IMS Collections

Vol. 0 (2010)

arXiv: math.PR/0000000

On the asymptotic distribution of the analytic center estimator

Keith Knight*

University of Toronto

Abstract: The analytic center estimator is defined as the analytic center of the so-called membership set. In this paper, we consider the asymptotics of this estimator under fairly general assumptions on the noise distribution.

1. Introduction

Consider the linear regression model

$$Y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \varepsilon_i \quad (i = 1, \cdots, n) \tag{1.1}$$

where $\boldsymbol{x}_i$ is a vector of covariates (of length p) whose first component is always 1, $\boldsymbol{\beta}$ is a vector of unknown parameters and $\varepsilon_1, \cdots, \varepsilon_n$ are i.i.d. random variables with $|\varepsilon_i| \le \gamma_0$ where it is assumed that γ_0 is known. We will not necessarily require that the bound γ_0 be tight although there are advantages in estimation if it is known that the noise is "boundary visiting" in the sense that $P(|\varepsilon_i| \le \gamma_0 - \epsilon) < 1$ for all $\epsilon > 0$.

Given the bound γ_0 on the absolute errors, we can define the so-called membership set (Schweppe, 1968; Bai *et al.*, 1998)

$$\mathcal{S}_n = \left\{ \boldsymbol{\phi} : -\gamma_0 \le Y_i - \boldsymbol{x}_i^T \boldsymbol{\phi} \le \gamma_0 \text{ for all } i = 1, \cdots, n \right\}, \tag{1.2}$$

which contains all parameter values consistent with the assumption that $|\varepsilon_i| \le \gamma_0$. There is a considerable literature on estimation based on the membership set in different settings; see, for example, Milanese and Belforte (1982), Mäkilä (1991), Tse *et al.* (1993), and Akçay *et al.* (1996).

The membership set $\mathcal{S}_n$ in (1.2) is a bounded convex polyhedron and we can use some measure of its center to estimate $\boldsymbol{\beta}$. The analytic center estimator $\widehat{\boldsymbol{\beta}}_n$ is defined to be the maximizer of the concave objective function

$$\begin{aligned} g_n(\boldsymbol{\phi}) &= \sum_{i=1}^n \ln\left(\gamma_0^2 - (Y_i - \boldsymbol{x}_i^T \boldsymbol{\phi})^2\right) \\ &= \sum_{i=1}^n \left\{ \ln(\gamma_0 - Y_i + \boldsymbol{x}_i^T \boldsymbol{\phi}) + \ln(\gamma_0 + Y_i - \boldsymbol{x}_i^T \boldsymbol{\phi}) \right\}. \end{aligned} \tag{1.3}$$

*This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada

[1]Department of Statistics, University of Toronto, 100 St. George St., Toronto, ON M5S 3G3 Canada e-mail: keith@utstat.toronto.edu

AMS 2000 subject classifications: Primary 62J05; secondary 62G20

Keywords and phrases: analytic center estimator, set membership estimation, Poisson processes

$\widehat{\boldsymbol{\beta}}_n$ is the analytic center (Sonnevend, 1985) of the membership set $\mathcal{S}_n$. The idea is that the logarithmic function essentially acts as a barrier function that forces the estimator away from the boundary of $\mathcal{S}_n$ and thus makes the constraint that the estimator must lie in $\mathcal{S}_n$ redundant. In certain applications, the analytic center estimator is computationally convenient since it can be computed efficiently in "on-line" applications, more so other estimators based on the membership set such as the Chebyshev center or the maximum volume inscribed ellipsoid estimators. Bai *et al.* (2000) derive some convergence results for the analytic center estimator but do not give its limiting distribution. In addition, Bai *et al.* (1998), Akçay (2004), and Kitamura *et al.* (2005) discuss properties of the membership set, showing under different conditions that the membership set shrinks to a single point as the sample size increases.

The maximizer of g_n in (1.3) lies in the interior of $\mathcal{S}_n$ and hence $\widehat{\boldsymbol{\beta}}_n$ satisfies

$$\sum_{i=1}^{n} \frac{Y_i - \boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}_n}{\gamma_0^2 - (Y_i - \boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}_n)^2} \boldsymbol{x}_i = \boldsymbol{0}. \tag{1.4}$$

The "classical" approach to asymptotic theory is to approximate (1.4) by a linear function of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ and derive the limiting distribution of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ via this approximation. However, expanding (1.4) in a Taylor series around $\boldsymbol{\beta}$, it is easy to see that if the distribution of $\{\varepsilon_i\}$ has a sufficiently large concentration of probability in a neighbourhood of $\pm\gamma_0$ then asymptotic normality will not hold. Intuitively, we should have a faster convergence rate in such cases but a different approach is needed to prove this.

In this paper, we will consider the asymptotic distributions of both the membership set and the analytic center estimator under the assumption that the noise distribution is regularly varying at the boundaries $\pm\gamma_0$ of the error distribution. In section 2, we provide some of the necessary technical foundation for section 3 where we derive the asymptotics of the membership set and the analytic center estimator.

2. Technical preliminaries

Define F to be the distribution function of $\{\varepsilon_i\}$; we then define non-decreasing functions G_1 and G_2 on $[0, 2\gamma_0]$ by

$$G_1(t) = 1 - F(\gamma_0 - t) \tag{2.1}$$

$$G_2(t) = F(-\gamma_0 + t). \tag{2.2}$$

We will assume that both G_1 and G_2 are regularly varying at 0 with the same parameter of regular variation α and that G_1 and G_2 are "balanced" in a neighbourhood of 0. More precisely, for each $x > 0$,

$$\lim_{t \downarrow 0} \frac{G_k(tx)}{G_k(t)} = x^{\alpha} \quad \text{for } k = 1, 2$$

and

$$\lim_{t \downarrow 0} \frac{G_1(t)}{G_1(t) + G_2(t)} = \kappa$$

where $0 < \kappa < 1$. Thus for some sequence of constants $\{a_n\}$ with $a_n \to \infty$ and some $\alpha > 0$, we have

$$\lim_{n\to\infty} nG_1(t/a_n) = \kappa t^{\alpha} \tag{2.3}$$

$$\lim_{n\to\infty} nG_2(t/a_n) = (1-\kappa) t^{\alpha} \tag{2.4}$$

where $0 < \kappa < 1$. The parameter α describes the concentration of probability mass close to the endpoints $\pm\gamma_0$; this concentration increases as α becomes smaller.

The type of convergence as well as the rate of convergence are determined by α. If $\alpha > 2$, we can approximate the left hand side of (1.4) by a linear function and obtain asymptotic normality using the classical argument. On the other hand, when $\alpha < 2$, the limiting distribution is determined by the the errors lying close to the endpoints $\pm\gamma_0$; in particular, given the conditions (2.1)–(2.4) on the distribution F of $\{\varepsilon_i\}$, it is straightforward to derive a point process convergence result for the number of $\{\varepsilon_i\}$ lying within $O(a_n^{-1})$ of $\pm\gamma_0$.

We will make the following assumptions about the errors $\{\varepsilon_i\}$ and the design $\{\boldsymbol{x}_i\}$:

(A1) $\{\varepsilon_i\}$ are i.i.d. random variables on $[-\gamma_0, \gamma_0]$ with distribution function F where G_1 and G_2 defined in (2.1) and (2.2) satisfy (2.3) and (2.4) for some sequence $\{a_n\}$, $\alpha > 0$, and $0 < \kappa < 1$.

(A2) There exists a probability measure μ on R^p such that for each set B with $\mu(\partial B) = 0$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I(\boldsymbol{x}_i \in B) = \mu(B).$$

Moreover, the mass of μ is not concentrated on a lower dimensional subspace of R^p.

Under conditions (A1) and (A2), it is easy to verify that the point process

$$M_n(A \times B) = \sum_{i=1}^{n} I\left\{a_n(\gamma_0 - \varepsilon_i) \in A, -\boldsymbol{x}_i \in B\right\} + \sum_{i=1}^{n} I\left\{a_n(\gamma_0 + \varepsilon_i) \in A, \boldsymbol{x}_i \in B\right\} \tag{2.5}$$

converges in distribution with respect to the vague topology on measures (Kallenberg, 1983) to a Poisson process M whose mean measure is given by

$$E[M(A \times B)] = \left\{\alpha \int_A t^{\alpha-1}\, dt\right\} \bar{\mu}(B) \tag{2.6}$$

where

$$\bar{\mu}(B) = \kappa\mu(-B) + (1-\kappa)\mu(B). \tag{2.7}$$

We can represent the points of the limiting Poisson process M in terms of two independent sequences of i.i.d. random variables $\{E_i\}$ and $\{\boldsymbol{X}_i\}$ where $\{E_i\}$ are exponential with mean 1 and $\{\boldsymbol{X}_i\}$ have the measure $\bar{\mu}$ defined in (2.7). For a given value of α, we then define

$$\Gamma_i = E_1 + \cdots + E_i \quad \text{for } i \geq 1. \tag{2.8}$$

The points of the Poisson process M in (2.5) (with mean measure given in (2.6)) are then represented by $\{(\Gamma_i^{1/\alpha}, \boldsymbol{X}_i) : i \geq 1\}$.

In the case where the support of $\{\boldsymbol{x}_i\}$ (and of the limiting measure μ) is unbounded, we need to make some additional assumptions; note that (A3) and (A4) below hold trivially (given (A1) and (A2)) if $\{\boldsymbol{x}_i\}$ are bounded.

(A3) G_1 and G_2 defined in (2.1) and (2.2) satisfy

$$n\{G_1(t/a_n) + G_2(t/a_n)\} = t^\alpha\{1 + r_n(t)\}$$

where for any $\boldsymbol{u}$,

$$\max_{1\le i\le n} |r_n(\boldsymbol{x}_i^T\boldsymbol{u})| \to 0.$$

(A4) For the measure μ defined in (A2),

$$\frac{1}{n}\sum_{i=1}^n \|\boldsymbol{x}_i\|^\alpha \to \int \|\boldsymbol{x}\|^\alpha\,\mu(d\boldsymbol{x}) < \infty.$$

Moreover,

$$\frac{1}{n}\max_{1\le i\le n} \|\boldsymbol{x}_i\|^\alpha \to 0.$$

As stated above, $\widehat{\boldsymbol{\beta}}_n$ maximizes a concave objective function or, equivalently, minimizes a convex objective function. The key tool that will be used in deriving the limiting distribution of $\widehat{\boldsymbol{\beta}}_n$ is the notion of epi-convergence in distribution (Geyer, 1994; Pflug, 2005; Knight, 2001; Chernozhukov, 2005; Chernozhukov and Hong, 2004) and point process convergence for extreme values (Kallenberg, 1983; Leadbetter *et al*, 1983).

3. Asymptotics

It is instructive to first consider the asymptotic behaviour of the membership set as a random set. Define a centered and rescaled version of $\mathcal{S}_n$ defined in (1.2):

$$\begin{aligned}\mathcal{S}_n' &= a_n(\mathcal{S}_n - \boldsymbol{\beta}) \\ &= \bigcap_{i=1}^n \left\{\boldsymbol{u} : a_n(\varepsilon_i - \gamma_0) \le \boldsymbol{u}^T\boldsymbol{x}_i \le a_n(\varepsilon_i + \gamma_0)\right\}. \end{aligned} \tag{3.1}$$

Note that $\mathcal{S}_n'$ is closely related to the point process M_n defined in (2.5).

The following result describes the asymptotic behaviour of $\{\mathcal{S}_n'\}$ as a sequence of random closed sets using the topology induced by Painlevé-Kuratowski convergence (Molchanov, 2005). Since we have a finite dimensional space, it follows that the Painlevé-Kuratowski topology coincides with the Fell (hit or miss) topology (Beer, 1993); thus $\mathcal{S}_n' \xrightarrow{d} \mathcal{S}'$ if

$$P\left(\mathcal{S}_n' \cap K \ne \emptyset\right) \to P\left(\mathcal{S}' \cap K \ne \emptyset\right)$$

for all compact sets K such that

$$P\left(\mathcal{S}' \cap K \ne \emptyset\right) = P\left(\mathcal{S}' \cap \operatorname{int} K \ne \emptyset\right).$$

It turns out that the convexity of the random sets $\{\mathcal{S}_n'\}$ provides a very simple sufficient condition for checking convergence in distribution.

Lemma 3.1. *Assume the model (1.1) and conditions (A1)–(A4). If $\mathcal{S}_n'$ is defined as in (3.1) then*

$$\mathcal{S}_n' \xrightarrow{d} \mathcal{S}' = \bigcap_{i=1}^\infty \left\{\boldsymbol{u} : \boldsymbol{u}^T\boldsymbol{X}_i \le \Gamma_i^{1/\alpha}\right\} \tag{3.2}$$

where $\{\Gamma_i\}$, $\{\boldsymbol{X}_i\}$ are independent sequences with Γ_i defined in (2.8) and $\{\boldsymbol{X}_i\}$ i.i.d. with distribution $\bar{\mu}$ defined in (2.7).

Proof. First, note that $\mathcal{S}'$ has an open interior with probability 1. To see this, define

$$\begin{aligned}
\mathcal{S}'' &= \bigcap_{i=1}^{\infty} \left\{ \boldsymbol{u} : \|\boldsymbol{u}\| \|\boldsymbol{X}_i\| \leq \Gamma_i^{1/\alpha} \right\} \\
&= \left\{ \boldsymbol{u} : \|\boldsymbol{u}\| \leq \min_i \frac{\Gamma_i^{1/\alpha}}{\|\boldsymbol{X}_i\|} \right\}
\end{aligned}$$

and note that $\mathcal{S}'' \subset \mathcal{S}'$. Using the properties of the Poisson process M whose mean measure is defined in (2.6), we have

$$P\left(\min_i \frac{\Gamma_i^{1/\alpha}}{\|\boldsymbol{X}_i\|} > r \right) = \exp\left(-r^{\alpha} \int \|\boldsymbol{x}\|^{\alpha} \, \bar{\mu}(d\boldsymbol{x}) \right)$$

for $r \geq 0$. Thus $\mathcal{S}''$ contains an open set with probability 1 and therefore so must $\mathcal{S}'$. The fact that $\mathcal{S}'$ contains an open set makes proof of convergence in distribution very simple; we simply need to show that

$$P\left(\boldsymbol{u}_1 \in \mathcal{S}_n', \cdots, \boldsymbol{u}_k \in \mathcal{S}_n'\right) \to P\left(\boldsymbol{u}_1 \in \mathcal{S}', \cdots, \boldsymbol{u}_k \in \mathcal{S}'\right)$$

for any $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_k$. Defining $x_+ = xI(x > 0)$ and $x_- = -xI(x < 0)$, we then have

$$\begin{aligned}
&P\left(\boldsymbol{u}_1 \in \mathcal{S}_n', \cdots, \boldsymbol{u}_k \in \mathcal{S}_n'\right) \\
&= \prod_{i=1}^{n} \left\{ 1 - G_2\left(a_n^{-1} \max_{1\leq j\leq k} (\boldsymbol{u}_j^T \boldsymbol{x}_i)_+ \right) - G_1\left(a_n^{-1} \min_{1\leq j\leq k} (\boldsymbol{u}_j^T \boldsymbol{x}_i)_- \right) \right\} \\
&\to \exp\left[-\int \left\{ (1-\kappa) \left(\max_{1\leq j\leq k} \boldsymbol{u}_j^T \boldsymbol{x} \right)_+^{\alpha} + \kappa \left(\min_{1\leq j\leq k} \boldsymbol{u}_j^T \boldsymbol{x} \right)_-^{\alpha} \right\} \mu(d\boldsymbol{x}) \right] \\
&= \exp\left\{ -\int \left(\max_{1\leq j\leq k} \boldsymbol{u}_j^T \boldsymbol{x} \right)_+^{\alpha} \bar{\mu}(d\boldsymbol{x}) \right\} \\
&= P\left(\boldsymbol{u}_1 \in \mathcal{S}', \cdots, \boldsymbol{u}_k \in \mathcal{S}'\right),
\end{aligned}$$

which completes the proof. □

Note that $\mathcal{S}'$ is bounded with probability 1; this follows since for any $\boldsymbol{u} \neq \boldsymbol{0}$, $P(\boldsymbol{X}_i^T \boldsymbol{u} > 0) \geq \min(\kappa, 1-\kappa) > 0$, hence $P(\boldsymbol{X}_i^T \boldsymbol{u} > 0$ infinitely often$) = 1$. Thus with probability 1, for each $\boldsymbol{u} \in \mathcal{S}'$ there exists j such that such that $0 < \boldsymbol{X}_j^T \boldsymbol{u} \leq \Gamma_j$ and so for t sufficiently large $t\boldsymbol{u} \notin \mathcal{S}'$.

Lemma 3.1 says that points in the membership set lie within $O_p(a_n^{-1})$ of $\boldsymbol{\beta}$ and therefore the analytic center estimator $\widehat{\boldsymbol{\beta}}_n$ (or indeed any estimator based on the membership set) must satisfy $\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} = O_p(a_n^{-1})$. Since $a_n = n^{1/\alpha} L(n)$ it follows that we have a faster than $O_p(n^{-1/2})$ convergence rate when $\alpha < 2$. On the other hand, if $\alpha > 2$ then $n^{1/2}/a_n \to \infty$; fortunately, in these cases, it is typically possible to achieve $O_p(n^{-1/2})$ convergence.

Theorem 3.1. *Assume the model (1.1) and conditions (A1)–(A4) for some $\alpha \geq 2$ and assume that*

$$E[(\gamma_0 - \varepsilon_i)^{-1}] = E[(\gamma_0 + \varepsilon_i)^{-1}].$$

Suppose that $\widehat{\boldsymbol{\beta}}_n$ maximizes (1.3).

(i) If $\alpha > 2$ *then* $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 C^{-1})$ *where*

$$\begin{aligned} \sigma^2 &= Var[\varepsilon_i/(\gamma_0^2 - \varepsilon_i^2)] \Big/ \left\{E[(\gamma_0^2 + \varepsilon_i^2)/(\gamma_0^2 - \varepsilon_i^2)^2]\right\}^2 \\ \text{and} \quad C &= \int \boldsymbol{x}\boldsymbol{x}^T \, \mu(d\boldsymbol{x}). \end{aligned}$$

(ii) If $\alpha = 2$ *then*

$$\frac{b_n^{(1)}}{b_n^{(2)}}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, C^{-1})$$

where $\{b_n^{(1)}\}$ *satisfies*

$$\frac{1}{b_n^{(1)}} \sum_{i=1}^n \frac{\gamma_0^2 + \varepsilon_i^2}{(\gamma_0^2 - \varepsilon_i^2)^2} \xrightarrow{p} 1$$

and $\{b_n^{(2)}\}$ *satisfies*

$$\frac{1}{b_n^{(2)}} \sum_{i=1}^n \frac{\varepsilon_i}{\gamma_0^2 - \varepsilon_i^2} \boldsymbol{x}_i \xrightarrow{d} \mathcal{N}(\mathbf{0}, C^{-1}).$$

The proof of Theorem 3.1 is standard and will not be given here. Note that conditions (A2)–(A4) are much stronger than necessary for Theorem 3.1 to hold. For example, we need only assume that

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_i^T &\to C \\ \text{and} \quad \frac{1}{n} \max_{1 \le i \le n} \|\boldsymbol{x}_i\|^2 &\to 0 \end{aligned}$$

for asymptotic normality to hold. More generally, Theorem 3.1 also holds in the case where the bounds $\pm\gamma_0$ are overly conservative in the sense that for some $\epsilon > 0$,

$$P(-\gamma_0 + \epsilon \le \varepsilon_i \le \gamma_0 - \epsilon) = 1.$$

In this case, if the model (1.1) contains an intercept (that is, one element of $\boldsymbol{x}_i$ is always 1) then we can rewrite the model (1.1) as

$$\begin{aligned} Y_i &= \theta + \boldsymbol{x}_i^T \boldsymbol{\beta} + (\varepsilon_i - \theta) \\ &= \boldsymbol{x}_i^T \boldsymbol{\beta}' + \varepsilon_i' \quad (i = 1, \cdots, n) \end{aligned}$$

where $\varepsilon_i' = \varepsilon_i - \theta$. Then there exists θ such that $\{\varepsilon_i'\}$ satisfies the moment conditions in Theorem 3.1 and so the proof of Theorem 3.1 will go through as before.

When $\alpha < 2$, the limiting behaviour of $\widehat{\boldsymbol{\beta}}_n$ is highly dependent on the limiting Poisson process M (with mean measure given in by (2.6)). In particular, the sequences of random variables $\{(\gamma_0 - \varepsilon_i)^{-1}\}$ and $\{(\gamma_0 + \varepsilon_i)^{-1}\}$ lie in the domain of a stable law with index α and so it is not surprising to have non-Gaussian limiting distributions.

Theorem 3.2. *Assume the model (1.1) and conditions (A1)–(A4) for some* $0 < \alpha < 2$ *and assume that* $\widehat{\boldsymbol{\beta}}_n$ *maximizes (1.3). Define* $\{\Gamma_i\}$ *and* $\{\boldsymbol{X}_i\}$ *as in Lemma 3.1 and* $\mathcal{S}'$ *as in (3.2).*

(a) If $\alpha < 1$ *then* $a_n(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} \boldsymbol{U}$ *where* $\boldsymbol{U}$ *maximizes*

$$\sum_{i=1}^{\infty} \ln\left(1 - \frac{\boldsymbol{X}_i^T \boldsymbol{u}}{\Gamma_i^{1/\alpha}}\right)$$

over $\boldsymbol{u} \in \mathcal{S}'$.

(b) If $\alpha = 1$ *and*

$$n a_n^{-1} E\left[\frac{\varepsilon_i}{\gamma_0^2 - \varepsilon_i^2} I\left(\left|\frac{\varepsilon_i}{\gamma_0^2 - \varepsilon_i^2}\right| \le a_n\right)\right] \to 0$$

then $a_n(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} \boldsymbol{U}$ *where* $\boldsymbol{U}$ *maximizes*

$$\sum_{i=1}^{\infty} \ell\left(\frac{\boldsymbol{X}_i^T \boldsymbol{u}}{\Gamma_i^{1/\alpha}}\right) - \sum_{i=1}^{\infty}\left\{\frac{\boldsymbol{X}_i^T \boldsymbol{u}}{\Gamma_i^{1/\alpha}} - E\left(\frac{\boldsymbol{X}_i^T \boldsymbol{u}}{\Gamma_i^{1/\alpha}} I(\Gamma_i^{1/\alpha} \ge 1)\right)\right\}$$

over $\boldsymbol{u} \in \mathcal{S}'$ *where* $\ell(x) = \ln(1-x) + x$.

(c) If $1 < \alpha < 2$ *and*

$$E[(\gamma_0 - \varepsilon_i)^{-1}] = E[(\gamma_0 + \varepsilon_i)^{-1}]$$

then $a_n(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} \boldsymbol{U}$ *where* $\boldsymbol{U}$ *maximizes*

$$\sum_{i=1}^{\infty} \ell\left(\frac{\boldsymbol{X}_i^T \boldsymbol{u}}{\Gamma_i^{1/\alpha}}\right) - \sum_{i=1}^{\infty}\left\{\frac{\boldsymbol{X}_i^T \boldsymbol{u}}{\Gamma_i^{1/\alpha}} - E\left(\frac{\boldsymbol{X}_i^T \boldsymbol{u}}{\Gamma_i^{1/\alpha}}\right)\right\}$$

over $\boldsymbol{u} \in \mathcal{S}'$ *where* $\ell(x) = \ln(1-x) + x$.

Proof. $a_n(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ maximizes the concave function

$$Z_n(\boldsymbol{u}) = \sum_{i=1}^{n}\left\{\ln\left(1 + \frac{\boldsymbol{x}_i^T \boldsymbol{u}}{a_n(\gamma_0 - \varepsilon_i)}\right) + \ln\left(1 - \frac{\boldsymbol{x}_i^T \boldsymbol{u}}{a_n(\gamma_0 + \varepsilon_i)}\right)\right\}$$

subject to $\boldsymbol{u} \in \mathcal{S}_n'$ defined in (3.1). Since the limiting objective function is finite on an open set (since $\mathcal{S}'$ contains an open set with probability 1), it suffices to show finite dimensional weak convergence of Z_n. Note that we can write (for $\boldsymbol{u} \in \mathcal{S}_n'$),

$$Z_n(\boldsymbol{u}) = \int \ln\left(1 - \frac{\boldsymbol{x}^T \boldsymbol{u}}{w}\right) M_n(dw \times d\boldsymbol{x})$$

where M_n is defined in (2.5). For $\alpha < 1$, we approximate $\ln(1 + \boldsymbol{x}^T \boldsymbol{u}/w)$ by a sequence of bounded functions $\{g_m(w, \boldsymbol{x}; \boldsymbol{u})\}$. Following Lepage *et al.* (1981), we have

$$\begin{aligned} \int g_m(w, \boldsymbol{x}; \boldsymbol{u}) \quad &\xrightarrow{d} \quad \sum_{i=1}^{\infty} g_m(\Gamma_i^{1/\alpha}, \boldsymbol{X}_i; \boldsymbol{u}) \quad \text{as } n \to \infty \\ &\to \quad \sum_{i=1}^{\infty} \ln(1 - \boldsymbol{X}_i^T \boldsymbol{u}/\Gamma_i^{1/\alpha}) \quad \text{with probability 1 as } m \to \infty \end{aligned}$$

and

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\left[\left|\int \left\{\ln(1 + \boldsymbol{x}^T \boldsymbol{u}/w) - g_m(w, \boldsymbol{x}; \boldsymbol{u})\right\} M_n(dw \times d\boldsymbol{x})\right| > \epsilon\right] = 0.$$

For $1 \le \alpha < 2$, a similar argument works by writing $\ln(1+\boldsymbol{x}^T\boldsymbol{u}/w) = \boldsymbol{x}^T\boldsymbol{u}/w + \ell(\boldsymbol{x}^T\boldsymbol{u}/w)$ and applying the argument used for $\alpha < 1$ to

$$\int \ell(\boldsymbol{x}^T\boldsymbol{u}/w)\, M_n(dw \times d\boldsymbol{x}) = \sum_{i=1}^n \left\{ \ell\left(-\frac{\boldsymbol{x}_i^T\boldsymbol{u}}{a_n(\gamma_0-\varepsilon_i)}\right) + \ell\left(\frac{\boldsymbol{x}_i^T\boldsymbol{u}}{a_n(\gamma_0+\varepsilon_i)}\right)\right\}.$$

The result now follows by noting that, in each case, the limiting objective function Z has a unique maximizer on the set $\mathcal{S}'$; to see this, note that Z is strictly concave on $\mathcal{S}'$ and that as $\boldsymbol{u} \to \partial\mathcal{S}'$, $Z(\boldsymbol{u}) \to -\infty$. □

In Theorem 3.2, note that no moment condition is needed when $\alpha < 1$. In this case, the limit of $a_n(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$, $\boldsymbol{U}$, can be interpreted as the analytic center of the random set $\mathcal{S}'$, and thus

$$P(\boldsymbol{U} \in \operatorname{int}\mathcal{S}') = 1.$$

In contrast, we require a moment condition for $1 \le \alpha < 2$ (such as

$$E[(\gamma_0-\varepsilon_i)^{-1}] = E[(\gamma_0+\varepsilon_i)^{-1}] \tag{3.3}$$

for $\alpha > 1$) in order to have $P(\boldsymbol{U} \in \operatorname{int}\mathcal{S}') = 1$. What happens if the moment condition, for example (3.3), fails? Theorem 3.3 below states that the limiting distribution of $a_n(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ is concentrated the vertices of the limiting membership set $\mathcal{S}'$.

Theorem 3.3. *Assume the model (1.1) and conditions (A1)–(A4) for some $\alpha \ge 1$ and assume that $\widehat{\boldsymbol{\beta}}_n$ maximizes (1.3). Define $\mathcal{S}'$ as in (3.2) with $\{\Gamma_i\}$ and $\{\boldsymbol{X}_i\}$ as in Lemma 3.1. If for some (non-negative) sequence $\{b_n\}$ ($b_n = n$ for $\alpha > 1$)*

$$b_n^{-1}\sum_{i=1}^n \left\{(\gamma_0-\varepsilon_i)^{-1} - (\gamma_0+\varepsilon_i)^{-1}\right\} \xrightarrow{p} \omega \ne 0$$

then $a_n(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} \boldsymbol{U}$ where $\boldsymbol{U}$ maximizes

$$\omega \int \boldsymbol{u}^T\boldsymbol{x}\,\mu(d\boldsymbol{x}) \quad \text{subject to } \boldsymbol{u} \in \mathcal{S}'.$$

Proof. $a_n(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ maximizes

$$Z_n(\boldsymbol{u}) = \frac{a_n}{b_n}\sum_{i=1}^n \left\{\ln\left(1+\frac{\boldsymbol{x}_i^T\boldsymbol{u}}{a_n(\gamma_0-\varepsilon_i)}\right) + \ln\left(1-\frac{\boldsymbol{x}_i^T\boldsymbol{u}}{a_n(\gamma_0+\varepsilon_i)}\right)\right\}$$

for $\boldsymbol{u} \in \mathcal{S}_n'$. Defining $\ell(x) = \ln(1-x)+x$ as before, we have (for $\boldsymbol{u} \in \mathcal{S}_n'$),

$$\begin{aligned} Z_n(\boldsymbol{u}) &= \frac{1}{b_n}\sum_{i=1}^n \boldsymbol{x}_i^T\boldsymbol{u}\left\{(\gamma_0-\varepsilon_i)^{-1} - (\gamma_0+\varepsilon_i)^{-1}\right\} \\ &\quad + \frac{a_n}{b_n}\sum_{i=1}^n \left\{\ell\left(-\frac{\boldsymbol{x}_i^T\boldsymbol{u}}{a_n(\gamma_0-\varepsilon_i)}\right) + \ell\left(\frac{\boldsymbol{x}_i^T\boldsymbol{u}}{a_n(\gamma_0+\varepsilon_i)}\right)\right\} \\ &= \frac{1}{b_n}\sum_{i=1}^n \boldsymbol{x}_i^T\boldsymbol{u}\left\{(\gamma_0-\varepsilon_i)^{-1} - (\gamma_0+\varepsilon_i)^{-1}\right\} + o_p(1) \\ &\xrightarrow{p} \omega\int \boldsymbol{u}^T\boldsymbol{x}\,\mu(d\boldsymbol{x}) \end{aligned}$$

noting that $a_n = o(b_n)$ and applying the results of Adler and Rosalsky (1991). Since $\mathcal{S}'$ is bounded, the linear function $\omega \int \boldsymbol{u}^T \boldsymbol{x}\, \mu(d\boldsymbol{x})$ has a finite maximum on $\mathcal{S}'$. Uniqueness follows from the assumption that the measure μ puts zero mass on lower dimensional subsets. □

For $\alpha > 1$, $\omega = E[(\gamma_0 - \varepsilon_i)^{-1} - (\gamma_0 + \varepsilon_i)^{-1}]$ while for $\alpha = 1$, ω is typically first moment of an appropriately truncated version of $(\gamma_0 - \varepsilon_i)^{-1} - (\gamma_0 + \varepsilon_i)^{-1}$ where the truncation depends on the slowly varying component of the distribution function F near $\pm\gamma_0$. Note that the limiting distribution of $a_n(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ depends on ω only via its sign. Like κ, ω is a measure of the relative weight of the distribution F near its endpoints $\pm\gamma_0$. However, they are not necessarily related in the sense that for a given value of κ, ω can be positive or negative; for example, $\kappa > 1/2$ does not imply that $\omega > 0$. The following implication of Theorem 3.3 is interesting: Even though $\widehat{\boldsymbol{\beta}}_n$ lies in the interior of $\mathcal{S}_n$ (and thus $a_n(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ lies in the interior of $\mathcal{S}'_n$), the limiting distribution is concentrated on the boundary of $\mathcal{S}'$.

It is also interesting to compare the limiting distribution of the analytic center estimator to those of other estimator, for example, the least squares estimator constrained to the membership set and the Chebyshev center estimator. The constrained least squares estimator $\widetilde{\boldsymbol{\beta}}_n$ minimizes

$$\sum_{i=1}^{n} (Y_i - \boldsymbol{x}_i^T \boldsymbol{\phi})^2$$

subject to $\boldsymbol{\phi} \in \mathcal{S}_n$. The asymptotics of $\widetilde{\boldsymbol{\beta}}_n$ depend on whether or not $E(\varepsilon_i) = 0$. If $E(\varepsilon_i) \neq 0$ then $a_n(\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ converges in distribution to the maximizer of $E(\varepsilon_i) \int \boldsymbol{x}^T \boldsymbol{u}\, \mu(d\boldsymbol{x})$ subject to $\boldsymbol{u} \in \mathcal{S}'$ similar to the result of Theorem 3.3 with ω defined differently; note that this result holds for any $\alpha > 0$. Moreover, for $\alpha \geq 1$, the analytic center estimator $\widehat{\boldsymbol{\beta}}_n$ and the constrained least squares estimator $\widetilde{\boldsymbol{\beta}}_n$ have the same limiting distribution if both ω and $E(\varepsilon_i)$ are non-zero and have the same sign. On the other hand, when $E(\varepsilon_i) = 0$, the type of limiting distribution depends on α, specifically whether or not $\alpha < 2$. If $\alpha \geq 2$ then $\widetilde{\boldsymbol{\beta}}_n$ has the same limiting distribution as the unconstrained least squares estimator; for example, for $\alpha > 2$, we have

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathrm{Var}(\varepsilon_i) C^{-1})$$

where C is defined as in Theorem 3.1. For $\alpha < 2$, $a_n(\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ converges in distribution to the maximizer of $\boldsymbol{W}^T \boldsymbol{u}$ subject to $\boldsymbol{u} \in \mathcal{S}'$ where $\boldsymbol{W} \sim \mathcal{N}(\mathbf{0}, C)$ and $\boldsymbol{W}$ is independent of the Poisson process defining $\mathcal{S}'$.

Similarly, we can derive the asymptotics for the Chebyshev center estimator, defined as the center of largest radius ball (in the L_r norm) contained within $\mathcal{S}_n$; $\widetilde{\boldsymbol{\beta}}_n$ maximizes δ subject to the constraints

$$\begin{aligned} \boldsymbol{x}_i^T \boldsymbol{\phi} + \|\boldsymbol{x}_i\|_q \delta &\leq Y_i + \gamma_0 \quad \text{for } i = 1, \cdots, n \\ -\boldsymbol{x}_i^T \boldsymbol{\phi} + \|\boldsymbol{x}_i\|_q \delta &\leq \gamma_0 - Y_i \quad \text{for } i = 1, \cdots, n \end{aligned}$$

where q is such that $r^{-1} + q^{-1} = 1$. If $(\widetilde{\boldsymbol{\beta}}_n, \Delta_n)$ is the solution of this linear program then $(a_n(\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}), a_n \Delta_n) \xrightarrow{d} (\boldsymbol{U}, \Delta_0)$ where the limit maximizes δ subject to

$$\boldsymbol{u}^T \boldsymbol{X}_i + \delta \|\boldsymbol{X}_i\|_q \leq \Gamma_i^{1/\alpha} \quad \text{for } i \geq 1.$$

Note that $P(\boldsymbol{U} \in \mathrm{int}\, \mathcal{S}') = 1$ without any moment conditions. The downside of the Chebyshev center estimator is that it is somewhat computationally more complex than the analytic center estimator.

Acknowledgements

The author would like to thank the referees for their very useful comments on this paper.

References

[1] Adler, A. and Rosalsky, A. (1991) On the weak law of large numbers for normed weighted sums of I.I.D. random variables. *International Journal of Mathematics and Mathematical Sciences.* **14**, 191-202.

[2] Akçay, H. (2004) The size of the membership-set in a probabilistic framework. *Automatica.* **40**, 253-260.

[3] Akçay, H., Hjalmarsson, H. and Ljung, L. (1996) On the choice of norm in system identification. *IEEE Transactions on Automatic Control.* **41**, 1367-1372.

[4] Bai, E.W., Cho, H. and Tempo, R. (1998) Convergence properties of the membership set. *Automatica.* **34**, 1245-1249.

[5] Bai, E.W., Fu, M., Tempo, R. and Ye, Y. (2000) Convergence results of the analytic center estimator. *IEEE Transactions on Automatic Control.* **45**, 569-572.

[6] Beer, G. (1993) *Topologies on Closed and Closed Convex Sets.* Kluwer, Dordrecht.

[7] Chernozhukov, V. (2005) Extremal quantile regression. *Annals of Statistics.* **33**, 806-839.

[8] Chernozhukov, V. and Hong, H. (2004) Likelihood estimation and inference in a class of non-regular econometric models. *Econometrica.* **77**, 1445-1480.

[9] Geyer, C.J. (1994) On the asymptotics of constrained M-estimation. *Annals of Statistics.* **22**, 1993-2010.

[10] Kallenberg, O. (1983) *Random Measures.* (third edition) Akademie-Verlag.

[11] Kitamura, W., Fujisaki, Y. and Bai, E.W. (2005) The size of the membership set in the presence of disturbance and parameter uncertainty. *Proceedings of the 44th IEEE Conference on Decision and Control.* 5698-5703.

[12] Knight, K. (2001) Limiting distributions of linear programming estimators. *Extremes.* **4**, 87-104.

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

[14] Lepage, R., Woodroofe, M. and Zinn, J. (1981) Convergence to a stable distribution via order statistics. *Annals of Probability.* **9**, 624-632.

[15] Mäkilä, P.M. (1991) Robust identification and Galois sequences. *International Journal of Control.* **54**, 1189-1200.

[16] Milanese, M. and Belforte, G. (1982) Estimation theory and uncertainty intervals evaluation in the presence of unknown but bounded errors – linear families of models and estimators. *IEEE Transactions on Automatic Control.* **27**, 408-414.

[17] Molchanov, I. (2005) *Theory of Random Sets.* Springer, London.

[18] Pflug, G. Ch. (1995) Asymptotic stochastic programs. *Mathematics of Operations Research.* **20**, 769-789.

[19] Schweppe, F.C. (1968) Recursive state estimation: unknown but bounded errors and system inputs. *IEEE Transactions on Automatic Control.* **13**, 22-28.

[20] Sonnevend, G. (1985) An analytic center for polyhedrons and new classes of global algorithms for linear (smooth convex) programmming. In *Lecture Notes in Control and Information Sciences.* **84**, 866-876.

[21] Tse, D.N.C., Daleh, M.A. and Tsitsiklis, J.N. (1993) Optimal asymptotic identification under bounded disturbances. *IEEE Transactions on Automatic Control.* **38**, 1176-1190.