Heavy-Tailed Distributions in Finance

An Empirical Study

Computing Prelim
Christopher G. Green
University of Washington
Supervisor: R. Doug Martin

Revision date: **November 13, 2005**

Contents

List of Notation

$F * G$	The convolution of F and G
$\mathbb{N}$	The natural numbers $1, 2, 3, \ldots$
$\mathbb{R}$	The real numbers
$\mathbb{C}$	The complex numbers
iid	independent and identically-distributed
$\xrightarrow{d}$	converges in distribution to
$s_{\alpha,\beta}(x; \sigma, \mu)$	density of a stable distribution with parameters α, β, σ, and μ
$S_{\alpha,\beta}(x; \sigma, \mu)$	distribution function of a stable distribution with parameters α, β, σ, and μ
$\mathscr{T}_\nu$	Student's t distribution with ν degrees of freedom

Chapter 1

Introduction

The term "heavy-tailed distribution" refers to statistical distributions with more mass in their tails than the standard normal (a.k.a. Gaussian) distribution. In essence, greater mass in the tails of these distribution means that, under these models, extreme events are more likely to occur than under the normal model. Heavy-tailed distributions have become important tools in modern financial work, where the failure of the normal model to adequately capture the observed frequency of extreme events (such as market crashes) is (or should be, by now) well-known [Tei71, MRP98, RM00].

One of the earliest uses of distributional modeling in finance was in the study of changes in price between transactions in speculative markets [Fam63]. The seminal works of Bachelier [Bac67] and Osborne [Osb59] gave theoretical arguments in support of the hypothesis of approximately normally distributed price changes. Empirical work by Kendall [KH53] and others supported this hypothesis while at the same time suggesting that the tails of the empirical distributions of price changes where heavier than those of the normal distribution [Fam63]. This lead to investigations by Mandelbrot [Man63a, Man63b] and Fama [Fam63] into alternative models for price change data.

Mandelbrot argued for distributional models that satisfied the Pareto law, at least in the tails of the distirbution [Man63a]. The stable distributions of Lévy [Lév24] were natural candidates for models due to their closure under certain types of finite sums of random variables and their status as the only possible limits of infinite sums of random variables. Unfortunately, the lack of closed-form formulas for the stable distributions hindered their widespread use. Numerical methods for stable distributions were available by the early 70's [DuM71], but computers at the time were not fast enough to permit large empirical studies. It was not until the 90's that faster algorithms and faster hardware made the use of stable distributions practical. In the meantime, other heavy-tailed distributions, such as Student's t and the various extreme-value distributions, served as computationally feasible substitutes.

Heavy-tailed models have found their way into all areas of finance. This work is far from a comprehensive investigation of all the uses of heavy-tailed distributions in finance. Rather, we have limited our focus to some of the more practical concerns when using heavy-tailed distributions.

- How well do estimators of stable distribution parameters perform in the presence of outliers?

- Is the extra effort required by stable distributions warranted? Would a simpler to use model, such as the t model, give results that are just as good?
- In the location-scale model, what is the correlation between the tail-fatness parameter and scale?
- Is tail fatness related to firm size? If so, how?

Chapter 2

Univariate Stable Distributions

2.1 Introduction

Stable distributions play an important role in the theory of probability—they are the only possible limiting distributions of infinite sums of independent, identially-distributed (iid) random variables. When the random variables all have finite variances, this result is more commonly known as the Central Limit Theorem, and the limiting distribution is the familar normal (or Gaussian) distribution. The other members of the stable family, the limiting distributions when the restriction of finite variance is removed, are just as interesting as the normal, but much harder to study due to the lack of closed-form formulas for their densities and distribution functions.

The concept of a stable distribution was first introduced by Lévy around 1924 [Lév24] in his studies of sums of independent random variables [Hal81]. Some of the basic details of stable distributions remained unclear, though, until the 1936 paper of Khintchine and Levy [KL36].

According to Lévy's original definition, a distribution F is "stable" if for each pair of positive real numbers a_1 and a_2 there exists another positive real number a such that

$$F(a_1x) * F(a_2x) \equiv F(ax). \tag{2.1.1}$$

(Here $*$ denotes convolution.) This definition, however, has a significant drawback—in certain cases, stability is not preserved under translation, i.e., $F(x+b)$ is not stable even though $F(x)$ is. This led Lévy to introduce the weaker notion of "quasi-stability": a distribution F is quasi-stable if for all real numbers a_1, b_1, a_2, and b_2, with a_1 and a_2 positive, there exist real numbers a and b, with a positive, such that

$$F(a_1x+b_1) * F(a_2x+b_2) \equiv F(ax+b). \tag{2.1.2}$$

In the modern literature, Lévy's quasi-stable distributions are known as stable distributions, while his "stable" distributions are referred to as strictly stable distributions[Hal81].

Stable distributions have popped up in numerous scientific fields over the years. The earliest known occurrence of a stable distribution is generally agreed to be the 1919 paper of the Danish astronomer Holtsmark [HC73, Zol86, Hol19]. In his studies of the gravitational field of stars he derived (via Fourier transform methods) a probability distribution for the gravitational force exerted by a group of stars at a point in space. This distribution, now known in astrophysics as the "Holtsmark" distribution, corresponds to a symmetric stable distribution with index $\alpha = 3/2$ (see below).[1]

Stable distributions have also proven useful in the study of Brownian motion [Fel71], in economics and finance [Man63a, Man63b, Fam63], in electrical engineering [HC73], and in telecommunications [Kur01]. We will discuss their applications in finance in a later chapter; the interested reader can refer to the citations for more information about other applications.

2.1.1 Other Types of Stability

In this work we shall only be concerned with stability under (nonrandom) summation, often called Paretian stability due to the Pareto-like behavior of the tails of such stable distributions. There are, however, other types of stability (and the resulting distributions) that could be discussed. For instance, we could consider stability with respect to the maximum operation, i.e., distributions for which $\max\{X_i : i = 1, \ldots, n\}$ is equal in distribution to X_1 (after suitable translation and rescaling) whenever the random variables X_i are iid. We could also consider stability under random summation, etc. We refer the reader to the work of Rachev and Mittnik [RM00] for more information on such matters.

2.2 Theoretical Background

2.2.1 Introduction and Basic Definitions

First, let us start with a definition of a stable distribution that is more modern than that of Lévy.

Definition 2.2.1. *A (non-degenerate) distribution F is **stable** if, for all $n \in \mathbb{N}$, there exist constants $c_n > 0$ and d_n such that, whenever $X_1, X_2, \ldots, X_n$, and X are independently and identically distributed with distribution function F, the sum $X_1 + \cdots + X_n$ is distributed as $c_n X + d_n$.[DuM71].*

A random variable X is termed stable if its distribution function is stable.

If $d_n = 0$ for all n, the distribution (random variable, etc.) is said be **strictly stable**.

We will show below that the coefficient c_n must take the form $n^{1/\alpha}$ for some $\alpha \in (0, 2]$. The exponent α is called the **index** or **characteristic exponent** of the distribution.

The easiest way to develop stable distributions is via characteristic functions [Fel71, Lam96]. Let $\phi_k(t)$ and $\phi(t)$ be the characteristic functions of X_k and X, respectively. The characteristic function of the sum $X_1 + \cdots + X_n$ is simply the product of the characteristic functions of the summands, i.e., $\prod_{k=1}^{n} \phi_k(t)$. Since the X_k are iid the characteristic functions $\phi_k(t)$ are all identical, so this product equals $\phi(t)^n$.

[1] Although Holtsmark's work predates that of Lévy, Holtsmark only studied the specific case mentioned. It was Lévy who first introduced the notion of stability and who did the first in depth work on the stable family; thus he is commonly credited with their invention.

The characteristic function of $c_nX + d_n$, on the other hand, is $e^{id_nt}\phi(c_nt)$. Since the two quantities $X_1 + \cdots + X_n$ and $c_nX + d_n$ are equal in distribution, their characteristic functions must agree everywhere, so we must have

$$\phi(t)^n = e^{id_n}\phi(c_nt). \tag{2.2.3}$$

Let us consider first the case of F symmetric about $x = 0$[2]. The characteristic function of a symmetric distribution is real-valued[3]. Furthermore, it is a continuous, even function of its argument. Finally, we always have $\phi(0) = 1$ for a characteristic function.

Next, notice that by symmetry, we have

$$-(X_1 + \cdots + X_n) \overset{d}{=} X_1 + \cdots X_n \overset{d}{=} c_nX + d_n \overset{d}{=} -c_nX + d_n,$$

which forces $d_n = 0$. Thus $\phi(t)^n = \phi(c_nt)$. If $c_n = 1$ for all n, then $\phi(t)$ is identically 1, and F is a degenerate distribution. Since we have excluded that case, $c_n \neq 1$ for at least one n.

We now show that $\phi(t)$ is supported on the entire real line. Suppose to the contrary that $\phi(t)$ vanishes somewhere in $\mathbb{R}$. Then because it is a continous function that attains the value 1 at 0, the zero set of $\phi(t)$, restricted to the positive real line, must have a smallest element t_0. Using our previous observation we see that

$$0 = \phi(t_0)^n = \phi(c_nt_0)$$

and

$$0 = \phi(t_0/c_n)^n = \phi(t_0).$$

Hence both c_nt_0 and t_0/c_n are also zeros of $\phi(t)$. But, since c_n is positive and not 1, one of these numbers is strictly smaller than t_0, which contradicts the choice of t_0. Hence, we conclude that $\phi(t) > 0$ for all $t \in \mathbb{R}$.

Now that we know $\phi(t)$ is a positive function, we may safely work with its logarithm, which we will call $\psi(t)$. Since $\phi(t)$ is real-valued and bounded above by 1, $\psi(t)$ is a real-valued continuous nonpositive function. In terms of $\psi(t)$, our functional equation (2.2.3) for the characteristic function of a stable distribution is

$$n\psi(t) = \psi(c_nt). \tag{2.2.4}$$

At this point, it is not entirely obvious that the constant c_n is unique for a fixed n. Suppose that (for a given n) there are two constants, c_n and c'_n, for which (2.2.4) holds. Without loss of generality we may assume that $c_n > c'_n$. It is clear that

$$\psi\left(\frac{c'_n}{c_n}t\right) = \psi(t).$$

[2]Our argument follows that of Lamperti[Lam96].

[3]The characteristic function of $-X$ is the complex conjugate of that of X. By symmetry, however, X and $-X$ are identically distributed and hence, have identical characteristic functions. Thus, imaginary part of the characteristic function of X must vanish. The same reasoning can be used to show that the characteristic function of X is an even function.

By repeated use of this relation, we can actually establish the relation

$$\psi\left(\left[\frac{c'_n}{c_n}\right]^k t\right) = \psi(t),$$

for *every* $k \geq 1$. Since $c_n > c'_n$, $c'_n/c_n < 1$, and as $k \to \infty$, $c'_n/c_n \to 0$. Since $\psi(t)$ is continuous, we must have $\psi(t) = \psi(0) = 0$ for all t. This means that F is a degenerate distribution, which contradicts our assumption about F. Thus $c_n = c'_n$, i.e., c_n is unique [Lam96].

Our functional relation (2.2.4) is multiplicative, i.e., it implies that for any two integers m,n

$$\psi(c_{mn}t) = mn\psi(t) = m\psi(c_n t) = \psi(c_m c_n t).$$

Since the constants c_n, c_m, and c_{mn} are unique, this forces $c_{mn} = c_m c_n$ for all pairs of positive integers m and n.

Our functional relation (2.2.4) will be easier to solve if we allow n to take values in the positive real-numbers. To this end, we must extend our coefficient identity $c_{mn} = c_m c_n$ to positive real indices. We switch to a more traditional notation $c(y)$ for the coefficients here to avoid confusion; for any integer n we will have $c(n) = c_n$. It is clear that our argument establishing the uniqueness of the coefficients still holds as well.

We can easily see that $c(1) = 1$. We define the extension of the coefficient identity to reciprocals of integers in the obvious way: $c(1/n) = 1/c(n)$. It then follows that for any rational number p/q, $c(p/q) = c(p)/c(q)$, so we have extended our identity to the positive rationals. We can check that our functional relationship still holds.

$$q\psi(t) = \psi(c(q)t) \implies \psi(t)/q = \psi(t/c(q)) = \psi(c(1/q)t).$$

Since the rationals are dense in $\mathbb{R}$, for every real number y there is a sequence y_j of rationals that converges to y. Using the functional relation (2.2.4), we have for any t

$$\lim_{j\to\infty} \psi(c(y_j)t) = \lim_{j\to\infty} y_j\psi(t) = y\psi(t).$$

If the sequence $\{c(y_j)\}$ had 0 as a limit point, then the left-hand side of the above equation vanishes for all t, and F is degenerate. Since $y_j > 0$, we can make the same argument on the analogous equation that results from considering $c(1/y_j)$ to see that $\{c(y_j)\}$ does not diverge. Thus the sequence $\{c(y_j)\}$ is bounded away from 0 and ∞, and hence has a subsequence that converges to some positive number y'. But in fact all subsequences of $\{c(y_j)\}$ must converge to y', by the uniqueness of the coefficients $c(\cdot)$. Thus the limit $\lim c(y_j)$ exists and is positive. Moreover, the limit is independent of the sequence taking to y [Lam96]. Therefore, it makes sense to *define* $c(y)$ as $\lim c(y_j)$. With this definition in hand, we can verify that the coefficient relationship $c(xy) = c(x)c(y)$ and the functional relationship (2.2.4) still hold. We claim that $c(y)$ is a continuous function.

Let y_j be a sequence of *positive real numbers* that approaches a limit y. From our functional relationship we have

$$\lim_{j\to\infty} \psi(c(y_j)t) = \lim_{j\to\infty} y_j\psi(t) = y\psi(t) = \psi(c(y)t).$$

By repeating the argument we used to justify the limit for rational y_j, we can show that the sequence $\{c(y_j)\}$ has a unique (positive) limit that is independent of the sequence used to reach y. By the uniqueness of the coefficients, we must have $\lim c(y_j) = c(y)$. Hence $c(y)$ is a continuous function.

We claim that $c(x)$ is strictly increasing.

Lemma 2.2.1. *The "coefficient function" $c(x)$ is strictly increasing.*

Proof. [Shorack [Sho00]] We first prove that the integer-indexed coefficients c_n are increasing. It suffices to show that $c_n/c_{n+1} \leq 1$, or equivalently, that $(c_n/c_{n+1})^k$ is bounded above for all n and k, independently of n and k.

From the coefficient relationship, we know that $(c_n)^k = c_{n^k}$, for all n and k. Let $a = n^k$, and let $a+b = (n+1)^k$, so that $(c_n/c_{n+1})^k = c_a/c_{a+b}$. By the definition of stability, we know that

$$X_1 + \cdots + X_a \stackrel{d}{=} c_a X$$
$$X_1 + \cdots + X_{a+b} \stackrel{d}{=} c_{a+b} X.$$

For any positive real number y, consider $P(X > \frac{c_a}{c_{a+b}} y)$. We have

$$\begin{aligned} P(X > \frac{c_a}{c_{a+b}} y) &= P(c_{a+b} X > c_a y) = P(X_1 + \cdots + X_{a+b} > c_a y) \\ &= P(c_a X + c_b X > c_a y) \geq P(X > y) P(X > 0). \end{aligned}$$

By symmetry, $P(X > 0) = 1/2$, so the RHS vanishes if and only if $P(X > y) = 0$, which by symmetry forces X to be degenerate. Hence, we conclude that $P(X > \frac{c_a}{c_{a+b}} y)$ is bounded away from zero for every y, which implies that $\frac{c_a}{c_{a+b}}$ cannot be arbitrarily large and thus, must be bounded above independently of a and b.

Next, we prove the claim for rational numbers. Suppose $m/n \leq p/q$. Then $mq \leq np$, so $c(mq) \leq c(np)$ by the proof for integers. But c is multiplicative, so we have $c(m)/c(n) \leq c(p)/c(q)$, which is equivalent to $c(m/n) \leq c(p/q)$. Thus, the lemma holds for rational numbers.

The extension to real numbers is now immediate—for any real x and y, there are sequences of rationals x_n and y_n respectively that increase to x and y. Since $c(x)$ is continuous and $c(x_n) \leq c(y_n)$ for each n, we conclude that $c(x) \leq c(y)$ for x and y.

To see that $c(x)$ is strictly increasing, suppose $c(x) = c(y)$ but $x \neq y$. Then by the properties of c, $c(x/y) = c(y/x) = 1$. Moreover, $c((x/y)^k) = c((y/x)^k) = 1$ for any integer k. Since c is increasing, though, c is 1 for any z that falls between two powers of x/y and y/x. The powers of x/y and y/x converge to 0 and ∞, respectively, so in fact c is identically 1, which implies that the distribution in question is degenerate. Hence $x = y$. $\square$

We are now in familiar territory—the functional equation $c(xy) = c(x)c(y)$ has only one family of solutions when $c(x)$ is required to be continuous and strictly increasing, namely, $c(x) = x^p$ for some exponent $p > 0$.

Lemma 2.2.2. *All solutions of the functional equation $c(xy) = c(x)c(y)$, $x, y > 0$, subject to the condition that $c(x)$ is continuous and strictly increasing, take the form $c(x) = x^p$ for some exponent $p > 0$.*

Proof. Clearly the function $c(x) = x^p$ satisfies the functional equation for any value of p. To see that these are the only solutions, suppose $\tilde{c}(x)$ is a solution. Clearly $\tilde{c}(1) = 1$, $\tilde{c}(x^n) = \tilde{c}(x)^n$ for any $x > 0$ and any integer n, $\tilde{c}(1/x) = 1/\tilde{c}(x)$ for any $x > 0$, and $\tilde{c}(x/y) = \tilde{c}(x)/\tilde{c}(y)$ for any $x, y > 0$.

$\tilde{c}(x)$ is a strictly positive function (when restricted to $(0,\infty)$), for if it had a zero x_0 in $(0,\infty)$ the functional relation would imply that $\tilde{c}$ is infinite at $1/x_0$, in violation of the assumed continuity of $\tilde{c}$. The logarithm of $\tilde{c}$ is thus well-defined.

Consider the function $f(x) = \log \tilde{c}(x)$. From the properties of $\tilde{c}(x)$ we know that $f(x)$ is a continuous function satifying the relations $f(xy) = f(x) + f(y)$, $f(x^n) = nf(x)$, and $f(1/x) = -f(x)$. We shall show that $f(x) = p \log x$ for some $p > 0$, which will prove the claim.

Consider the difference quotient $(f(x+h) - f(x))/h$. From the properties of f, we have

$$\frac{f(x+h) - f(x)}{h} = \frac{f\left(\frac{x+h}{x}\right)}{h} = \frac{1}{x}\frac{f(1+h/x)}{h/x}.$$

If we take the limit of both sides as $h \to 0$, we see that the derivative $f'(x)$, satisfies

$$f'(x) = \frac{1}{x}\lim_{h/x \to 0}\frac{f(1+h/x)}{h/x} = \frac{1}{x}f'(1),$$

provided both sides exist.

From the relation $f(x^n) = nf(x)$, we can deduce that $f(x)/n = f(x^{1/n})$. This implies that $f'(1)$ can be written as

$$\lim_{k \to 0} f((1+k)^{1/k}).$$

Since f is continuous, we can take the limit inside of f, where we immediately recognize $\lim_{k \to 0}(1+k)^{1/k} = e$. Hence we have established that $f'(1)$ exists and is finite (and in fact, equals $f(e)$). Therefore we know that $f'(x) = f(e)/x$ for all $x > 0$, which means that $f(x) = f(e)\log x$. Hence, $\tilde{c}(x) = x^p$, where $p = \log \tilde{c}(e)$; since $\tilde{c}(x)$ is strictly increasing, we must have $p > 0$ as claimed. $\square$

With this new bit of information, our functional relationship becomes

$$\psi(y^p t) = y\psi(t)$$

where p is still unspecified. If we let $t = 1$ and make the change of variable $x = y^{1/p}$, where $y > 0$, we obtain the representation

$$\psi(x) = \psi(1)x^{1/p}, x > 0.$$

Since ϕ is even, we know that $\psi(-x) = \psi(x)$, hence the complete definition of $\psi(x)$ is

$$\psi(x) = \psi(1)|x|^{1/p}.$$

Hence, the characteristic function $\phi(t)$ of a symmetric stable distribution F has the form

$$\phi(t) = e^{-c|t|^\alpha}, \tag{2.2.5}$$

where $c = -\psi(1) > 0$ is a positive constant and $\alpha = 1/p$.

The exponent α is not 0, for that would result in a degenerate distribution. It cannot be negative, for then $\phi(t)$ would be infinite at the origin (instead of 1). To see that α is at most 2, suppose that $\alpha > 2$, and consider the second derivative of $\phi(t)$ at the origin. Recall that the characteristic function is defined as

$$\phi(t) = \int e^{itx} dF(x).$$

The second derivative of $\phi(t)$ at 0 is given by $-\int x^2 dF(x)$, i.e., it is the negative of the second central moment of F. On the other hand, directly differentiating our expression for $\phi(t)$ shows that $\phi''(0) = 0$ if $\alpha > 2$. Hence, the second moment (and hence, the variance) of F not only exists but is 0. This forces F to be a degenerate distribution, with a constant characteristic function. For $\alpha > 2$ our form of $\phi(t)$ is clearly not constant, so we conclude that $\alpha \leq 2$. (In addition, we see that $e^{-c|t|^\alpha}$ is not the characteristic function of any distribution when $\alpha > 2$.) This finishes the derivation of the form of the characteristic function of a symmetric stable distribution.

In general the characteristic function of a stable distribution takes the form

$$\phi(t) = \exp\left(i\mu t - \sigma^\alpha |t|^\alpha \left[1 - i\beta\, \mathrm{sgn}(t)\omega(t,\alpha)\right]\right), \tag{2.2.6}$$

where

$$\omega(t,\alpha) = \begin{cases} \tan(\frac{\pi}{2}\alpha), & \alpha \neq 1, \\ -\frac{2}{\pi}\log|t|, & \alpha = 1. \end{cases} \tag{2.2.7}$$

Here $\mu \in \mathbb{R}$ is a location parameter, $\sigma > 0$ is a scale parameter, and $\beta \in [-1,1]$ is a "skewness" parameter (not be confused with the coefficient of skewness of a general distribution)[4]. We will denote the corresponding stable density and stable distribution function as $s_{\alpha,\beta}(x;\sigma,\mu)$ and $S_{\alpha,\beta}(x;\sigma,\mu)$, respectively, and we will often use shorthand such as "$X \sim S_{\alpha,\beta}(x;\sigma,\mu)$" to indicate stable random variables.

The derivation of the characteristic function of an asymmetric stable distribution is harder. First, if X is an asymmetric stable random variable, then, letting X_1 and X_2 denote iid copies of X, we can consider the symmetric stable random variable $X_1 - X_2$. If $\phi(t)$ denotes the characteristic function of X, then the characteristic function of $X_1 - X_2$ will be $\phi(t)\overline{\phi(t)}$, which is equal to $|\phi(t)|^2$. Since the resulting random variable is symmetric, we know that its characteristic function has the form given in equation (2.2.5). Hence, we know that

$$|\phi(t)|^2 = e^{-c|t|^\alpha}, \tag{2.2.8}$$

for some $c > 0$ and some $\alpha \in (0,2]$. This fact alone does not get us very far.

The classical approach to proving the representation (2.2.6) uses Lévy's representation of the characteristic function of an infinitely divisible law [Hal81]. This material is more technical than what we have discussed so far; the ambitious reader may consult the text of Gnedenko and Kolmogorov [GK68] or the second volume of Feller's classic text [Fel71]. We now instead discuss some of the interesting properties of stable distributions.

---

[4]The paper of Khintchine and Lévy [KL36] discusses why β must lie in the interval $[-1,1]$

2.2.2 Interpretations of the Parameters

The μ parameter appearing in Equation (2.2.6) is a location parameter: if X is any random variable and r is any real number, the characteristic function of $X+r$ is e^{irt} times the characteristic function of X. It is easy to see that when X has the $S_{\alpha,\beta}(x;\sigma,\mu)$ distribution, this factor readily folds into the $i\mu t$ term of (2.2.6) to give the characteristic function of a random variable with a $S_{\alpha,\beta}(x;\sigma,\mu+r)$ distribution. Thus in our studies of stable distributions we can often restrict our attention to those with $\mu=0$.

The σ parameter behaves like a scale parameter—most of the time. Consider the random variable aX, where $a \neq 0$ is a real number. If $\phi(t)$ is the characteristic function of X, then $\phi(at)$ is the characteristic function of aX. For a stable random variable X, the characteristic function of aX will be

$$\begin{aligned}\phi(at) &= \exp\left(ia\mu t - \sigma^{\alpha}|at|^{\alpha}\left[1 - i\beta\,\mathrm{sgn}(at)\omega(at,\alpha)\right]\right). \\ &= \exp\left(i(a\mu)t - (|a|\sigma)^{\alpha}|t|^{\alpha}\left[1 - i\,\mathrm{sgn}(a)\beta\,\mathrm{sgn}(t)\omega(at,\alpha)\right]\right).\end{aligned}$$

When $\alpha \neq 1$, the resulting characteristic function is that of a $S_{\alpha,\mathrm{sgn}(a)\beta}(x;|a|\sigma,a\mu)$ random variable. Hence, for $a=\frac{1}{\sigma}$, $\frac{1}{\sigma}X$ has a $S_{\alpha,\beta}\left(x;1,\frac{\mu}{\sigma}\right)$ distribution.

When $\alpha=1$, however, the ω function involves t and hence contributes an extra term that folds into the location term $ia\mu t$.

$$\begin{aligned}\phi(at) &= \exp\left(ia\mu t - \sigma|at|\left[1 + i\beta\,\mathrm{sgn}(at)\frac{2}{\pi}\log|at|\right]\right) \\ &= \exp\left(ia\mu t - \sigma|at|\left[1 + i\beta\,\mathrm{sgn}(at)\frac{2}{\pi}(\log|a| + \log|t|)\right]\right).\end{aligned}$$

We recognize the resulting characteristic function as that of a random variable with a $S_{\alpha,\mathrm{sgn}(a)\beta}\left(x;|a|\sigma,a\mu-\frac{2}{\pi}a\sigma\beta\log|a|\right)$ distribution. Therefore, rescaling a stable random variable of index 1 by $(1/\sigma)$ shifts the location parameter nonlinearly [ST94].

The β parameter is known as a "skewness" parameter since it controls the (a)symmetry of the distribution[5]. It can be shown [DuM71] that β satisfies the relation

$$\lim_{x\to\infty}\frac{1-S_{\alpha,\beta}(x)-S_{\alpha,\beta}(-x)}{1-S_{\alpha,\beta}(x)+S_{\alpha,-\beta}(-x)} = \beta. \tag{2.2.9}$$

The numerator in Equation (2.2.9) is the amount by which the upper tail mass exceeds the lower tail mass, while the denominator is the total mass in the tails.

When $\beta=0$, the density of the distribution is symmetric about μ, since in that case the characteristic function of $X-\mu$ is real-valued. (So conversely symmetry about μ forces $\beta=0$.) Furthermore,

1. If $\alpha \in [1,2]$, the support of a stable density is all of $\mathbb{R}$;

2. If $\alpha \in (0,1)$ and $|\beta|=1$, the support of the density is $\mathrm{sgn}(\beta)(0,\infty)$. [ST94, DuM71]

[5] It does not, however, correspond to the statistical concept of skewness.

2.2.3 Properties of Stable Distributions

We claimed above that the constant c_n in the definition of a stable distribution must have the form $n^{1/\alpha}$. We now prove that claim. First, for a symmetric distribution, we know the characteristic function has the form $e^{-c|t|^{\alpha}}$. From the relation (2.2.3), we see than $e^{-nc|t|^{\alpha}} = e^{-c|c_n t|^{\alpha}}$, and hence $n = c_n^{\alpha}$. The result for an asymmetric distribution then follows from relation (2.2.3) and its complex conjugate, together with observation (2.2.8).

Asymmetric stable random variables enjoy a "mirror-image" property that often simplifies numerical computations: if X has the distribution $S_{\alpha,\beta}(x;\sigma,0)$, then $-X$ has the distribution $S_{\alpha,-\beta}(x;\sigma,0)$. This is easy to see from the representation of the characteristic function. The stable distribution function and density thus satisfy

$$S_{\alpha,\beta}(x) = 1 - S_{\alpha,-\beta}(-x) \tag{2.2.10}$$

and

$$s_{\alpha,\beta}(x) = s_{\alpha,-\beta}(-x) \tag{2.2.11}$$

respectively.

The stable density $s_{\alpha,\beta}(x;\sigma,\mu)$ is a very well-behaved function—it is a bounded continous function that is analytic throughout its support.

Proposition 2.2.1. *The density of a stable random variable is a bounded continous function that is analytic at all points of its support.*

Proof. First, to see that the density even exists, recall that by the Fourier inversion theorem, the density of a random variable can be expressed by

$$f(x) = \frac{1}{2\pi}\int e^{-itx}\phi(t)\,dt$$

if its characteristic function $\phi(t)$ is (L^1) integrable [Bre68]. For a stable random variable, $|\phi(t)| = \exp(-\sigma^{\alpha}|t|^{\alpha})$, which is integrable, so the density exists. Boundedness is immediate, since $|s_{\alpha,\beta}(x;\sigma,\mu)|$ is bounded by the L^1 norm of the integrand. Continuity follows from the Lebesgue dominated convergence theorem [Fol99]. Integrals of the derivatives of the integrand with respect to x are easily seen to be bounded by

$$\int |t|^n \exp(-\sigma^{\alpha}|t|^{\alpha})\,dt = \frac{\Gamma\left(\frac{n+1}{\alpha}\right)}{\pi\alpha\sigma^n}.$$

Hence, we may justifiably take derivatives of $s_{\alpha,\beta}(x;\sigma,\mu)$ under the integral sign, so all derivatives of $s_{\alpha,\beta}(x;\sigma,\mu)$ exist and are continuous.

To see that the density is actually analytic, and not just C^{∞}, we need to show that the Taylor series development of $s_{\alpha,\beta}(x;\sigma,\mu)$ converges to $s_{\alpha,\beta}(x;\sigma,\mu)$ everywhere it is defined. The coefficient of the n-th term of the Taylor expansion is bounded above by

$$\frac{\Gamma\left(\frac{n+1}{\alpha}\right)}{n!\pi\alpha\sigma^n} = \frac{1}{\pi\alpha\sigma^n}\frac{\Gamma\left(\frac{n+1}{\alpha}\right)}{\Gamma(n+1)}.$$

For $\alpha \geq 1$, the n-th root of this term is

$$\frac{1}{\sigma\sqrt[n]{\pi\alpha}}\sqrt[n]{\frac{\Gamma\left(\frac{n+1}{\alpha}\right)}{\Gamma(n+1)}}$$

For $\alpha = 1$, the above quantity converges to $1/\sigma$ as $n \to \infty$, so by Hadamard's formula [Alh79] the Taylor series has radius of convergence $1/\sigma > 0$ at all x. For $\alpha > 1$, an argument using Stirling's formula shows that the above quantity converges to 0 as $n \to \infty$, so the radius of convergence is infinite [GK68].

The above proof of analyticity breaks down for $\alpha < 1$. An alternative proof is given in [Zol86]. □

Unfortunately, the stable density and distribution function do not have closed-form representations in general. In the next section we will say more about this.

The stable density is also unimodal [IC59] and has all absolute moments of orders $\leq \alpha$ [DuM71]. Hence when $\alpha > 1$, the mean of the distribution exists—it just happens to be μ—but the variance is infinite unless $\alpha = 2$.

The tails of a stable density (with $\alpha \neq 2$) are Pareto-like, in the sense that,

$$\lim_{x\to\infty} x^{\alpha}\left(1 - S_{\alpha,\beta}(x;\sigma,\mu)\right) = C_{\alpha}\frac{1+\beta}{2}\sigma^{\alpha}$$
$$\lim_{x\to\infty} x^{\alpha}\left(S_{\alpha,\beta}(-x;\sigma,\mu)\right) = C_{\alpha}\frac{1-\beta}{2}\sigma^{\alpha}$$

for some positive constant C_{α}, the form of which is not important here[6][ST94, DuM71].

Finally, we mention a result on the distribution of the sum of a finite number of stable random variables with the same index.

Proposition 2.2.2. *If $X_1,\ldots,X_n$ are independent random variables with $X_k \sim S_{\alpha,\beta_k}(x;\sigma_k,\mu_k)$, then for any $a_k > 0$, the weighted sum $X = \sum_k a_k X_k$ is distributed as $S_{\alpha,\beta}(x;\sigma,\mu)$, where*

$$\beta = \frac{\sum a_k^{\alpha}\sigma_k^{\alpha}\beta_k}{\sum a_k^{\alpha}\sigma_k^{\alpha}}, \quad \sigma^{\alpha} = \left(\sum a_k^{\alpha}\sigma_k^{\alpha}\right), \quad \textit{and} \quad \mu = \sum a_k\mu_k.$$

Proof. By independence, the characteristic function of the sum is the product of the characteristic functions.

$$E\exp\left(i\left(a_1X_1+\cdots+a_nX_n\right)t\right) = E\exp\left(iX_1(a_1t)\right)\cdots E\exp\left(iX_n(a_nt)\right).$$

From the canonical form of the characteristic function (2.2.6), we see that the right hand side of the above equation is

$$\exp\left(i(a_1\mu_1+\cdots+a_n\mu_n)t - |t|^{\alpha}\sum_{k=1}^{n}a_k^{\alpha}\sigma_k^{\alpha}\left[1 - i\beta_k\operatorname{sgn}(t)\omega(t,\alpha)\right]\right).$$

Grouping terms now proves the proposition. □

[6] Explicitly, the constant C_{α} is defined by

$$C_{\alpha}^{-1} = \int_0^{\infty} x^{-\alpha}\sin x\,dx.$$

A closed-form formula is given in [ST94].

2.2.4 Closed-Form Formulas for Stable Distributions

As we stated in the last section, the inverse Fourier transform of the characteristic function of a stable distribution cannot, in general, be expressed in closed-form. There are some special cases for which simple expressions are available. First, for $\alpha = 2$, $\omega(t,2) = 0$, so the characteristic function reduces to

$$\phi(t) = \exp(i\mu t - \sigma^2 t^2).$$

This is the characteristic function of a normal distribution with mean μ and variance $2\sigma^2$. (Take note of the extra factor of 2.) The β parameter here does not come into play; typically we just use $\beta = 0$ for the normal distribution to be consistent (since it's a symmetric distribution).

Next, when $\alpha = 1$ and $\beta = 0$ (so that the logarithm term isn't present), the characteristic function is

$$\phi(t) = \exp(i\mu t - \sigma|t|).$$

The inverse Fourier transform of this function is precisely the Cauchy density (with location parameter μ and scale parameter σ.)

There are two additional closed-forms available. When $\alpha = 1/2$ and $\beta = 1$ we obtain the density of the so-called Lévy distribution:

$$s_{1/2,1}(x;\sigma,\mu) = \sqrt{\frac{\sigma}{2\pi}} \frac{1}{(x-\mu)^{3/2}} \exp\left(-\frac{\sigma}{2(x-\mu)}\right).$$

This distribution (which also happens to be the distribution of $1/Z^2$ where $Z \sim N(0,1)$) pops up in random walk theory. The other closed-form distribution is the "mirror-image" of this one, i.e., the distribution with $\alpha = 1/2$ and $\beta = -1$.

While other stable densities do not have nice closed-form formulas, some are representable in terms of special functions. Zolotarev [Zol86] derives representations of stable densities with certain rational values of α in terms of special functions (Whittaker functions) via differential equations. These representations are too complex, however, for day-to-day work, and too specific to be of general computational use.

2.2.5 Plots of Stable Densities and Distributions

Some plots of stable densities and distribution functions for various values of α and β are shown in Figures 2.1-2.6.

2.2.6 Other Parameterizations

There are many ways of expressing the (log) characteristic function of a stable distribution, and this has led to considerable confusion in the literature. (In fact, Hall [Hal81] published a paper solely on this matter.) It also makes the use of tables and numerical routines for stable distributions tricky, as one has to transform the desired parameters to the form used by that particular piece of code.[7]

[7] For instance, Holt and Crow [HC73] flip the sign of β for $\alpha \neq 1$.

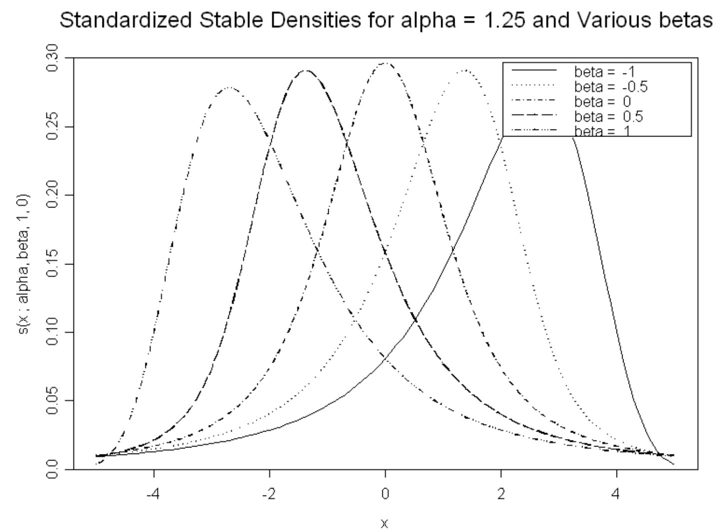

Figure 2.1: Stable PDF's for $\alpha = 1.25$ and various values of β.

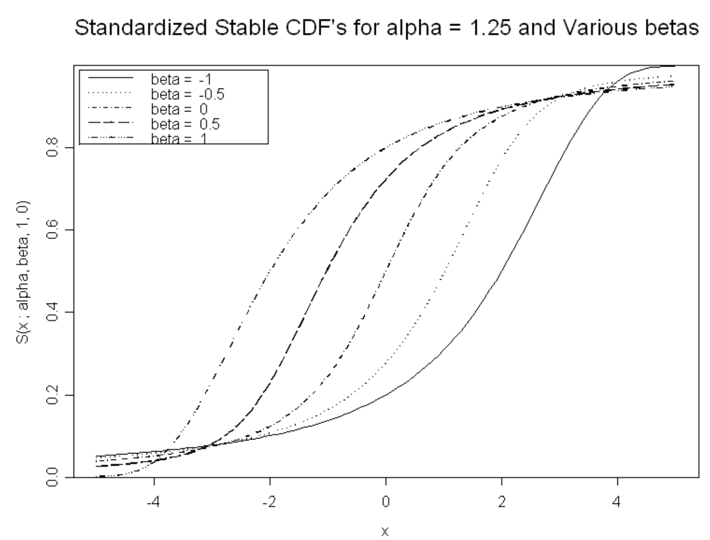

Figure 2.2: Stable CDF's for $\alpha = 1.25$ and various values of β.

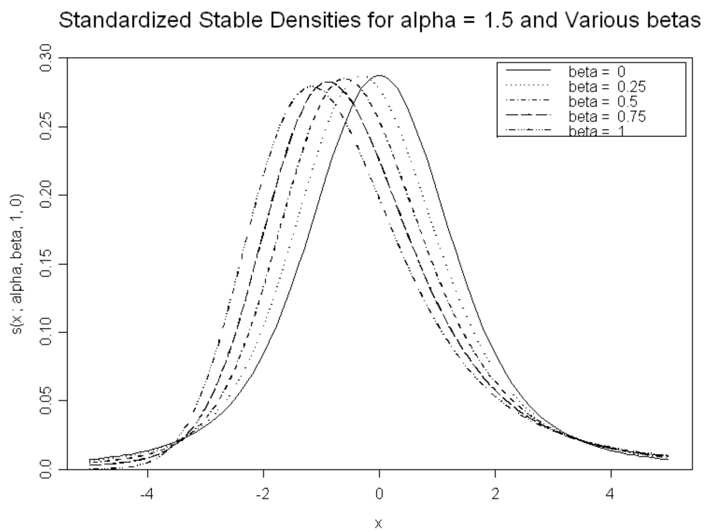

Figure 2.3: Stable PDF's for $\alpha = 1.5$ and various values of β.

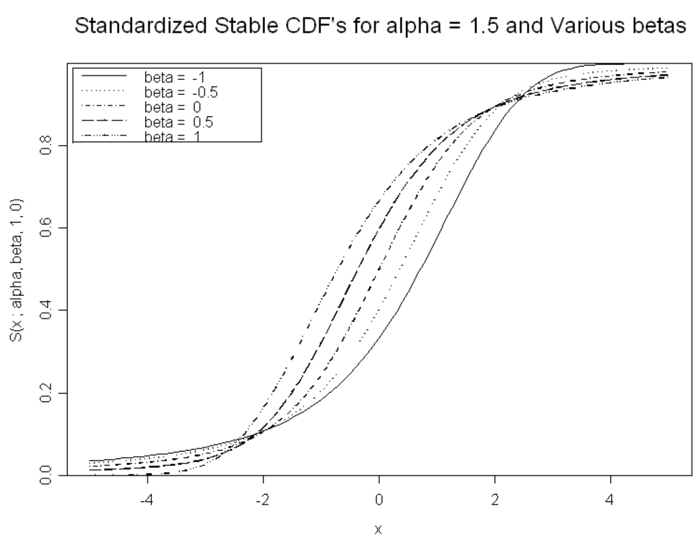

Figure 2.4: Stable CDF's for $\alpha = 1.50$ and various values of β.

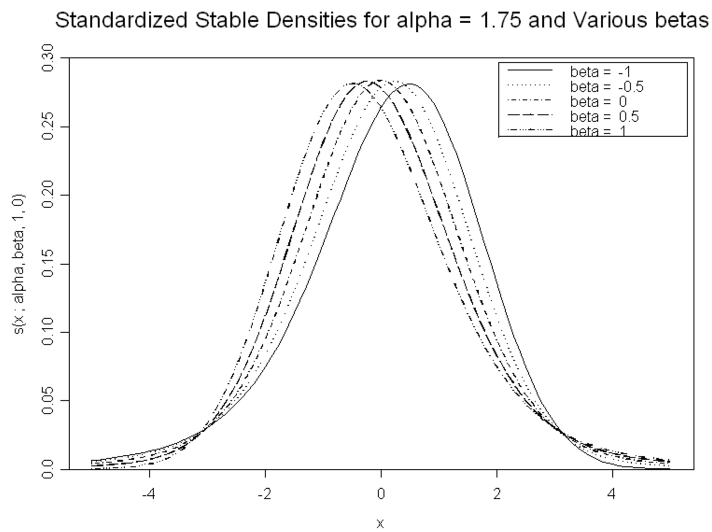

Figure 2.5: Stable PDF's for $\alpha = 1.75$ and various values of β.

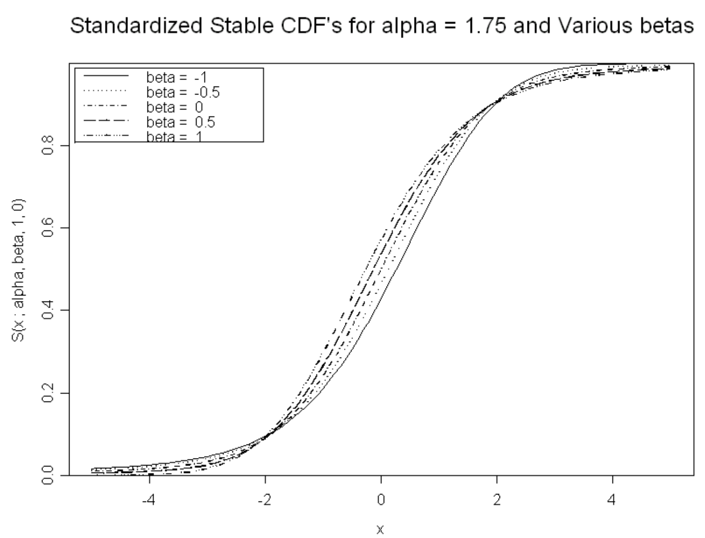

Figure 2.6: Stable CDF's for $\alpha = 1.75$ and various values of β.

The most common reason to change the parameterization is the discontinuity in the representation (2.2.6) at $\alpha = 1$: the ω function is not continuous at $\alpha = 1$. Zolotarev [Zol86] noted that this can be fixed by defining a "shifted" location parameter $\zeta = \mu - \beta \tan(\frac{\pi}{2}\alpha)$. The resulting characteristic function is then

$$\phi(t) = \exp\left(i\zeta t - \sigma^\alpha \left(|t|^\alpha + i\beta t \omega_M(t,\alpha)\right)\right), \tag{2.2.12}$$

where

$$\omega_M(t,\alpha) = \begin{cases} (|t|^{\alpha-1} - 1)\tan(\frac{\pi}{2}\alpha), & \alpha \neq 1, \\ -\frac{2}{\pi}\log|t|, & \alpha = 1. \end{cases} \tag{2.2.13}$$

(In Zolotarev's treatise [Zol86] this is known as the "M" parameterization.) If we use this parameterization and define the characteristic function for $\alpha = 1$ by continuity, we can actually get the stable distributions to convergence in distribution as $\alpha \to \alpha_0$, $\beta \to \beta_0$, $\sigma \to \sigma_0$, $\zeta \to \zeta_0$ [ST94].

It is sometimes useful to put the log-characteristic function for $\alpha \neq 1$ into "polar" form. Define β_2 by the relation

$$\beta \tan(\frac{\pi}{2}\alpha) = \tan\left(\frac{\pi}{2}\beta_2 K(\alpha)\right),$$

where $K(\alpha) = 1 - |1-\alpha| = \min(\alpha, 2-\alpha)$, and σ_2 by the relation

$$\sigma_2 = \sigma\left(\cos\beta_2 \frac{\pi}{2} K(\alpha)\right)^{1/\alpha}.$$

Then the log of the characteristic function takes the form [ST94]

$$-\sigma_2^\alpha |t|^\alpha e^{-i\beta_2 \operatorname{sgn}(t)\frac{\pi}{2}K(\alpha) + it\mu}. \tag{2.2.14}$$

(In Zolotarev's treatise [Zol86] this is known as the "B" parameterization, albeit with a slightly different definition of $K(\alpha)$.)

There are additional parameterizations out there, but we will not have any use for them here. The interested reader may consult [Zol86] to learn more.

2.3 A Brief Review of Numerical Techniques for Stable Distributions

2.3.1 Computation of the Density

There are two main techniques at this time: asymptotic series and numerical evaluation of the inverse Fourier transform of the characteristic function. The latter can be split into FFT-based integrators and more direct integrators.

Asymptotic Methods

Bergström [Ber52, DuM71] derived several infinite series expansions for $s_{\alpha,\beta}(x;1,0)$. These expressions are stated in the "polar" parameterization discussed in Section 2.2.6. If we denote the stable density in the polar parameterization as $s^*_{\alpha,\beta^*}(x;1,0)$, then the stable density in the canonical parameterization can be expressed as

$$s_{\alpha,\beta}(x;1,0) = c^* s^*_{\alpha,\beta^*}(c^* x;1,0),$$

where $c^* = [1+\beta^2(\tan(\frac{\pi}{2}\alpha))^2]^{-1/2\alpha}$ [DuM71].

Bergström's results state that

$$s^*_{\alpha,\beta^*}(x;1,0) = \frac{1}{\pi}\sum_{k=1}^{\infty}\frac{\Gamma(1+k/\alpha)}{k!}(-x)^{k-1}\sin\left[\frac{k\pi}{2\alpha}(\alpha+\beta^*\min(\alpha,2-\alpha))\right], \quad (2.3.15)$$

and

$$s^*_{\alpha,\beta^*}(x;1,0) = \frac{1}{\pi x^{1+\alpha}}\sum_{k=1}^{\infty}\frac{\Gamma(1+k\alpha)}{k!}(-x^{-\alpha})^{k-1}\sin\left[\frac{k\pi}{2}(\alpha+\beta^*\min(\alpha,2-\alpha))\right]. \quad (2.3.16)$$

The first series is absolutely convergent for all x when $\alpha \in (1,2]$; for $\alpha \in (0,1)$ it is only valid asymptotically as $x \to 0$. The second series is absolutely convergent for $x > 0$ when $\alpha \in (0,1)$ and valid asymptotically as $x \to \infty$ for $\alpha \in (1,2)$ [DuM71].

For the always problematic $\alpha = 1$ case, neither series converges unless $\beta = 0$. In that case, the first series converges for $|x| < 1$ while the second converges for $|x| > 1$. For $\alpha = 2$, the first series is just a Taylor series for the density of a normal distribution with mean 0 and variance 2. The second series vanishes entirely when $\alpha = 2$, and $\alpha \in (0,1)$ and $\beta^* = -1$, or $\alpha \in (1,2)$ and $\beta^* = 1$ due to the extremely fast decay of the right (resp., left) tails of the distributions with these parameters [DuM71].

Methods based on Numerical Integration

These methods all utilise the representation of the density as the inverse Fourier transform of the characteristic function:

$$f(x) = \frac{1}{2\pi}\int e^{-ixt}\phi(t)\,dt. \quad (2.3.17)$$

The integrand is not especially pleasant to deal with, and this has limited the number of practical implemenations of stable models.

The two majors approaches to numerical integration are the direct approach, in which the integral (or some fixed-up version of it) is computed using standard quadrature techniques, and the FFT approach, in which the integrand is made to look like the discrete Fourier transform of some sequence. Typically, these approaches are further combined with asymptotic methods to provide more accuracy for values deep in the tails of the distributions.

Direct Numerical Integration Methods Standard quadrature methods, when used with the canonical parameterization of the characteristic function (Equation (2.2.6)), have difficulty with Equation (2.3.17) when α is near 1. Nolan's approach [Nol97] instead uses Zolotarev's "M" parameterization of a standardized stable distribution (see Section 2.2.6). He applies an adaptive quadrature method to an integral representation for the stable density due to Zolotarev [Zol86]. To express these formulas in the "M" parameterization, he introduces the following notation.

$$\zeta = \begin{cases} -\beta \tan\left(\frac{\pi}{2}\alpha\right), & \alpha \neq 1 \\ 0, & \alpha = 1 \end{cases} \tag{2.3.18}$$

$$\theta_0 = \begin{cases} \frac{1}{\alpha} \arctan(\beta \tan\left(\frac{\pi}{2}\alpha\right)), & \alpha \neq 1 \\ \frac{\pi}{2}, & \alpha = 1 \end{cases} \tag{2.3.19}$$

$$c_1(\alpha,\beta) = \begin{cases} \frac{1}{\pi}\left(\frac{\pi}{2} - \theta_0\right), & \alpha < 1, \\ 0, & \alpha = 1, \\ 1, & \alpha > 1, \end{cases} \tag{2.3.20}$$

$$V(\theta;\alpha,\beta) = \begin{cases} (\cos\alpha\theta_0)^{1/\alpha}\left(\frac{\cos\theta}{\sin\alpha(\theta_0+\theta)}\right)^{\frac{\alpha}{\alpha-1}} \frac{\cos(\alpha\theta_0+(\alpha-1)\theta)}{\cos\theta}, & \alpha \neq 1 \\ \frac{2}{\pi}\left(\frac{\frac{\pi}{2}+\beta\theta}{\cos\theta}\right)\exp\left(\frac{1}{\beta}\left(\frac{\pi}{2}+\beta\theta\right)\tan\theta\right), & \alpha = 1, \beta \neq 0. \end{cases} \tag{2.3.21}$$

The integral representations for the density are then defined as follows [Nol97].

$$s_{\alpha,\beta}(x;1,0) = \begin{cases} \frac{\alpha(x-\zeta)^{\frac{1}{\alpha-1}}}{\pi|\alpha-1|}\int_{\theta_0}^{\frac{\pi}{2}} V(\theta;\alpha,\beta)\exp(-(x-\zeta)^{\frac{\alpha}{\alpha-1}}V(\theta;\alpha,\beta)\,d\theta, & x > \zeta \\ \frac{\Gamma(1+\frac{1}{\alpha})\cos\theta_0}{\pi(1+\zeta^2)^{1/(2\alpha)}}, & x = \zeta \\ s_{\alpha,-\beta}(-x;1,0), & x < \zeta \end{cases} \tag{2.3.22}$$

when $\alpha \neq 1$; and

$$s_{1,\beta}(x;1,0) = \begin{cases} \frac{1}{2|\beta|}e^{-\frac{\pi}{2}\frac{x}{\beta}}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} V(\theta;1,\beta)\exp\left(-e^{-\frac{\pi}{2}\frac{x}{\beta}}V(\theta;1,\beta)\right), & \beta \neq 0 \\ \frac{1}{\pi(1+x^2)}, & \beta = 0. \end{cases} \tag{2.3.23}$$

The integrand starts at 0 when $\theta = \theta_0$, increases monotonically to a maximum ($1/e$), then decreases monotonically back to 0 at $\theta = \frac{\pi}{2}$ [Buc95, Nol97]. For some combinations of (α,β) (e.g., when α is near 1, when β is near 1) the mass of the integrand can be highly concentrated, and thus, easily missed by the quadrature routine. To remedy this, Nolan finds the location θ_2 of the maximum (via a standard root-finding algorithm), then computes the integrals in two pieces. This approach also works for the $\alpha = 1$ case.

Fast Fourier Transform Methods When the discretization of the integral is done a certain way, it can be computed faster using the Fast Fourier Transform (FFT) than using standard quadrature methods [Che91, DM98, MDC99, RM00]. For simplicity, let us assume the characteristic function is in the canonical form (given in Equation (2.2.6)). After standardization (subtracting μ and dividing by σ), the stable density can be expressed as

$$s_{\alpha,\beta}(x;1,0) = \int e^{-2\pi i x\omega} \phi(2\pi\omega;\alpha,\beta,1,0)\, d\omega, \tag{2.3.24}$$

where the standardized characteristic function is

$$\phi(t;\alpha,\beta,1,0) = \exp\left(-|t|^{\alpha} + i\beta t|t|^{\alpha-1}\tan(\frac{\pi}{2}\alpha)\right). \tag{2.3.25}$$

We will evaluate the integral in Equation (2.3.24) using N equally-spaced x points, centered at 0, with mesh size h:

$$x_k = \left(k-1-\frac{N}{2}\right)h, \quad k = 1,\ldots,N. \tag{2.3.26}$$

The density can be computed for other values inside the grid via interpolation. For values outside the grid we can use Bergström's asymptotic method.

The resulting integrals can be discretized using the "right-hand" rule for N equally-spaced points

$$\omega_n = \left(n-1-\frac{N}{2}\right)s \tag{2.3.27}$$

with mesh size s.

$$s_{\alpha,\beta}(x_k;1,0) \approx s\sum_{n=1}^{N} e^{-2\pi i x_k\omega_n}\phi(2\pi\omega_n;\alpha,\beta,1,0). \tag{2.3.28}$$

For the choice $s = (hN)^{-1}$, some algebra reduces the approximating sum to

$$s_{\alpha,\beta}(x_k;1,0) \approx s(-1)^{k-1-\frac{N}{2}} \sum_{n=1}^{N} (-1)^{n-1} e^{-\frac{2\pi i(n-1)(k-1)}{N}} \phi(2\pi\omega_n;\alpha,\beta,1,0). \tag{2.3.29}$$

The sum is precisely the discrete Fourier transform of the sequence

$$(-1)^{n-1}\phi(2\pi\omega_n), \quad n = 1,\ldots,N, \tag{2.3.30}$$

so it can be computed in $\Omega(N\log N)$ time using the FFT. After multiplying each of the resulting values by the quantity $s(-1)^{k-1-\frac{N}{2}}$, we obtain the values of the (standardized) stable PDF at each x_k.

The above procedure will obviously have trouble near the discontinuity at $\alpha = 1$. In theory, switching to the "M" parameterization will fix this, since it removes the discontinuity in the characteristic function. Numerically, however, this is not the case: the characteristic function contains a term of the form $(|t|^{\alpha}-1)\tan\left(\frac{\pi}{2}\alpha\right)$, which is ill-behaved near $\alpha = 1$. A Taylor series about $\alpha = 1$ can be used to remedy the problem; such a series is tedious to calculate, though, and (in what limited experiments we have perfomed) seems to require too many terms of be of practical value. We are presently working on a better solution to this problem. (We also limited many of our later experiments to $\alpha \geq 1.25$ for this reason.)

Implementation

We have implemented Nolan's method in the S-PLUS™ function `dstable.int`. The FFT method, along with Bergström's approximations for the tails, is implemented in the S-PLUS™ function `dstable.fft`. Both methods have been verified against the tabulation of Holt and Crow [HC73].

2.3.2 Computation of the Distribution Function

Computation of the stable distribution function follows roughly the same paths as computation of the density function. One can integrate Berström's asymptotic series for the density termwise. One can also use the computed density values to approximate the distribution via quadrature.

Nolan [Nol97] applies quadrature methods to Zolotarev's integral representations for the distribution functions (expressed in the "M" parameterization).

$$S_{\alpha,\beta}(x;1,0) = \begin{cases} c_1(\alpha,\beta) + \frac{\operatorname{sgn}(1-\alpha)}{\pi}\int_{\theta_0}^{\frac{\pi}{2}} \exp(-(x-\zeta)^{\frac{\alpha}{\alpha-1}} V(\theta;\alpha,\beta)\, d\theta, & x > \zeta \\ \frac{1}{\pi}\left(\frac{\pi}{2} - \theta_0\right), & x = \zeta \\ 1 - S_{\alpha,-\beta}(-x;1,0), & x < \zeta \end{cases} \tag{2.3.31}$$

when $\alpha \neq 1$; and

$$S_{1,\beta}(x;1,0) = \begin{cases} \frac{1}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \exp\left(-e^{-\frac{\pi}{2}\frac{x}{\beta}} V(\theta;1,\beta)\right), & \beta > 0 \\ \frac{1}{2} + \frac{1}{\pi}\arctan x, & \beta = 0 \\ 1 - S_{1,-\beta}(-x;1,0), & \beta < 0 \end{cases} \tag{2.3.32}$$

In this case the integrands are well-behaved functions, so it is not necessary to split the domain of integration.

Implementation

We have implemented Nolan's method in the S-PLUS™ function `pstable.int`. Another S-PLUS™ function, `pstable.fft`, uses the density values computed by `dstable.fft` along with quadrature to approximate the CDF.

2.3.3 Computation of Stable Quantiles

Quantiles for stable distributions are usually computed via one of two ways: inverse interpolation of the distribution function $S_{\alpha,\beta}(x;\sigma,\mu)$ or solution of the equation $S_{\alpha,\beta}(x;\sigma,\mu) = p$ using Newton's method (or other similar techniques). For the special cases of the Gaussian, the Cauchy, and the Lévy distributions, there are more straightforward algorithms.

When the distribution function is computed using the FFT, values of the CDF are computed over a regularly spaced grid (on the characteristic function side), so inverse interpolation is the easiest method to use. When a direct integration method is used to compute the density and distribution functions, it is usually more straightforward to find quantiles use Newton's method. Initial values can be found using the FFT method on a coarse grid.

Implementation

We have implemented both methods in S-PLUS™ as the functions `qstable.fft` and `qstable.int`, respectively.

For values of $p < 0.5$, the computation is often more difficult to do directly when the distribution is skewed to the right, so we use the following relation. Suppose $X \sim S_{\alpha,\beta}(x;1,0)$; then using property 2.2.10 we have

$$\begin{aligned} p &= P(X \le x_p) \\ 1-p &= 1 - P(X \le x_p) = \int_{x_p}^{\infty} s_{\alpha,\beta}(x;1,0)\,dx \\ &= \int_{-\infty}^{-x_p} s_{\alpha,\beta}(-y;1,0)\,dy = \int_{-\infty}^{-x_p} s_{\alpha,-\beta}(y;1,0)\,dy \\ p &= 1 - S_{\alpha,-\beta}(-x_p;1,0). \end{aligned}$$

2.3.4 Generation of Stable Random Numbers

The problem of generating random numbers from a stable distribution was first approached using the probability integral tranformation—independent random numbers were generated from a uniform distribution, then transformed using an approximation to the distribution function $S_{\alpha,\beta}(x;\sigma,\mu)$ [FR68, DuM71, PHL75]. The amount of effort needed to compute even an approximation to the distribution function for a stable random variable made this approach impractical for small samples.

Chambers, Mallows, and Stuck [CMS76] derived a faster algorithm for random numbers by deducing a representation of a stable random variable as a (complicated) function of a uniform random variable and an exponential variable. Using an integral representation for the distribution function $S_{\alpha,\beta}(x;1,0)$ derived by Zolotarev [Zol66], it can be shown that

$$\begin{aligned} S_{\alpha,\beta}(x;1,0) &= \frac{\sin\alpha(\Phi-\Phi_0)}{(\cos\Phi)^{1/\alpha}}\left(\frac{\cos(\Phi-\alpha(\Phi-\Phi_0))}{W}\right)^{(1-\alpha)/(\alpha)}, \quad \alpha \neq 1, \\ S_{1,\beta}(x;1,0) &= \frac{2}{\pi}\left(\frac{1}{2}\pi + \beta\Phi\tan(\Phi) - \beta\ln\left(\frac{\frac{\pi}{2}W\cos\Phi}{\frac{\pi}{2}+\beta\Phi}\right)\right), \end{aligned}$$

where W is an Exponential(1) variable, Φ is Uniform on $(-\frac{\pi}{2},\frac{\pi}{2})$, and the constant Φ_0 equals $-\frac{\pi}{2}\beta\frac{1-|1-\alpha|}{\alpha}$. This representation suffers from the usual discontinuity at $\alpha = 1$ and an additional discontinuity at $\beta = 1$, so instead they use a modified distribution, $S'_{\alpha,\beta}(x;1,0)$, defined as

$$S'_{\alpha,\beta}(x;1,0) = \tan\alpha\Phi_0 + (\cos\alpha\Phi_0)^{-1/\alpha}S_{\alpha,\beta}(x;1,0),$$

and a modified skewness β', defined as

$$\beta'(\alpha,\beta) = -\tan(\frac{\pi}{2}(1-\alpha))\tan(\alpha\Phi_0), \quad \alpha \neq 1$$
$$\beta'(1,\beta) = \beta.$$

The modified distribution is continuous near $\alpha = 1$ and has the correct limiting behavior as $\beta \to 1$. To compute the distribution accurately, they define $\varepsilon = 1-\alpha$ and

$$z = \frac{\cos\varepsilon\Phi - \tan\alpha\Phi_0 \sin\varepsilon\Phi}{W\cos\Phi},$$

then express $S'_{\alpha,\beta}(x;1,0)$ as

$$S'_{\alpha,\beta'}(x;1,0) = \left(\frac{\sin\alpha\Phi}{\cos\Phi} - \tan\alpha\Phi\left(\frac{\cos\alpha\Phi}{\cos\Phi} - 1\right)\right)z^{\varepsilon/(1-\varepsilon)} + \tan\alpha\Phi_0\left(1 - z^{\varepsilon/(1-\varepsilon)}\right).$$

Implementation

In their paper Chambers, Mallows, and Stuck provide a FORTRAN implementation of their method, an updated version of which is included in S-PLUS™ as the function `rstab`. Since `rstab` expects a transformed skewness parameter β' as input, we have written a wrapper function, `rstable`, that permits the use of the standard β. Our wrapper function also supports $\sigma \neq 1$ and $\mu \neq 0$ (by rescaling and translation of the output of `rstab`).

2.4 Estimation of Stable Parameters

The usefulness of stable distributions would be quite limited if there were not means of estimating the parameters of a stable distribution from a sample. Numerous estimators have been proposed for use with stable distributions, but to date no single estimator has been widely accepted. This is largely due to the same factors that make numerical computation with stable distributions difficult—the lack of closed forms, the confusing plurarity of parameterizations, etc.

In this section we will examine some of the more popular methods for estimating the parameters of a stable distribution. We will also examine a new estimator derived from using quantile-quantile plots to choose the best parameters.

2.4.1 Estimation via Maximum Likelihood

The first estimation method that springs to the mind of the statistician is maximum likelihood estimation. Since the density of a stable distribution is generally not available in closed form, we immediately hit a snag. In the last chapter, though, we discussed some numerical methods for computing the density of a stable distribution fairly accurately, so we are not entirely sunk.

We can use the representation of the density of a stable distribution as the inverse Fourier transform of its characteristic function to build a likelihood function: given a sample $X_1, \ldots, X_n$ presumed to be from a stable distribution $s_{\alpha,\beta}(x;\sigma,\mu)$, the likelihood of the model can be expressed as

$$L(\theta|X_1,\ldots X_n) = \prod_{k=1}^{n} f(x_k) = \prod_{k=1}^{n} \int e^{-ix_k t} \phi(t;\alpha,\beta,\sigma,\mu)\,dt,$$

where $\theta = (\alpha,\beta,\sigma,\mu)$. The log-likelihood is then given by

$$\log L(\theta|X_1,\ldots X_n) = \sum_{k=1}^{n} \log \int e^{-ix_k t} \phi(t;\alpha,\beta,\sigma,\mu)\,dt.$$

In theory, we can solve the equation

$$\log L(\hat{\theta}|X_1,\ldots,X_n) = 0$$

for $\hat{\theta}$, the MLE of the parameters $\theta = (\alpha,\beta,\sigma,\mu)$. In practice, however, the optimization is complicated by fact that exact derivatives are just as difficult to obtain numerically as the densities. Derivatives can be approximated using finite differences, but this is just as expensive.

In his dissertation (and subsequent papers) DuMouchel demonstrated that the MLE is a consistent estimator with an asymptotic normal distribution [DuM71, DuM73a, DuM73b, DuM75].

2.4.2 Quantile-Based Methods

The quantile-based methods for estimating the parameters of a stable distribution have the advantages of being fast and easy-to-use, but are often ill-suited for financial work, as they ignore data values deep in the tails of the empirical distribution of the sample [RM00].

The method of McCulloch

McCulloch [McC86] proposed a modification to the method of Fama and Roll [FR71]. The parameters of the stable distribution $S_{\alpha,\beta}(x;\sigma,\mu)$ are estimated using five predetermined sample quantiles. Unlike Fama and Roll's method, it works for $\alpha \in [0.6, 2.0]$ and $\beta \in [-1,1]$, and all its estimators are consistent.

The method assumes the standard parameterization for a stable distribution (see Equation (2.2.6)). McCulloch defines the population quantities

$$\upsilon_\alpha = \frac{x_{.95}-x_{.05}}{x_{.75}-x_{.25}}, \upsilon_\beta = \frac{x_{.95}+x_{.05}-2x_{.5}}{x_{.95}-x_{.05}}, \upsilon_\sigma = \frac{x_{.75}+x_{.25}}{\sigma}, \upsilon_\zeta = \frac{\zeta - x_{.5}}{\sigma},$$

where x_q is the q-th quantile of a stable distribution, and ζ is the shifted location parameter:

$$\zeta = \begin{cases} \mu + \beta\sigma\tan(\frac{\pi}{2}\alpha), & \alpha \neq 1 \\ \mu, & \alpha = 1. \end{cases}$$

The first two quantities υ_α and υ_β are independent of the location and scale parameters, so they can be tabulated as functions $\phi_1(\alpha,\beta)$ and $\phi_2(\alpha,\beta)$ of α and β. The sample versions of υ_α and υ_β are then computed by substituting sample quantiles (corrected for continuity) for population quantiles. The sample versions are consistent estimators of the population versions (since the sample quantiles are consistent estimators of the population quantiles and υ_α and υ_β are continuous functions of the quantiles). Inversion of the ϕ_1 and ϕ_2 functions (numerically) yields consistent estimators of α and β.

The estimator of σ is obtained in similar fashion using υ_σ. μ, however, cannot be directly estimated using an analogous quantity due to a singularity in the resulting ϕ function. (The singularity arises from the discontinuity in the canonical parameterization at $\alpha = 1$.) Zolotarev's shifted location parameter ζ, however, does not suffer from this problem, so McCulloch estimates it by inverting υ_ζ, then unshifting.

2.4.3 Methods based on Fitting the Characteristic Function

Koutrouvelis [Kou80, Kou81] proposed fitting the characteristic function to the observations as a means of estimating the parameters of a stable distribution. Here the characteristic function is expressed in its canonical form (Equation (2.2.6)). Koutrouvelis considers the logarithm $\psi(t)$ of the characteristic function

$$\psi(t) = \log\phi(t) = -|\sigma t|^\alpha + i\left(\mu t + |\sigma t|^\alpha \beta w(t,\alpha,\sigma)\right).$$

The logarithm of the (negative of the) real part of this expression yields a linear relation in which the unknown index α is the slope, and σ is a shift:

$$\log(-\Re(\psi(t))) = \alpha(\log|t| + \log\sigma).$$

Likewise, the imaginary part of the expression yields a nonlinear relationship with β and μ as coefficients:

$$\Im(\psi(t)) = \mu t + |\sigma t|^\alpha \beta \operatorname{sgn}(t)\tan(\frac{\pi}{2}\alpha).$$

Here α and σ are assumed to have been determined from Equation (2.4.3).

The location and scale parameters μ and σ have the largest effect on the estimation [PHL75], so Koutrouvelis estimates them first using the method of Fama and Roll [FR71], then standardizes the observations using these initial estimates. Next, he forms the sample characteristic function

$$\phi(\hat{t_k}) = \frac{1}{N}\sum_{n=1}^{N}\exp(it_k x_n)$$

at predetermined points t_k, $k = 1,\ldots,K$ using the standardized observations x_n. The linear relation specified in Equation (2.4.3) is then fit using generalized least squares (GLS) to determine the estimate $\hat{\alpha}$ of the index α and an updated estimate $\hat{\sigma}_1$ of the scale.

To find β and μ, Koutrouvelis plugs the estimates of α and σ into Equation (2.4.3). The sample characteristic function is once again computed at predetermined points t_l, $l = 1,\ldots L$, which has the nice side-effect of linearizing Equation (2.4.3). The estimate $\hat{\beta}$ of the symmetry parameter β and the updated estimate $\hat{\mu}_1$ of the location parameter μ can then be found by GLS.

Finally, the scale and location estimates $\hat{\sigma}_1$ and $\hat{\mu}_1$ are rescaled to match the scale and location of the original observations (since $\hat{\sigma}_1 \approx 1$ and $\hat{\mu}_1 \approx 0$).

Koutrouvelis's method performs well [KW98], but at a high computational cost. Kogon and Williams [KW98] proposed a modified version of Koutrouvelis's method. In their version, the characteristic function is parameterized using Zolotarev's shifted location parameter ζ, which eliminates the always-problematic discontinuity at $\alpha = 1$. This again gives regression equations similar to those obtained by Koutrouvelis. This time the initial values for σ and ζ are obtained using (the more accurate) McCulloch's method [McC86]. The regression can then be done using ordinary least squares, which reduces the complexity of the algorithm.

In simulation studies Kogon and Williams found that their estimator performed at least as well as Koutrouvelis's estimator for estimation of the stable index, with the performance being much better near $\alpha = 1$. The skewness parameter β was better estimated by the Koutrouvelis estimator, but this can be remedied by using GLS to perfom the regressions in the Kogon-Williams estimator instead of OLS [KW98].

2.4.4 An Estimator of the Stable Parameters Based on Q-Q Plots

Introduction

In this section we discuss a novel method for estimating the parameters of a stable distribution. The method optimizes the correlation between quantiles of the data and those of the target distribution. We also present some asymptotic results for the estimator.

Theoretical Background

Let $X_1, \ldots, X_n$ denote the observed data, and let $X_{n:1}, \ldots, X_{n:n}$ denote their order statistics.

Suppose we wanted to create a quantile-quantile plot of the observed data against some hypothesized distribution $F(y;\theta)$. We would generate the $0.5/n, \ldots, (n-0.5)/n$ quantiles of the theoretical distribution, and plot the points $(X_{n:i}, Y_{n:i})$, where $Y_{n:i} = F^{-1}(\frac{i-0.5}{n};\theta)$. The theoretical distribution would be declared a good fit if the points lie approximately on a 45 degree line. Stated mathematically, the distribution is a good fit if the correlation between the X's and the Y's is approximately 1.

We exploit this idea to find the parameters of stable distribution. The (squared) sample correlation coefficient is

$$r^2 = \frac{\left(\sum(x_i - \bar{x})(y_i - \bar{y})\right)^2}{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2},$$

which we can rewrite using the order statistics as

$$r^2 = \frac{\left(\sum(x_{n:i} - \bar{x})(y_{n:i} - \bar{y})\right)^2}{\sum(x_{n:i} - \bar{x})^2 \sum(y_{n:i} - \bar{y})^2}.$$

Replacing y's with F's

$$r^2(\theta) = \frac{\left(\sum(x_{n:i} - \bar{x})(F^{-1}(\frac{i-0.5}{n};\theta) - \sum_j F^{-1}(\frac{j-0.5}{n};\theta))\right)^2}{\sum(x_{n:i} - \bar{x})^2 \sum(F^{-1}(\frac{i-0.5}{n};\theta) - \sum_j F^{-1}(\frac{j-0.5}{n};\theta))^2}.$$

where we use the notation $r^2(\theta)$ to emphasize the dependence of the sample correlation coefficient on the parameters θ of the theoretical distribution. This quantity is then maximized over $\theta = (\alpha, \beta, \sigma, \mu)$.

In our experiments we found that the Q-Q estimator as described above was incredibly slow (since derivatives were not easily available) and suffered from convergence problems. Specifically, tracing the intermediate values produced by the optimizer showed that the algorithm had difficulty balancing α and σ. We found via trial and error that the following algorithm had much better performance.

1. Obtain initial estimates $(\alpha_0, \beta_0, \sigma_0, \mu_0)$ of $(\alpha, \beta, \sigma, \mu)$ via McCulloch's quantile estimator.

2. Normalize the observations using σ_0 and μ_0.

3. Fit α and β using the Q-Q method with σ and μ fixed at 1 and 0, respectively.

4. Report these values of α and β, along with σ_0 and μ_0, as the estimate.

This algorithm converges much faster than the previous method (but it is still slower than the other estimators on average). We considered an extension to this algorithm, in which the estimates of σ and μ are improved (perhaps by using the values of α and β computed in Step 4 above, along with σ_0 and μ_0, as initial values to the previous method), but, given the already slow convergence of the Q-Q estimator, the resulting estimator would be far too slow to be of any practical use.

2.4.5 Empirical Influence Functions

The empirical influence function (EIF) of an estimator $\hat{\theta}_n$ of θ at the sample $\mathbf{x}$ is defined as

$$\mathrm{EIF}(x; \hat{\theta}, \mathbf{x}) = (n+1)\left(\hat{\theta}_n(x, \mathbf{x}) - \hat{\theta}_n(\mathbf{x})\right). \tag{2.4.33}$$

The empirical influence function is a finite-sample version of the (asymptotic) influence function [SM05, HRRS86]. It measures how much an estimator changes in the presence of a single contamination point (the x). For example, suppose $\hat{\theta}_n$ is the sample mean, $\bar{x}$. Then a simple calculation shows that

$$\mathrm{EIF}(x; \bar{x}, \mathbf{x}) = (n+1)\left(\frac{1}{n+1}\sum_{i=1}^{n+1} x_i - \frac{1}{n}\sum_{i=1}^{n} x_i\right) = x - \hat{\theta}_n(\mathbf{x}).$$

Since this function is unbounded in x, we see that a contamination point can have arbitrarily large influence on the quantity being estimated by $\bar{x}$. That is, a single outlier can pull the estimate very far away from the true value.

The EIF's for the estimators of the parameters of a stable distribution are not simple enough to work out by hand, but we can still compute them numerically. In Figures 2.9-2.21 we show some EIF's for the above estimators (McCulloch's quantile estimator, the Kogon-Williams estimator, the MLE, and the Q-Q estimator). In order to show the amount of influence directly, we have removed the normalization factor of $n+1$. In all the plots

α	β	$P(X \leq -100)$	$P(X > 100)$	α	β	$P(X \leq -100)$	$P(X > 100)$
1.25	−1.0	0.001353	0.000227	1.25	−0.5	0.001078	0.000515
1.50	−1.0	0.000341	0.000043	1.50	−0.5	0.000267	0.000118
1.75	−1.0	0.000063	0.000006	1.75	−0.5	0.000049	0.00002
2.00	−1.0	0	0	2.00	−0.5	0	0
1.25	0.0	0.000799	0.000799	1.25	0.5	0.000515	0.001078
1.50	0.0	0.000192	0.000192	1.50	0.5	0.000118	0.000267
1.75	0.0	0.000035	0.000035	1.75	0.5	0.00002	0.000049
2.00	0.0	0	0	2.00	0.5	0	0
1.25	1.0	0.000227	0.001353				
1.50	1.0	0.000043	0.000341				
1.75	1.0	0.000006	0.000063				
2.00	1.0	0	0				

Table 2.1: Tail probabilities of stable distributions with various α and β.

we used a sample size of $n = 100$, the n quantiles $q_{.005}, \ldots, q_{.995}$ as the sample $\mathbf{x}$, and 100 contamination points spaced evenly in the interval $[-100, 100]$. The interval was chosen so that the resulting tail probabilities were smaller than ≈ 0.001. (A table of tail probabilities is shown in Table 2.1.) The true values of α used were $1.25, 1.50, 1.75$, and 2.00 (arranged from top to bottom), while the true values of β used were $-1, -.5, 0, .5$, and 1 (arranged from left to right). The scale and location parameters were fixed at 1 and 0, respectively. In some of the EIF computations, the numerical accuracy of the underlying density, etc. routines has a noticeable effect on the results. (How noticeable the effect is depends on the range over which the plot is made.) The two most frequently occurring situations are (a) when α is near 1 and the distribution is maximally skewed ($|\beta| \approx 1$) and when (b) $\alpha = 2$. In the former situation, the approximation to the density is not accurate enough in the "short" tail of the density (e.g., for $\beta = 1$, the distribution is right-skewed, so the approximation is lacking near the left tail). This is illustrated in Figure 2.7, in which we have plotted contours of the log-likelihood at the sample (left plot) and at the contaminated sample (right plot). Clearly, the true parameters do not maximize the log-likelihood as they should. This effect is only drastic when the contamination point is close to the boundary of the region in which the approximation holds; if the contamination point occurs further out in the tail, the magnitude of the effect is much smaller, and can only be seen on a suitably local scale.

In the second scenario, the difficulty is that the skewness parameter β is meaningless. There is no real way to estimate it, and hence, the estimator can be "completely wrong". The log-likelihood surface is shown in Figure 2.8; the left plot is the log-likelihood at the sample, and the right plot is the log-likelihood at the contaminated sample. The log-likelihood of the sample is relatively constant along the plane $\alpha = 2$, while the log-likelihood of the contaminated sample is flat near the boundary of the parameter space, dropping off sharply near the edge. Thus the computed value of $\hat{\theta}$ at the (uncontaminated) sample can be wrong, which leads to EIF's that appear to span an incorrect range (e.g., $[-1,1]$ instead of $[0,2]$).

Discussion of the EIF plots

The first pair of plots (Figs. 2.9 and 2.10) depicts the EIF's for McCulloch's quantile estimator. The spike near 0 is an numerical artefact—the EIF is actually a step function.

The second pair of plots (Figs. 2.11 and 2.12) depicts the EIF's for the regression-based estimator of Kogon and Williams. The EIF's show periodic behavior, which we believe is due to the separation of the real and imaginary parts in the regression.

The third pair of plots (Figs. 2.13 and 2.14) depicts the EIF's for the maximum likelihood estimator. The jaggedness of the EIF for α when $\alpha = 1.25$ is due to the numerical problem we discussed earlier. The abnormal range for the EIF of β when $\alpha = 2$ is due to the previously mentioned difficulty with estimating β when α is near 2.

The next plot (Figure 2.15) depicts the EIF for the maximum likelihood estimator when only α is unknown. In this computation, we assumed $\beta = 0$, $\sigma = 1$, and $\mu = 0$.

The next pair of plots (Figs. 2.16 and 2.17) shows the EIF's for the maximum likelihood estimator when only α and β are unknown. In this computation we assumed $\sigma = 1$ and $\mu = 0$.

The next pair of plots (Figs. 2.18 and 2.19) shows the EIF's for the maximum likelihood estimator when only μ is known (assumed to be 0).

The last pair of plots (Figs. 2.20 and 2.21) depicts the EIF's for the (modified) Q-Q estimator. The convergence problems associated with the estimator are responsible for the odd appearance of the EIF's.

To summarize,

- In general, outliers bias the estimates of α towards $\alpha = 1$ (i.e., a heavier-tailed distribution;
- with respect to maximum likelihood estimation, *a priori* knowledge of some of the parameters does not have a significant effect on how the MLE behaves in the presence of outliers;
- β becomes more difficult to estimate as α nears 2 (which is to be expected, since β matters less; and
- the Q-Q estimator, even in its modified form, is still too unstable to be useful.

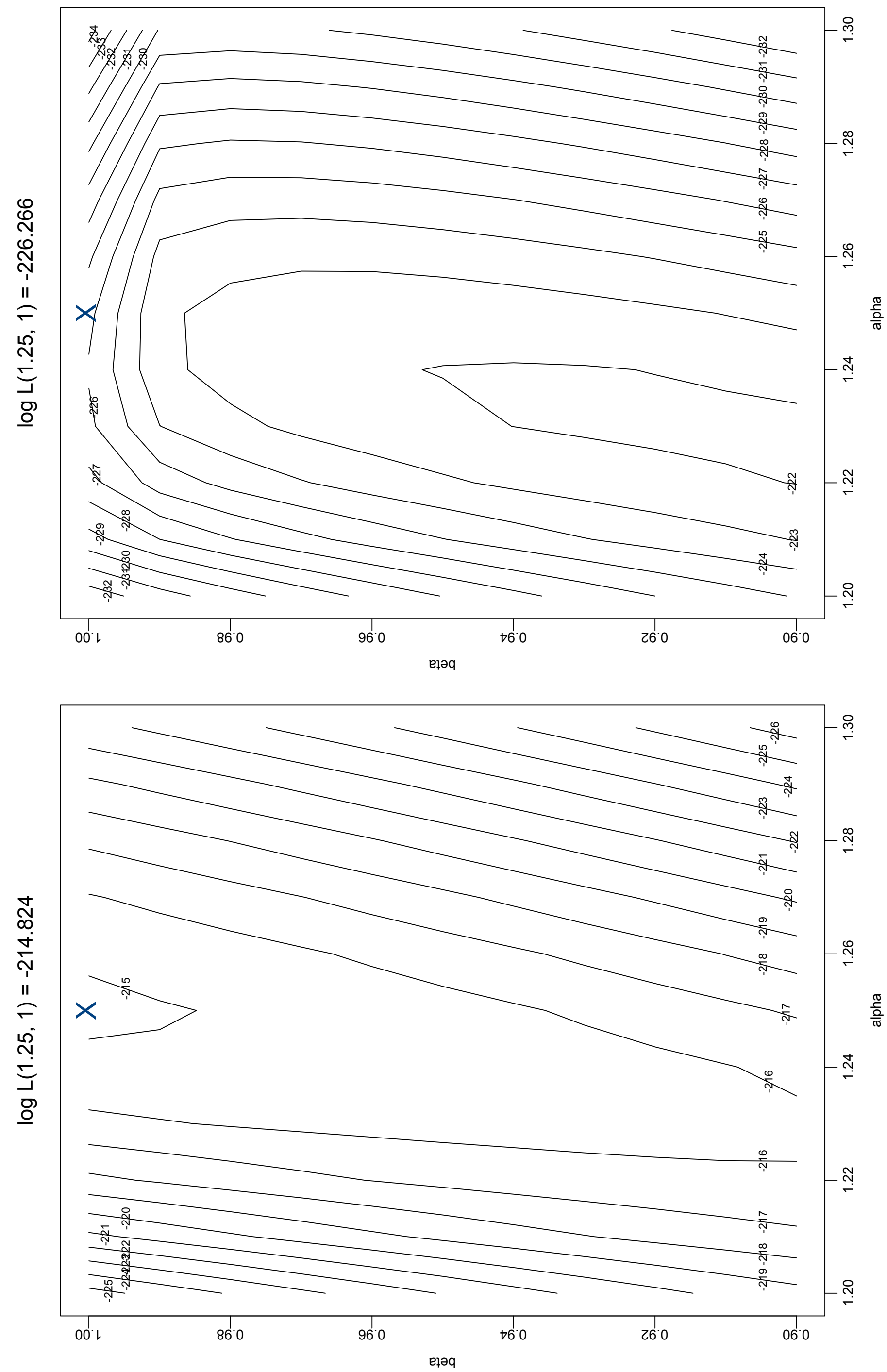

Figure 2.7: Contour plots of the likelihood at the sample (left) and at the contaminated sample (right). The "X" indicates the location of the true parameters $\alpha = 1.25$ and $\beta = 1$. The contamination point is -5.8, which is just outside the 0.005 quantile of the distribution with those parameters.

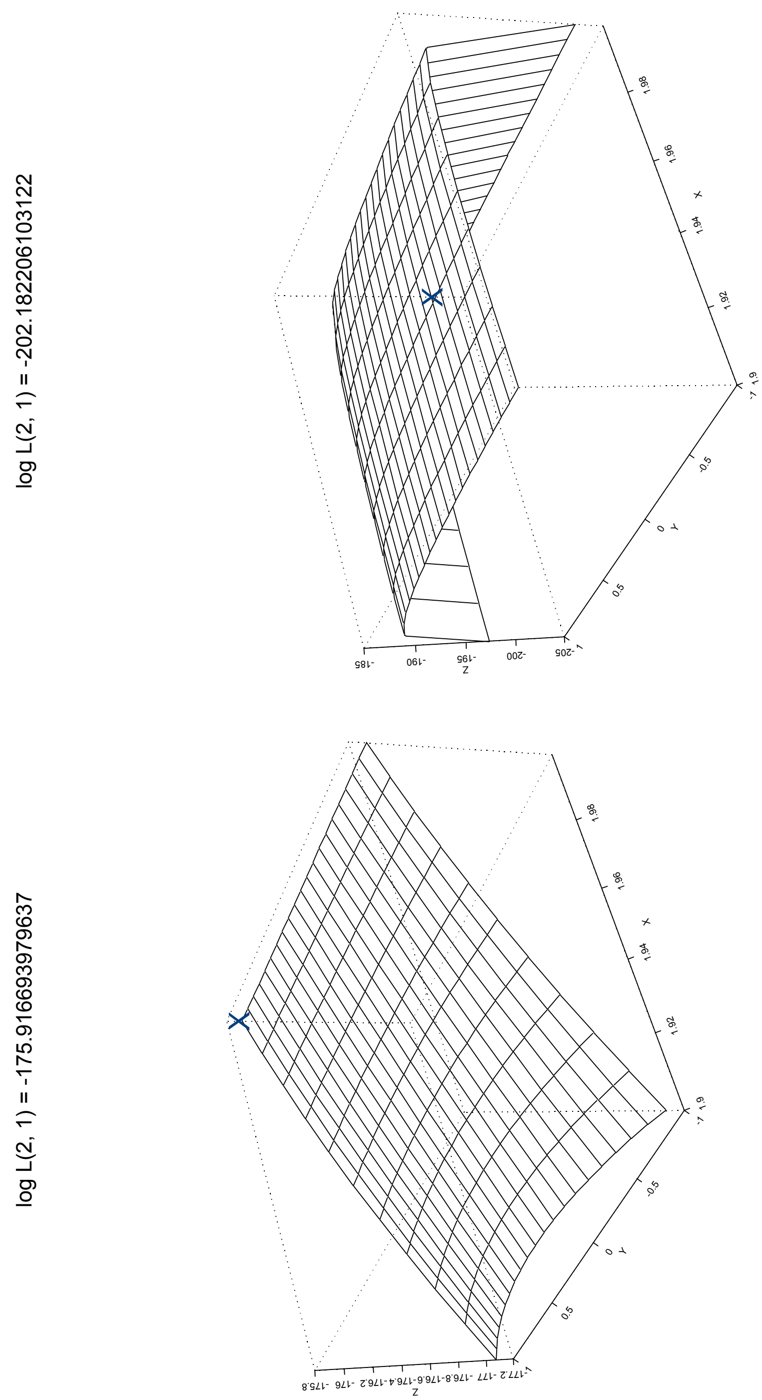

Figure 2.8: 3-D plot of the log-likelihood at the sample (left) and at the contaminated sample (right). The "X" indicates the location of the true maximum. The contamination point here is -10.

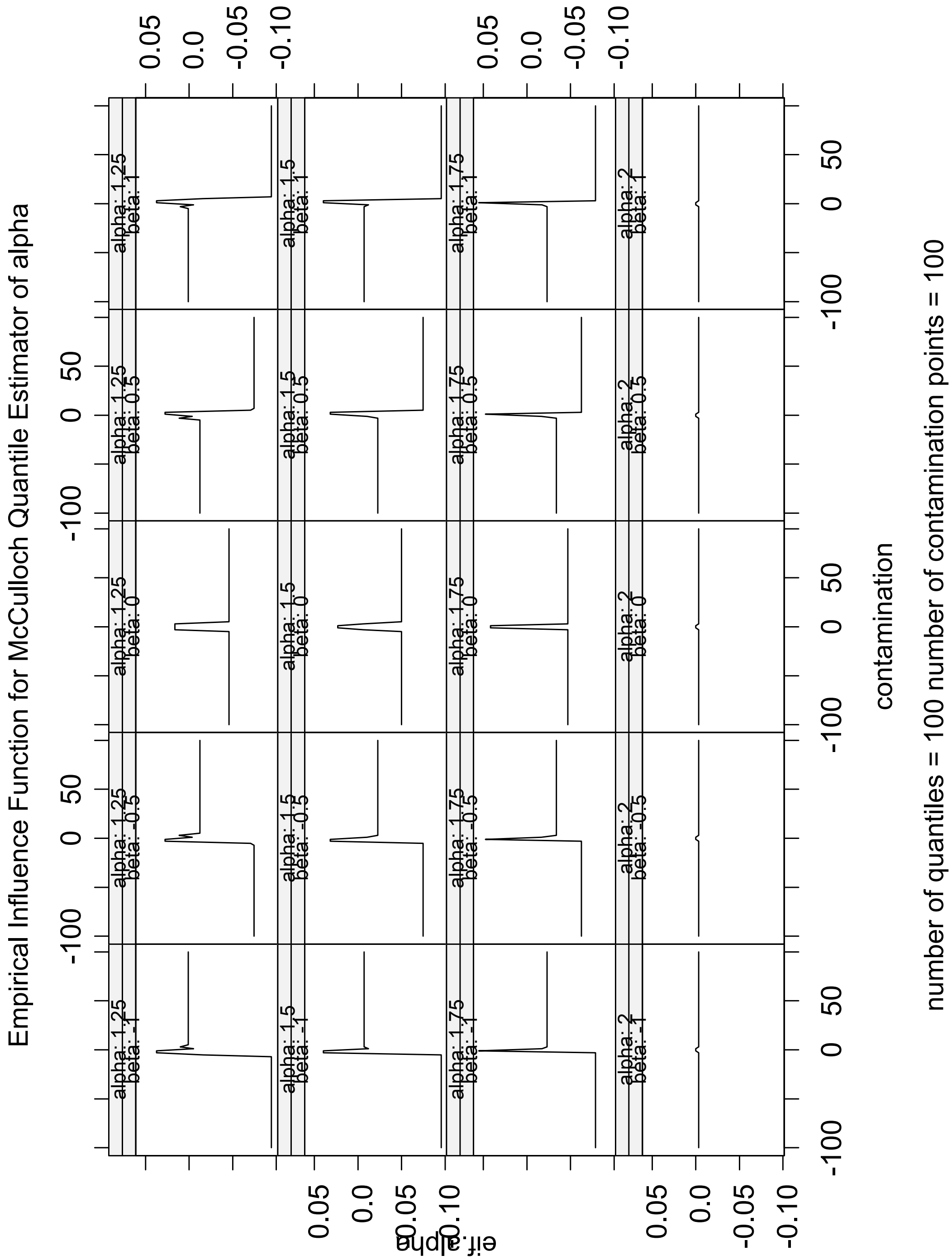

Figure 2.9: Empirical influence functions for McCulloch's quantile estimator of α.

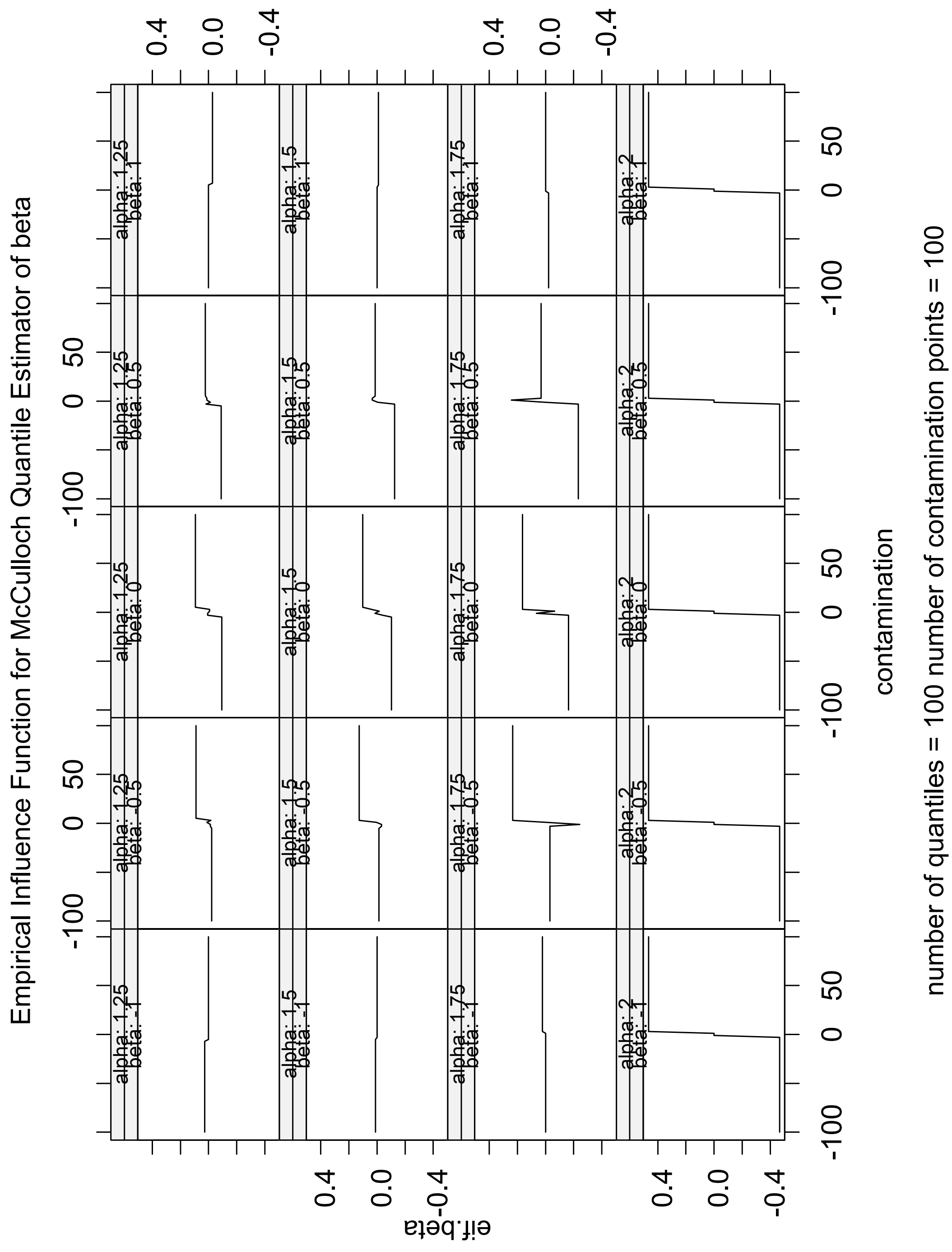

Figure 2.10: Empirical influence functions for McCulloch's quantile estimator of β.

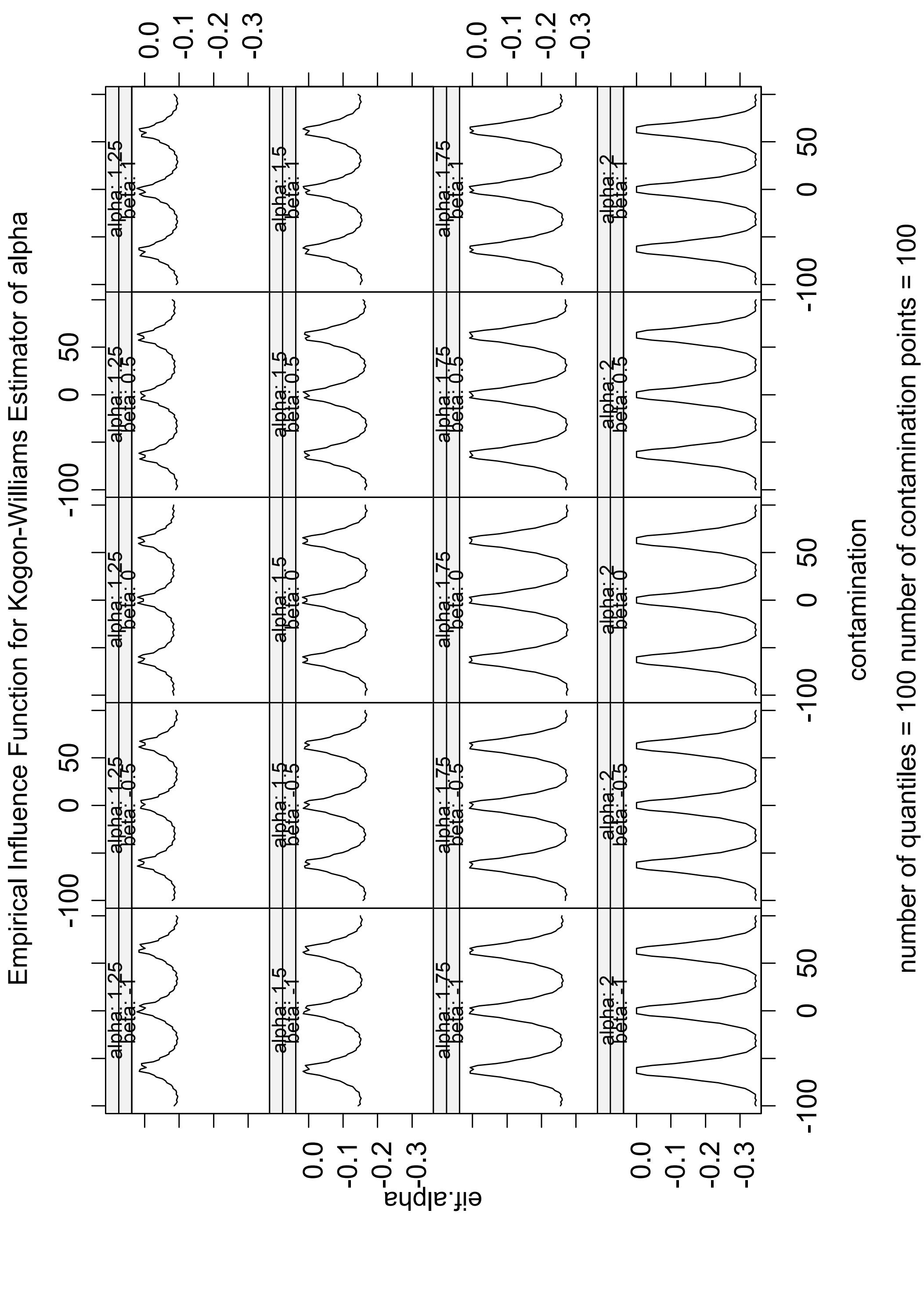

Figure 2.11: Empirical influence functions for the Kogon-Williams estimator of α.

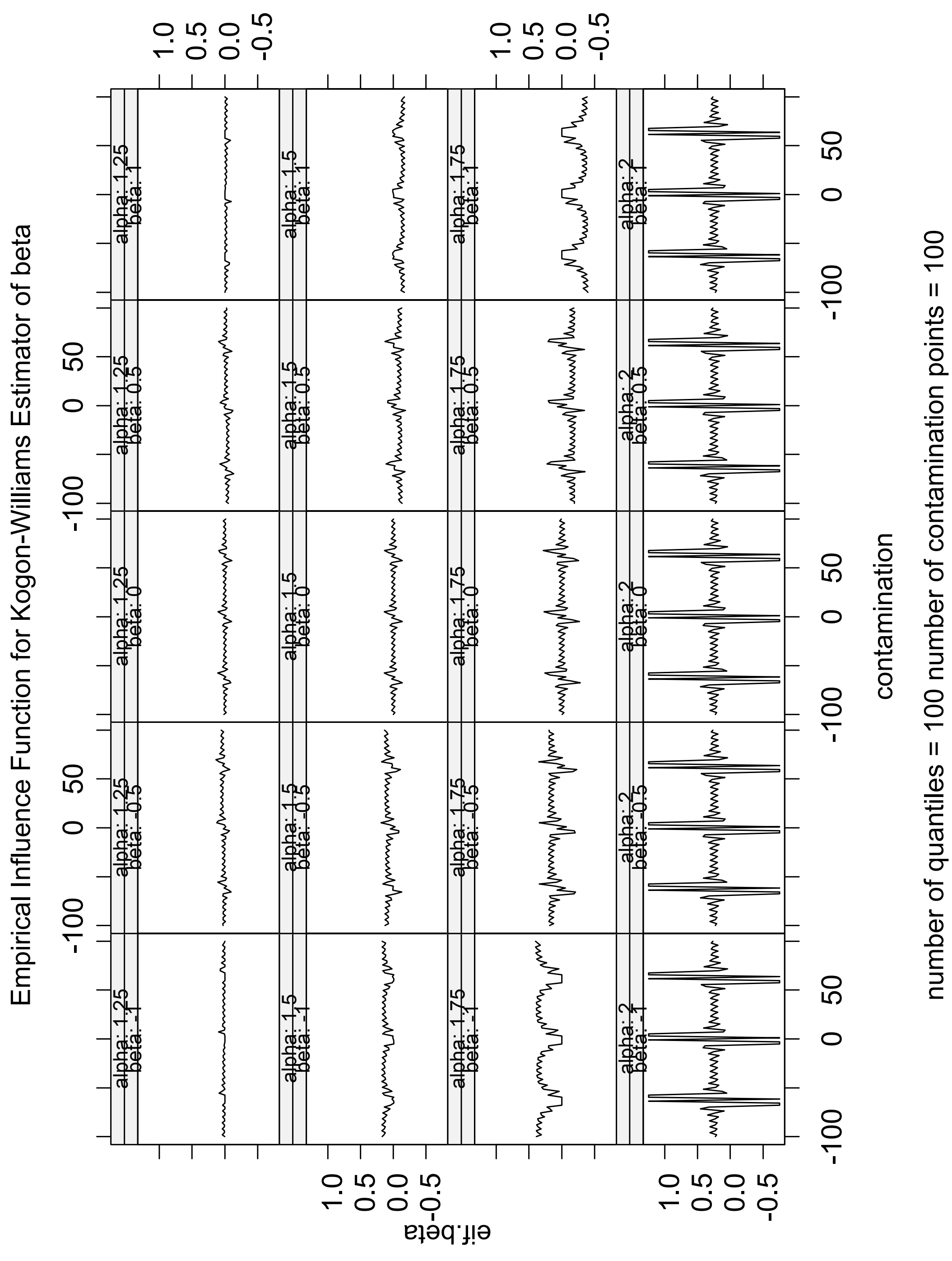

Figure 2.12: Empirical influence functions for the Kogon-Williams estimator of β.

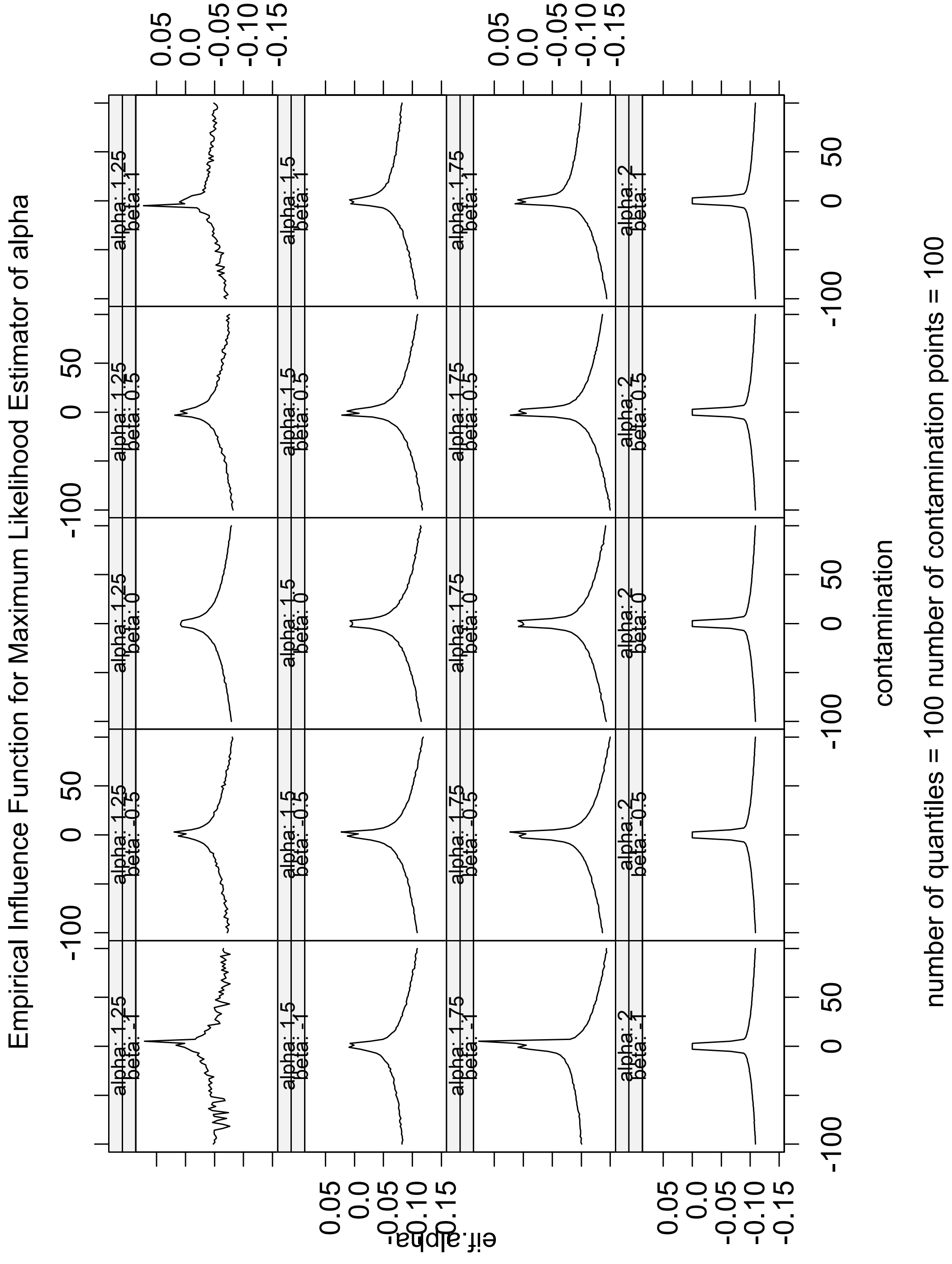

Figure 2.13: Empirical influence functions for the maximum likelihood estimator of α.

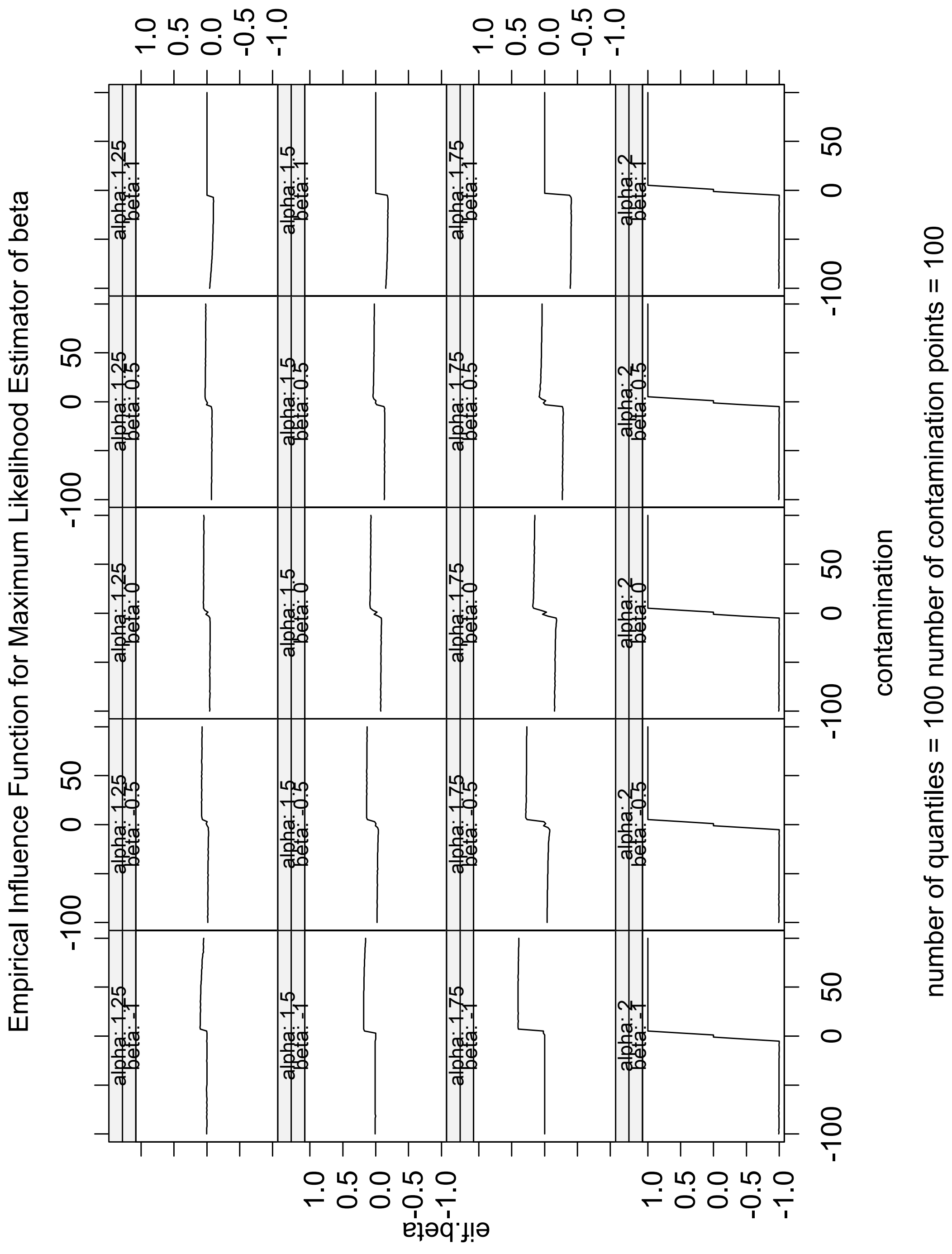

Figure 2.14: Empirical influence functions for the maximum likelihood estimator of β.

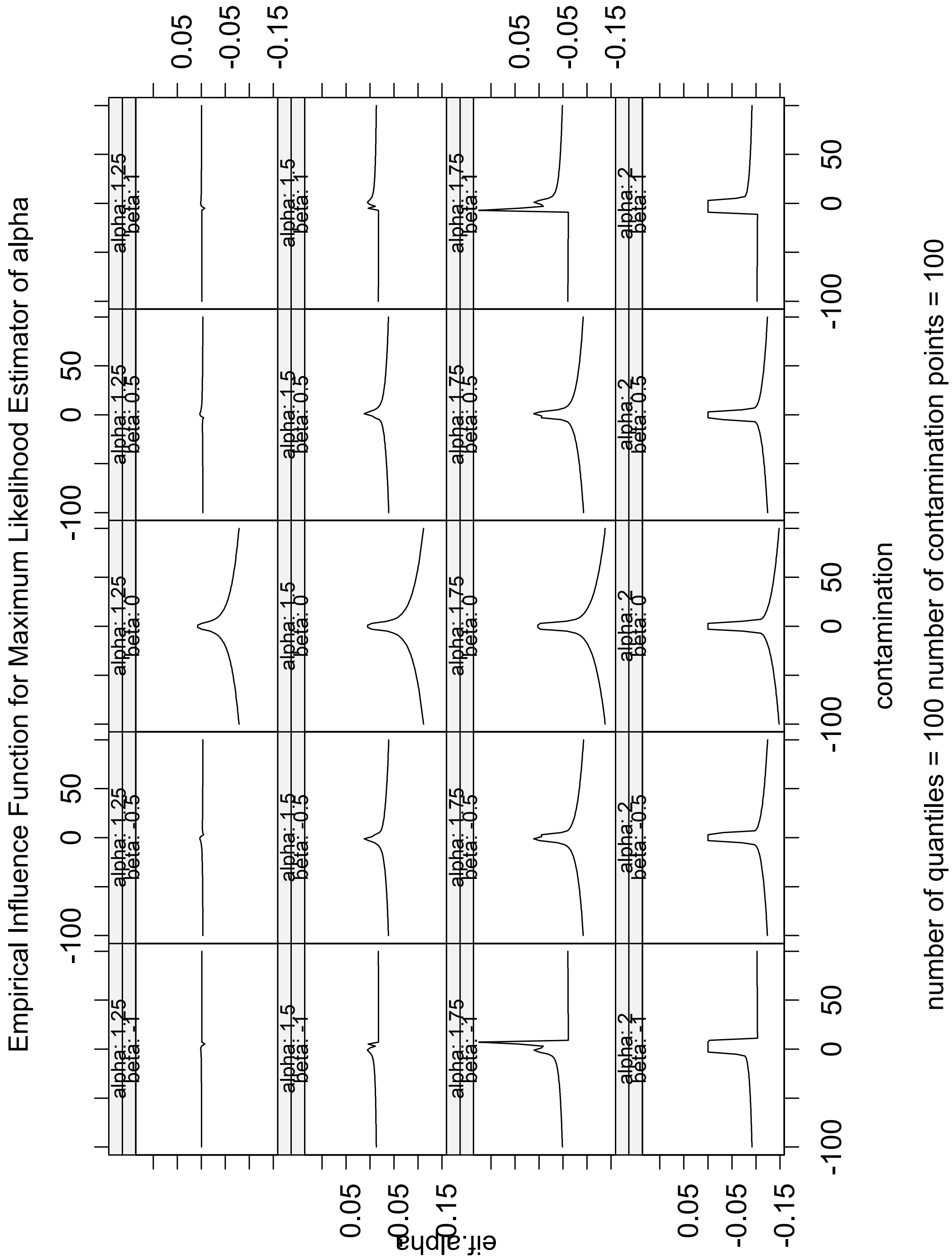

Figure 2.15: Empirical influence functions for the maximum likelihood estimator of α when only α is unknown.

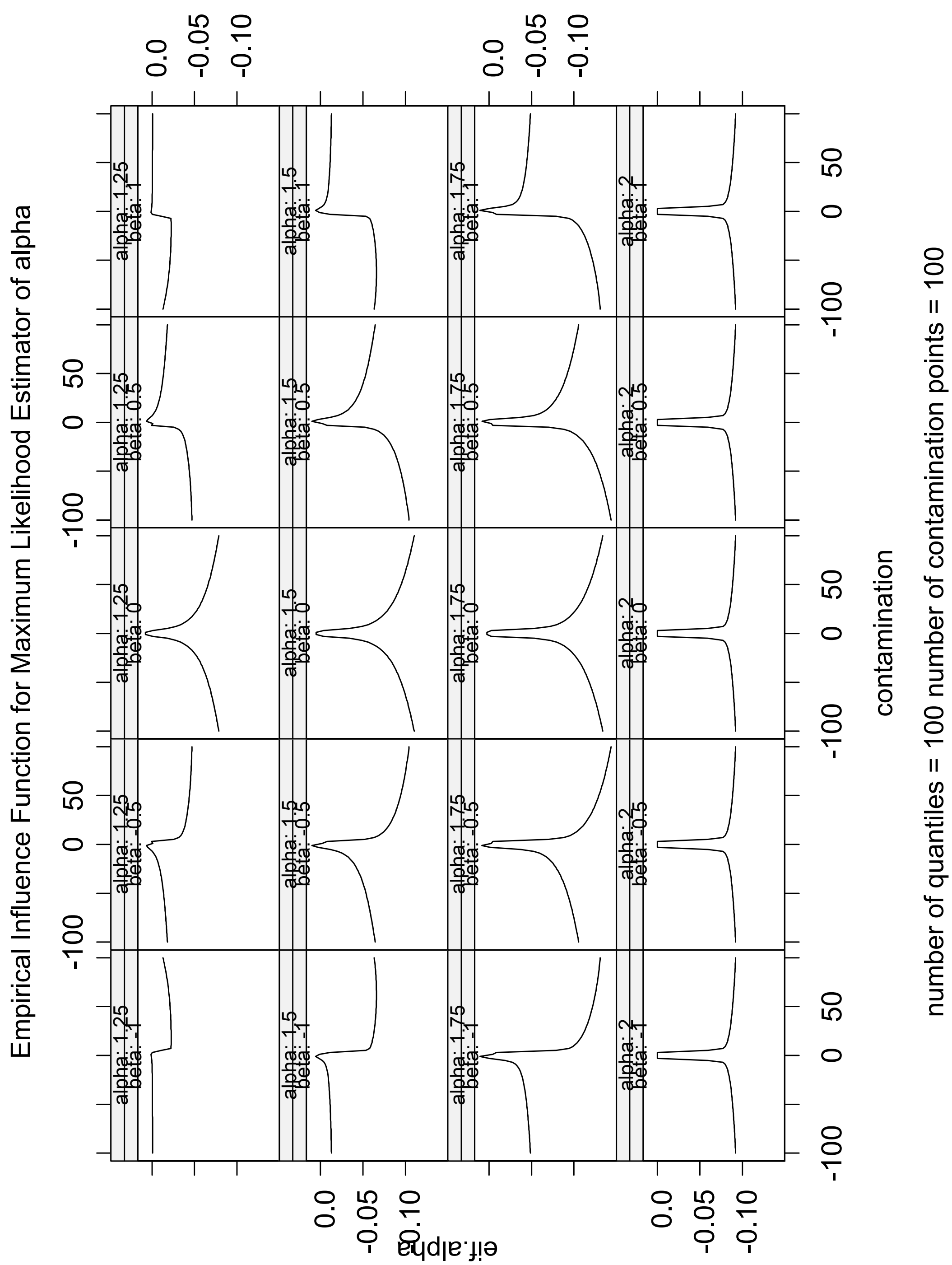

Figure 2.16: Empirical influence functions for the maximum likelihood estimator of α when only α and β are unknown.

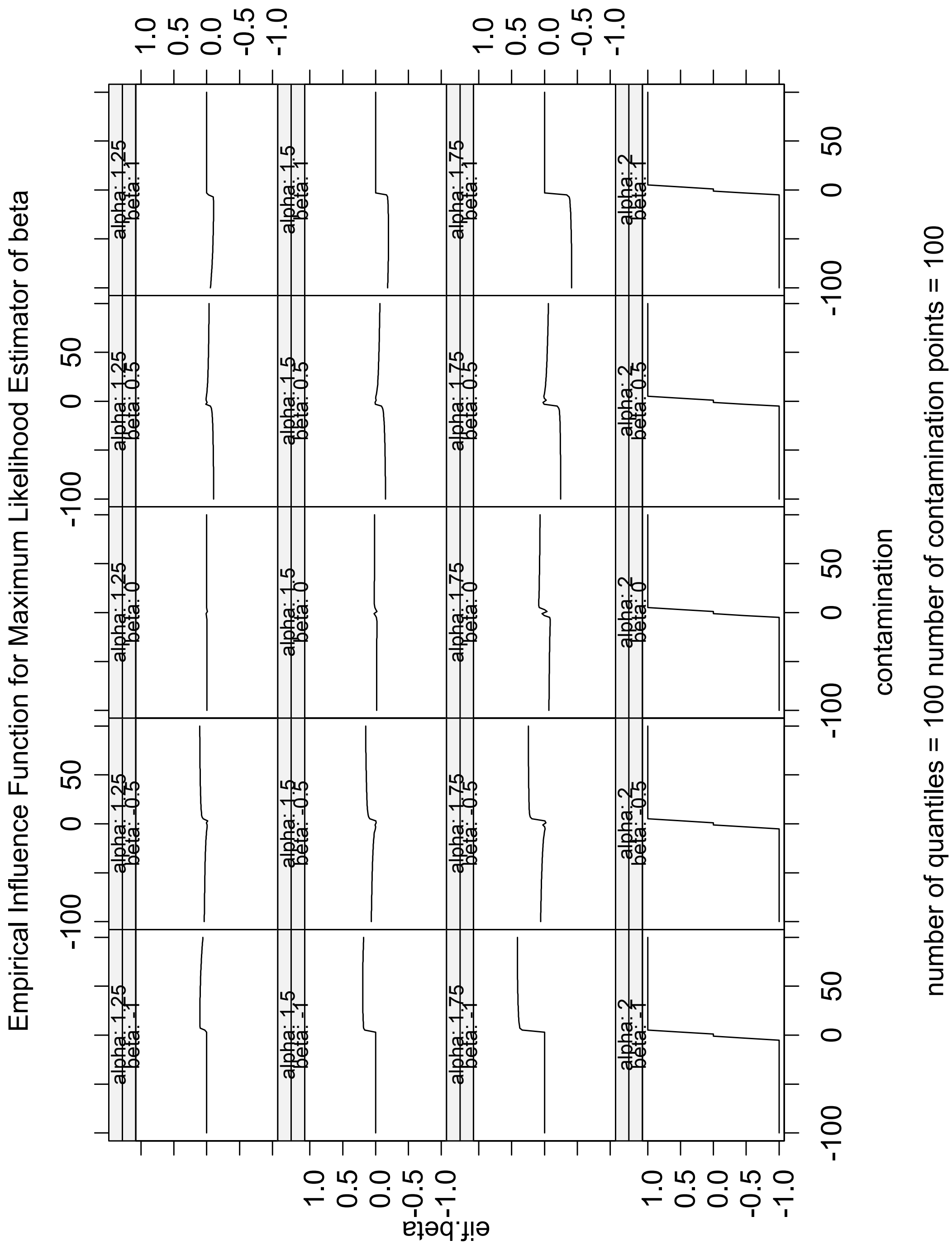

Figure 2.17: Empirical influence functions for the maximum likelihood estimator of β when only α and β are unknown.

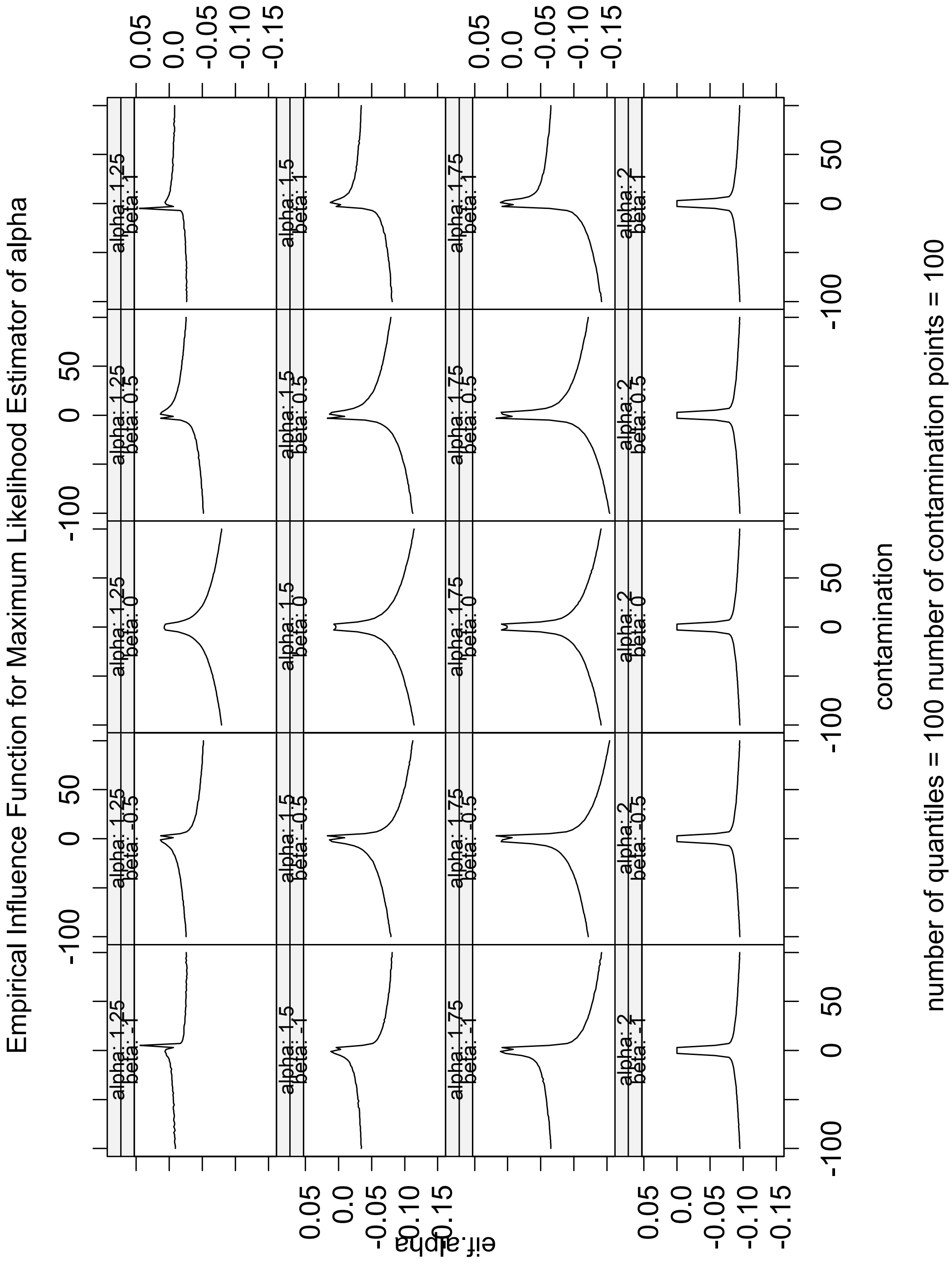

Figure 2.18: Empirical influence functions for the maximum likelihood estimator of α when only μ is known.

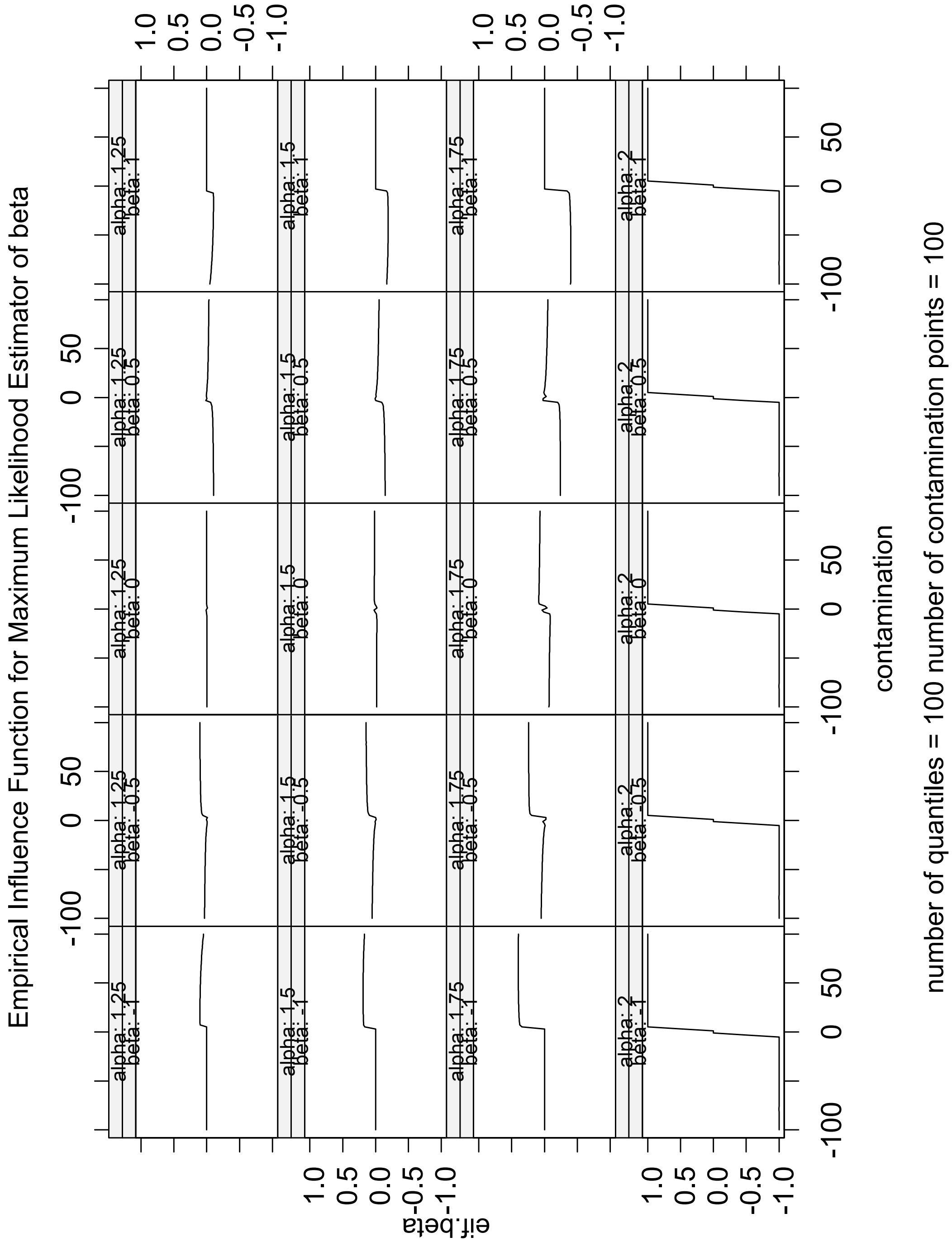

Figure 2.19: Empirical influence functions for the maximum likelihood estimator of β when only μ is known.

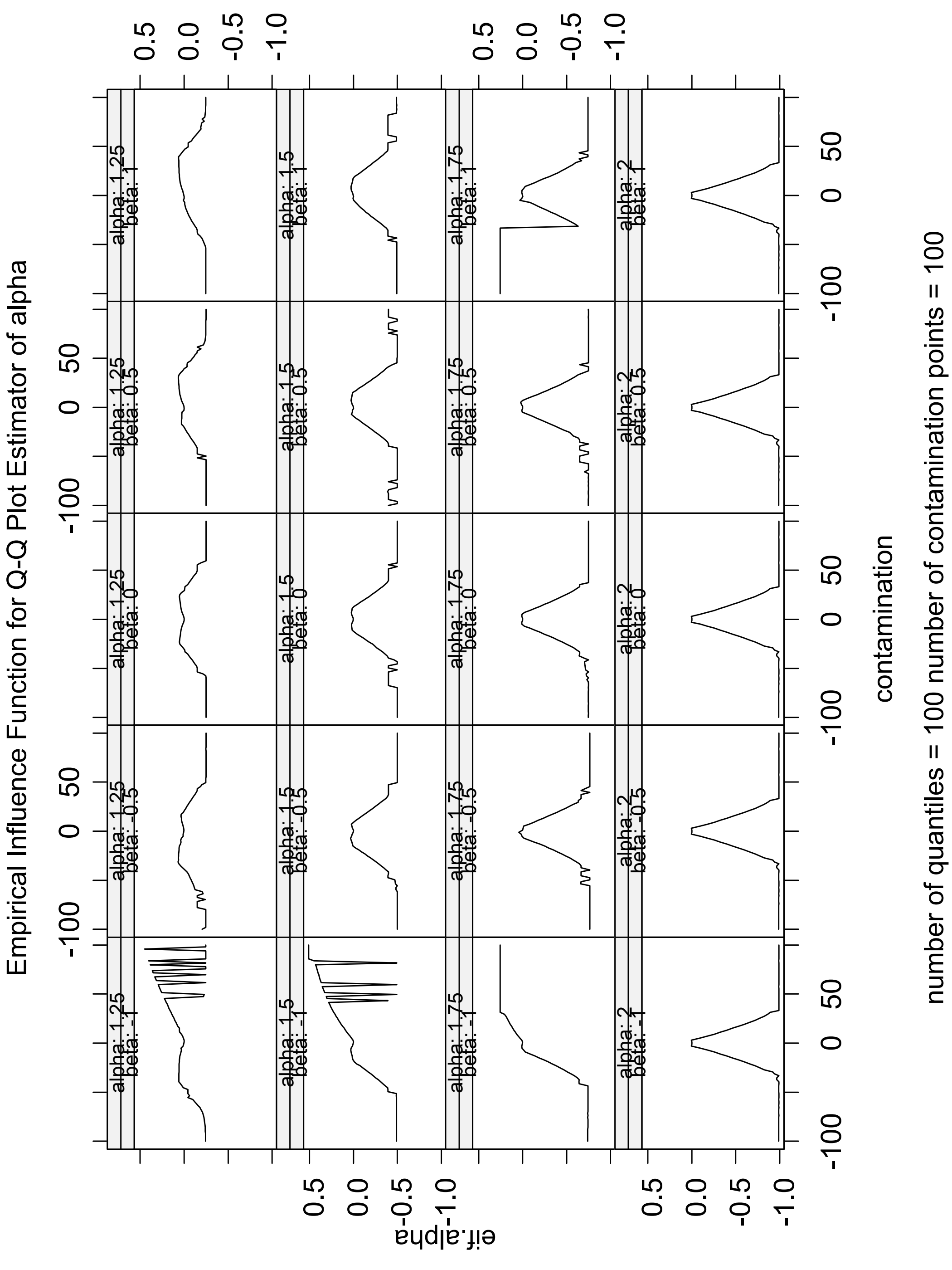

Figure 2.20: Empirical influence functions for the Q-Q plot estimator of α.

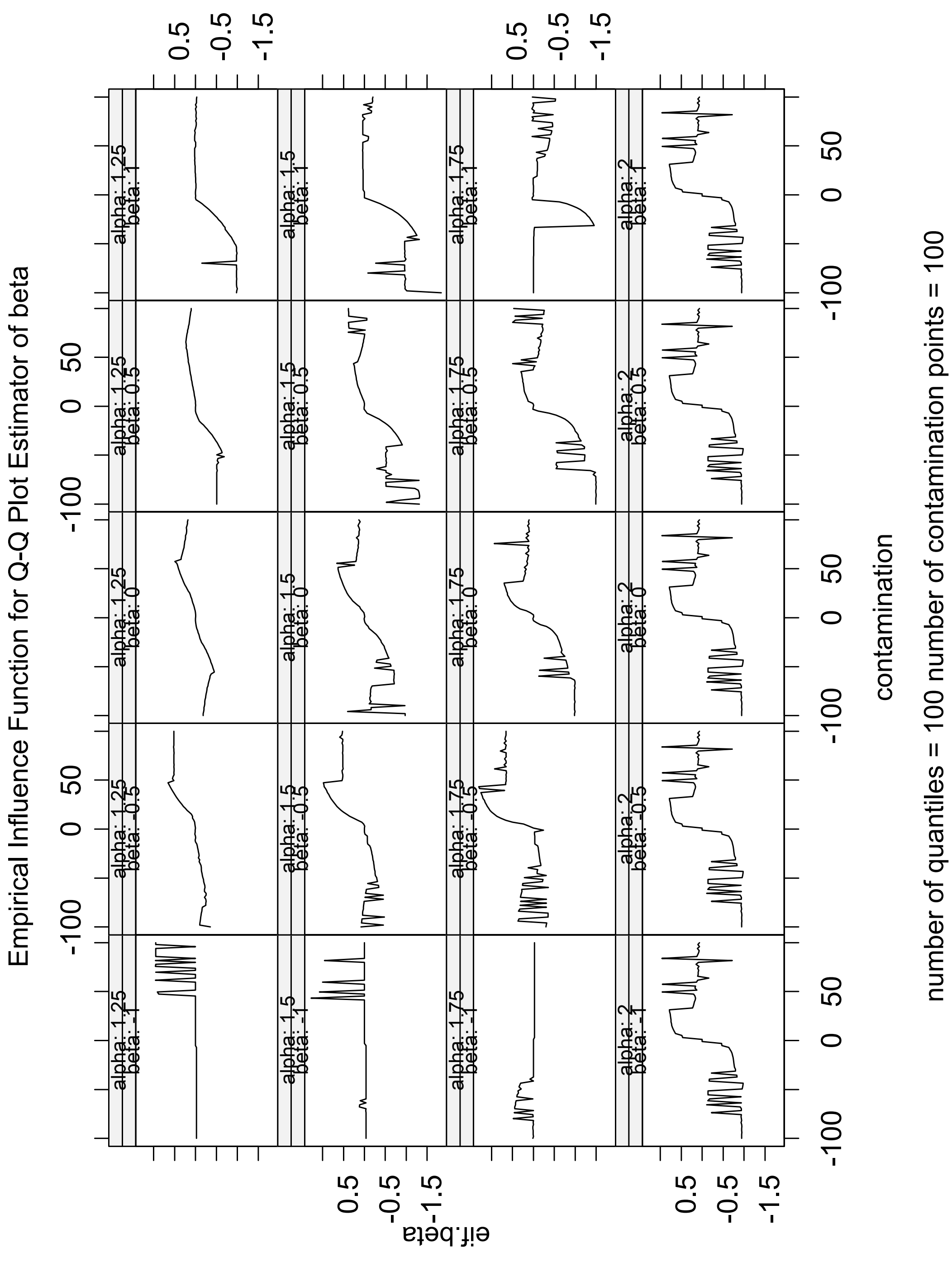

Figure 2.21: Empirical influence functions for the Q-Q plot estimator of β.

2.5 Score and Fisher Information Matrix for Stable Laws

Suppose the random variable X has a $S_{\alpha,\beta}(x;\sigma,\mu)$ distribution. As we have seen, its density can be expressed as the inverse Fourier transform of its characteristic function:

$$s_{\alpha,\beta}(x;\sigma,\mu) = \int e^{-ixt}\phi(t;\alpha,\beta,\sigma,\mu)\,dt,$$

where

$$\phi(t;\alpha,\beta,\sigma,\mu) = \exp\left(-\sigma^{\alpha}|t|^{\alpha}\left(1+i\beta\,\mathrm{sgn}(t)\omega(t;\alpha)\right)+i\mu t\right).$$

For n iid samples $X_1,\dots,X_n$ of this sort, the log-likelihood of the stable model is given by

$$\log L(\alpha,\beta,\sigma,\mu|x_1,...,x_n) = \sum_{k=1}^{n}\log s_{\alpha,\beta}(x_k;\sigma,\mu)$$

It is not immediately obvious that the stable density can be differentiated with respect to its parameters. Rachev and Mittnik [RM00] state that the characteristic function of a stable density, in the shifted or “M” parameterization, is twice-differentiable with respect to its parameters. Recall that in this parameterization, the logarithm of this characteristic function is given by

$$\log\phi(t) = \begin{cases} it\zeta - c|t|^{\alpha} + ict(|t|^{\alpha-1}-1)\beta\tan\left(\frac{\pi}{2}\alpha\right), & \alpha\neq 1\\ it\zeta - c|t| - ic\beta\frac{2}{\pi}\log|t|, & \alpha=1,\end{cases}$$

where ζ is the shifted location parameter, and $c=\sigma^{\alpha}$ is an alternative parameterization of the scale.

The process of computing the derivatives is greatly simplified by the observations that

$$\frac{\partial}{\partial\theta_1}\phi(t) = \frac{\partial}{\partial\theta_1}\exp\left(\log\phi(t)\right) = \left(\frac{\partial}{\partial\theta_1}\log\phi(t)\right)\phi(t)$$

and

$$\frac{\partial^2}{\partial\theta_1\partial\theta_2} = \left[\frac{\partial^2}{\partial\theta_1\partial\theta_2}\log\phi(t) + \left(\frac{\partial}{\partial\theta_1}\log\phi(t)\right)\left(\frac{\partial}{\partial\theta_2}\log\phi(t)\right)\right]\phi,$$

where $\theta_1,\theta_2\in\{\alpha,\beta,\sigma,\mu\}$.

The derivatives of $\log\phi(t)$ are shown below.

$$\frac{\partial\log\phi(t)}{\partial\alpha} = \begin{cases} -c(\log|t|)|t|^{\alpha} + ict\beta\left[(\log|t|)|t|^{\alpha-1}\tan\left(\frac{\pi}{2}\alpha\right) + (|t|^{\alpha-1}-1)\frac{\pi}{2}\sec^2\frac{\pi}{2}\alpha\right], & \alpha\neq 1\\ -c(\log|t|)|t| + ict\beta\left[-\frac{2}{\pi}(\log|t|)^2 + -\frac{2}{\pi}(\log|t|)\right], & \alpha=1\end{cases}$$

$$\frac{\partial\log\phi(t)}{\partial\beta} = \begin{cases} ict(|t|^{\alpha-1}-1)\tan\left(\frac{\pi}{2}\alpha\right), & \alpha\neq 1\\ -ict\frac{2}{\pi}\log|t|, & \alpha=1\end{cases}$$

$$\frac{\partial\log\phi(t)}{\partial c} = \begin{cases} -|t|^{\alpha} + it(|t|^{\alpha-1}-1)\beta\tan\left(\frac{\pi}{2}\alpha\right), & \alpha\neq 1\\ -|t| - i\beta\frac{2}{\pi}\log|t|, & \alpha=1,\end{cases}$$

$$\frac{\partial\log\phi(t)}{\partial\zeta} = it$$

The $\alpha = 1$ cases are computed using the limiting forms of the derivatives for $\alpha \neq 1$ and the computations

$$\lim_{\alpha \to 1} (|t|^{\alpha-1} - 1) \tan\left(\frac{\pi}{2}\alpha\right) = -\frac{2}{\pi} \log(|t|)$$

$$\lim_{\alpha \to 1} (\log|t|)|t|^{\alpha-1} \tan\left(\frac{\pi}{2}\alpha\right) + (|t|^{\alpha-1} - 1)\frac{\pi}{2} \sec^2 \frac{\pi}{2}\alpha = -\frac{1}{\pi}(\log(|t|))^2.$$

For what it's worth, we have the identities

$$\frac{\partial s_{\alpha,\beta}(x;c,\zeta)}{\partial x} = -\frac{\partial s_{\alpha,\beta}(x;c,\zeta)}{\partial \zeta}$$

$$\frac{\partial \log \phi(t)}{\partial c} = -|t|^{\alpha} + \frac{\beta}{c} \frac{\partial \log \phi(t)}{\partial \beta}.$$

Second derivatives for the $\alpha \neq 1$ case can now be computed without much additional difficulty.

$$\begin{aligned}
\frac{\partial^2 \log \phi(t)}{\partial \alpha^2} &= -c(\log|t|)^2 |t|^{\alpha} \\
&\quad + ict\beta \left\{ (\log|t|) \left[(\log|t|)|t|^{\alpha-1} \tan\left(\frac{\pi}{2}\alpha\right) + |t|^{\alpha-1} \frac{\pi}{2} \sec^2(\frac{\pi}{2}\alpha) \right] \right. \\
&\quad \left. + \left[(\log|t|)|t|^{\alpha-1} \frac{\pi}{2} \sec^2(\frac{\pi}{2}\alpha) + \frac{\pi^2}{2}(|t|^{\alpha-1} - 1) \tan\left(\frac{\pi}{2}\alpha\right) \sec^2(\frac{\pi}{2}\alpha) \right] \right\} \\
\frac{\partial^2 \log \phi(t)}{\partial \alpha \partial \beta} &= ict \left[(\log|t|)|t|^{\alpha-1} \tan\left(\frac{\pi}{2}\alpha\right) + (|t|^{\alpha-1} - 1)\frac{\pi}{2} \sec^2(\frac{\pi}{2})\alpha \right] \\
\frac{\partial^2 \log \phi(t)}{\partial \alpha \partial c} &= -(\log|t|)|t|^{\alpha} \\
&\quad + it\beta \left[(\log|t|)|t|^{\alpha-1} \tan\left(\frac{\pi}{2}\alpha\right) + (|t|^{\alpha-1} - 1)\frac{\pi}{2} \sec^2(\frac{\pi}{2})\alpha \right] \\
\frac{\partial^2 \log \phi(t)}{\partial \beta \partial c} &= it(|t|^{\alpha-1} - 1) \tan\left(\frac{\pi}{2}\alpha\right) \\
\frac{\partial^2 \log \phi(t)}{\partial \alpha \partial \zeta} &= \frac{\partial^2 \log \phi(t)}{\partial \beta^2} = \frac{\partial^2 \log \phi(t)}{\partial \beta \partial \zeta} = \frac{\partial^2 \log \phi(t)}{\partial c^2} = \frac{\partial^2 \log \phi(t)}{\partial c \partial \zeta} = \frac{\partial^2 \log \phi(t)}{\partial \zeta^2} = 0
\end{aligned}$$

The second derivatives for the $\alpha = 1$ case can again be found by taking limits.

The derivatives of the density can be found by differentiating under the integral sign and using the above formulas. To find the score equations we can use the relation

$$\frac{\partial \log f}{\partial \theta} = \frac{\frac{\partial f}{\partial \theta}}{f}$$

while for the information matrix we need the relation

$$\frac{\partial^2 \log f}{\partial \theta_1 \partial \theta_2} = \frac{\frac{\partial^2 f}{\partial \theta_1 \partial \theta_2}}{f} - \frac{\frac{\partial f}{\partial \theta_1} \frac{\partial f}{\partial \theta_2}}{f^2}.$$

Using the first equation, the score equations can be written as

$$\frac{\partial}{\partial \theta_1} \sum_{k=1}^{n} \log s_{\alpha,\beta}(x_k; \sigma, \mu) = \sum_{k=1}^{n} \frac{\int e^{-ix_k t} \phi(t) \frac{\partial}{\partial \theta_1} \log \phi(t)\, dt}{\int e^{-ix_k t} \phi(t)\, dt},$$

where as before θ_1 is one of $\{\alpha, \beta, \sigma, \mu\}$. Likewise, the terms of the information matrix (for $n = 1$) can be expressed as

$$-E\left[\frac{\partial^2}{\partial \theta_1 \partial \theta_2} \log s_{\alpha,\beta}(x; \sigma, \mu)\right] = -\int\int e^{-ixt} \phi(t) \frac{\partial^2}{\partial \theta_1 \partial \theta_2} \log \phi(t)\, dt\, dx$$
$$+ \int \frac{\left(\int e^{-ixt} \phi(t) \frac{\partial}{\partial \theta_1} \log \phi(t)\, dt\right) \left(\int e^{-ixt} \phi(t) \frac{\partial}{\partial \theta_2} \log \phi(t)\, dt\right)}{\left(\int e^{-ixt} \phi(t)\, dt\right)}\, dx.$$

In particular, even though many of the second derivatives of $\log \phi(t)$, it is not obvious if the corresponding terms of the information matrix vanish (which would simply matters greatly, perhaps allowing us to compute terms of the inverse matrix directly).

The direct numerical computation of the information matrix is not at all straightforward—many integrals must be approximated. The *observed* information matrix can be obtained numerically through finite-difference approximations, and is typically provided as a part of the output of an optimizer (e.g., `optim` in the MASS package). This method, however, is also very computationally-intensive. In order to reduce the computational time to a reasonable amount, we limited our calculations to a single value of σ ($\sigma = 1$), five values of β ($\beta \in \{\pm 1, \pm 0.5, 0\}$), and nine values of α ($\alpha \in \{1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0\}$). We computed the empirical correlation between α and σ by inverting the numerical Hessian obtained from a ML fit of the stable law parameters to a random sample of size 1000 from the hypothesized true distribution. To deal with the fact that the answers will vary from experiment to experiment due to various numerical factors (round-off error, etc.), we performed 500 replications for each set of parameters. The .025, .5, and .975 quantiles[8] of the values obtained are plotted in Figure 2.22 as a function of α for each value of β.

[8] For α near 1.0 the empirical correlation sometimes falls outside of $[-1, 1]$. These values were discarded in our estimation of the correlation.

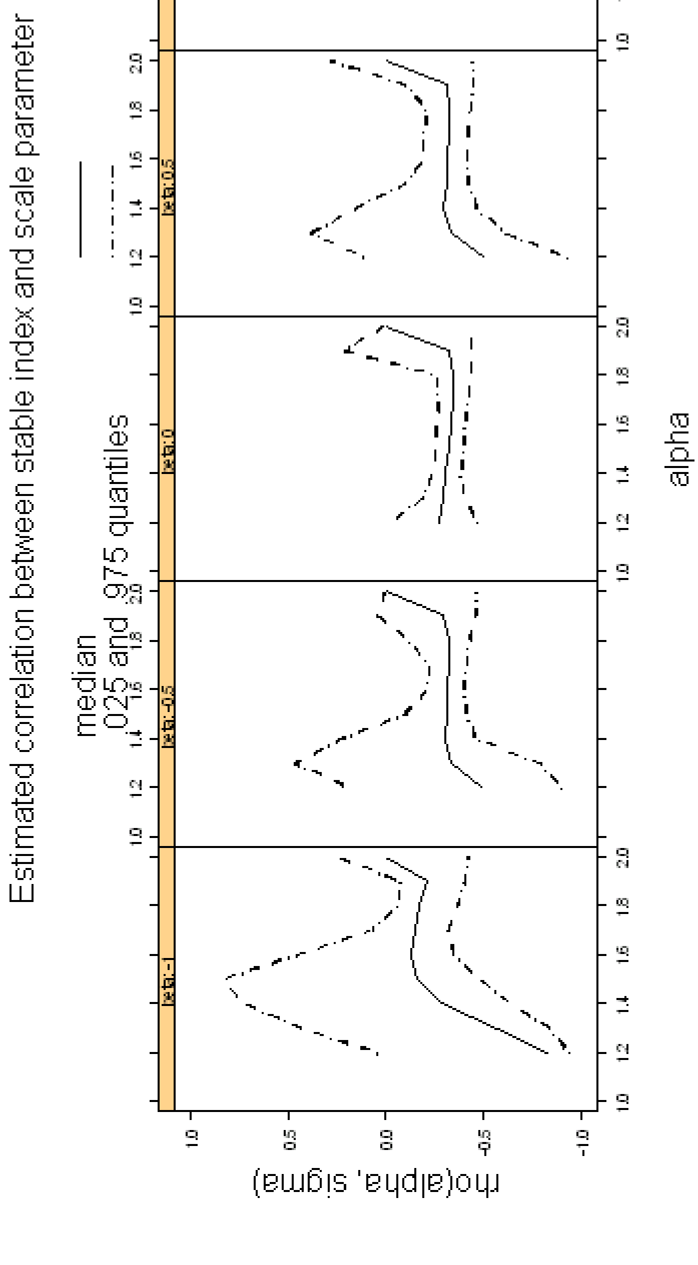

Figure 2.22: Estimated Asymptotic Correlation of α and σ.

Chapter 3

Univariate *t* Distributions

3.1 Introduction

Student's t distribution is well-known in classical statistics. It arises as the distribution of the ratio of the average of n independent observations drawn from a normal distribution, minus their true mean, to their standard deviation (scaled by $\sqrt{n}$). That is, the statistic

$$\frac{\bar{x}-\mu}{(s/\sqrt{n})} \tag{3.1.1}$$

has a t distribution. The t distribution has a shape parameter (denoted ν) that is known as the **degrees of freedom** of the distribution. The distribution of the ratio above is said to have $n-1$ degrees of freedom. By standard probabilistic methods, we can show that the density function of the t distribution is

$$f(x|\nu) = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\left(1+\frac{x^2}{\nu}\right)^{(\nu+1)/2}}. \tag{3.1.2}$$

We can then define the t distribution for nonintegral ν by this formula. We will henceforth adopt the notation $\mathscr{T}_\nu$ to denote a t distribution with ν degrees of freedom.

The t distribution was first described in 1908 by William S. Gossett, an employee of the Guinness Brewery. Company restrictions prohibited him from publishing his results under his real name, so he published his work under the pseudonym "Student" [Stu08]. Thus, his distribution came to be known as "Student's t".

The Cauchy distribution corresponds to the t distribution with 1 degree of freedom. The standard normal distribution is the limiting distribution as $\nu \to \infty$.

The usual derivation of the distribution of the ratio (3.1.1) uses the facts that (1) the numerator is a $N(0,\sigma^2)$ random variable; (2) the denominator is the square root of a χ^2_n/n random variable; and (3) the numerator and denominator are independent random variables. Since the χ^2_n distribution is a special case of the Gamma distribution (a $Gamma(n/2,n/2)$ distribution, to be exact), this derivation also shows that the t distribution is an inverse-Gamma variance mixture of normal distributions.

The mean of the t distribution is 0 (when $\nu > 1$) while the variance is $\frac{\nu}{\nu-2}$ (when $\nu > 2$). If $\nu \leq 1$ the mean does not exist, and if $\nu \leq 2$ the variance of the t distribution does not exist.

3.2 Score Function and Information Matrix for t Law

Suppose the random variable X is a shifted, rescaled version of a standard Student's t random variable T with ν degrees of freedom, i.e.,

$$\frac{X-\mu}{\sigma} \equiv T \sim \mathscr{T}_\nu.$$

The density of T was derived in Equation (3.1.2). The density of X is therefore

$$f(x|\nu,\mu,\sigma) = \frac{1}{\sigma}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\frac{1}{\sqrt{\nu\pi}}\frac{1}{\left(1+\frac{(x-\mu)^2}{\sigma^2\nu}\right)^{\frac{\nu+1}{2}}}.$$

For n iid samples $X_1,\ldots,X_n$ of this sort, the loglikelihood of the t model is given by

$$l(\nu,\mu,\sigma|x_1,\ldots,x_n) = \log L(\nu,\mu,\sigma|x_1,...,x_n) = \sum_{k=1}^{n}\log f(x_k|\nu,\mu,\sigma).$$

The logarithm of the density is

$$\log f(x|\nu,\mu,\sigma) = -\log\sigma + \log\Gamma\left(\frac{\nu+1}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\log\nu - \frac{1}{2}\log\pi \\ - \frac{\nu+1}{2}\log\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right).$$

The score equations for ν,μ, and σ are hence given by

$$\frac{\partial l}{\partial \nu} = \frac{n}{2}\left[\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right] - \frac{n}{2\nu} - \frac{1}{2}\sum_{k=1}^{n}\log\left(1+\frac{(x_k-\mu)^2}{\nu\sigma^2}\right) + \frac{\nu+1}{2\nu}\sum_{k=1}^{n}\frac{(x_k-\mu)^2}{\nu\sigma^2+(x_k-\mu)^2}$$

$$\frac{\partial l}{\partial \mu} = (\nu+1)\sum_{k=1}^{n}\frac{x_k-\mu}{\nu\sigma^2+(x_k-\mu)^2}$$

$$\frac{\partial l}{\partial \sigma} = -\frac{1}{\sigma}\left[n-(\nu+1)\sum_{k=1}^{n}\frac{(x_k-\mu)^2}{\nu\sigma^2+(x_k-\mu)^2}\right],$$

where $\psi(z)$ is the digamma function:

$$\psi(z) = \frac{d\log\Gamma(z)}{dz}.$$

The density $f(x;\nu,\mu,\sigma)$ is supported on the entire real line and can be differentiated with respect to each of the parameters. Furthermore, the integral $\int f(x;\nu,\mu,\sigma)\,dx$ can be differentiated under the integral sign with respect to each of the parameters. Therefore, the expected value of each score function is zero. Applying this observation to $\frac{\partial l}{\partial \nu}$, $\frac{\partial l}{\partial \mu}$, and $\frac{\partial l}{\partial \sigma}$, we obtain the results

$$\begin{aligned}
E\left(\frac{(x-\mu)^2}{\nu\sigma^2+(x-\mu)^2}\right) &= \frac{1}{\nu+1}\\
E\left(\log\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)\right) &= \psi\left(\frac{\nu+1}{2}\right)-\psi\left(\frac{\nu}{2}\right)\\
E\left(\frac{(x-\mu)}{\nu\sigma^2+(x-\mu)^2}\right) &= 0.
\end{aligned}$$

The second derivatives can all be calculated in similar fashion, albeit with some tedious algebra.

$$\begin{aligned}
\frac{\partial^2 l}{\partial \nu^2} &= \frac{n}{4}\left[\psi'\left(\frac{\nu+1}{2}\right)-\psi'\left(\frac{\nu}{2}\right)\right]+\frac{n}{2\nu^2}+\frac{\nu-1}{2\nu^2}\sum_{k=1}^{n}\frac{(x_k-\mu)^2}{\nu\sigma^2+(x_k-\mu)^2}\\
&\quad -\sigma^2\frac{\nu+1}{2\nu}\sum_{k=1}^{n}\frac{(x_k-\mu)^2}{\left(\nu\sigma^2+(x_k-\mu)^2\right)^2}\\
\frac{\partial^2 l}{\partial \mu^2} &= (\nu+1)\sum_{k=1}^{n}\frac{-\nu\sigma^2+(x_k-\mu)^2}{\left(\nu\sigma^2+(x_k-\mu)^2\right)^2}\\
\frac{\partial^2 l}{\partial \sigma^2} &= \frac{1}{\sigma^2}\left[n-(\nu+1)\sum_{k=1}^{n}\frac{(x_k-\mu)^2}{\nu\sigma^2+(x_k-\mu)^2}\right]-2\nu(\nu+1)\sum_{k=1}^{n}\frac{(x-\mu)^2}{\left(\nu\sigma^2+(x_k-\mu)^2\right)^2}\\
\frac{\partial^2 l}{\partial \sigma\partial \mu} &= -2\nu(\nu+1)\sigma\sum_{k=1}^{n}\frac{x_k-\mu}{\left(\nu\sigma^2+(x_k-\mu)^2\right)^2}\\
\frac{\partial^2 l}{\partial \nu\partial \mu} &= \sum_{k=1}^{n}\frac{x_k-\mu}{\nu\sigma^2+(x_k-\mu)^2}-(\nu+1)\sigma^2\sum_{k=1}^{n}\frac{x_k-\mu}{\left(\nu\sigma^2+(x_k-\mu)^2\right)^2}\\
\frac{\partial^2 l}{\partial \sigma\partial \nu} &= \frac{1}{\sigma}\sum_{k=1}^{n}\frac{(x_k-\mu)^2}{\nu\sigma^2+(x_k-\mu)^2}-\sigma(\nu+1)\sum_{k=1}^{n}\frac{(x_k-\mu)^2}{\left(\nu\sigma^2+(x_k-\mu)^2\right)^2}
\end{aligned}$$

The derivative $\psi'(z)$ of the digamma function is known as the trigamma function.

Since the logarithm of the density is twice differentiable with respect to each of parameters, the information matrix $I(\nu,\mu,\sigma)$ is the negative of the expectation of the matrix of second derivatives. By using our previous observations, and the calculations

$$
\begin{aligned}
E\left(\frac{1}{(\nu\sigma^2+(x-\mu)^2)^2}\right) &= \frac{1}{\nu\sigma^4}\frac{\nu+2}{(\nu+1)(\nu+3)} \\
E\left(\frac{x-\mu}{(\nu\sigma^2+(x-\mu)^2)^2}\right) &= 0 \\
E\left(\frac{(x-\mu)^2}{(\nu\sigma^2+(x-\mu)^2)^2}\right) &= E\left(\frac{1}{\nu\sigma^2+(x-\mu)^2}\right) - E\left(\frac{\nu\sigma^2}{(\nu\sigma^2+(x-\mu)^2)^2}\right) \\
&= \frac{1}{\sigma^2(\nu+1)} - \nu\sigma^2\frac{1}{\nu\sigma^4}\frac{\nu+2}{(\nu+1)(\nu+3)} = \frac{1}{\sigma^2(\nu+1)(\nu+3)}.
\end{aligned}
$$

we can explicitly calculate the expected information matrix.

$$
I_n(\nu,\mu,\sigma) = n\begin{pmatrix} \frac{1}{4}\left[\psi'\left(\frac{\nu}{2}\right)-\psi'\left(\frac{\nu+1}{2}\right)\right]-\frac{1}{\nu}\left[\frac{1}{\nu+1}-\frac{1}{2(\nu+3)}\right] & 0 & \frac{1}{\sigma}\left[\frac{1}{\nu+3}-\frac{1}{\nu+1}\right] \\ 0 & \frac{1}{\sigma^2}\left[1-\frac{2}{\nu+3}\right] & 0 \\ \frac{1}{\sigma}\left[\frac{1}{\nu+3}-\frac{1}{\nu+1}\right] & 0 & \frac{2}{\sigma^2}\frac{\nu}{\nu+3} \end{pmatrix}.
$$

For notational convenience, let us denote the diagonal terms of this matrix by A, B, and C, in that order, and the lone nonzero off-diagonal term by D. In this notation, the determinant of the information matrix is $B(AC-D^2)$, and the inverse is given by

$$
I_n^{-1}(\nu,\mu,\sigma) = \frac{1}{n}\frac{1}{B(AC-D^2)}\begin{pmatrix} BC & 0 & -DB \\ 0 & AC-D^2 & 0 \\ -DB & 0 & AB \end{pmatrix}
$$

Figures 3.1-3.4 show the dependence of the asymptotic variances and correlations of the MLE's on ν for various values of σ.

The shape of the variance of $\hat{\nu}$ as a function of ν is somewhat surprising, so we verified that the variance increased with ν via (a) Monte Carlo simulation and (b) computing the asymptotic variance in the case when σ and μ are known. We believe that the variance increases with ν because, for ν large, the densities are very closely approximated by the normal distribution. Hence, for ν large, the densities are harder to distinguish from one another—data generated from a $\mathscr{T}_{10}$ look similar to data generated from a $\mathscr{T}_{11}$.

3.3 Estimators

Figures 3.5 and 3.6 show empirical influence functions for the maximum likelihood estimator and Q-Q estimator, respectively, of ν. The EIF of the maximum likelihood estimator shows that large outliers lead to smaller estimates of ν (see the lower panels of Figure 3.5), while inliers (near 0) lead to larger estimates of ν (i.e., the fitted distribution is more normal). The Q-Q estimator shows similar behavior, but is more severely influenced by outliers as ν increases.

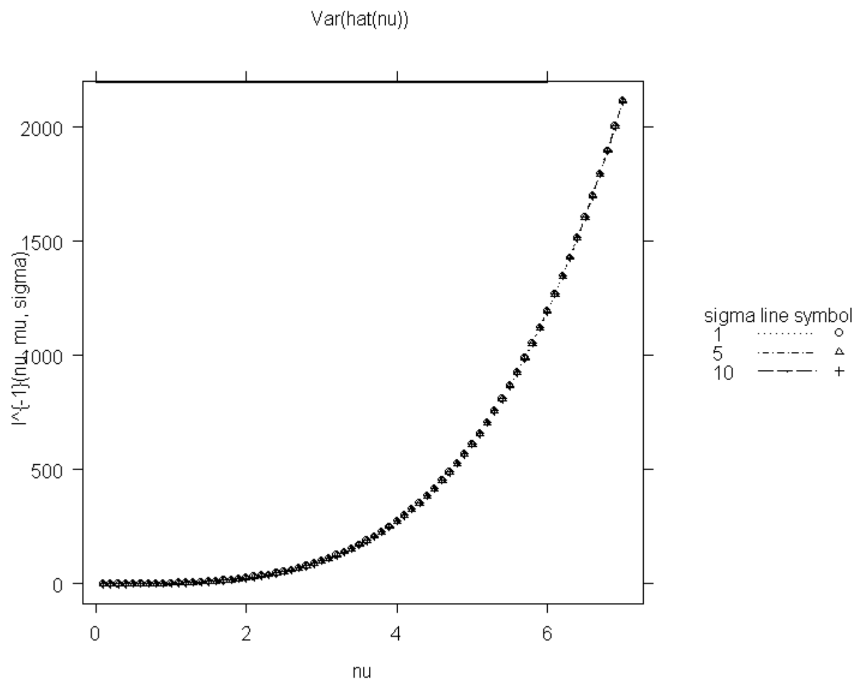

Figure 3.1: Asymptotic Variance of ν.

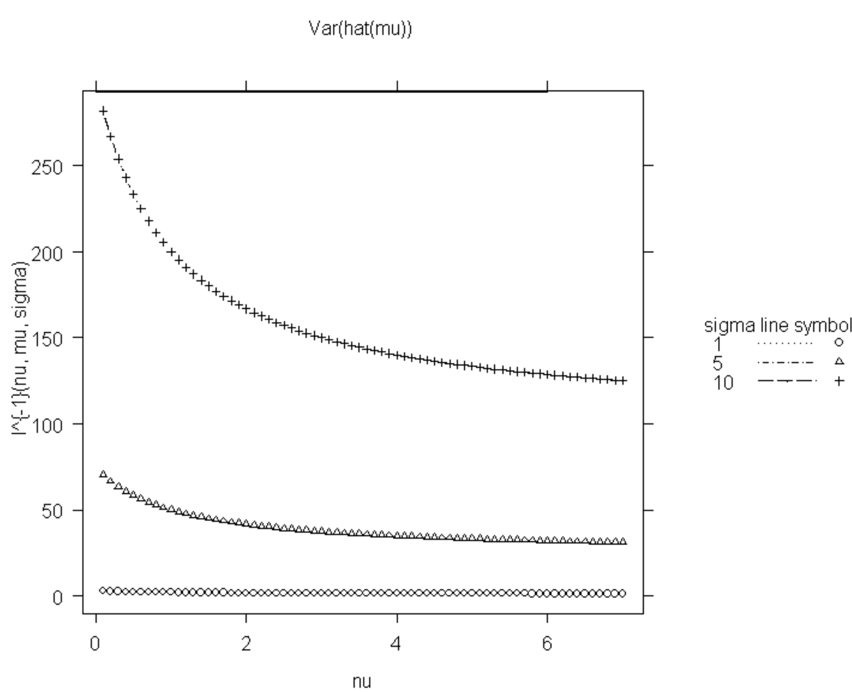

Figure 3.2: Asymptotic Variance of μ.

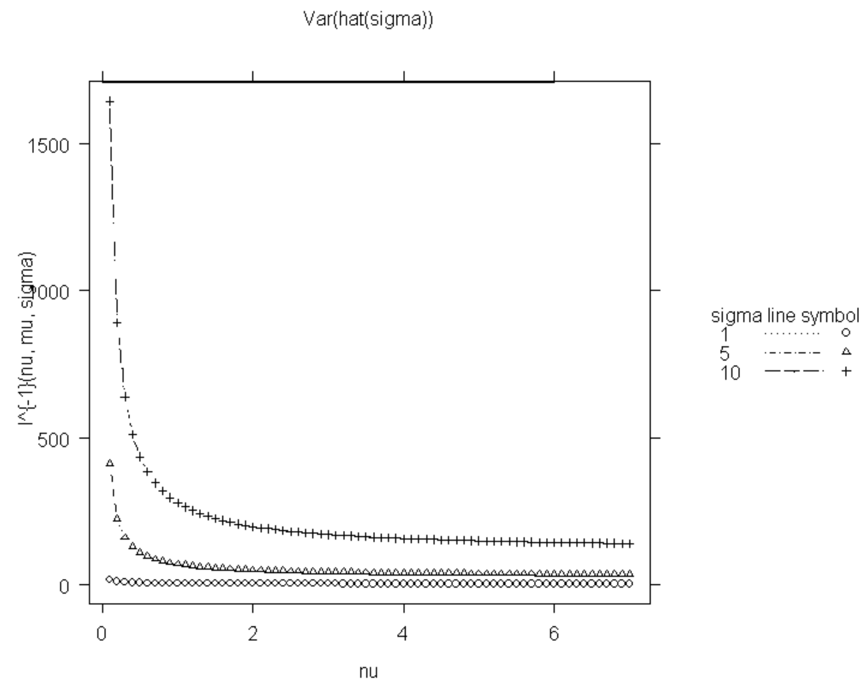

Figure 3.3: Asymptotic Variance of σ.

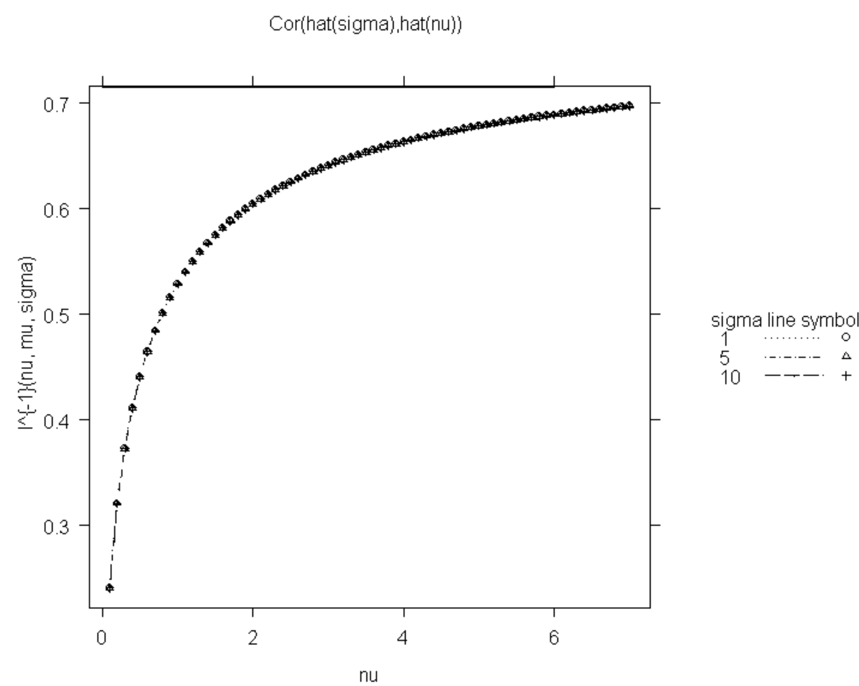

Figure 3.4: Asymptotic Correlation of σ and ν.

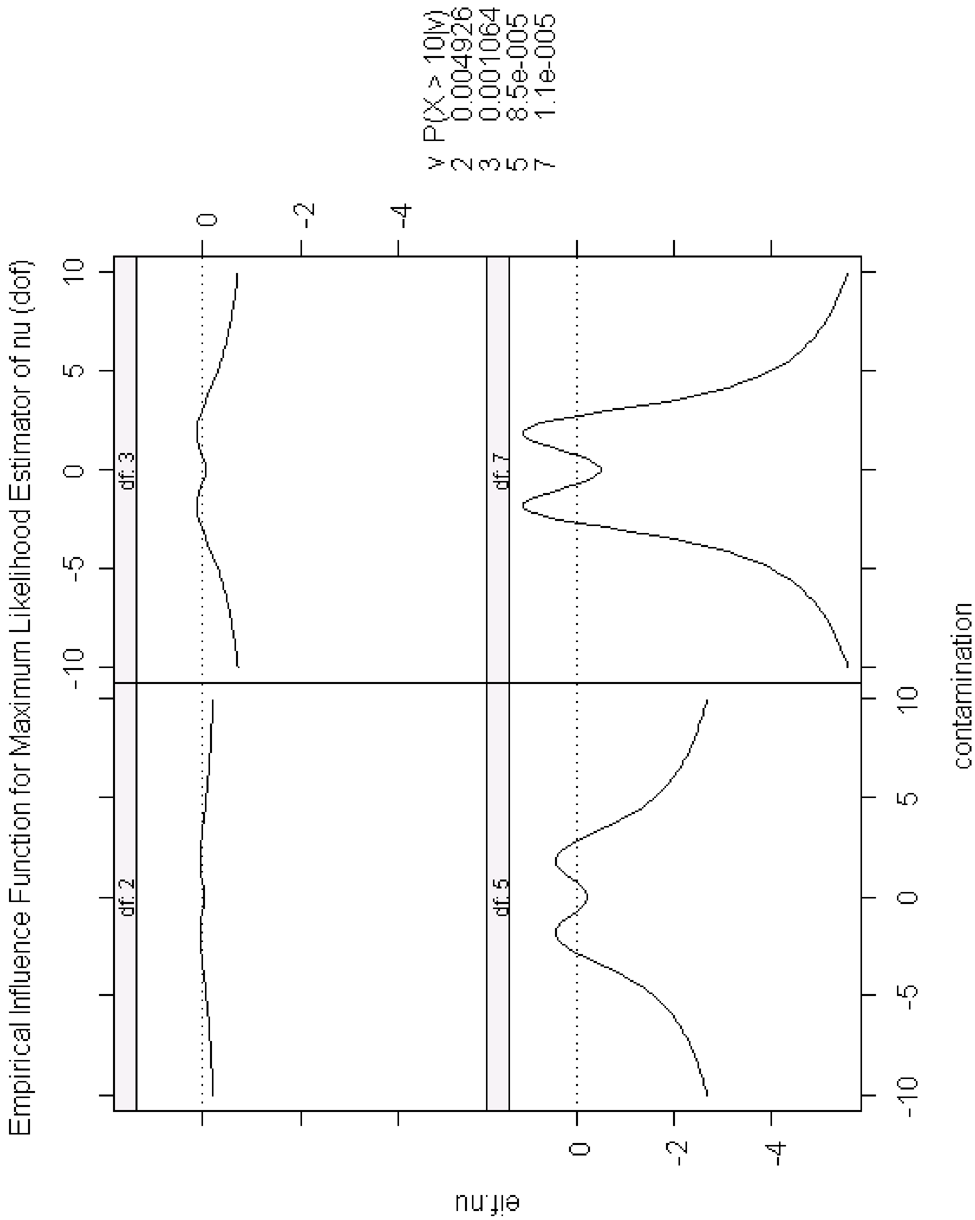

Figure 3.5: Empirical influence functions for maximum likelihood estimator of ν.

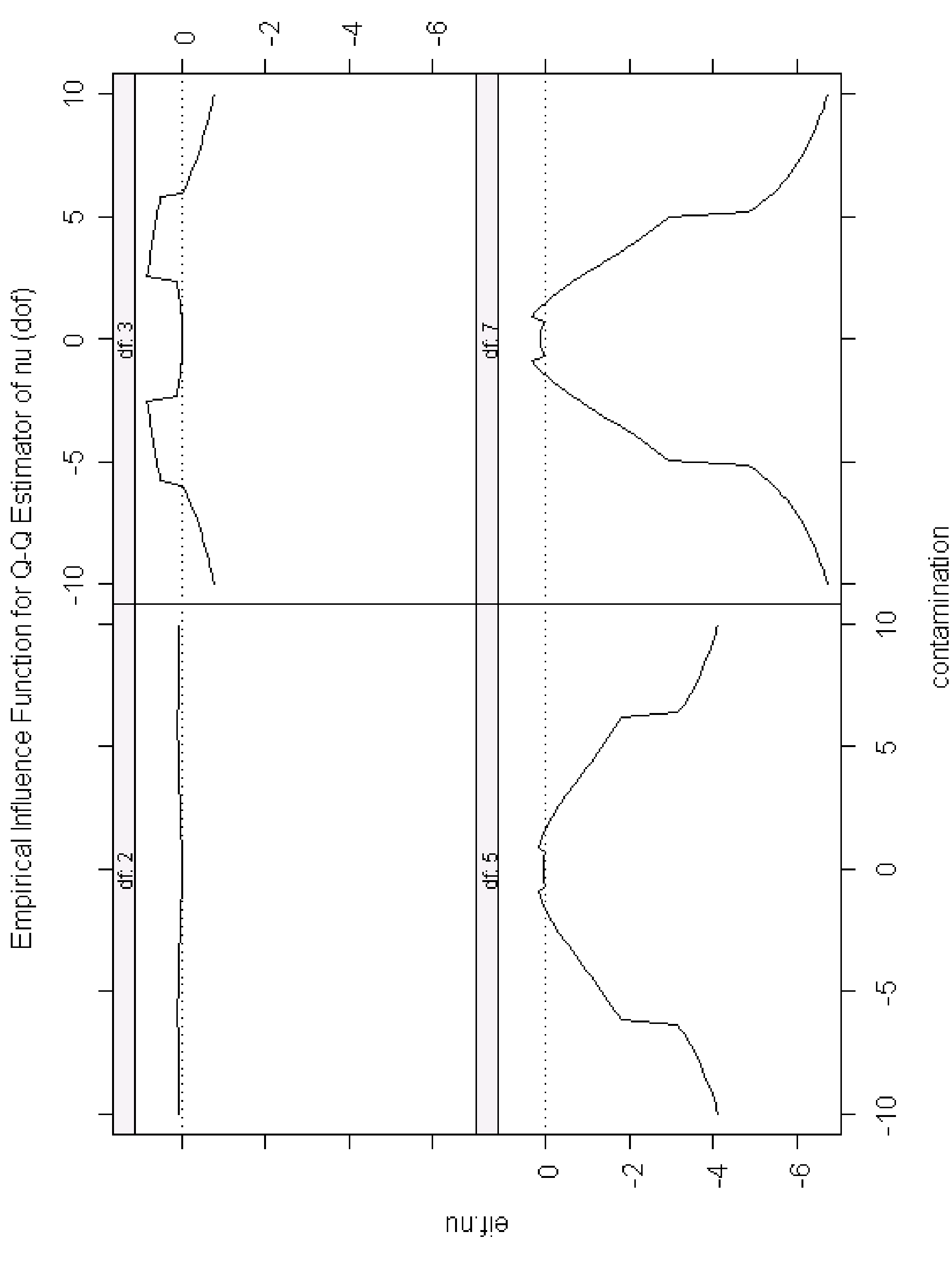

Figure 3.6: Empirical influence functions for Q-Q estimator of ν.

Chapter 4

Performance of Estimators

4.1 Estimator comparisons

In order to compare the performance (bias and variance) of the estimators we have discussed in Chapter 2 and 3, we performed a Monte Carlo study similar to that performed by Rachev and Mittnik [RM00]. We used a sample size of $n = 1000$ and 400 replications for each set of parameters. The results are displayed as boxplots below.

4.1.1 Estimators of stable distribution parameters

We studied the four estimators of the stable law parameters we discussed in Chapter 2—McCulloch's quantile estimator, Kogon and Williams's regression-based estimator, the maximum likelihood estimator, and the modified Q-Q estimator. We considered all combinations of the parameter sets $\alpha \in \{1.25, 1.50, 1.75, 2.00\}$, $\beta \in \{-1, -0.5, 0, 0.5, 1\}$. The scale and location parameters were standardized to 1 and 0. The resulting boxplots are shown in Figures 4.1-4.20.

Results

- As an estimator of α, the MLE was (in all scenarios) the best estimator (in terms of the median and the spread of the estimator). The modified Q-Q estimator was generally the worst—it was often biased and almost always showed large variability. The Kogon-Williams estimator was better than McCulloch's quantile estimator but slightly more variable than the MLE. This suggests that, in situations where computational speed is a factor, the Kogon-Williams estimator is the best substitute for the MLE (as an estimator of α).

- For estimation of β, there are two main points to take away from the boxplots 4.6-4.10: (a) away from $\alpha = 2$, the MLE and the Kogon-Williams estimator are the best estimators; and (b) near $\alpha = 2$, all estimators have difficulty estimating β (which makes sense, since the skewness parameter becomes meaningless at $\alpha = 2$).

- While all estimators are relatively unbiased for the scale parameter, they are all very variable.

- All estimators are good at estimating the location parameter for α sufficiently far away from 1.

4.1.2 Estimators of t distribution parameters

We studied the two estimators of the t law parameters we discussed in Chapter 3—the maximum likelihood estimator and the (unmodified) Q-Q estimator. We considered the parameter set $\nu \in \{2,3,5,7\}$. The scale and location parameters were standardized to 1 and 0. The resulting boxplots are shown in Figures 4.21-4.23.

Results

- The MLE and the Q-Q estimator perform similarly as estimators of ν.
- The Q-Q estimator is biased as an estimator of the scale parameter, but this bias decreases as ν increases.
- Both estimators have no trouble estimating the location parameter; the Q-Q estimator is very variable at smaller values of ν, though.

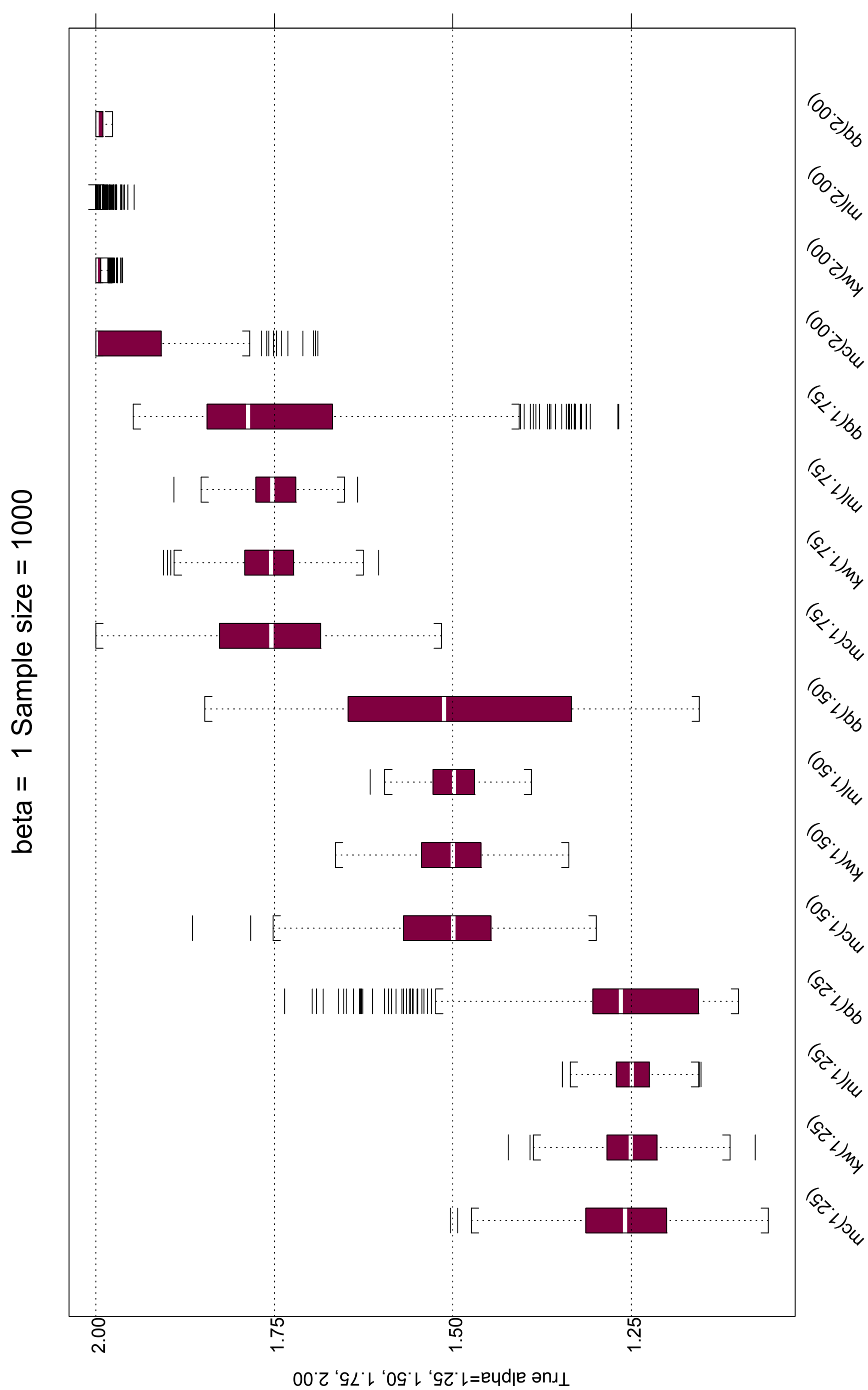

Figure 4.1: Boxplot of stable law parameter estimators of α for $\beta = 1$.

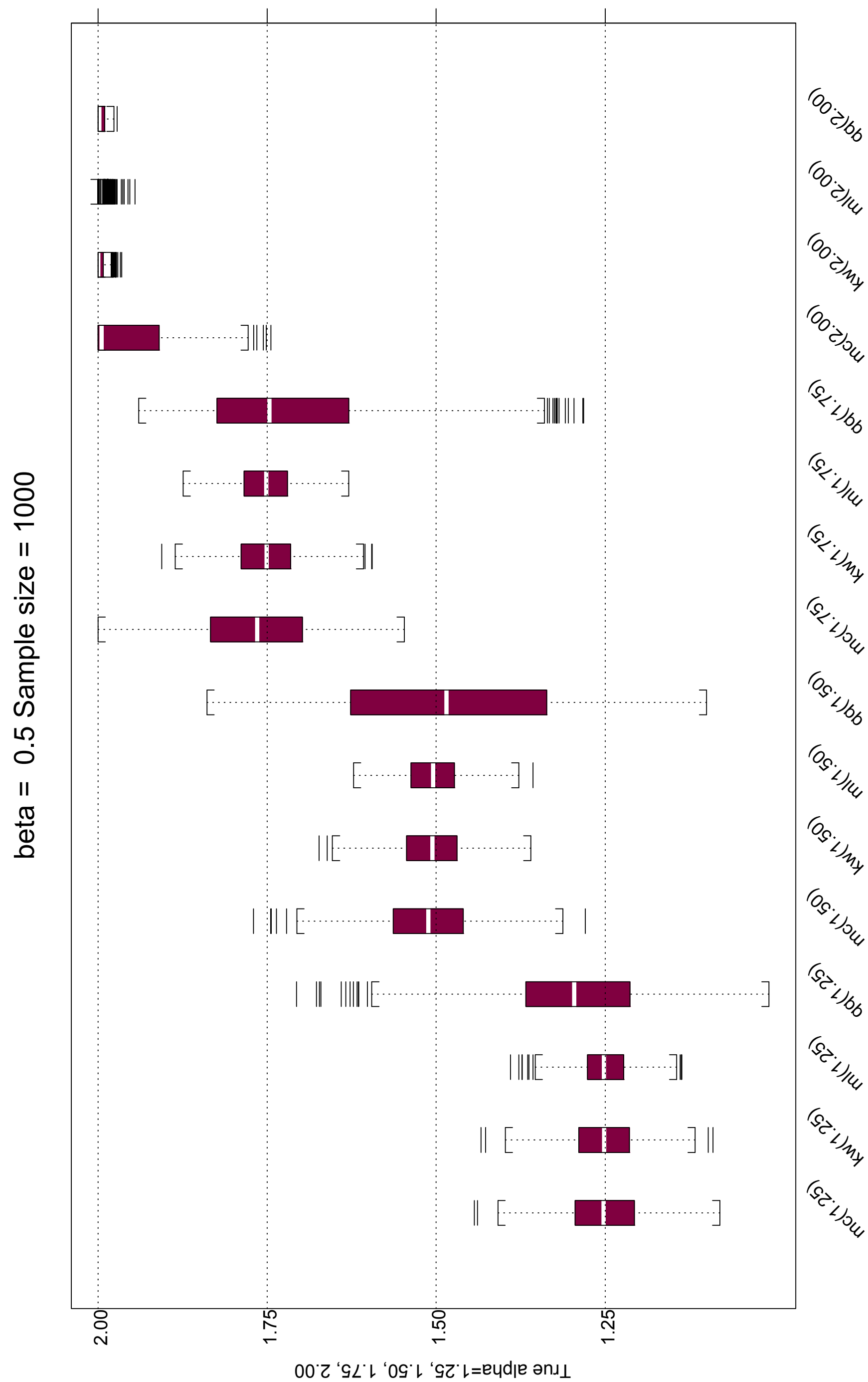

Figure 4.2: Boxplot of stable law parameter estimators of α for $\beta = 0.5$.

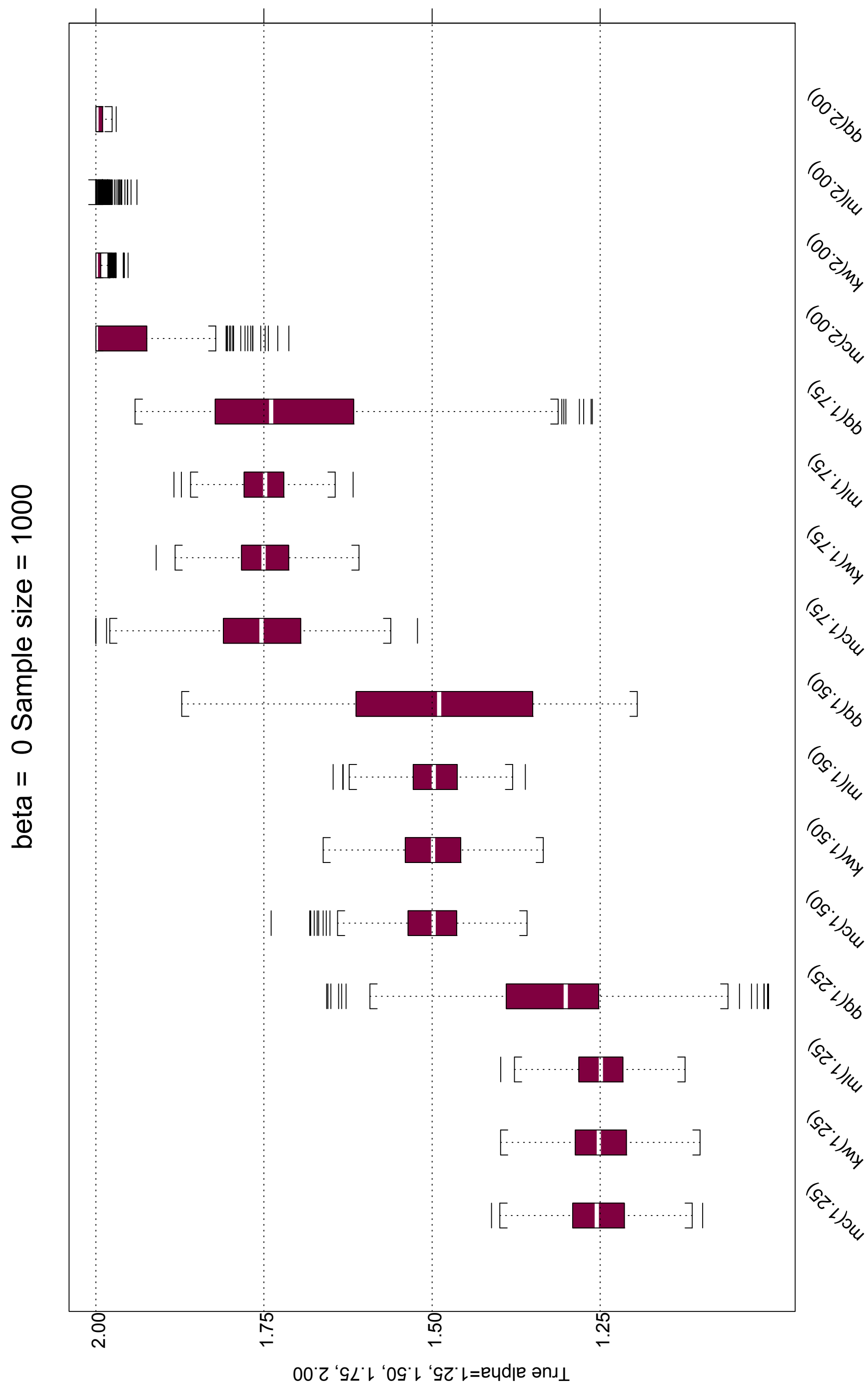

Figure 4.3: Boxplot of stable law parameter estimators of α for $\beta = 0$.

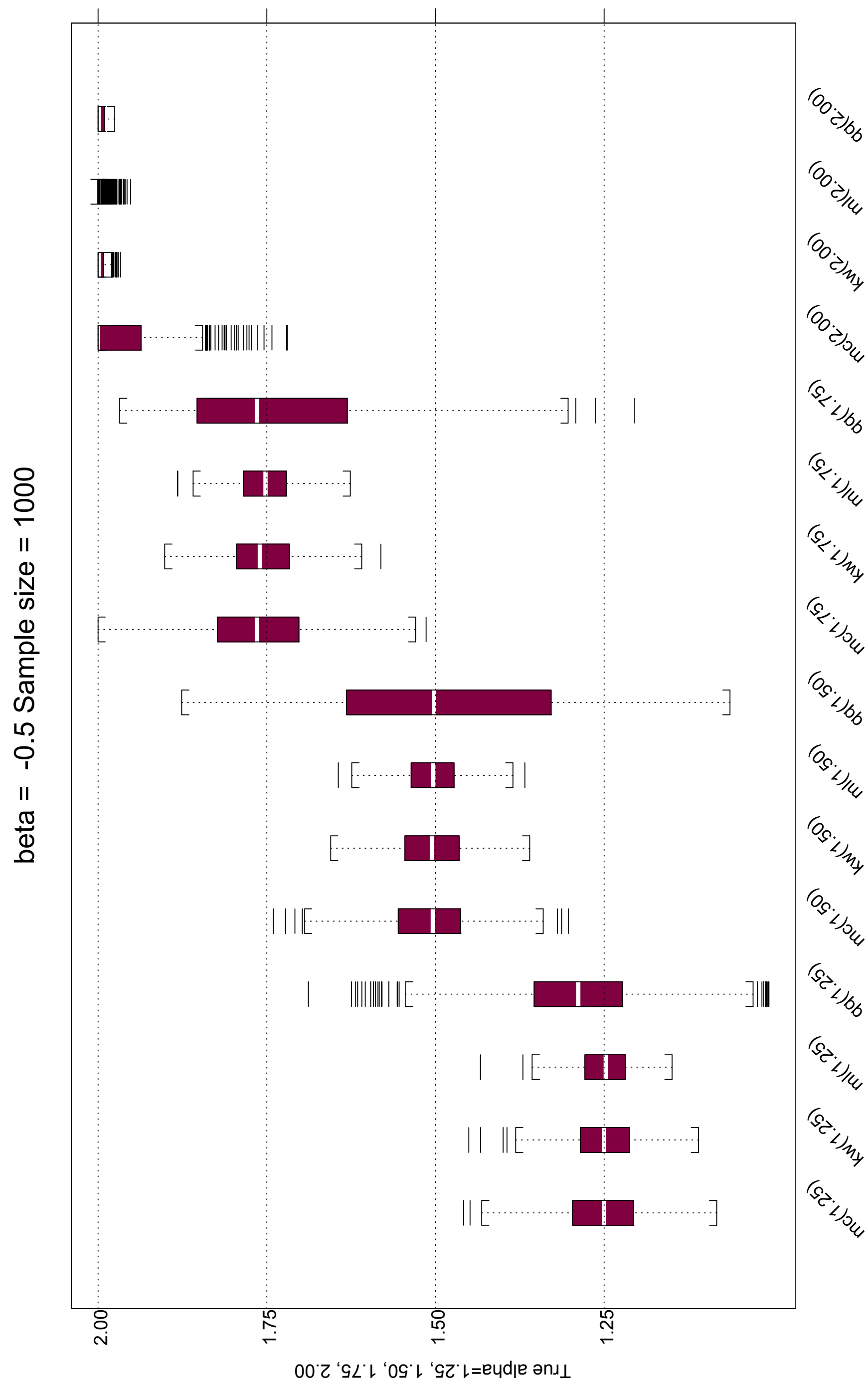

Figure 4.4: Boxplot of stable law parameter estimators of α for $\beta = -0.5$.

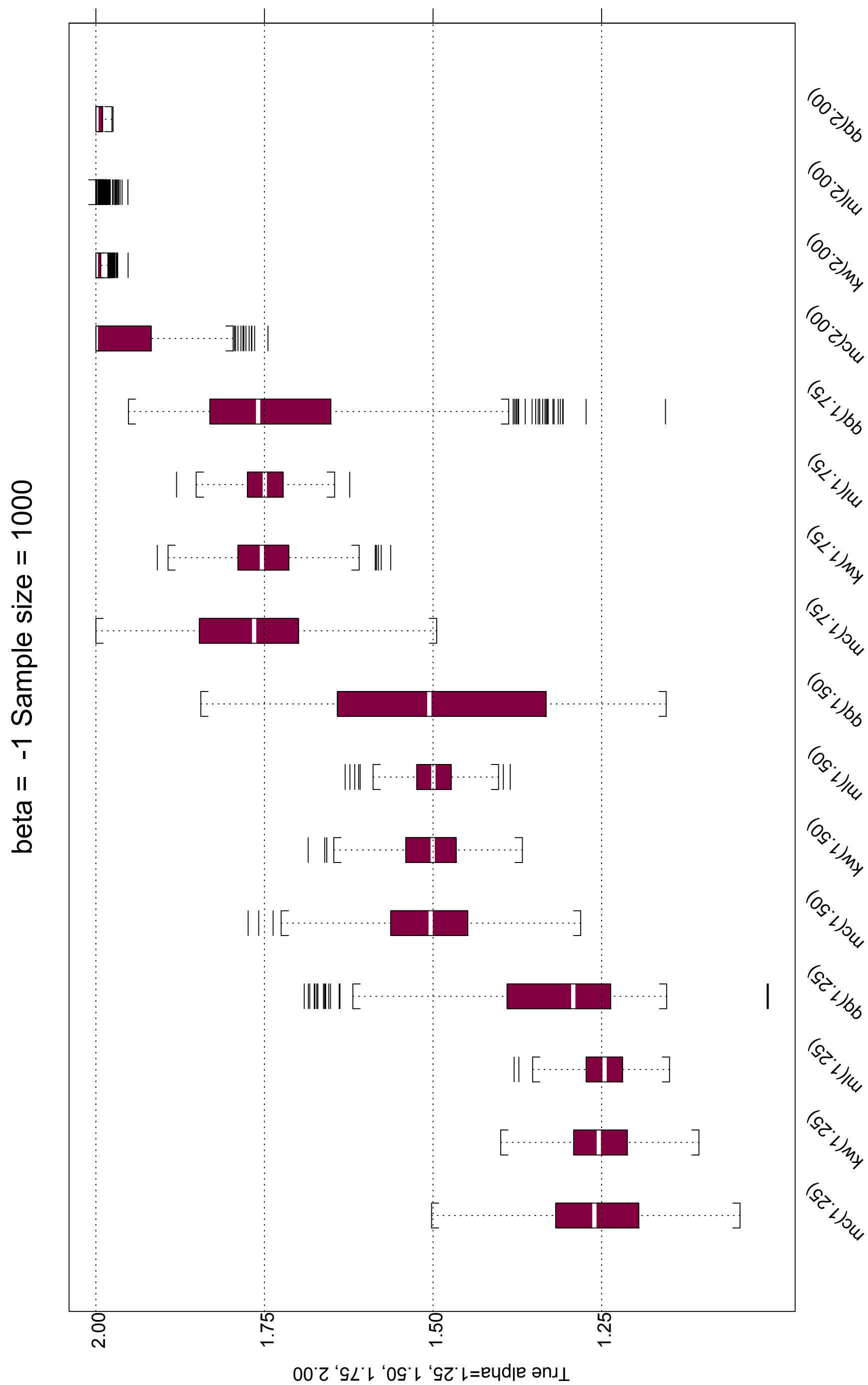

Figure 4.5: Boxplot of stable law parameter estimators of α for $\beta = -1$.

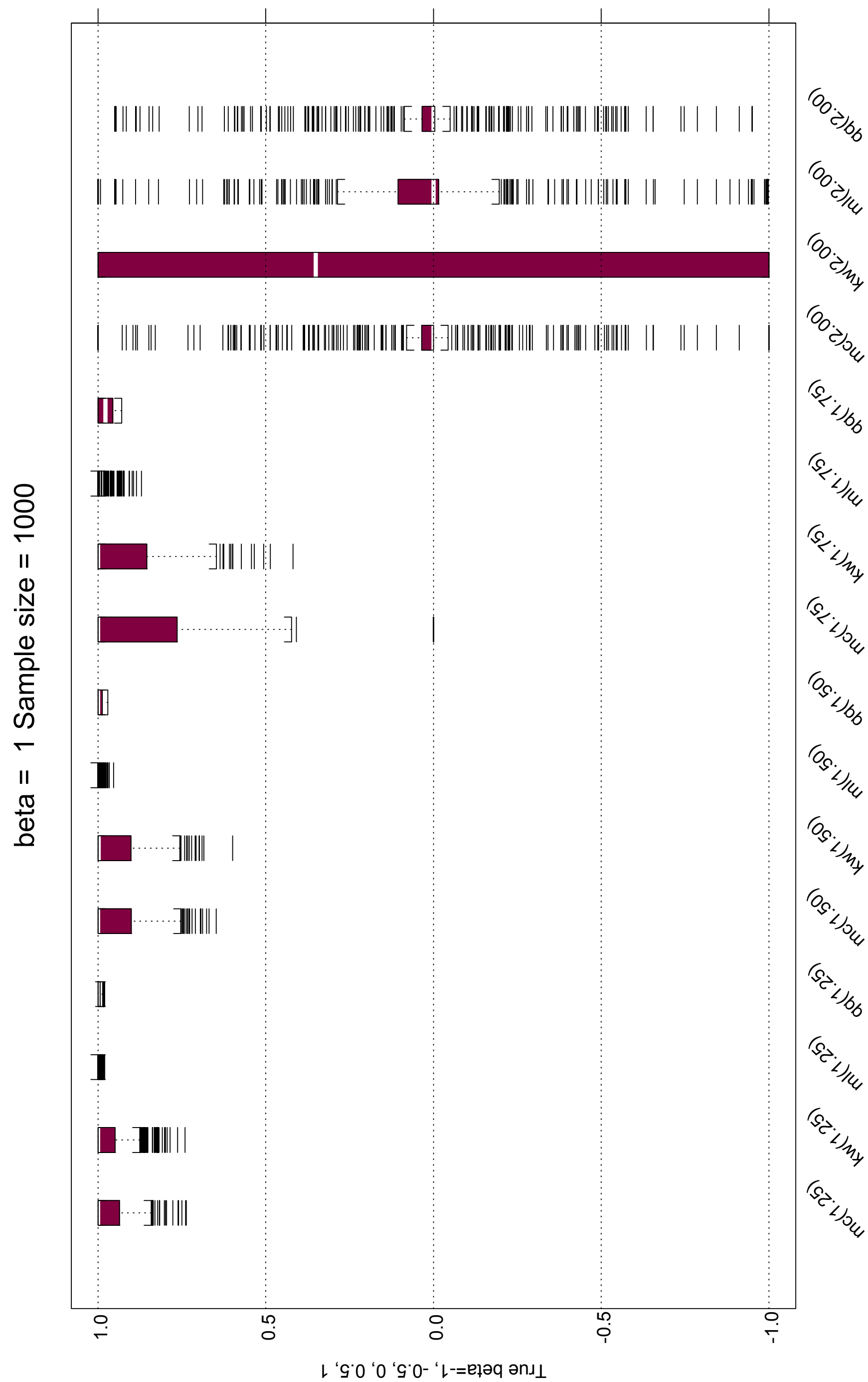

Figure 4.6: Boxplot of stable law parameter estimators of β for $\beta = 1$.

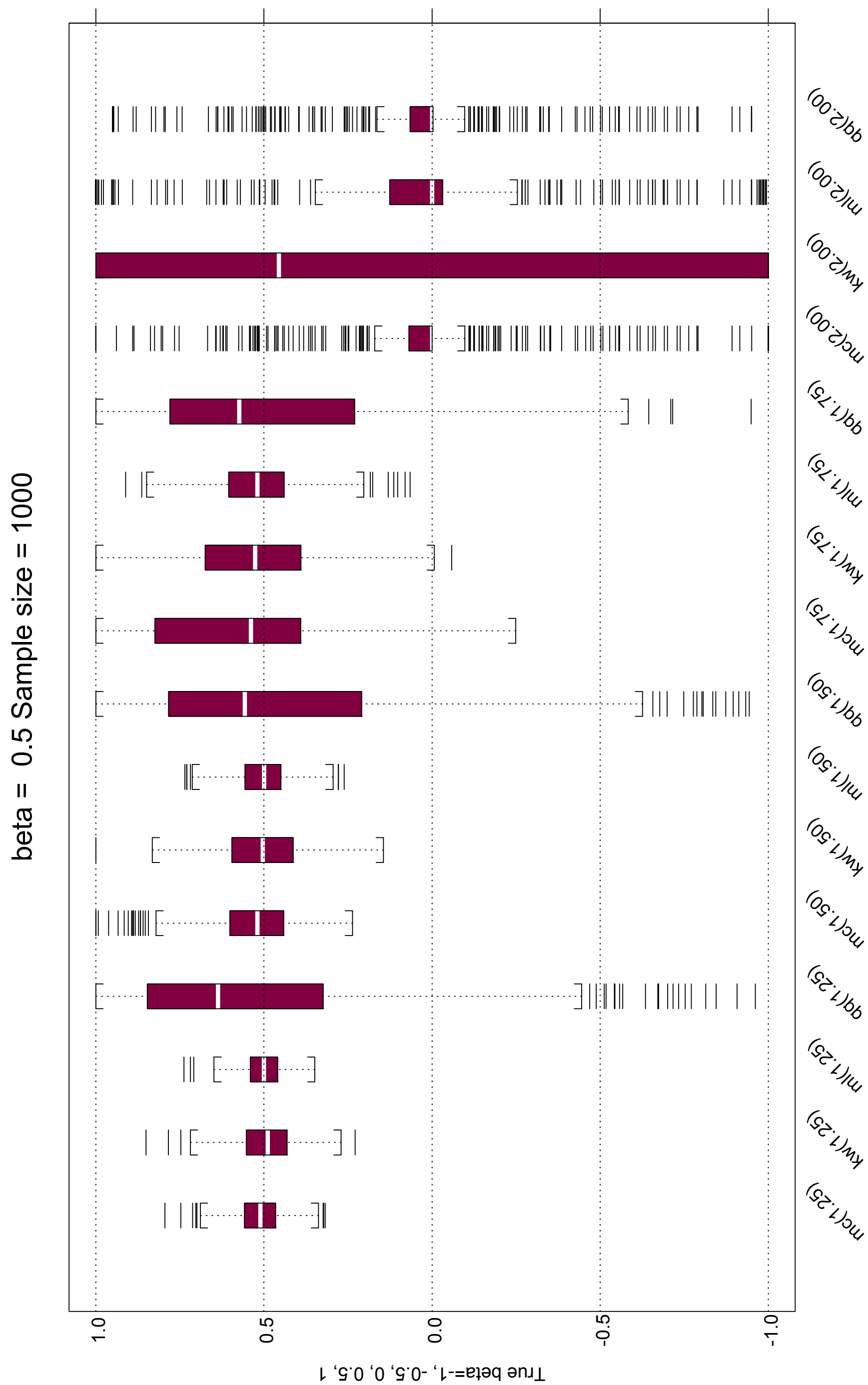

Figure 4.7: Boxplot of stable law parameter estimators of β for $\beta = 0.5$.

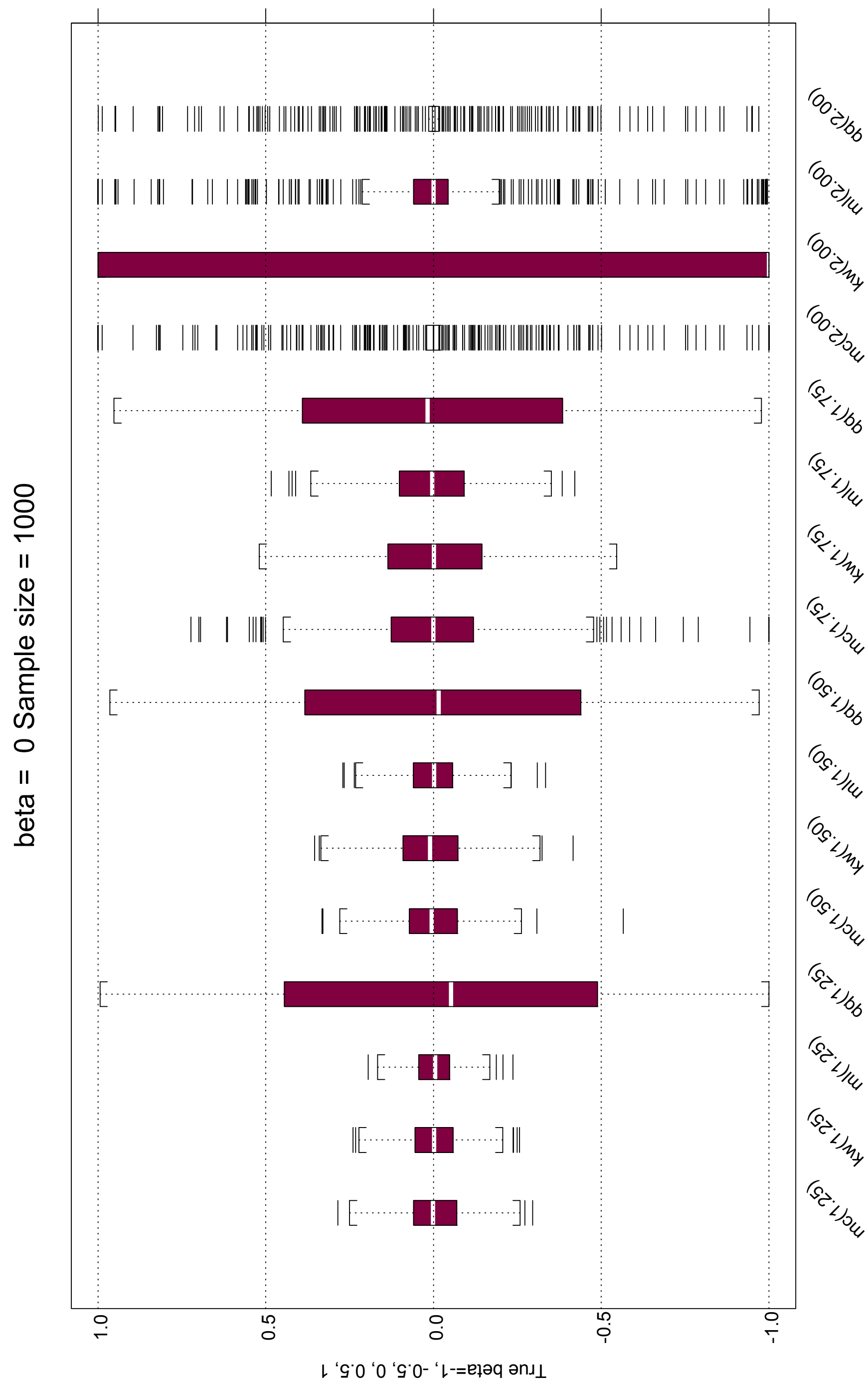

Figure 4.8: Boxplot of stable law parameter estimators of β for $\beta = 0$.

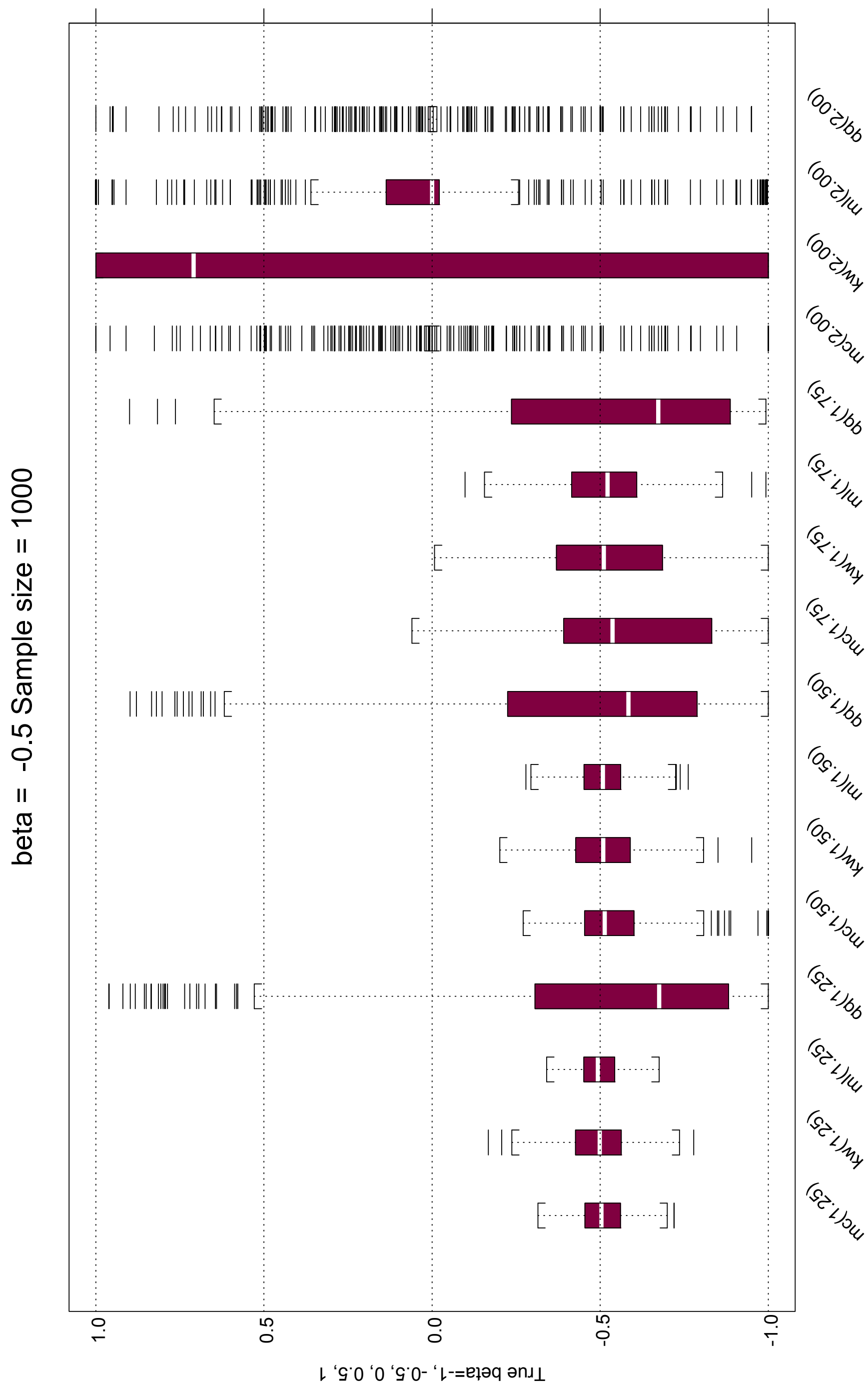

Figure 4.9: Boxplot of stable law parameter estimators of β for $\beta = -0.5$.

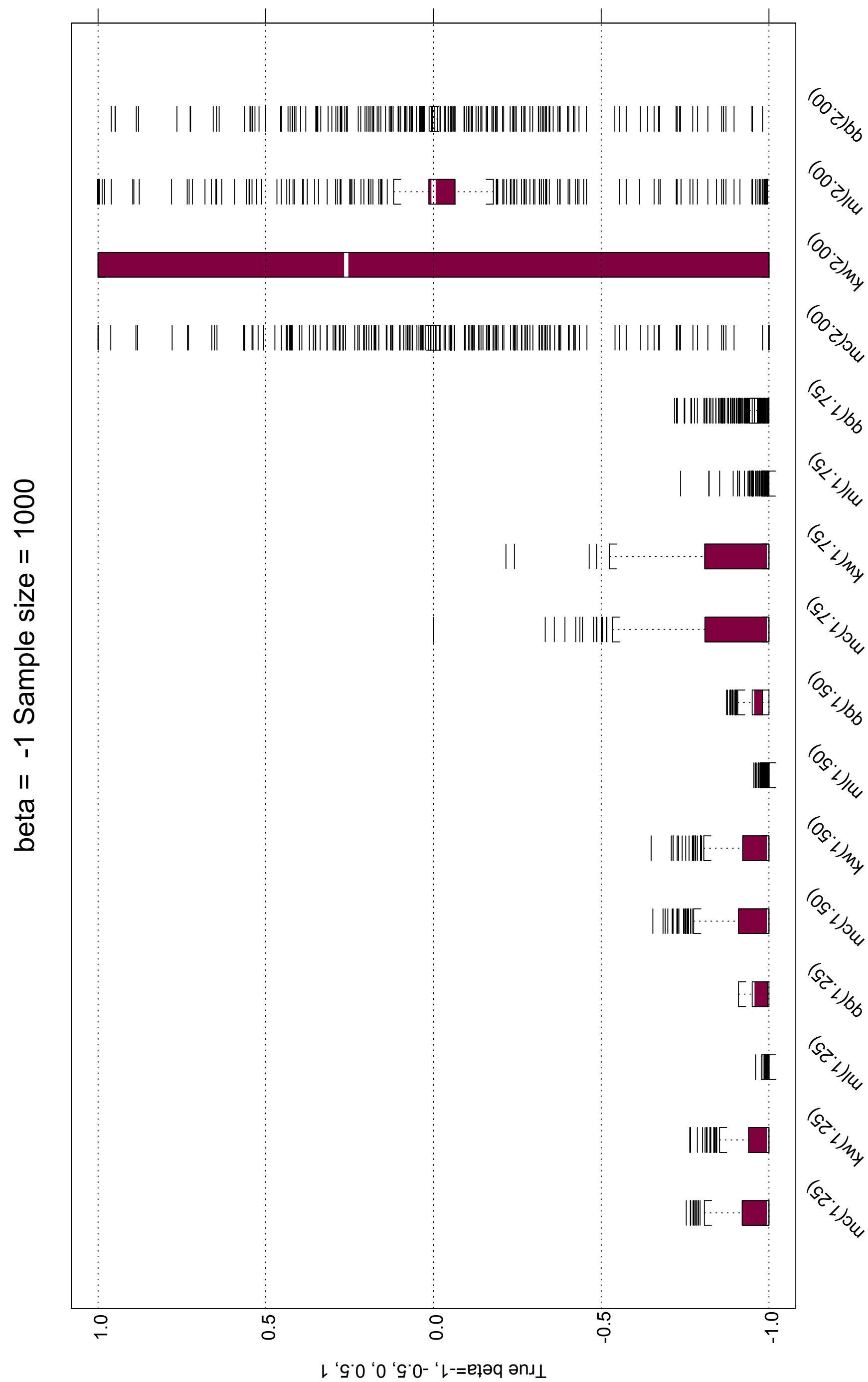

Figure 4.10: Boxplot of stable law parameter estimators of β for $\beta = -1$.

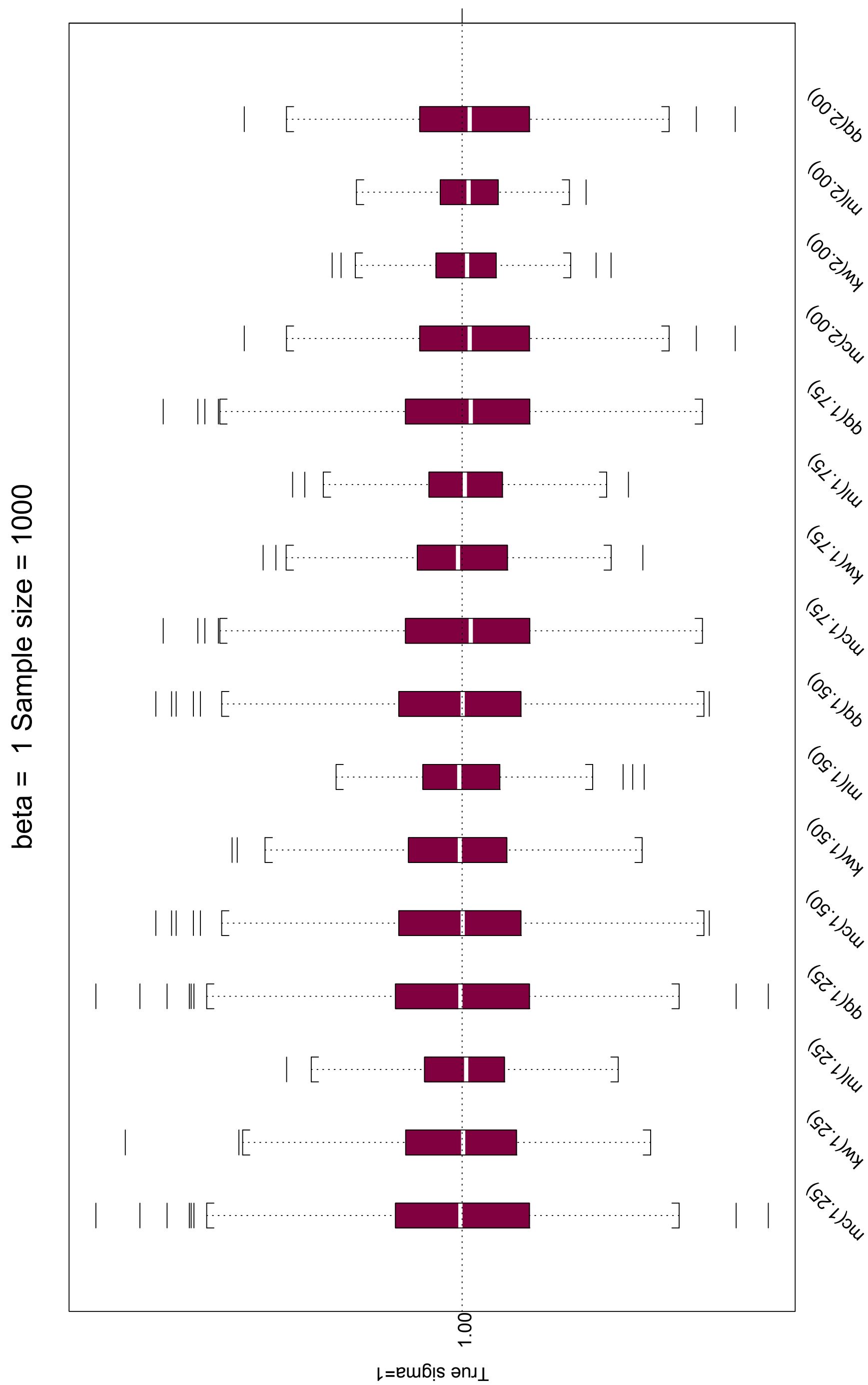

Figure 4.11: Boxplot of stable law parameter estimators of σ for $\beta = 1$.

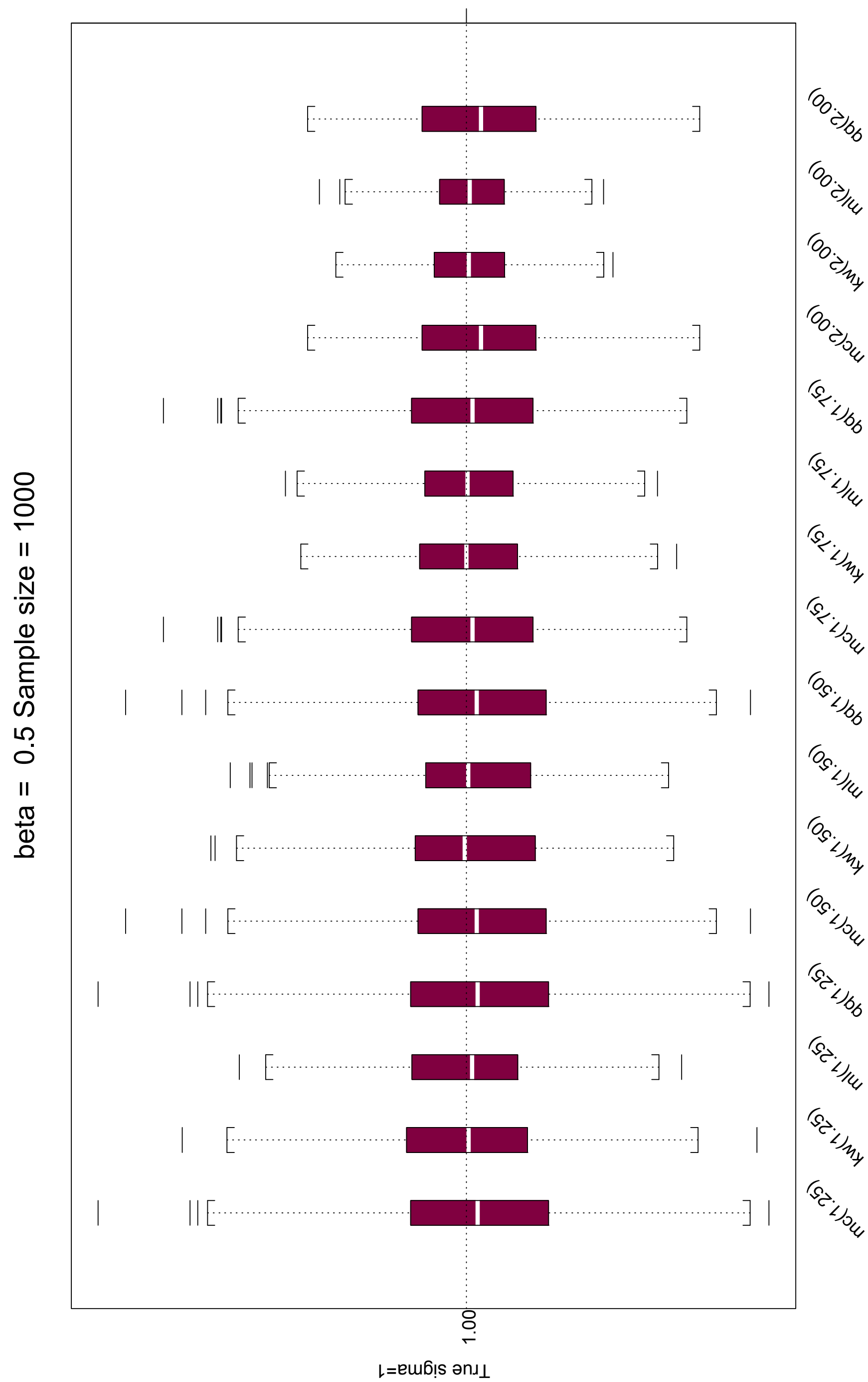

Figure 4.12: Boxplot of stable law parameter estimators of σ for $\beta = 0.5$.

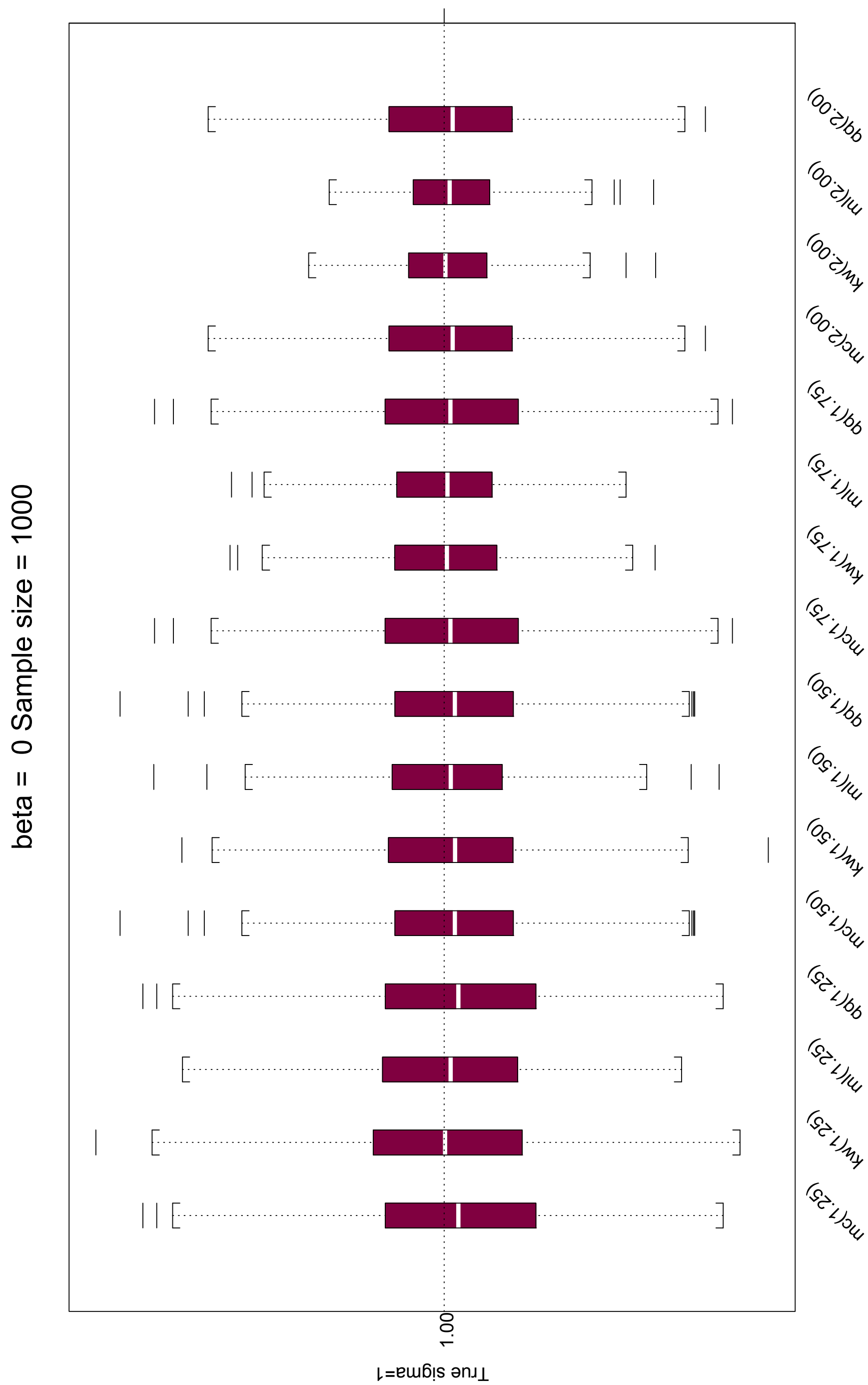

Figure 4.13: Boxplot of stable law parameter estimators of σ for $\beta = 0$.

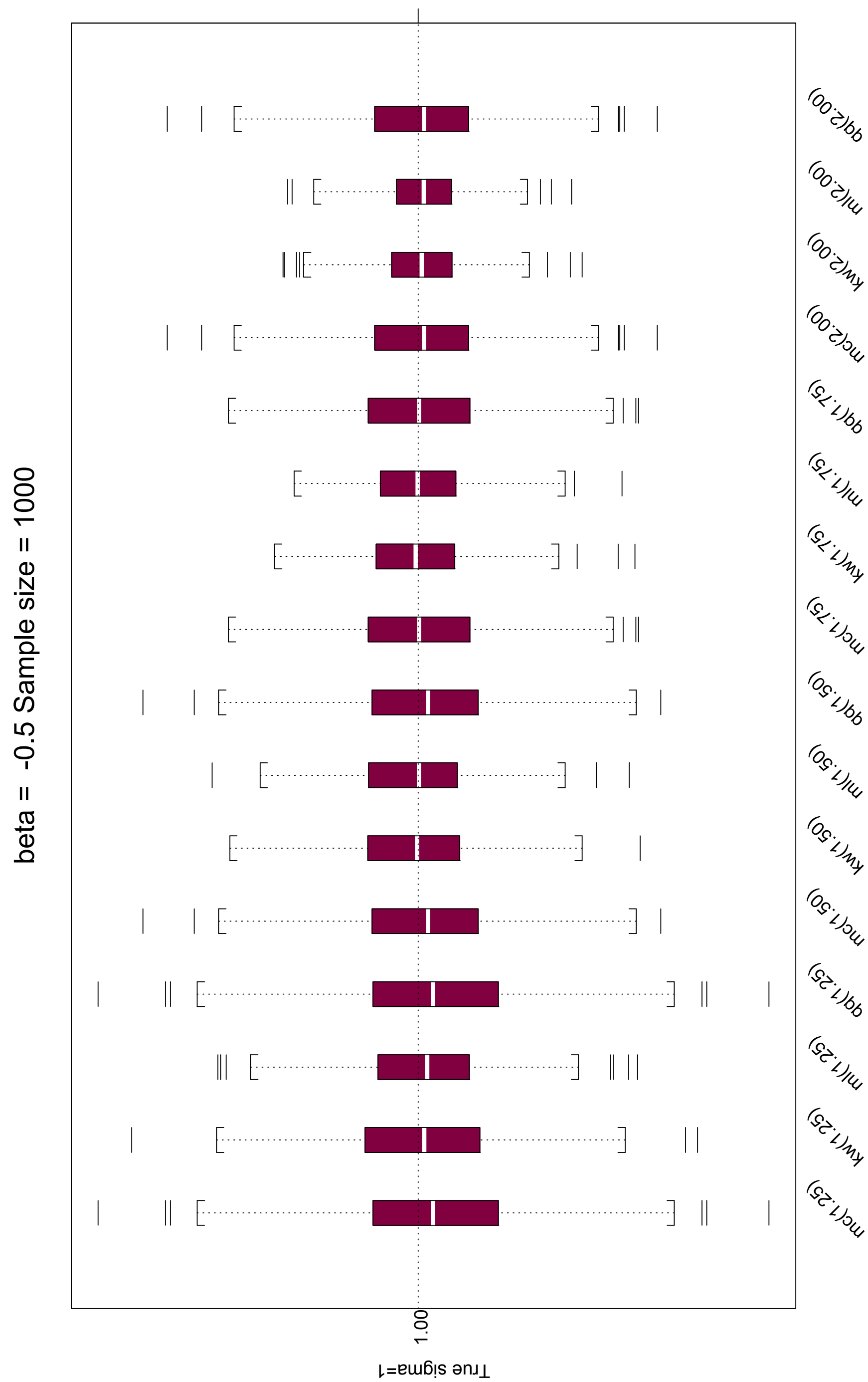

Figure 4.14: Boxplot of stable law parameter estimators of σ for $\beta = -0.5$.

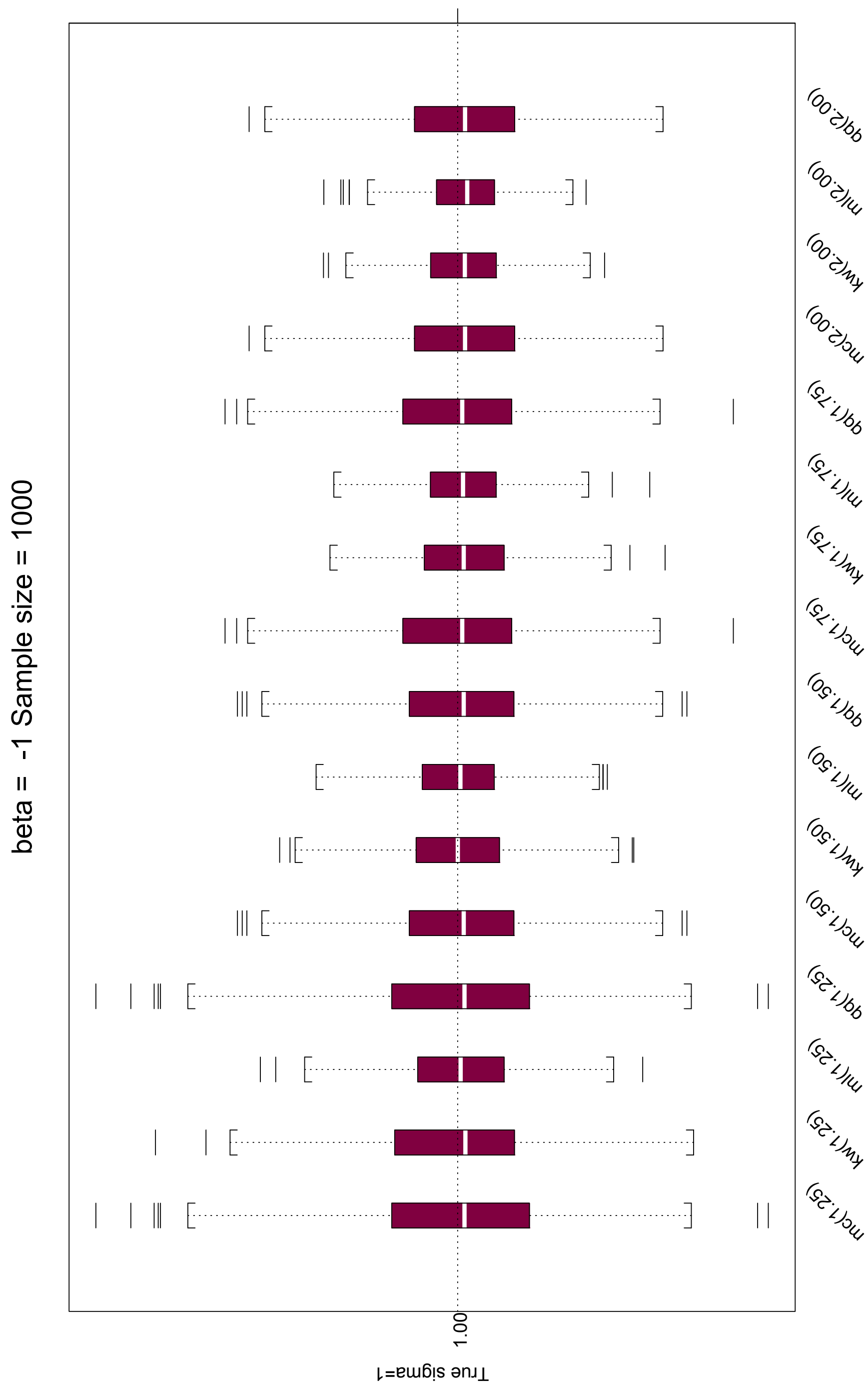

Figure 4.15: Boxplot of stable law parameter estimators of σ for $\beta = -1$.

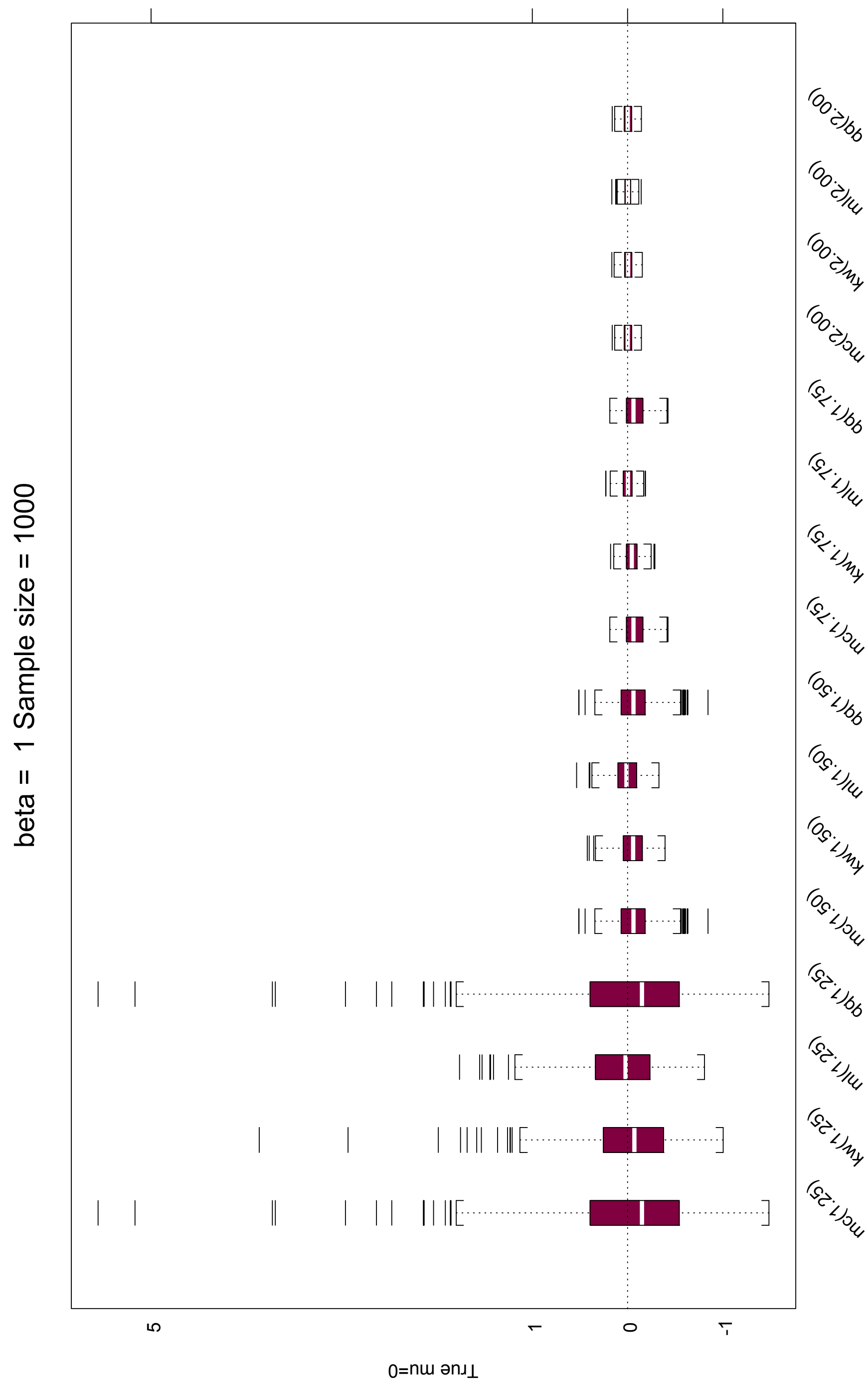

Figure 4.16: Boxplot of stable law parameter estimators of μ for $\beta = 1$.

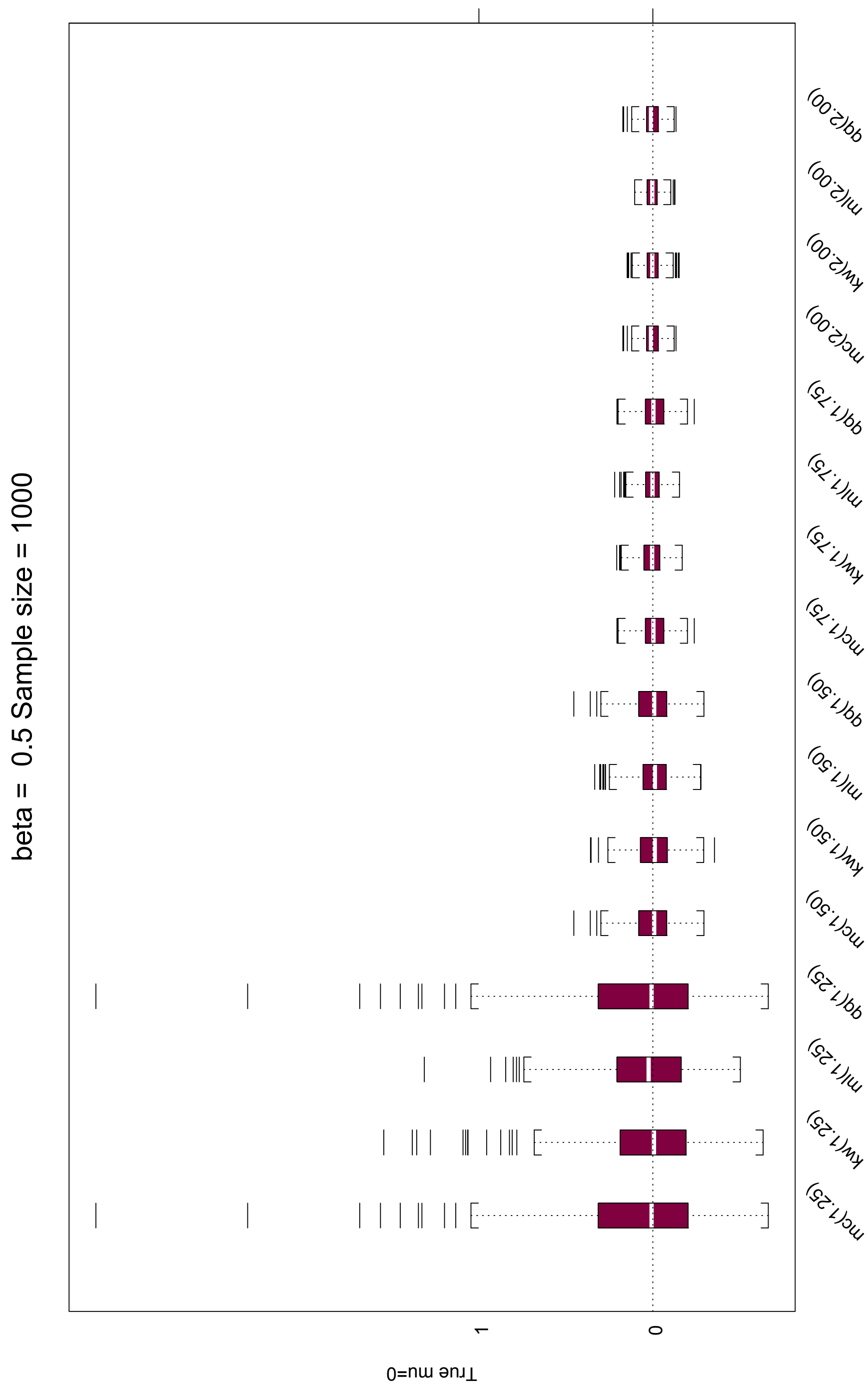

Figure 4.17: Boxplot of stable law parameter estimators of μ for $\beta = 0.5$.

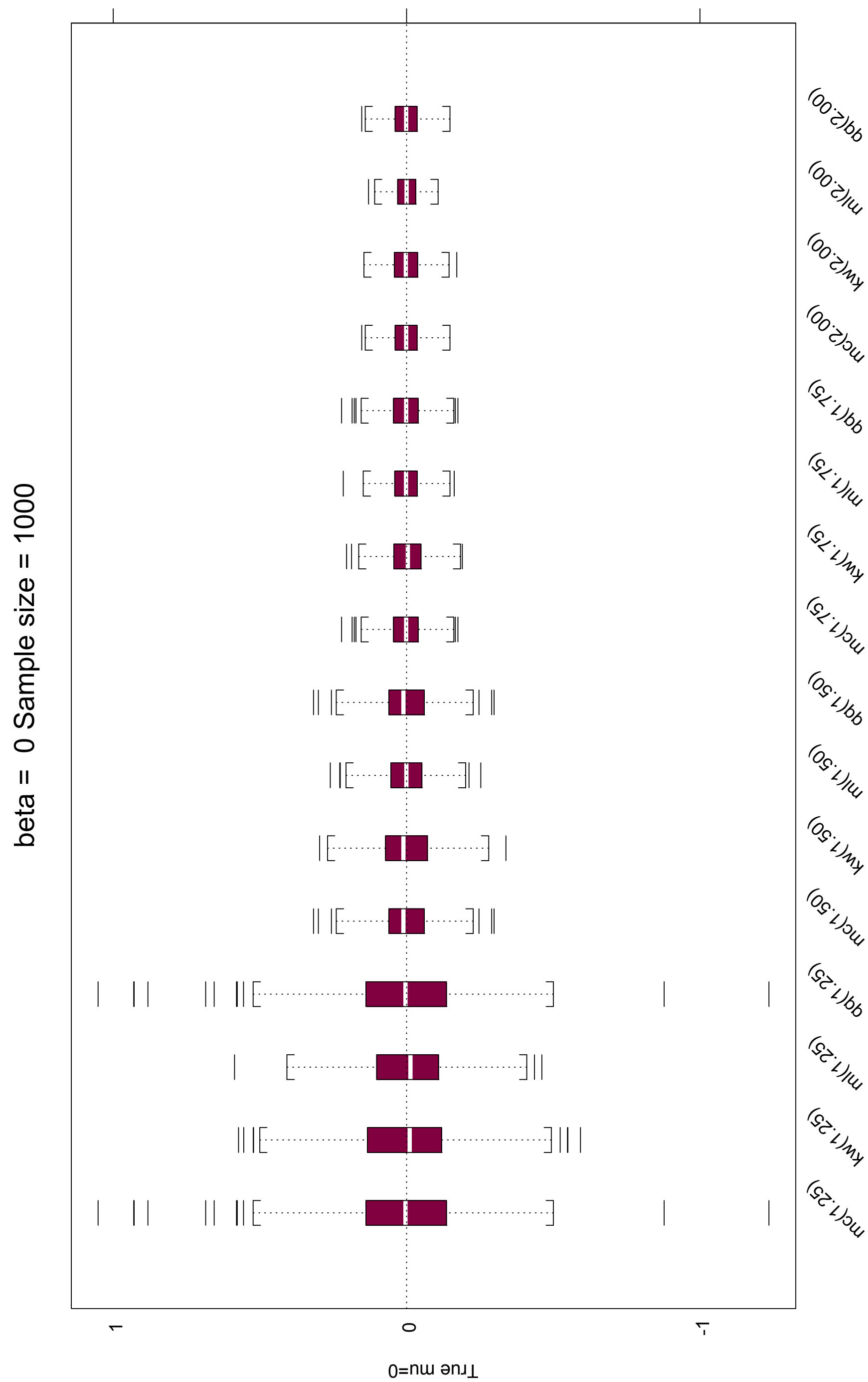

Figure 4.18: Boxplot of stable law parameter estimators of μ for $\beta = 0$.

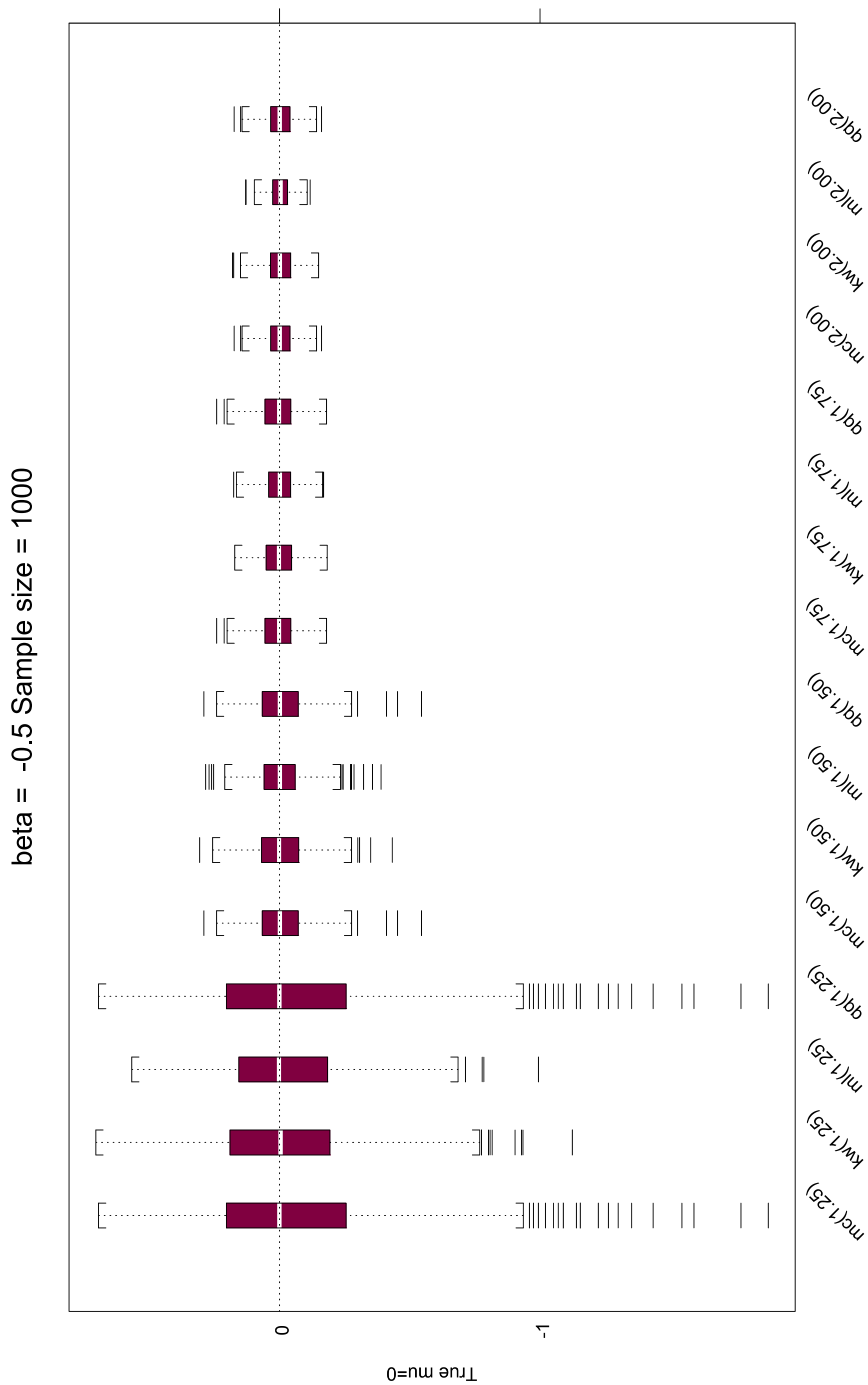

Figure 4.19: Boxplot of stable law parameter estimators of μ for $\beta = -0.5$.

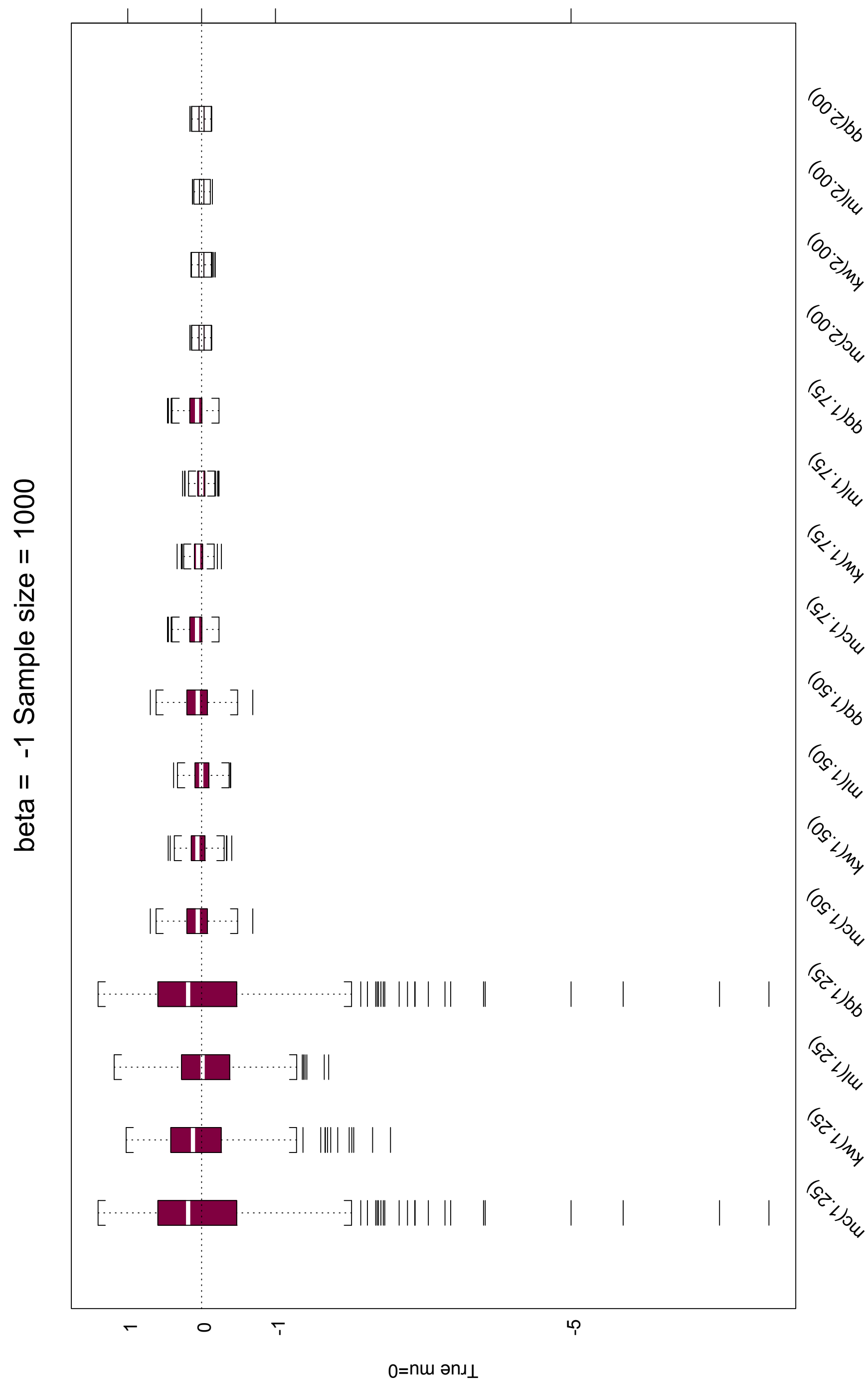

Figure 4.20: Boxplot of stable law parameter estimators of μ for $\beta = -1$.

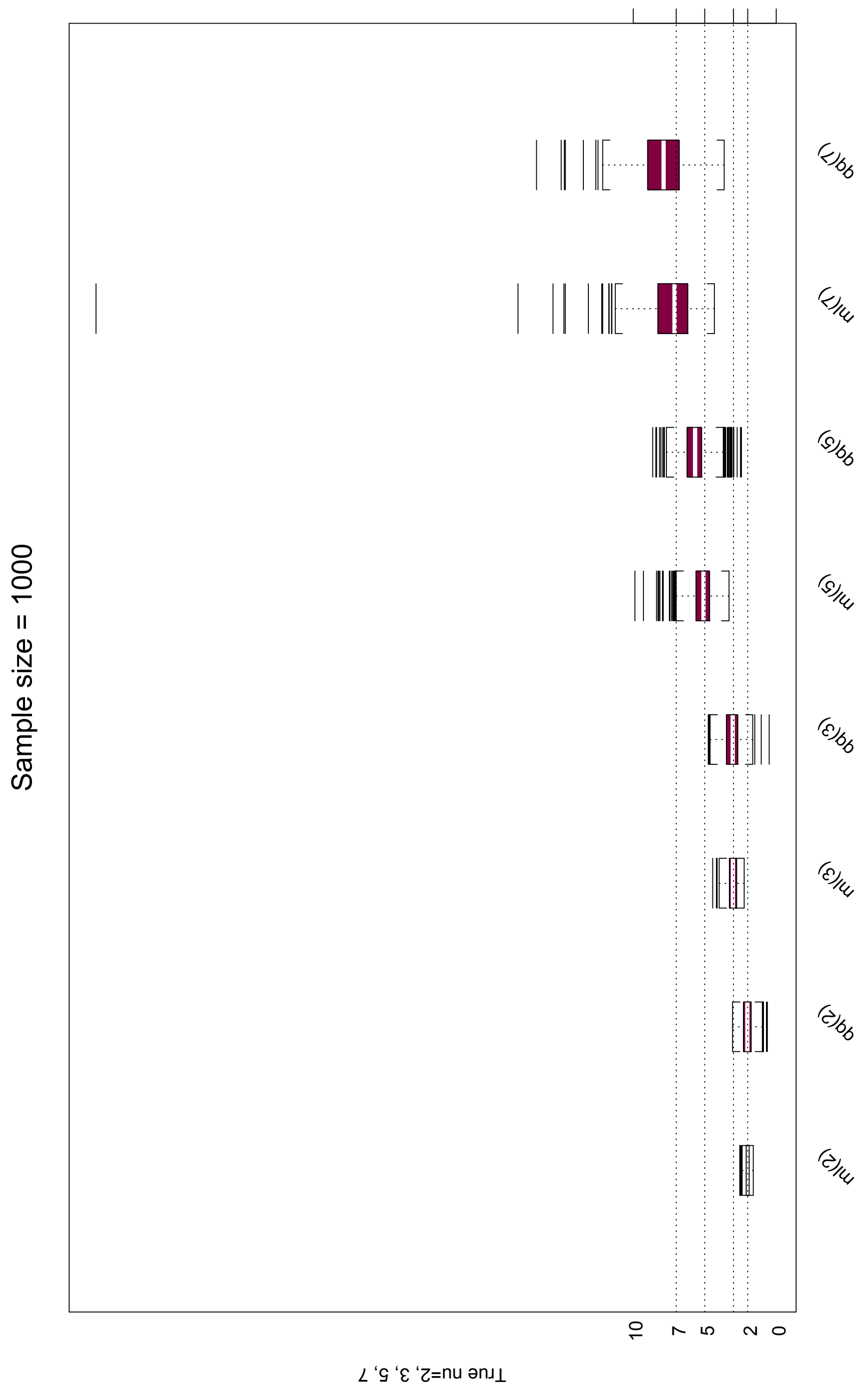

Figure 4.21: Boxplot of t law parameter estimators of ν.

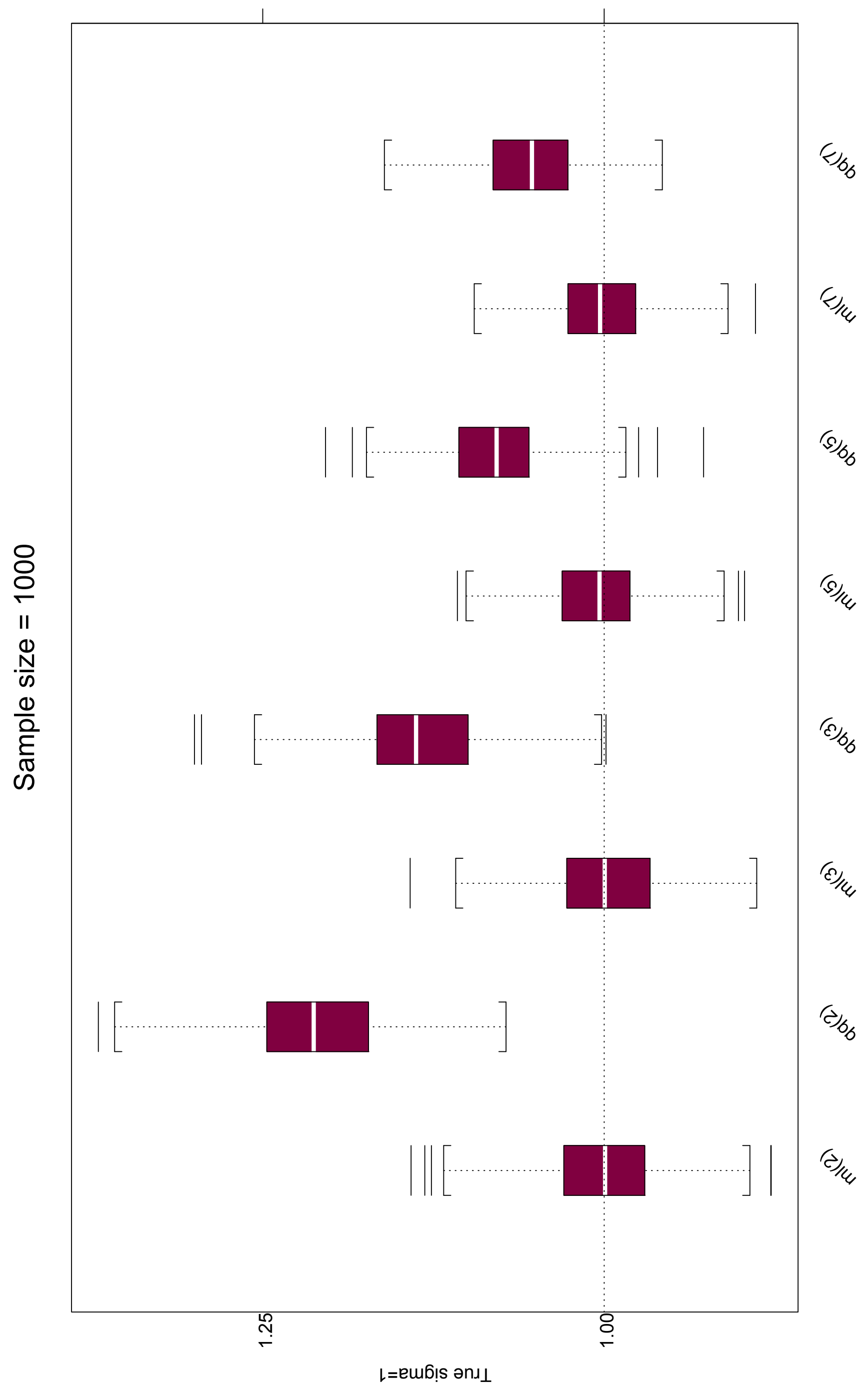

Figure 4.22: Boxplot of t law parameter estimators of σ.

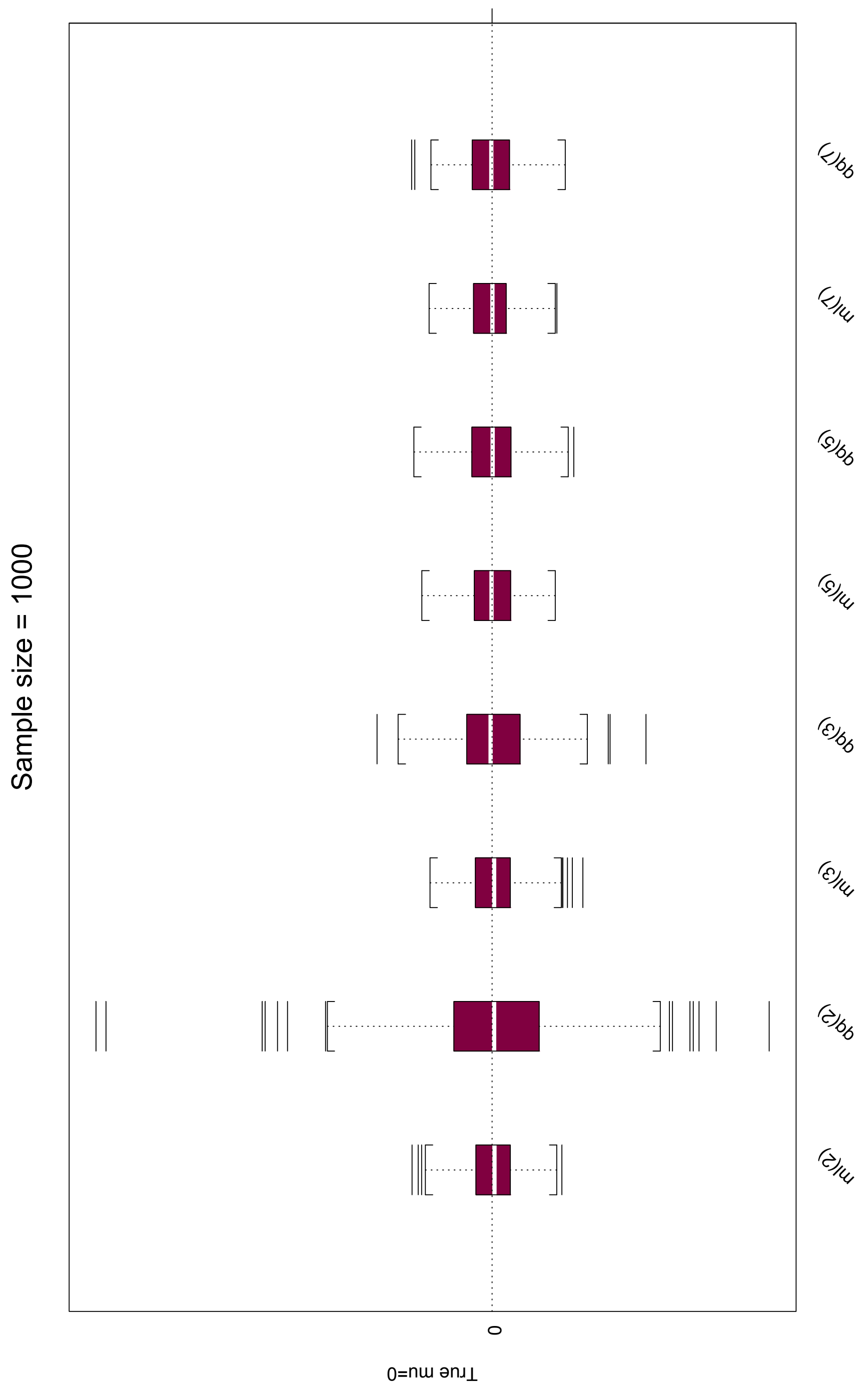

Figure 4.23: Boxplot of t law parameter estimators of μ.

Chapter 5

The Relationship Between Tail-Fatness and Firm Size

5.1 Introduction

It has been empirically observed that firm size is related to tail fatness; specifically, the distribution of returns for large firms tends to look fairly normal, while small firms tend to have return distributions with heavier tails than a normal distribution.

5.2 Stable Law Parameters and Firm Size

5.2.1 Data Setup

We obtained four years of daily returns and size[1] data on firms (with common stock) listed on the three major exchanges (NYSE, AMEX, Nasdaq) from the CRSP database [CRS]. Due to computational limitations, we split each firm's data by year (2001-2004). We fit a stable law to each firm's data using the maximum likelihood estimator. Financial returns are usually assumed to have a mean, so we constrain α to lie in $[1,2]$ during the fit. Since there are convergence issues when small samples with the MLE for the stable laws (due to the approximations), we eliminated firms with fewer than 100 observations or any missing return values. This left us with approximately 7000 firms each year.

Our density approximation is known to be inaccurate near $\alpha = 1$. Therefore, we do not trust any computed values that are less than 1.25, and we have removed all such firms from our sample (approximately 500 firms each year).

5.2.2 Statistics

Summary statistics for the MLE's of the stable law parameters are shown in Tables 5.1-5.4. Histograms of the estimates by year are shown in Figures 5.1-5.4.

[1]Here firm size is measured by the logarithm of market equity—the product of the price per share and the number of shares outstanding of the firm's stock.

Estimates of α

	2001	2002	2003	2004
Min	1.250	1.250	1.250	1.250
25%	1.520	1.560	1.580	1.600
Median	1.640	1.690	1.720	1.730
Mean	1.640	1.680	1.700	1.720
75%	1.760	1.810	1.840	1.850
Max	2.000	2.000	2.000	2.000
Std	0.167	0.174	0.176	0.171
MAD	0.178	0.189	0.189	0.185
NA	0.000	0.000	0.000	0.000

Table 5.1: Summary statistics of estimates of α by year.

Estimates of β

	2001	2002	2003	2004
Min	−0.999	−0.999	−0.999	−1.000
25%	0.029	−0.002	0.030	−0.091
Median	0.185	0.170	0.239	0.139
Mean	0.211	0.213	0.241	0.128
75%	0.390	0.407	0.469	0.377
Max	1.000	1.000	1.000	1.000
Std	0.353	0.383	0.420	0.452
MAD	0.260	0.287	0.323	0.347
NA	0.000	0.000	0.000	0.000

Table 5.2: Summary statistics of estimates of β by year.

Estimates of σ

	2001	2002	2003	2004
Min	0.00116	0.00078	0.00069	0.00063
25%	0.01230	0.01220	0.00934	0.00844
Median	0.02040	0.01860	0.01380	0.01230
Mean	0.02430	0.02200	0.01650	0.01410
75%	0.03280	0.02890	0.02130	0.01830
Max	0.12900	0.14400	0.10600	0.06280
Std	0.01580	0.01410	0.01050	0.00772
MAD	0.01390	0.01150	0.00826	0.00676
NA	0.00000	0.00000	0.00000	0.00000

Table 5.3: Summary statistics of estimates of σ by year.

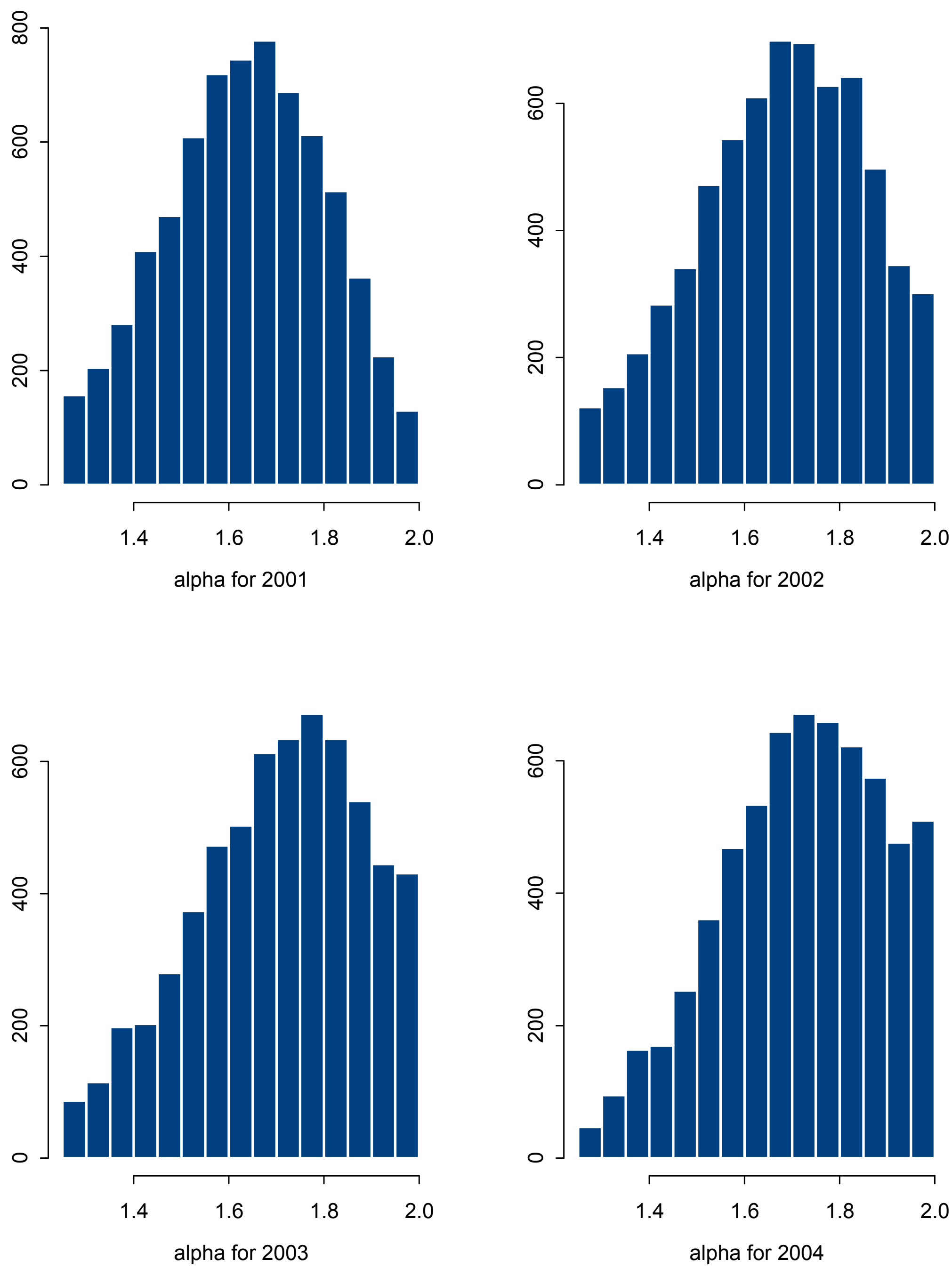

Figure 5.1: Histogram of ML-estimates of the stable index parameter.

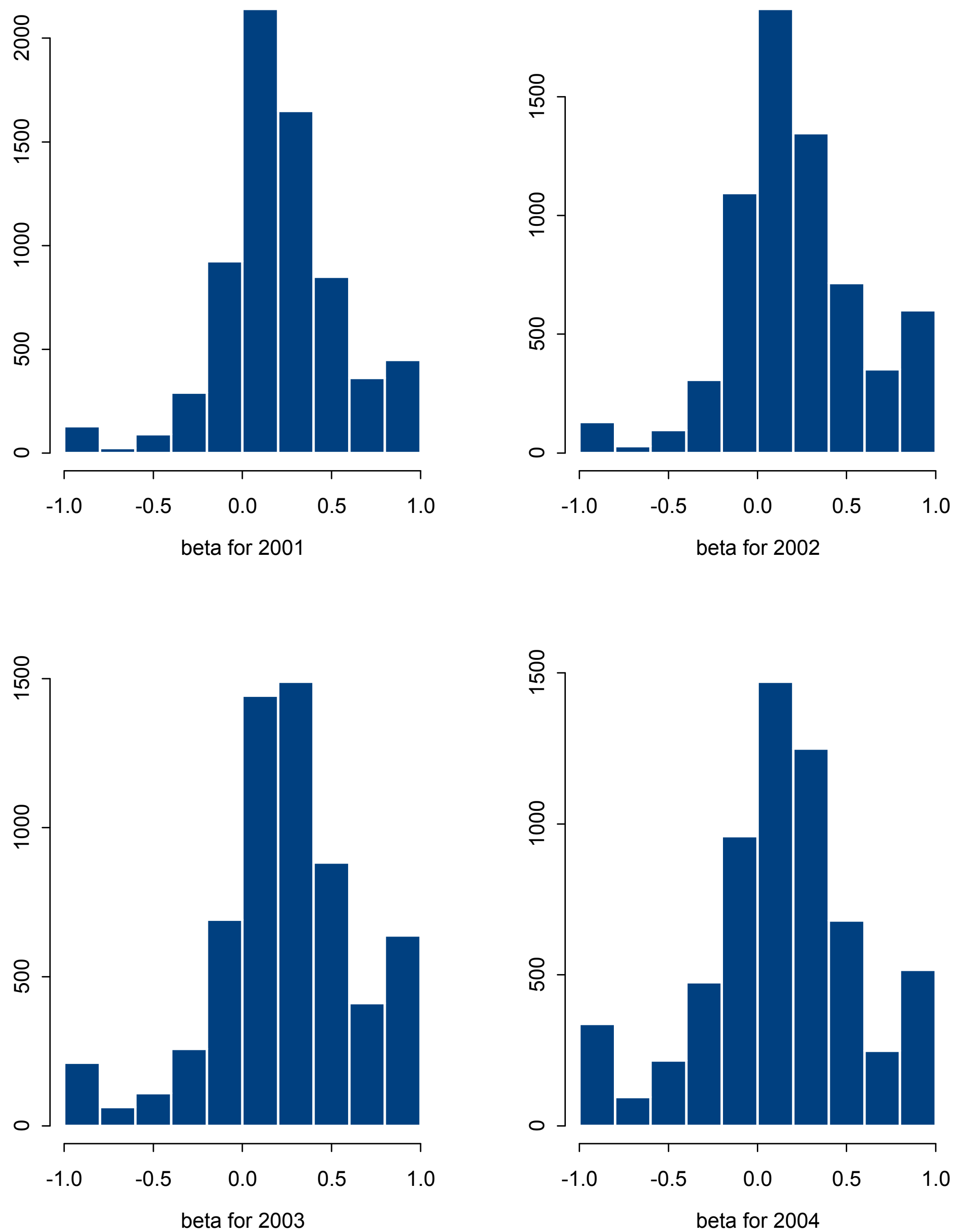

Figure 5.2: Histogram of ML-estimates of the skewness parameter.

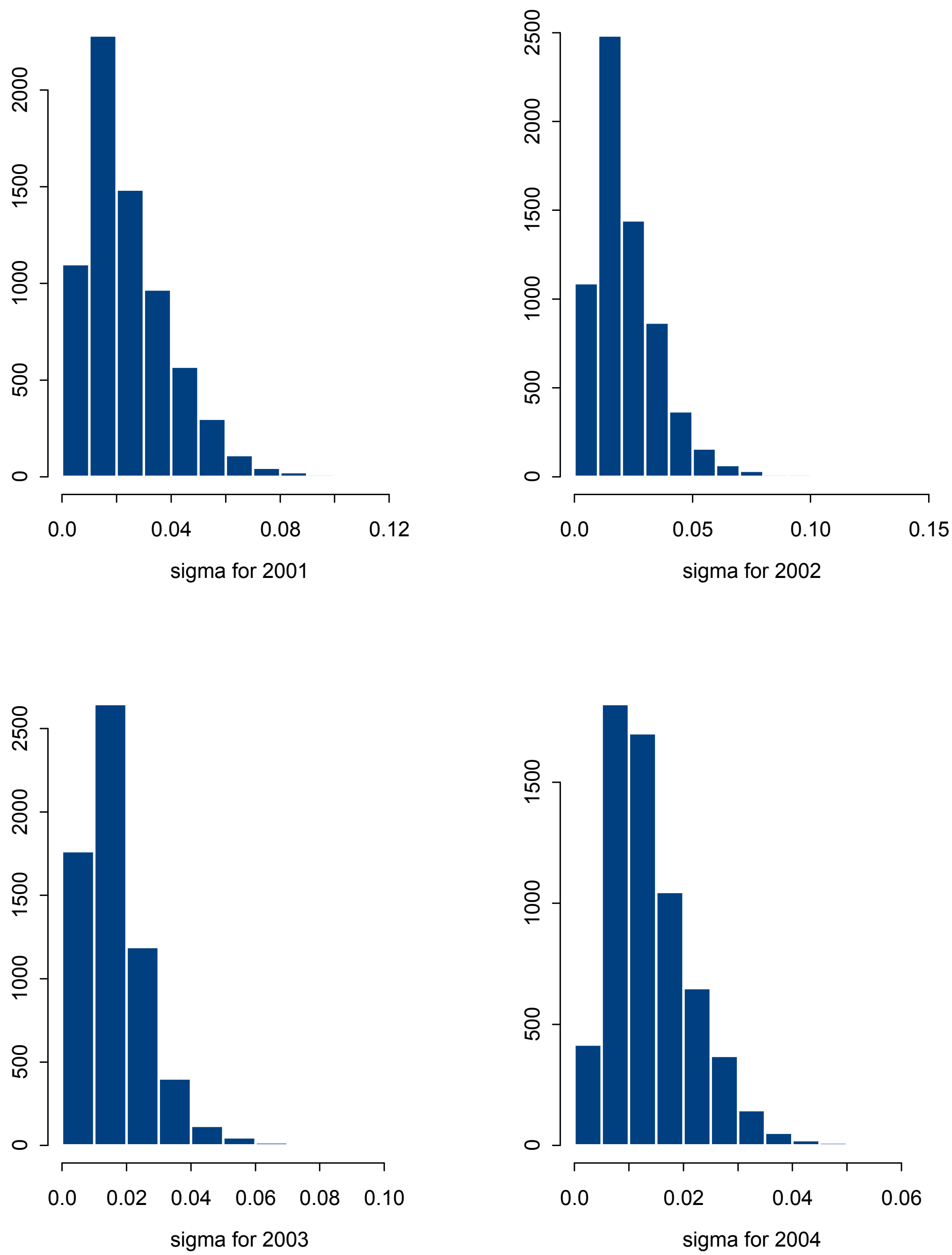

Figure 5.3: Histogram of ML-estimates of the scale parameter.

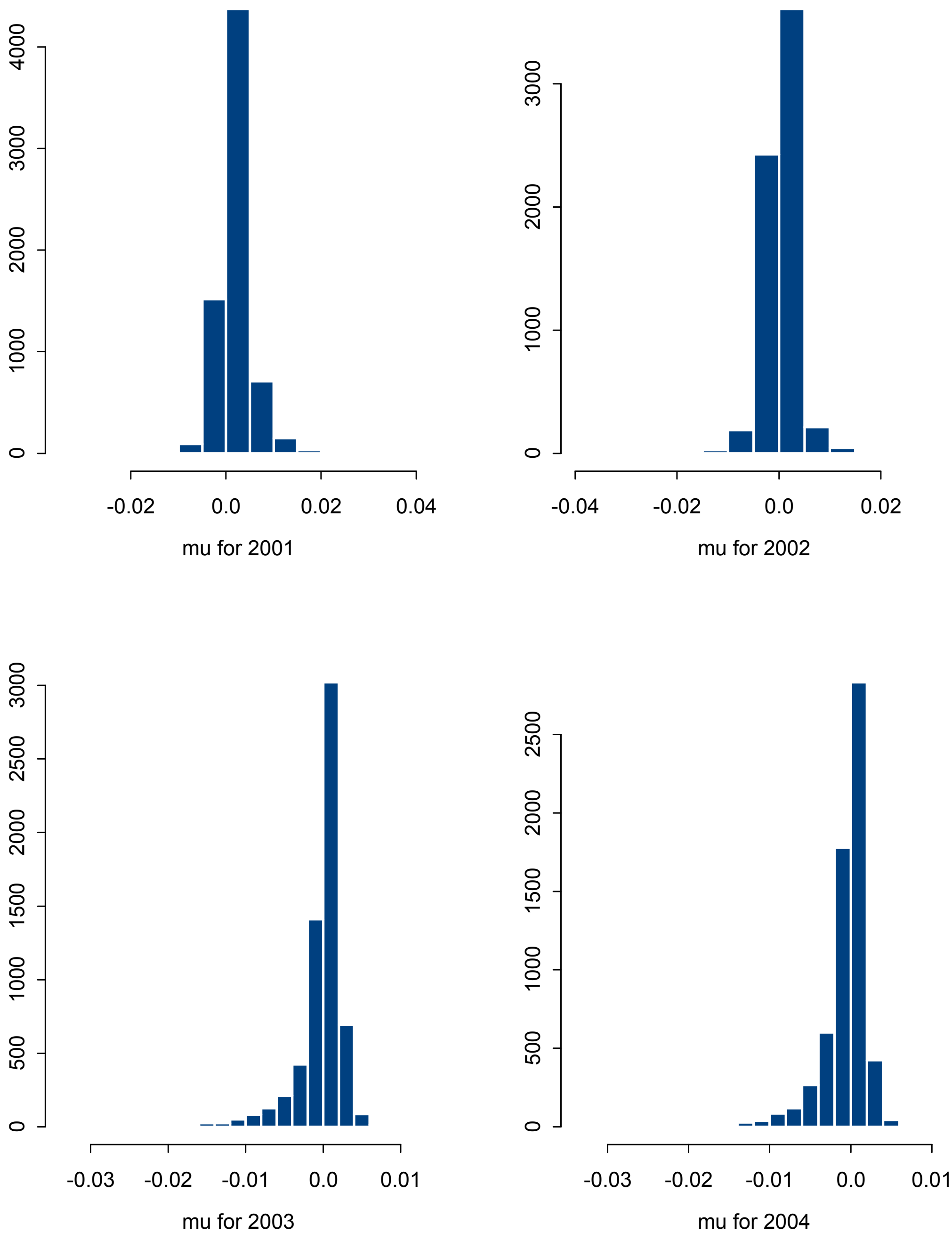

Figure 5.4: Histogram of ML-estimates of the location parameter.

Estimates of μ

	2001	2002	2003	2004
Min	-0.03130	-0.03620	-0.03260	-0.03200
25%	0.00007	-0.00078	-0.00082	-0.00129
Median	0.00122	0.00031	0.00045	0.00012
Mean	0.00179	0.00025	-0.00025	-0.00056
75%	0.00305	0.00131	0.00133	0.00097
Max	0.03790	0.02980	0.01090	0.00908
Std	0.00358	0.00306	0.00309	0.00276
MAD	0.00208	0.00154	0.00152	0.00152
NA	0.00000	0.00000	0.00000	0.00000

Table 5.4: Summary statistics of estimates of μ by year.

	2001	2002	2003	2004
α	0.065	0.034	0.020	0.035
β	-0.030	-0.007	0.002	0.018
σ	-0.060	-0.046	-0.032	-0.045
μ	-0.009	0.004	0.008	0.030

Table 5.5: Correlation of stable parameters with median size, by year.

5.2.3 Plots

As a first step in our investigation of the relationship between firm size and the stable law parameters, we have produced hexbin plots[2] of each firm's parameter estimates versus its median size (Figs. 5.5-5.8). We also computed the correlation between each of the parameters and the median size for each year; these results are shown in Table 5.5. From the hexbin plots and the correlation coefficients we surmise

- there is a weak linear relationship between the stable index α and size;
- no apparent relationship between the skewness parameter β and size;
- a nonlinear relationship between the scale parameter σ and size; and
- the location parameter is near zero most of the time, but there were a large number of negative returns in small firms in 2003 and 2004.

[2] A hexbin plot is a modification of the usual scatterplot, in which hexagons are used to represent groups of nearby points. The size of a hexagon reflects the density of points nearby. See [CLNL87] for further discussion of this technique.

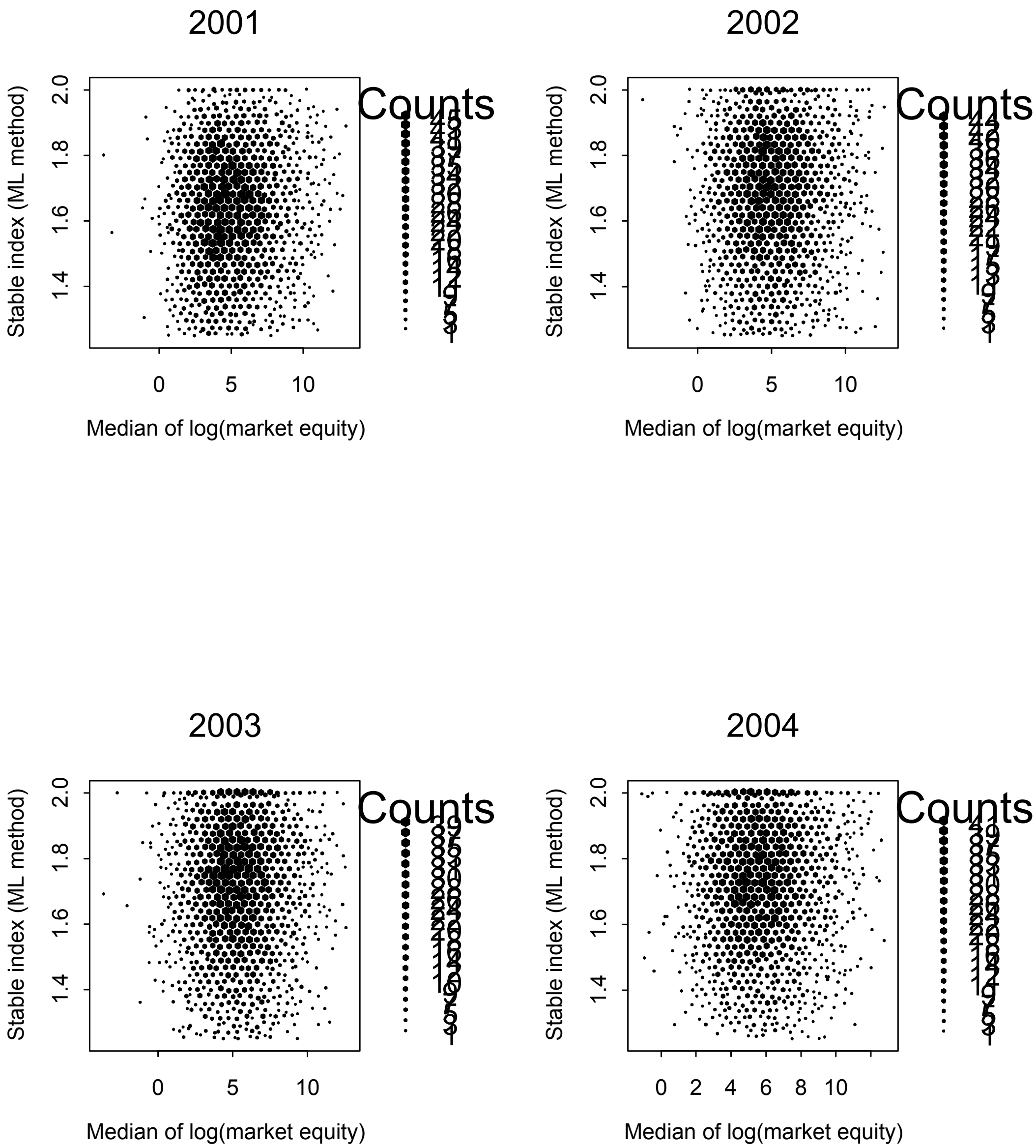

Figure 5.5: Hexbin plots of ML-estimated stable index versus median size.

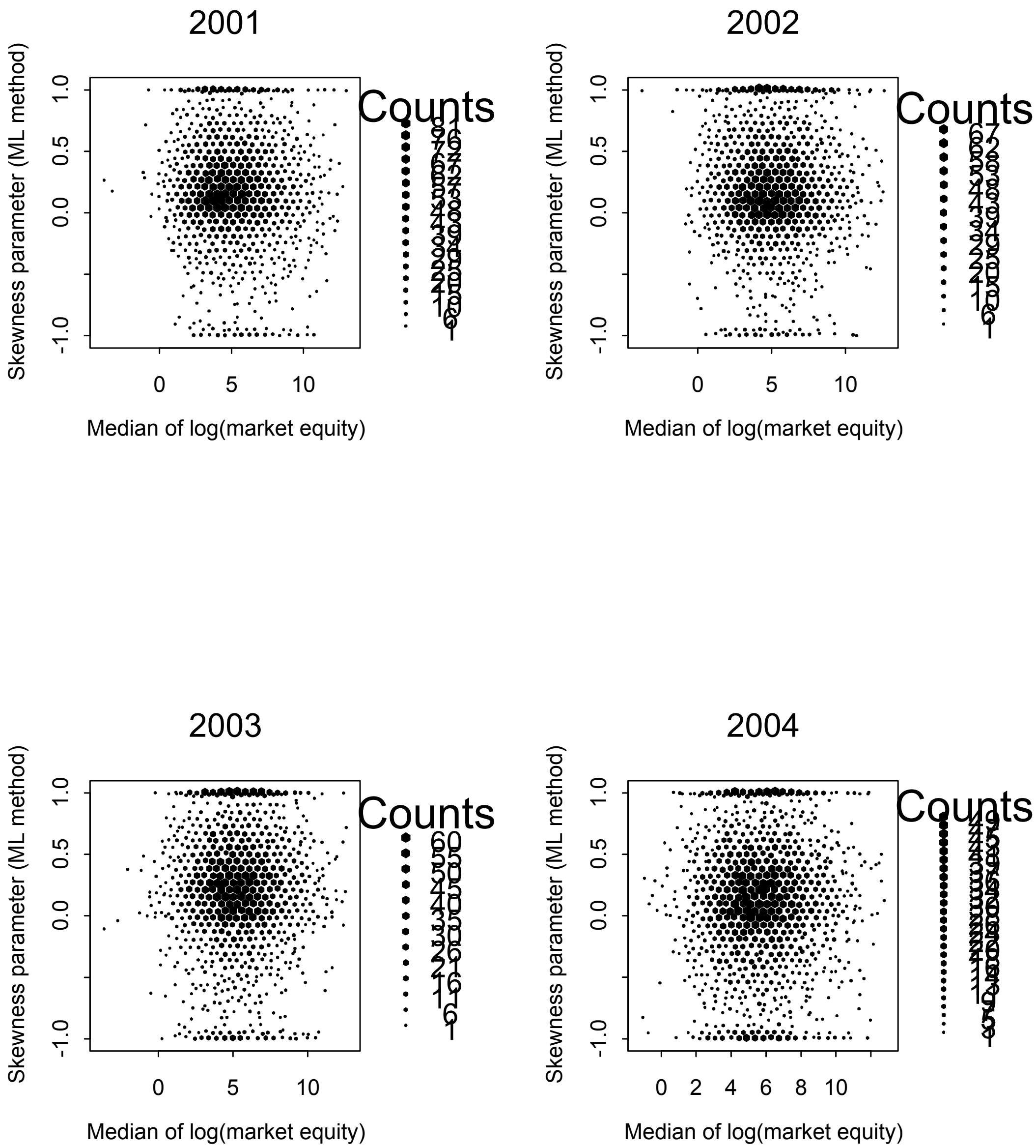

Figure 5.6: Hexbin plots of ML-estimated skewness parameter versus median size.

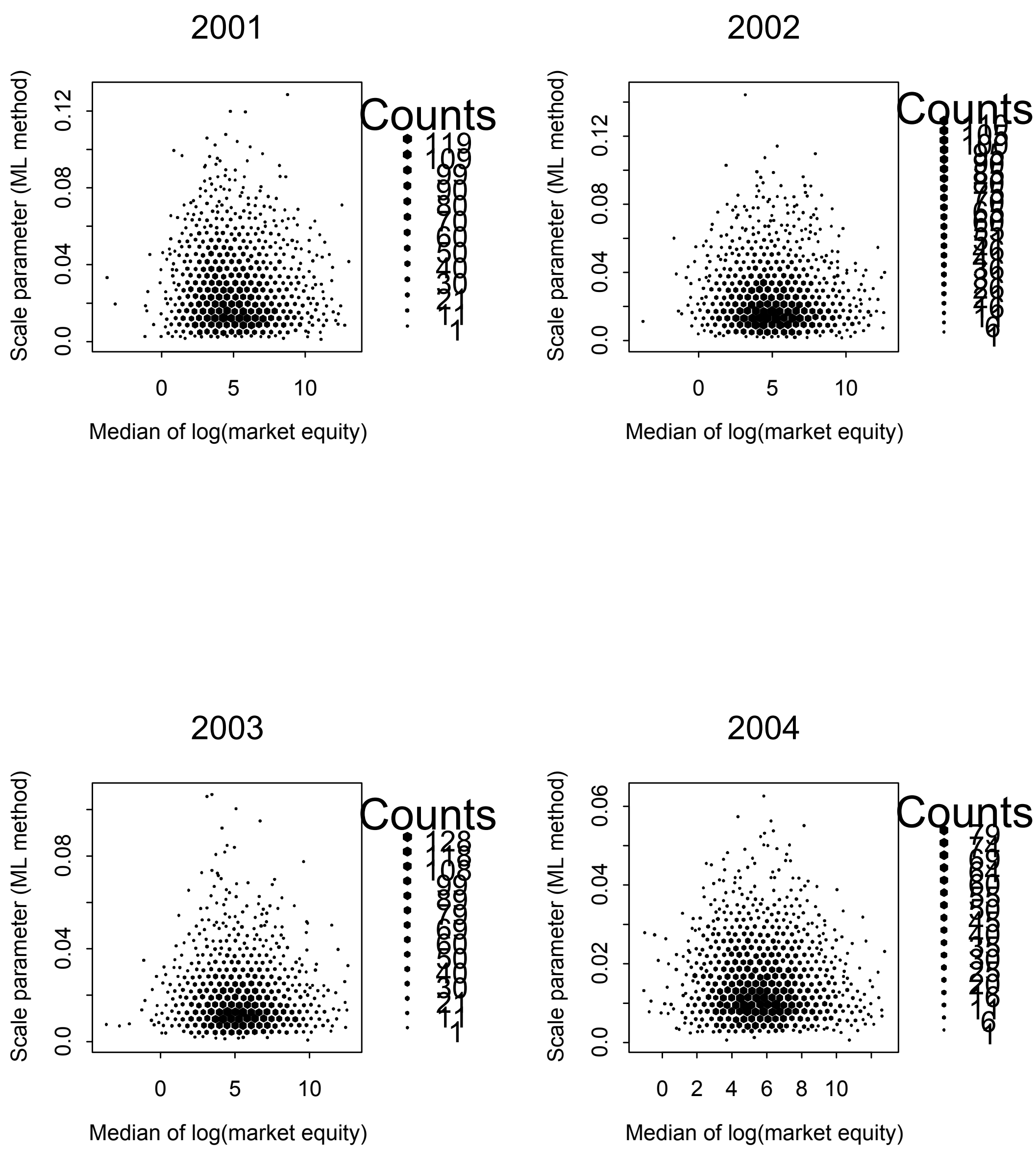

Figure 5.7: Hexbin plots of ML-estimated scale parameter versus median size.

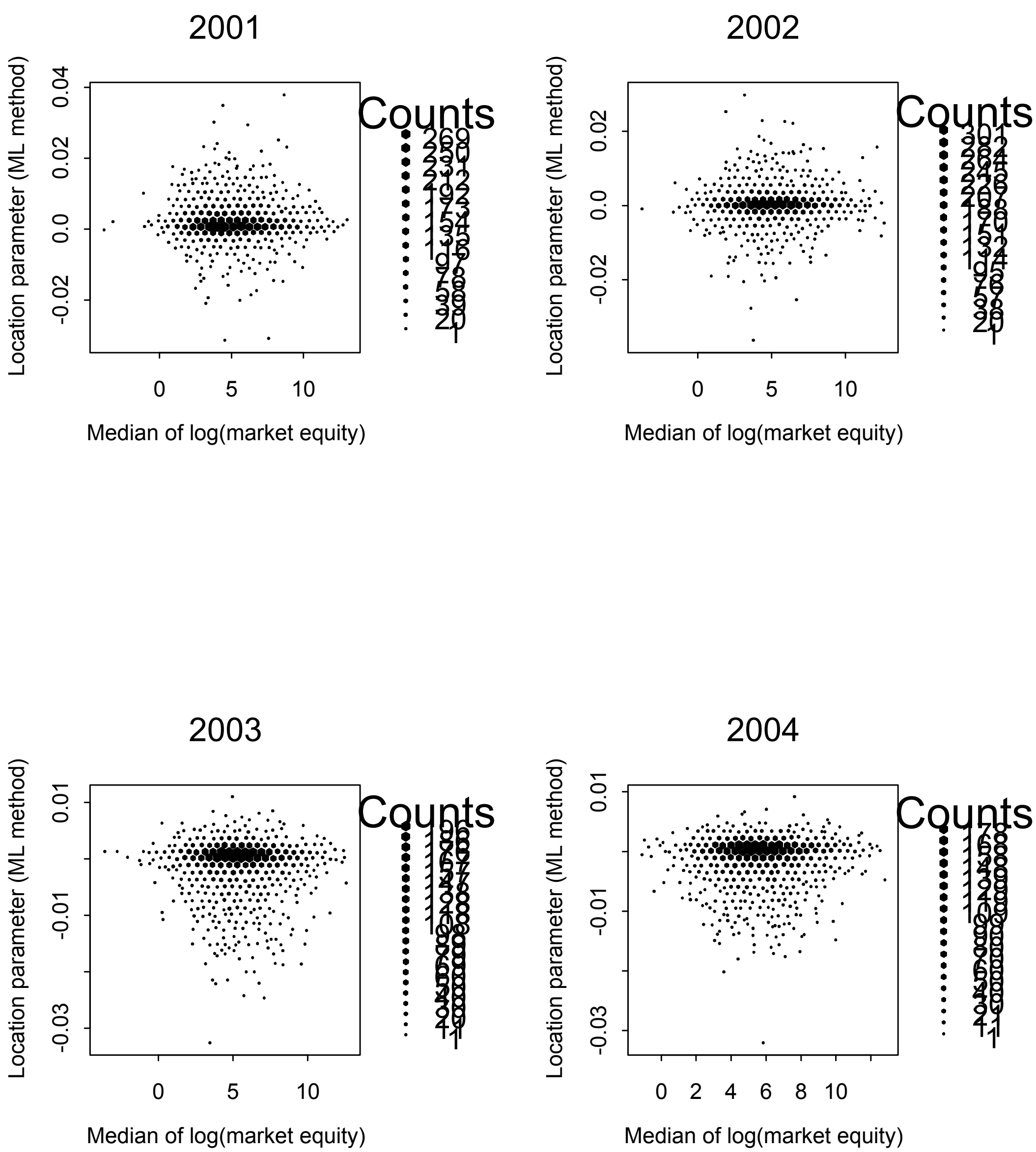

Figure 5.8: Hexbin plots of ML-estimated location parameter versus median size.

In order to clarify the relationship between the stable index and size, we produced a Trellis version of our scatterplot (Figs. 5.9-5.12, in which the data are separated by standard size categories ("NANO", "MICRO", "SMALL", "MEDIUM", "LARGE", and "MEGA"). We also computed the correlation of each the index parameter with size within the levels of size (Table 5.6). These plots and the correlation coefficients again suggest a weakly linear relationship between size and the stable index, except for the "MEGA" class, where there isn't enough data to really see anything, and the "NANO" class, where no relationship is obvious.

	NANO	MICRO	SMALL	MEDIUM	LARGE	MEGA
2001	0.011	0.009	0.048	0.103	0.106	0.086
2002	-0.026	-0.039	0.067	0.113	0.014	0.486
2003	-0.024	-0.007	0.003	-0.053	-0.018	-0.216
2004	-0.038	0.008	0.016	0.082	0.063	-0.535

Table 5.6: Correlation of stable index with median size, by year and market capitalization class.

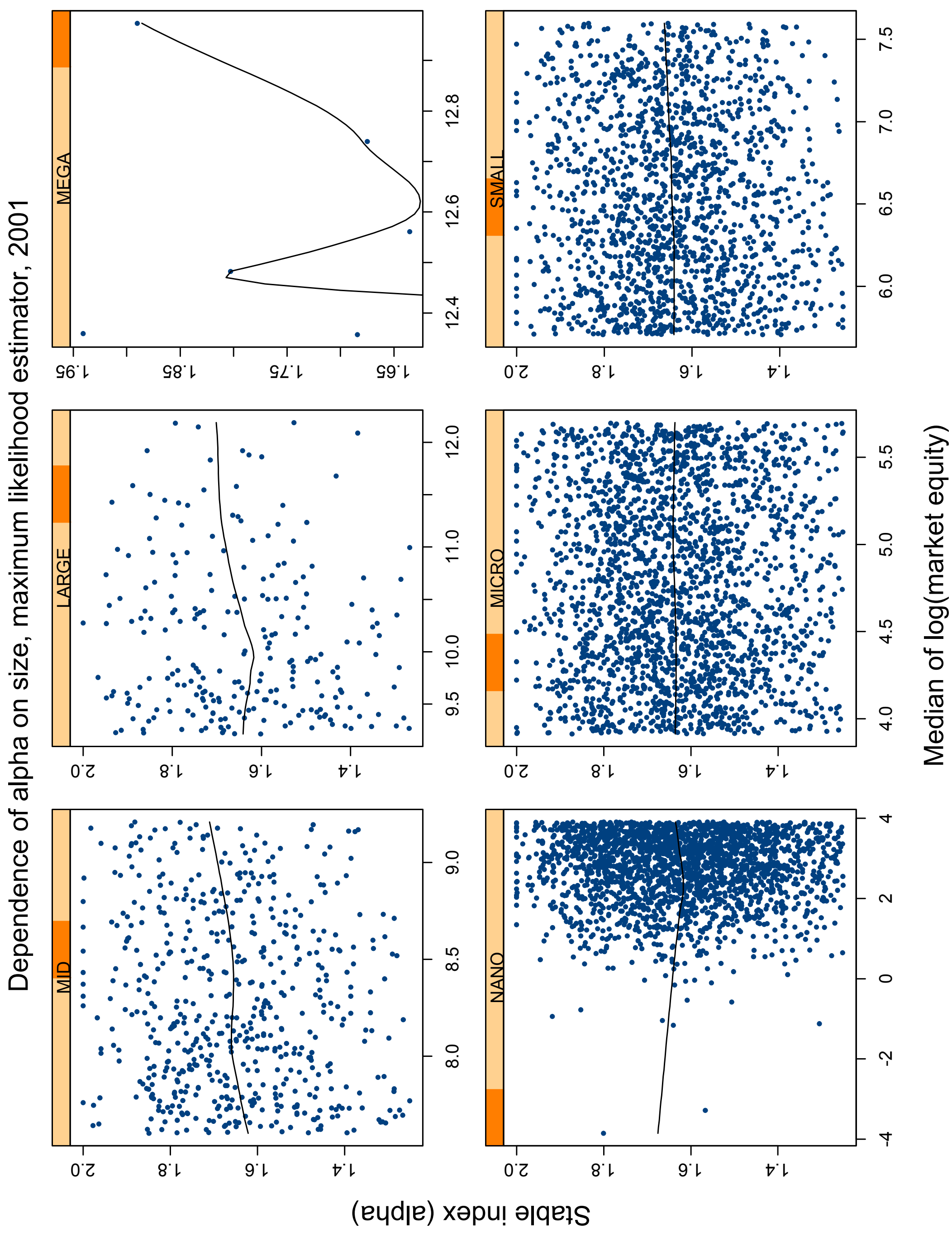

Figure 5.9: Scatterplot of ML-estimated stable index parameter against median size, within levels of size, for 2001 data. A loess line has been added to each panel.

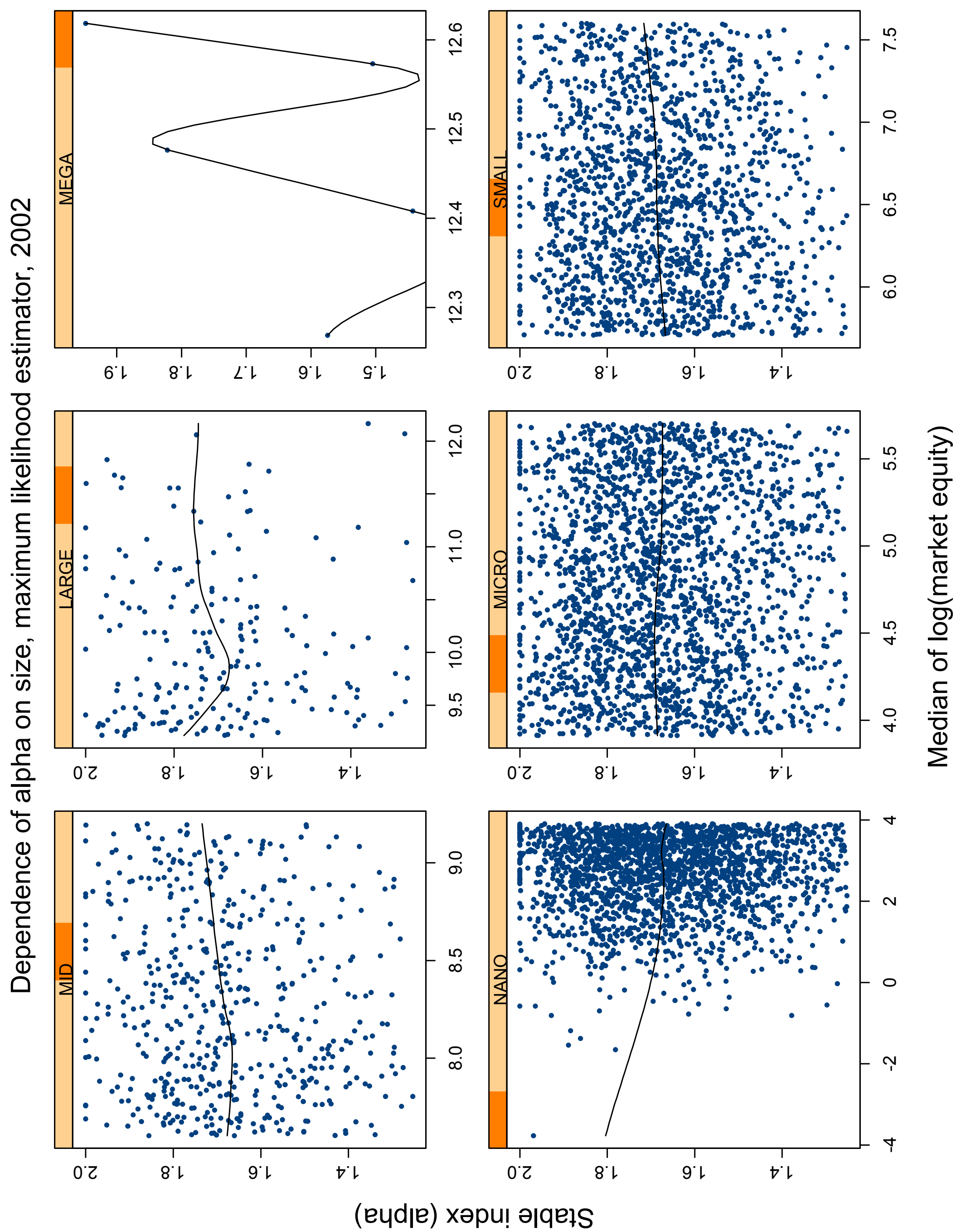

Figure 5.10: Scatterplot of ML-estimated stable index parameter against median size, within levels of size, for 2002 data. A loess line has been added to each panel.

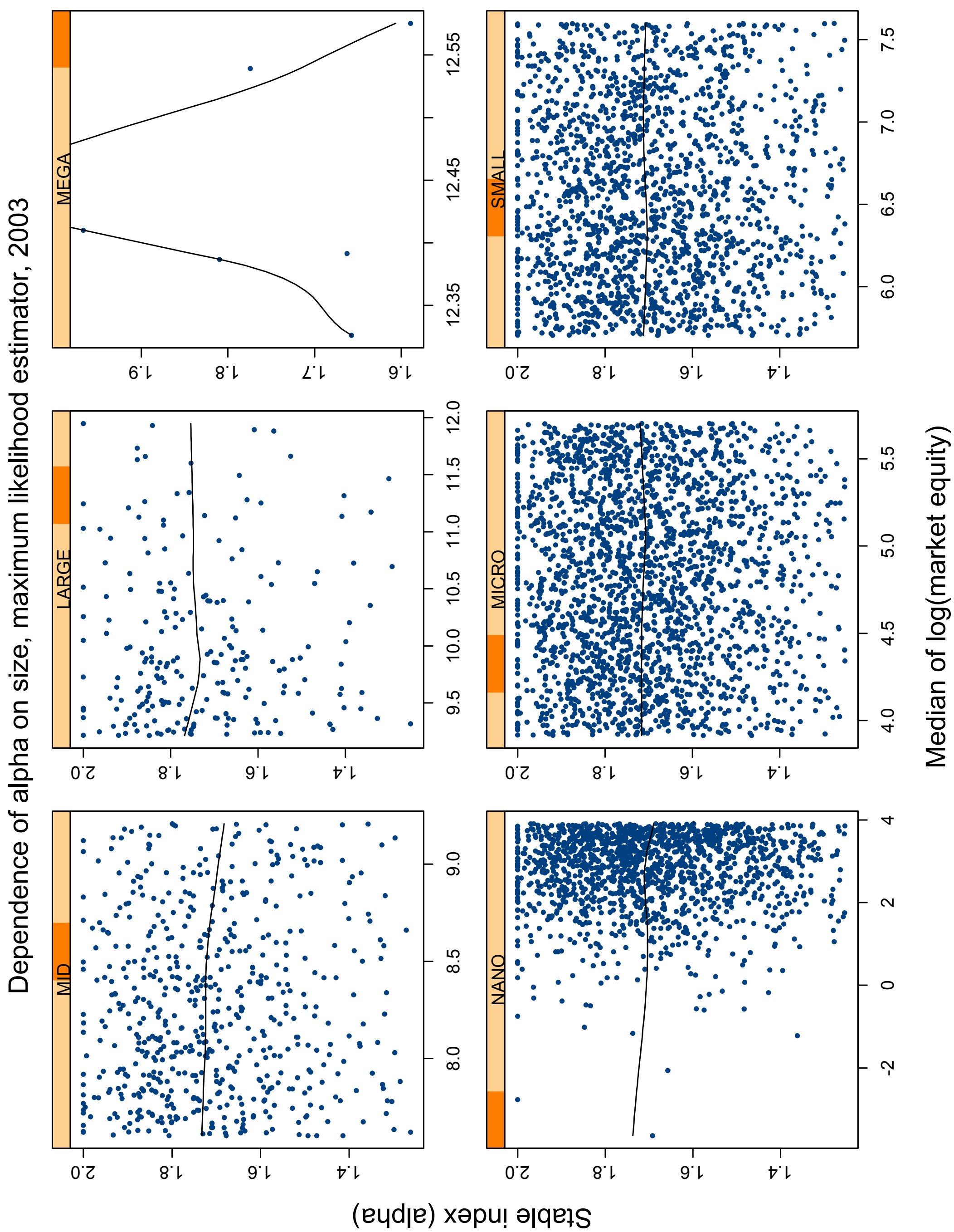

Figure 5.11: Scatterplot of ML-estimated stable index parameter against median size, within levels of size, for 2003 data. A loess line has been added to each panel.

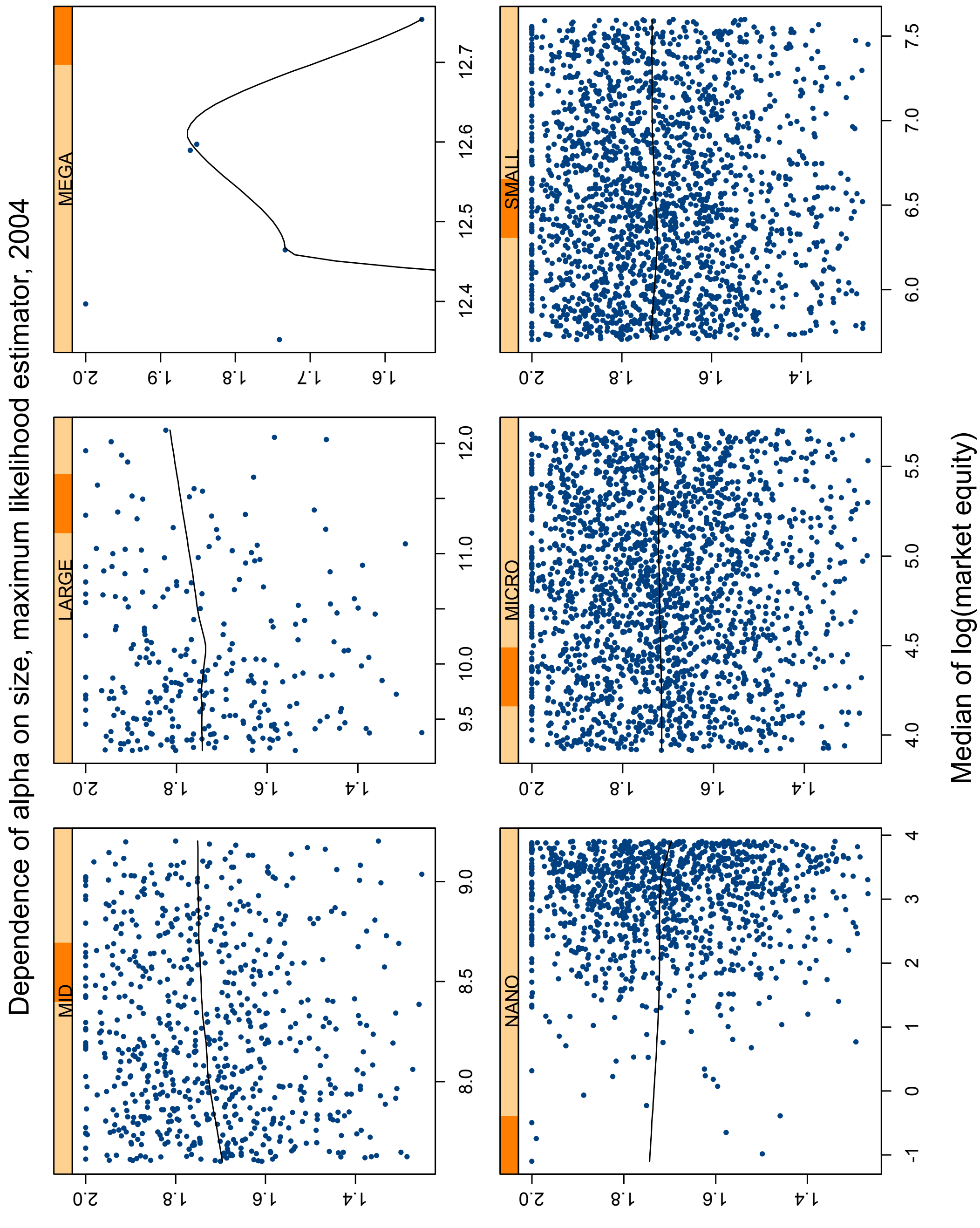

Figure 5.12: Scatterplot of ML-estimated stable index parameter against median size, within levels of size, for 2004 data. A loess line has been added to each panel.

5.2.4 Models

Based upon our scatterplots, we try to model the stable index as a linear function of size:

$$\alpha_i = \beta_0 + \beta_1 \log(ME_i) + \varepsilon_i \tag{5.2.1}$$

using both classical and robust linear regression. The results of this fit are shown in Tables 5.7-5.10. The fitted coefficients are significant (with one exception), but the model explains very little of the observed variation. The classical and robust regressions agree, so there is little reason so suspect the classical results were adversely affected by outliers.

2001		Coef.	Std.Error	t value	Pr(>|t|)
(Intercept)	classical	1.610	0.005	313.023	0.000
(Intercept)	robust	1.609	0.005	298.290	0.000
median.size.2001	classical	0.005	0.001	5.395	0.000
median.size.2001	robust	0.005	0.001	5.288	0.000

	classical	robust
Residual scale estimate	0.166	0.178
Percentage of variation explained	0.004	0.004

Table 5.7: Results of classical and robust linear regression of stable index on median size for year 2001.

2002		Coef.	Std.Error	t value	Pr(>|t|)
(Intercept)	classical	1.663	0.006	301.415	0.000
(Intercept)	robust	1.664	0.006	271.240	0.000
median.size.2002	classical	0.003	0.001	2.776	0.006
median.size.2002	robust	0.003	0.001	2.590	0.010

	classical	robust
Residual scale estimate	0.174	0.184
Percentage of variation explained	0.001	0.001

Table 5.8: Results of classical and robust linear regression of stable index on median size for year 2002.

Next we again try to model the stable index as a linear function of size, but with a different intercept for each size class.

$$\alpha_i = \beta_{SIZE} + \beta_1 \log(ME_i) + \varepsilon_i \tag{5.2.2}$$

using both classical and robust linear regression. The results of this fit are shown in Tables 5.11-5.14. The fitted intercepts are now significant, but the slope coefficents are not significant in 2002, 2003, and 2004. The model explains much of the observed variation, however, even though the residual scale is still rather large.

2003		Coef.	Std.Error	t value	Pr(>|t|)
(Intercept)	classical	1.694	0.006	268.849	0.000
(Intercept)	robust	1.697	0.007	235.918	0.000
median.size.2003	classical	0.002	0.001	1.575	0.115
median.size.2003	robust	0.002	0.001	1.486	0.137

	classical	robust
Residual scale estimate	0.176	0.185
Percentage of variation explained	0.000	0.000

Table 5.9: Results of classical and robust linear regression of stable index on median size for year 2003.

2004		Coef.	Std.Error	t value	Pr(>|t|)
(Intercept)	classical	1.700	0.007	255.191	0.000
(Intercept)	robust	1.701	0.007	232.128	0.000
median.size.2004	classical	0.003	0.001	2.749	0.006
median.size.2004	robust	0.003	0.001	2.747	0.006

	classical	robust
Residual scale estimate	0.170	0.181
Percentage of variation explained	0.001	0.001

Table 5.10: Results of classical and robust linear regression of stable index on median size for year 2004.

2001		Coef.	Std.Error	t value	Pr(>|t|)
cap.classes.median.2001NANO	classical	1.606	0.009	178.637	0.000
cap.classes.median.2001NANO	robust	1.606	0.009	178.637	0.000
cap.classes.median.2001MICRO	classical	1.604	0.015	108.667	0.000
cap.classes.median.2001MICRO	robust	1.604	0.015	108.667	0.000
cap.classes.median.2001SMALL	classical	1.603	0.020	80.108	0.000
cap.classes.median.2001SMALL	robust	1.603	0.020	80.108	0.000
cap.classes.median.2001MID	classical	1.602	0.026	62.186	0.000
cap.classes.median.2001MID	robust	1.602	0.026	62.186	0.000
cap.classes.median.2001LARGE	classical	1.579	0.032	49.065	0.000
cap.classes.median.2001LARGE	robust	1.579	0.032	49.065	0.000
cap.classes.median.2001MEGA	classical	1.691	0.078	21.785	0.000
cap.classes.median.2001MEGA	robust	1.691	0.078	21.785	0.000
median.size.2001	classical	0.006	0.003	2.145	0.032
median.size.2001	robust	0.006	0.003	2.145	0.032

	classical	robust
Residual scale estimate	0.166	0.166
Percentage of variation explained	0.990	0.990

Table 5.11: Results of classical and robust linear regression of stable index on median size for year 2001 .

2002		Coef.	Std.Error	t value	Pr(>|t|)
cap.classes.median.2002NANO	classical	1.671	0.009	184.122	0.000
cap.classes.median.2002NANO	robust	1.671	0.009	184.122	0.000
cap.classes.median.2002MICRO	classical	1.680	0.015	110.133	0.000
cap.classes.median.2002MICRO	robust	1.680	0.015	110.133	0.000
cap.classes.median.2002SMALL	classical	1.686	0.021	81.501	0.000
cap.classes.median.2002SMALL	robust	1.686	0.021	81.501	0.000
cap.classes.median.2002MID	classical	1.685	0.027	63.165	0.000
cap.classes.median.2002MID	robust	1.685	0.027	63.165	0.000
cap.classes.median.2002LARGE	classical	1.710	0.033	51.201	0.000
cap.classes.median.2002LARGE	robust	1.710	0.033	51.201	0.000
cap.classes.median.2002MEGA	classical	1.665	0.087	19.155	0.000
cap.classes.median.2002MEGA	robust	1.665	0.087	19.155	0.000
median.size.2002	classical	-0.001	0.003	-0.181	0.856
median.size.2002	robust	-0.001	0.003	-0.181	0.856

	classical	robust
Residual scale estimate	0.174	0.174
Percentage of variation explained	0.989	0.989

Table 5.12: Results of classical and robust linear regression of stable index on median size for year 2002 .

2003		Coef.	Std.Error	t value	Pr(>ltl)
cap.classes.median.2003NANO	classical	1.708	0.011	159.990	0.000
cap.classes.median.2003NANO	robust	1.708	0.011	159.990	0.000
cap.classes.median.2003MICRO	classical	1.723	0.017	100.824	0.000
cap.classes.median.2003MICRO	robust	1.723	0.017	100.824	0.000
cap.classes.median.2003SMALL	classical	1.725	0.023	74.812	0.000
cap.classes.median.2003SMALL	robust	1.725	0.023	74.812	0.000
cap.classes.median.2003MID	classical	1.741	0.030	58.933	0.000
cap.classes.median.2003MID	robust	1.741	0.030	58.933	0.000
cap.classes.median.2003LARGE	classical	1.766	0.037	48.220	0.000
cap.classes.median.2003LARGE	robust	1.766	0.037	48.220	0.000
cap.classes.median.2003MEGA	classical	1.789	0.084	21.344	0.000
cap.classes.median.2003MEGA	robust	1.789	0.084	21.344	0.000
median.size.2003	classical	-0.004	0.003	-1.062	0.288
median.size.2003	robust	-0.004	0.003	-1.062	0.288

	classical	robust
Residual scale estimate	0.176	0.176
Percentage of variation explained	0.989	0.989

Table 5.13: Results of classical and robust linear regression of stable index on median size for year 2003 .

2004		Coef.	Std.Error	t value	Pr(>ltl)
cap.classes.median.2004NANO	classical	1.705	0.012	139.202	0.000
cap.classes.median.2004NANO	robust	1.705	0.012	139.202	0.000
cap.classes.median.2004MICRO	classical	1.702	0.018	92.282	0.000
cap.classes.median.2004MICRO	robust	1.702	0.018	92.282	0.000
cap.classes.median.2004SMALL	classical	1.704	0.025	68.616	0.000
cap.classes.median.2004SMALL	robust	1.704	0.025	68.616	0.000
cap.classes.median.2004MID	classical	1.703	0.032	54.007	0.000
cap.classes.median.2004MID	robust	1.703	0.032	54.007	0.000
cap.classes.median.2004LARGE	classical	1.716	0.039	44.163	0.000
cap.classes.median.2004LARGE	robust	1.716	0.039	44.163	0.000
cap.classes.median.2004MEGA	classical	1.760	0.084	20.987	0.000
cap.classes.median.2004MEGA	robust	1.760	0.084	20.987	0.000
median.size.2004	classical	0.002	0.004	0.623	0.533
median.size.2004	robust	0.002	0.004	0.623	0.533

	classical	robust
Residual scale estimate	0.170	0.170
Percentage of variation explained	0.990	0.990

Table 5.14: Results of classical and robust linear regression of stable index on median size for year 2004 .

5.3 Cross-Sectional Analysis

In the second portion of our experiment, we attempted to find the "best" proxy for the tail index out of a small set of indicators of firm size—the logarithm of market equity, the ratio of the book equity to the market equity, and the price-earnings ratio.

We obtained daily prices, returns, and market equity on firms listed on the three major exchanges from the CRSP database [CRS]. Prices were adjusted for splits using the modified adjustment factor (`cfacpr`) provided by Wharton Research Data Services (WRDS). The daily prices and market equity were converted to monthly values by taking the value on the closing day of each month.

Due to time constraints, we fit a t-distribution to each firms returns on a monthly basis using the previous 12 months of data. To remove the occasional gross outlier in the fitted values (presumably due to periods where the returns looked very Gaussian), the fitted values were smoothed using a robust-version of the exponentially-weighted moving average

$$\begin{aligned} S_1 &= X_1 \\ S_n &= \lambda \min(X_n, a * X_{n-1}) + (1-\lambda) S_{n-1}. \end{aligned}$$

We used $a = 10$ and $\lambda = 0.93$.

Next, we gathered accounting data (book equity and earnings per share (excluding extraordinary items) and industry data (GICS codes) from the Compustat [Sta] North American Annual database. These values were converted to monthly values via linear interpolation.

We then performed cross-sectional regression on the joined data.

Bibliography

[AFT98] Robert J. Adler, Raisa E. Feldman, and Murad S. Taqqu (eds.), *A practical guide to heavy tails: Statistical techniques and applications*, Birkhäuser, Boston, MA, 1998.

[Alh79] Lars V. Alhfors, *Complex analysis : an introduction to the theory of analytic functions of one complex variable*, third ed., McGraw-Hill, New York, 1979.

[Bac67] Louis Bachelier, *Theory of speculation*, The Random Character of Stock Market Prices (Paul H. Cootner, ed.), MIT Press, revisied edition ed., 1967, Originally published as a thesis in Ann. Sci. École Norm. Sup. 3 (1018), 1900.

[Ber52] Harald Bergström, *On some expansions of stable distribution functions*, Ark. Mat. **2** (1952), 375–378. MR MR0065053 (16,377i)

[Bre68] Leo Breiman, *Probability*, Addison-Wesley, Reading, MA, 1968.

[Buc95] D. J. Buckle, *Bayesian inference for stable distributions*, J. Amer. Statist. Assoc. **90** (1995), no. 430, 605–613. MR MR1340513 (96d:62038)

[Che91] Yao Chen, *Distributions for asset returns*, Ph.D. thesis, State University of New York at Stony Brook, 1991.

[CLNL87] D. B. Carr, R. J. Littlefield, W. L. Nicholson, and J. S. Littlefield, *Scatterplot matrix techniques for large n*, Journal of the American Statistical Association **83** (1987), 424–436.

[CMS76] J. M. Chambers, C. L. Mallows, and B. W. Stuck, *A method for simulating stable random variables*, Journal of the Americal Statistical Association **71** (1976), no. 354, 340–344.

[CRS] CRSP, Center for Research in Security Prices (via Wharton Research Data Services). Graduate School of Business, The University of Chicago 2001-2004, Used with permission. All rights reserved. `www.crsp.uchicago.edu`. Accessed October 2005.

[DM98] Toker Doganoglu and Stefan Mittnik, *An approximation procedure for asymmetric stable paretian densities*, Computational Statistics **13** (1998), 463–475.

[DuM71] William H. DuMouchel, *Stable distributions in statistical inference*, Ph.D. thesis, Yale University, 1971.

[DuM73a] ______, *On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution*, Ann. Statist. **1** (1973), 948–957. MR MR0339376 (49 #4135)

[DuM73b] ______, *Stable distributions in statistical inference. I. Symmetric stable distributions compared to other symmetric long-tailed distributions*, J. Amer. Statist. Assoc. **68** (1973), 469–477. MR MR0378190 (51 #14359)

[DuM75] W. H. DuMouchel, *Stable distributions in statistical inference. II. Information from stably distributed samples*, J. Amer. Statist. Assoc. **70** (1975), 386–393. MR MR0378191 (51 #14360)

[Fam63] Eugene F. Fama, *Mandelbrot and the stable paretian hypothesis*, Journal of Business **36** (1963), 420–429.

[Fel71] William Feller, *An introduction to probability theory and its applications, volume ii*, second edition ed., John Wiley and Sons, New York, NY, 1971.

[Fol99] Gerald B. Folland, *Real analysis: Modern techniques and applications*, second edition ed., John Wiley & Sons, New York, 1999.

[FR68] Eugene F. Fama and Richard Roll, *Some properties of symmetric stable distributions*, J. Amer. Statist. Assoc. **63** (1968), 817–836. MR MR0233452 (38 #1774)

[FR71] Eugene Fama and Richard Roll, *Parameter estimates for symmetric stable distributions*, J. Amer. Statist. Assoc. **63** (1971), 817–836.

[GK68] B. V. Gnedenko and A. N. Kolmogorov, *Limit distributions for sums of independent random variables*, Addison-Wesley, Reading, MA, 1968.

[Hal81] Peter Hall, *A comedy of errors: The canonical form for a stable characteristic function*, Bull. London Math. Soc. **13** (1981), 23–27.

[HC73] D. Holt and E. Crow, *Tables and graphs of the stable probability density functions*, J. Research Nat. Burearu of Standards, B. Math. Sci. **77B** (1973), 143–198.

[Hol19] J Holtsmark, *Uber die verbreiterung von spektrallinien*, Ann. Physik **58** (1919), no. 363, 577–630.

[HRRS86] Frank R. Hampel, Elvezio M. Ronchetti, Peter J. Rousseeuw, and Werner A. Stahel, *Robust statistics: the approach based on influence functions*, John Wiley & Sons, New York, 1986.

[IC59] I. A. Ibragomiev and K. E. Chernin, *On the unimodality of stable laws*, Theory of Probability and its applications **IV** (1959), 417–419.

[KH53] M. G. Kendall and A. Bradford Hill, *The analysis of economic time series, i: Prices*, Journal of the Royal Statistical Society **XCVI** (1953), 11–25.

[KL36] A. Ya. Khintchine and Paul Lévy, *Sur les lois stables*, C. R. Acad. Sci. Paris **202** (1936), 374–376.

[Kou80] Ioannis A. Koutrouvelis, *Regression-type estimation of the parameters of stable laws*, J. American Statistical Association **75** (1980), no. 372, 919–928.

[Kou81] ______, *An iterative procedure for the estimation of the parameters of stable laws*, Comm. in Statistics B—Simulation and Computation **10** (1981), no. 1, 17–28.

[Kur01] Ercan Engin Kuruoğlu, *Density parameter estimation of skewed α-stable distributions*, IEEE Transactions on Signal Processing **49** (2001), no. 10, 2192–2201.

[KW98] Stephen M. Kogon and Douglas B. Williams, *Characteristic function based estimation of stable distribution parameters*, in Adler et al. [AFT98], pp. 310–335.

[Lam96] J. Lamperti, *Probability: A survey of the mathematical theory*, second edition ed., John Wiley and Sons, 1996.

[Lév24] Paul Lévy, *Théorie des erreurs. la loi de gauss les lois exceptionelles*, Bull. Soc. Math. France **52** (1924), 49–85.

[Man63a] Benoit Mandelbrot, *New methods in statistical economics*, The Journal of Political Economy **LXXI** (1963), no. 5, 421–440.

[Man63b] ______, *The variation of certain speculative prices*, Journal of Business **36** (1963), 394–419.

[McC86] J. Huston McCulloch, *Simple consistent estimators of stable distribution parameters*, Comm. Statist. B—Simulation Comput. **15** (1986), no. 4, 1109–1136. MR MR876783

[MDC99] Stefan Mittnik, Toker Doganoglu, and D. Chenyao, *Computing the probability density function of the stable paretian distribution*, Mathematical and Computer Modeling **29** (1999), 235–240.

[MRP98] Stefan Mittnik, Svetlozar Rachev, and M. S. Paolella, *Stable paretian modeling in finance: some empirical and theoretical aspects*, in Adler et al. [AFT98], pp. 79–110.

[Nol97] John P. Nolan, *Numerical calculation of stable densities and distribution functions*, Communications in statistics—stochastic models **13** (1997), no. 4, 759–774.

[Osb59] M. F. M. Osborne, *Brownian motion in the stock market*, Operations Research **7** (1959), no. 2, 145–173.

[PHL75] A. S. Paulson, E. W. Holcomb, and R. A. Leitch, *The estimation of the parameters of the stable laws*, Biometrika **62** (1975), no. 1, 163–170.

[RM00] Svetlozar Rachev and Stefan Mittnik, *Stable paretian models in finance*, John Wiley and Sons, New York, 2000.

[Sho00] Galen R. Shorack, *Probability for statisticians*, Springer-Verlag, New York, 2000.

[SM05] Bernd Scherer and R. Douglas Martin, *Introduction to modern portfolio optimization with* NUOPT*,* S-PLUS™ , and S+Bayes, Springer-Verlag, New York, 2005.

[ST94] Gennady Samorodnitsky and Murad S. Taqqu, *Stable non-gaussian random processes*, Chapman & Hall / CRC, Boca Raton, 1994.

[Sta] Standard & Poor's Compustat®data (via Wharton Research Data Services), Accessed October 2005.

[Stu08] Student, *The probable error of a mean*, Biometrika **6** (1908), 1–25.

[Tei71] J Teichmoeller, *A note on the distribution of stock price changes*, Journal of the American Statistical Association **66** (1971), 282–284.

[Zol66] Vladimir M. Zolotarev, *On representation of stable laws by integrals*, Selected Translations in Mathematical Statistics and Probability, vol. 6, pp. 84–88, American Mathematical Society, Providence, RI, 1966.

[Zol86] ______, *One-dimensional stable distributions*, Translations of Mathematical Monographs, vol. 65, American Mathematical Society, Providence, RI, 1986, Translated from the Russian by H. H. McFaden, Translation edited by Ben Silver. MR MR854867 (87k:60002)