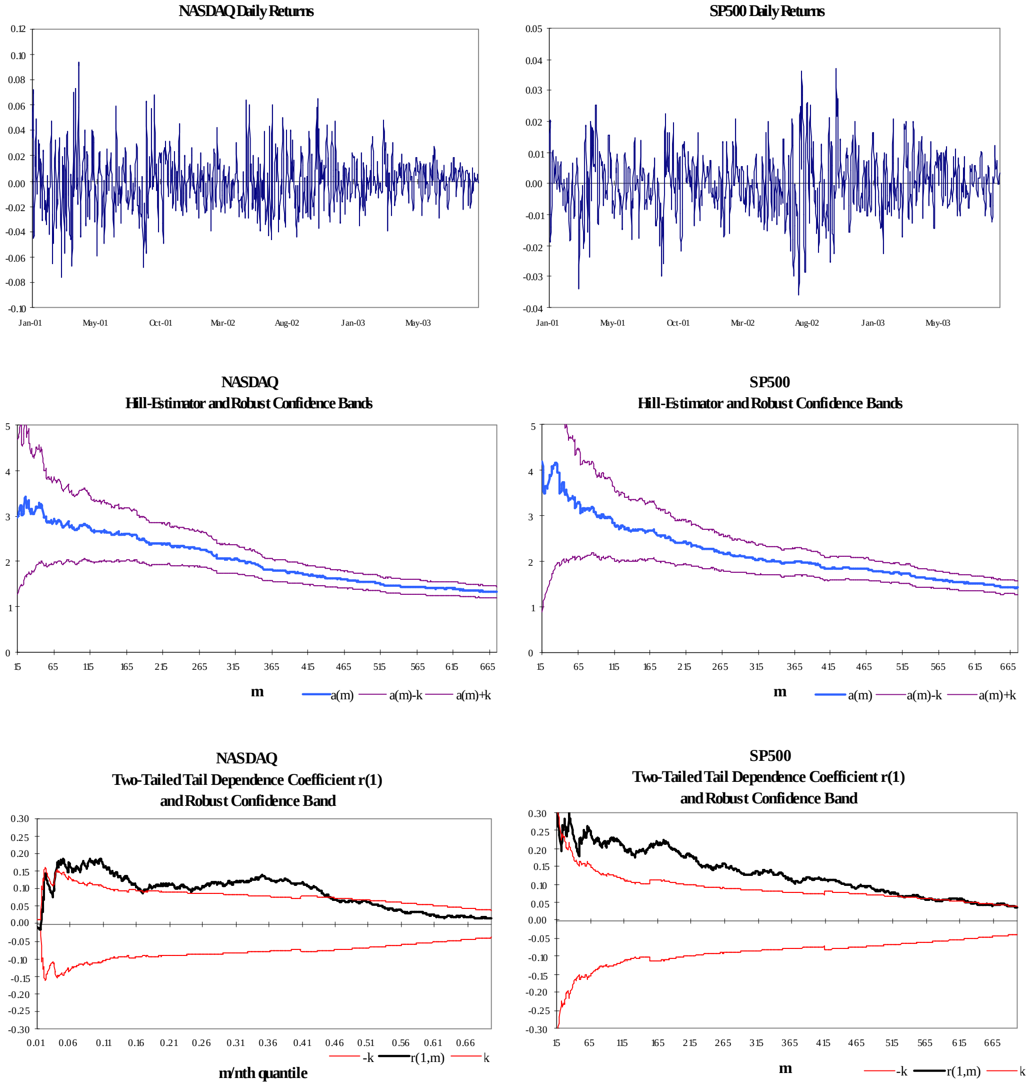
Figure 2
Cramér-von Mises Tests of Standard Normality on $\hat{\alpha}_{m_n}$





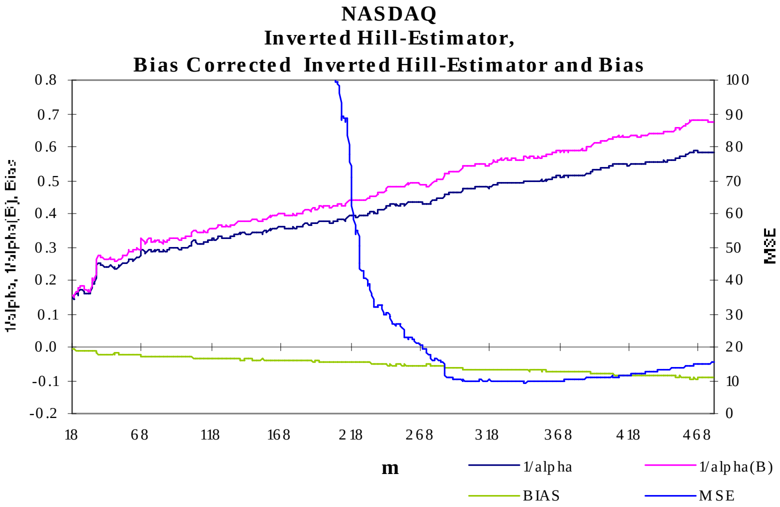
CvM statistic 1%, 5% and 10% critical values = .34, .22, .173.

Figure 3
NASDAQ and S&P500 Daily Log-Returns



Notes: The median two-tailed tail dependence coefficients $\hat{r}_{m_n}(1)$ over all m are NASDAQ: .10 ± .07, and SP500: .17 ± .09. As long as each fractile m belongs to a sequence in S_m, the set of all proportional sequences, the median dependence coefficient is consistent and asymptotically normal under Assumptions A-C. See Hill (2006) for related theory and computation of the above plotted consistent kernel confidence bands.

Figure 4


NASDAQ
Inverted Hill-Estimator,
Bias Corrected Inverted Hill-Estimator and Bias
1/alpha, 1/alpha(B), Bias
MSE
m
1/alpha
1/alpha(B)
BIAS
MSE

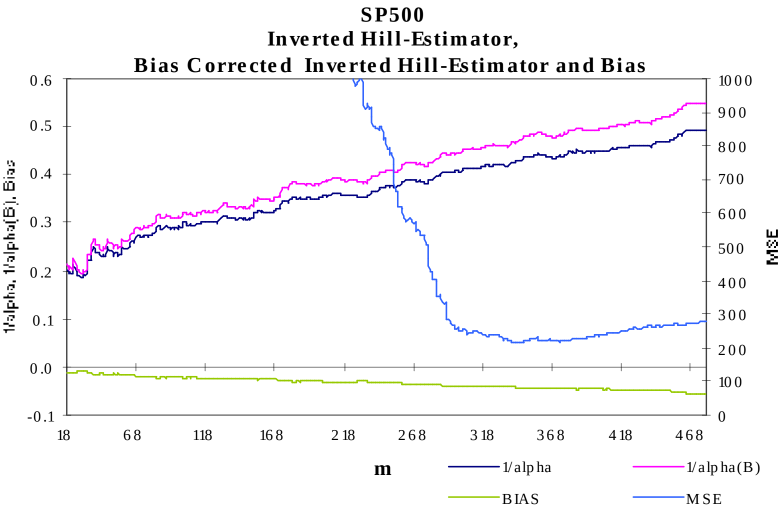

SP500
Inverted Hill-Estimator,
Bias Corrected Inverted Hill-Estimator and Bias
1/alpha, 1/alpha(B), Bias
MSE
m
1/alpha
1/alpha(B)
BIAS
MSE

Table 1.1: Stable, $\alpha = 1.7$, $n = 1000$

	AR(1)				G(1,1)[a]
ϕ	0	.2	.6	.9	
m	$\hat{\alpha}_{m_n} \pm k_{m_n}$	$\hat{\alpha}_{m_n} \pm k_{m_n}$	$\hat{\alpha}_{m_n} \pm k_{m_n}$	$\hat{\alpha}_{m_n} \pm k_{m_n}$	$\hat{\alpha}_{m_n} \pm k_{m_n}$
20	2.36±2.3[b]	2.32±2.5	1.38±1.3	5.14±4.4	2.17±2.4
50	2.73±2.1	2.38±1.4	1.42±.98	2.41±2.0	2.02±1.8
100	2.05±.73	2.41±1.3	1.55±.84	3.60±1.8	1.95±.83
150	2.32±.64	2.42±1.3	1.91±.80	3.21±1.3	2.03±.71
200	2.15±.39	2.19±.45	2.11±.66	2.88±.79	1.97±.42
300	1.89±.26	1.92±.31	1.89±.43	2.26±.48	1.81±.32
400	1.69±.16	1.71±.24	1.79±.30	1.72±.40	1.62±.23
500	1.47±.13	1.61±.15	1.48±.22	1.54±.20	1.41±.16
$\underline{m}_2$	369[c]	381	411	422	347
$\bar{m}_\alpha$	460[d]	477	518	510	454
$\hat{\bar{\sigma}}_{400}$	1.80	2.41	3.09	4.04	2.35_{400}
$KS_{.05}$	415-430[e,f]	420-435	422-445	390-440	310-380
$KS_{.10}$	390-460	400-445	405-470	350-470	260-410

Notes: a. GARCH(1,1)

b. Bandwidth $k_{m_n} = 1.96 \times \hat{\bar{\sigma}}_{m_n} / m_n^{1/2}$

c. Minimum m_n at which 2 does not occur in the 90% interval.

d. Maximum m_n at which $\alpha = 1.7$ occurs in the 95% interval.

e. KS_θ: Fractile range over which the Kolmogorov-Smirnov test of normality on $m_n^{1/2}(\hat{\alpha}_{m_n}\text{-}\alpha)/\hat{\bar{\sigma}}_{m_n}$ is not rejected at the θ-level.

f. Excluding Paretian iid and GARCH data, in all cases the Cramér-von Mises statistic is bi-modal over the fractile range. We only display the upper range of fractile values at which we fail to reject normality. See Figure 2.

Table 1.2: Minimum MSE Estimates

	AR(1)				G(1,1)
ϕ	0	.2	.6	.9	
$\hat{\alpha}_{m_n^*} \pm k_{m_n^*}$	1.66±.165	1.67±.202	1.63±.222	1.62±.274	1.77±.301
$m_n^*\{\hat{\sigma}^2_{m_n}\}^a$	443	462	517	444	318 (.83)
$\hat{\alpha}_{m_n^*} \pm k_{m_n^*}$	1.87±.319	2.34±.117	1.84±.183	1.69±.202	1.76±.386
$m_n^*\{\hat{\sigma}^2_{m_n}(\hat{B})\}^b$	331	565	540	496	297 (.82)

Notes: a. Non-"bias corrected" $m_n^*\{\hat{\sigma}^2_{m_n}\} = \arg\min_{m_n \in M_n}\{\hat{\sigma}^2_{m_n}\}$.

b. "Bias corrected" $m_n^*\{\hat{\sigma}^2_{m_n}(\hat{B})\} = \arg\min_{m_n \in M_n}\{\hat{\sigma}^2_{m_n}(\hat{B})\}$.

Table 2.1: Paretian[a], $\alpha = 1.7$, $n = 1000$

	AR(1)				**G(1,1)**
ϕ	0	.2	.6	.9	
m	$\hat{\alpha}_{m_n} \pm k_{m_n}$	$\hat{\alpha}_{m_n} \pm k_{m_n}$	$\hat{\alpha}_{m_n} \pm k_{m_n}$	$\hat{\alpha}_{m_n} \pm k_{m_n}$	$\hat{\alpha}_{m_n} \pm k_{m_n}$
20	1.83±1.15	1.85±1.20	2.14±1.53	3.18±2.62	2.09±2.1
50	1.79±.703	1.87±.783	2.04±.962	2.81±.146	1.88±1.7
100	1.85±.509	1.90±.584	2.12±.799	2.59±1.11	1.61±.72
150	1.89±.421	1.95±.485	2.15±.662	2.46±.842	1.59±.57
200	1.94±.372	1.99±.428	2.17±.573	2.32±.670	1.58±.39
300	2.00±.308	2.08±.360	2.08±.427	2.05±.453	1.61±.32
400	2.06±.268	2.13±.311	1.87±.302	1.78±.317	1.64±.28
500	2.11±.239	2.14±.270	1.61±.208	1.52±.221	1.65±.25
550	2.13±.228	2.13±.254	1.47±.180	1.39±.206	1.69±.31
$\underline{m}_2$	-	-	448	423	180
$\bar{m}_\alpha$	-	-	536	510	575
$\hat{\hat{\sigma}}_{400}$	2.73	3.17	3.08	3.23	2.59_{50}
$KS_{.05}$	15-48	15-50	430-485	400-450	15-85
$KS_{.10}$	15-110	15-95	430-500	360-480	15-575

Notes: The "Paretian Tail" is $P(X_t > z) = z^{-\alpha}(1 + z^{-\alpha})$.

Table 2.2: Minimum MSE Estimates

	AR(1)				**G(1,1)**
ϕ	0	.2	.6	.9	
$\hat{\alpha}_{m_n^*} \pm k_{m_n^*}$	1.78±.684	1.74±.796	1.63±.254	1.62±.264	1.59±.743
$m_n^*\{\hat{\sigma}^2_{m_n}\}^a$	51	46	452	446	48 (.56)
$\hat{\alpha}_{m_n^*} \pm k_{m_n^*}$	1.81±.674	1.66±.810	1.67±.245	1.63±.208	1.60±.327
$m_n^*\{\hat{\sigma}^2_{m_n}(\hat{B})\}$	56	46	496	484	355 (.85)

Table 4. Equity Mininimum MSE Estimates

	NASDAQ	SP500
$\hat{\alpha}_{m_n^*} \pm k_{m_n^*}$	1.71±.204	2.02±.257
$m_n^*(\hat{\sigma}^2_{m_n}) = m_n^*(\hat{\sigma}^2_{m_n}(\hat{B}))$	466	473

Notes: Sample size = 1422.