for $\epsilon > 0$, by Markov's inequality. It follows readily from (4) that

$$\sum_{n=1}^{\infty} P[\,|S_n - np\,| \geq n\epsilon] < \infty \qquad \text{for every } \epsilon > 0.$$

This fact is well known (e.g., Chow and Teicher 1978, p. 40), and in view of the Borel–Cantelli lemma, it is a stronger result than (1).

[Received June 1983. Revised February 1984.]

REFERENCES

CHOW, Y.S., and TEICHER, HENRY (1978), *Probability Theory: Independence, Interchangeability, Martingales*, New York: Springer-Verlag.

CHUNG, K.L. (1968), *A Course in Probability Theory*, New York: Harcourt, Brace and World.

LOÈVE, M. (1977), *Probability Theory* (4th ed.), New York: Springer-Verlag.

ROHATGI, V.K. (1976), *An Introduction to Probability Theory and Mathematical Statistics*, New York: John Wiley.

TOMKINS, R.J. (1979), "General Approaches to Strong Limit Theorems," *Selecta Statistica Canadiana*, 5, 173–200.

Four (Pathological) Examples in Asymptotic Statistics

ROGER W. KOENKER and GILBERT W. BASSETT*

Four simple examples illustrating varieties of pathological asymptotic behavior are presented. The examples are based on some recent work on l_1 asymptotics. The examples have some pedagogical value in clarifying the role of certain standard regularity conditions.

KEY WORDS: Asymptotics; Linear model; Median; Least absolute value estimation.

INTRODUCTION

This article offers four simple examples of unusual asymptotic statistical behavior based on recent work on the large-sample theory of l_1 (least absolute error) estimation. The first illustrates a case in which an estimator converges in probability but fails to converge with probability one. The second example illustrates a case in which n^{-1} times the objective function used to define an estimator converges to zero in neighborhoods of the true value, so the usual (Cesaro) identifiability condition fails; nevertheless the estimator converges in probability. The third example provides an unusual case of convergence in distribution in which the limiting distribution is bimodal and the convergence rate is $(\log \sqrt{n})^{-1/2}$. Finally, the fourth example illustrates a case in which the design is uninformative and $\sum x_i^2$ stays bounded, but weak consistency is salvaged by making the error distribution sufficiently peaked.

The article is intended to be heuristic, motivated by the observation that one can often gain insight by studying simple examples in which standard conditions fail and behavior verges on the pathological.

* Roger W. Koenker is Professor, Department of Economics, University of Illinois, Urbana-Champaign, IL 61820. Gilbert W. Bassett is Associate Professor, Department of Economics, University of Illinois, Chicago, IL 60680. The authors wish to express their appreciation to Peter Fishburn for helpful discussions and to both referees and the Associate Editor for their constructive suggestions.

1. WEAK, BUT NOT STRONG, CONSISTENCY

Suppose that

$$y_i = x_i\beta + u_i, \qquad i = 1, 2, \ldots, \tag{1.1}$$

where $\{x_i\}$ is a known sequence of scalars, β is an unknown parameter, and the u_i's are iid F with F symmetric about 0. We will consider the behavior of the l_1 estimator $\hat{\beta}$, which minimizes

$$V_n(b) = \sum_{i=1}^{n} |\, y_i - x_i b\,|. \tag{1.2}$$

When the x's grow very rapidly, there are cases in which the l_1 estimator converges weakly (i.e., in probability), but not strongly (i.e., almost surely), as the following example illustrates.

Let $x_i = 2^i$, so $\sum_{i=1}^{n-1} x_i = x_n - 2 < x_n$; thus

$$\hat{\beta}_n = y_n/x_n, \tag{1.3}$$

since for all $\delta \neq 0$,

$$\begin{aligned} &\sum_{i=1}^{n} |\, y_i - x_i(\hat{\beta}_n + \delta)\,| - |\, y_i - x_i\hat{\beta}_n\,| \\ &= |\, x_n\delta\,| + \sum_{i=1}^{n-1} |\, y_i - x_i(\hat{\beta}_n + \delta)\,| - |\, y_i - x_i\hat{\beta}_n\,| \\ &\geq |\, x_n\delta\,| - \sum_{i=1}^{n-1} |\, x_i\delta\,| \\ &= |\,\delta\,|\,[x_n - \sum_{i=1}^{n-1} x_i] \\ &> 0. \end{aligned} \tag{1.4}$$

For weak convergence we require that for all $\epsilon > 0$,

$$\begin{aligned} \Pr[\,|\,\hat{\beta}_n - \beta\,| < \epsilon] &= \Pr[\,|\,u_n/x_n\,| < \epsilon] \\ &= 1 - F(\epsilon x_n) + F(-\epsilon x_n) \\ &\to 0 \end{aligned} \tag{1.5}$$

as $n \to \infty$, which is obviously satisfied for any proper F.

But for strong convergence it is necessary that for any $\epsilon > 0$,

$$\sum_{n=1}^{\infty} \Pr[\,|u_n/x_n| > \epsilon] < \infty \tag{1.6}$$

(see Feller 1968, Vol. 1, p. 201, lemma 2). Note that the events in brackets are mutually independent by the iid hypothesis on the u's and the design assumption.

Now suppose F is symmetric and that in the tails,

$$F(u) = 1 - 1/\log_2(u). \tag{1.7}$$

Then (1.6) becomes, for $\epsilon = 1$ and some n_0,

$$2\sum_{n=1}^{\infty}(1 - F(x_n)) \geq 2\sum_{n=n_0}^{\infty}\frac{1}{n}, \tag{1.8}$$

which diverges. Thus for F satisfying (1.7) and $x_i = 2^i$, we have weak consistency, but not strong.

Since if either side converges,

$$\int_{-\infty}^{\infty} |u|\,F\{du\} = \int_{-\infty}^{\infty}[1 - F(u) + F(-u)]du, \tag{1.9}$$

existence of a first absolute moment for u is sufficient to assure strong convergence. Note, however, that (1.7) is worse than mere lack of a first moment. The Cauchy, for example, has tails like $1 - 1/u$. Here the l_1 estimator, which is generally immune to the tail behavior of u, is nearly undone because of the explosive design. Fixing the tail behavior (1.7) and assuming a strictly positive density for u at the median, the results of Bassett and Koenker (1978) imply weak convergence of $\hat{\beta}$ for the less explosive class of designs satisfying

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 \to D$$

for some $0 < D < \infty$. More explosive designs than $x_i = 2^i$ will also serve to restore strong consistency for a fixed error distribution. Take $F(u)$ with tails given by (1.7). By making the design more explosive, we can obviously make the left side of (1.8) converge. This illustrates the rather delicate interaction of design and error conditions in l_1 consistency arguments.

Examples of weak but not strong convergence may also be constructed for the least squares estimator. These examples are also somewhat pathological, since if $0 < Eu_i^2 < \infty$, then $\tilde{\beta}_n = \sum x_i y_i / \sum x_i^2$ is weakly consistent if and only if it is strongly consistent, and the latter obtains if and only if $\sum x^2 \to \infty$ (see Lai, Robbins, and Wei 1979). Relaxing the finite variance condition creates some scope for excitement. Take the sample mean: If tails are sufficiently fat so that $E\,|u| = \infty$, then strong consistency fails by the Kolmogorov strong law, but weak consistency may still be salvaged. Let $F'(u) = 1/(u^2 \log u)$ in the tails. Then $\tilde{\beta}_n = \bar{y} \to 0$ in probability (see Chung 1974, p. 111). As in the l_1 case, consistency of the least-squares estimator may be salvaged by making the design sufficiently explosive. For example, Chen, Lai, and Wei (1981) showed that if F is Cauchy in (1.1) and

$$n(\log n)^{\delta}/\left(\sum x_i^2\right) = O(1)$$

for some $\delta > 3$, then $\tilde{\beta} \to 0$ almost surely.

2. CONSISTENCY WITHOUT CESARO IDENTIFIABILITY

Again consider the model

$$y_i = x_i\beta + u_i, \tag{2.1}$$

with u_i iid F and F symmetric about 0, but now let

$$x_i = \sqrt{k} \quad \text{for } i = k(k-1)/2 \text{ for some integer } k$$
$$= 0 \quad \text{otherwise.}$$

Thus, $(1/n)\sum_{i=1}^{n} x_i^2 \to 1$, and if F has median 0 and continuous, positive density $f(0)$ at the median, then by Bassett and Koenker (1978),

$$\sqrt{n}(\hat{\beta} - \beta) \to N(0, (4f^2(0))^{-1}). \tag{2.3}$$

Thus a fortiori, $\hat{\beta}$ is consistent for β.

Now consider, however, the limiting behavior of the objective function $V(b)$. In particular, we wish to consider the limiting behavior of $n^{-1}[V(b) - V(0)]$ or, equivalently, setting $\delta = b - \beta$,

$$\begin{aligned} v_n(\delta) &= (1/n)\sum |u_i - x_i\delta| - |u_i| \\ &\leq (1/n)\sum |x_i\delta|. \end{aligned}$$

Typically, in nonlinear estimation problems, we would require that $Ev_n(\delta)$ have a unique minimum at $\delta = 0$. Such a condition is required, for example, in Oberhofer's (1982) paper on l_1 consistency; see Gallant, Burguete, and Souza (1983) for a detailed discussion of this general approach. But in the present example, letting $n = k(k-1)/2$ for some integer k,

$$\begin{aligned} Ev_n(\delta) &\leq (|\delta|/n)\sum_{i=1}^{n}|x_i| \\ &= (2|\delta|/k(k-1))[1 + \sqrt{2} + \cdots + \sqrt{k}\,] \\ &\leq (2|\delta|k\sqrt{k}\,)/k(k-1) \\ &\to 0. \end{aligned}$$

Thus the mean of the objective function converges to zero in any neighborhood of 0, and thus the typical identifiability condition fails. Nevertheless, $\sqrt{n}\,\hat{\beta}$ converges in distribution! This example differs from similar examples involving the least-squares estimator (e.g., see Wu 1981, where least-squares consistency is salvaged without Cesaro identifiability, but convergence in distribution occurs at rates slower than the conventional $1/\sqrt{n}$).

3. PATHOLOGICAL CONVERGENCE IN DISTRIBUTION

We turn to an example in which the design in (1.1) is simply $x_i \equiv 1$, so $\hat{\beta}$ is simply the sample median from a random sample from F. When F has a continuous and strictly positive density, say $f(0)$, at the median, it is well known that

$$\sqrt{n}\,(\hat{\beta} - \beta) \to N(0, (4f^2(0))^{-1}). \tag{3.1}$$

What happens when $f(0)=0$? Suppose, for $\alpha = (\log 2)^{-1/2}$,

$$
\begin{aligned}
F(x) &= 1 & \alpha < x \\
&= \tfrac{1}{2} + e^{-x^{-2}} & 0 < x \le \alpha \\
&= \tfrac{1}{2} & x = 0 \\
&= \tfrac{1}{2} - e^{-x^{-2}} & -\alpha \le x < 0 \\
&= 0 & x < -\alpha, \qquad (3.2)
\end{aligned}
$$

so F has a unique median of zero, and not only is $f(0)=0$, but *all* higher order derivatives of F are also zero at zero.

The sample median $\hat{\beta}$ minimizes

$$V_n(b) = \sum_{i=1}^{n} |y_i - b|, \tag{3.3}$$

so $\hat{\beta} < z$ if and only if the gradient

$$g_n(z) \equiv -\sum_{i=1}^{n} \operatorname{sgn}(y_i - z) > 0, \tag{3.4}$$

where $\operatorname{sgn}(u) = 1$ if $u \ge 0$ and $= -1$ otherwise. Note that since F is absolutely continuous, $y = z$ with probability 0. Let $\{\lambda_n\}$ be a sequence such that $\lambda_n \to \infty$ as $n \to \infty$. Then

$$
\begin{aligned}
\Pr[\lambda_n \hat{\beta} < z] &= \Pr[g_n(z/\lambda_n) > 0] \\
&= \Pr\big[(g_n(z/\lambda_n) - Eg_n)/\sqrt{\operatorname{var} g_n} \\
&\qquad > -Eg_n/\sqrt{\operatorname{var} g_n}\big], \qquad (3.5)
\end{aligned}
$$

where

$$Eg_n = n[1 - 2F(z/\lambda_n)] \tag{3.6}$$

and

$$\operatorname{var} g_n = 4n[F(z/\lambda_n)(1 - F(z/\lambda_n))]. \tag{3.7}$$

Now $(g_n - Eg_n)/(\operatorname{var} g_n)^{1/2}$ is a standardized binomial random variable with probability of success $p_n \to 1/2$ as $z/\lambda_n \to 0$, and therefore it converges in distribution to a standard Gaussian law.

We would like to choose $\{\lambda_n\}$ so that $-Eg_n/(\operatorname{var} g_n)^{1/2}$ converges to something that is bounded away from 0. Since $\operatorname{var} g_n \to n$ for any $\lambda_n \to \infty$,

$$\sqrt{n}\,[1 - 2F(z/\lambda_n)] = \sqrt{n}\,[-2e^{-|z|^{-2}} \cdot e^{-\lambda_n^2}] \tag{3.8}$$

must stay away from 0. Choosing $\lambda_n = (\log \sqrt{n})^{1/2}$ yields

$$\Pr[\lambda_n \hat{\beta} < z] \to \Pr[Z > 2e^{-z^{-2}}], \tag{3.9}$$

where Z is a standard Gaussian random variable. The density function of the limiting distribution is

$$f(x) = \phi(2e^{-x^{-2}})\,|4e^{-x^{-2}} x^{-3}|, \tag{3.10}$$

where ϕ denotes the standard Gaussian density. The density is depicted in Figure 1.

Thus not only is the *rate* of convergence considerably slower than the typical rate $1/\sqrt{n}$, the limiting distribution is also bizarre.

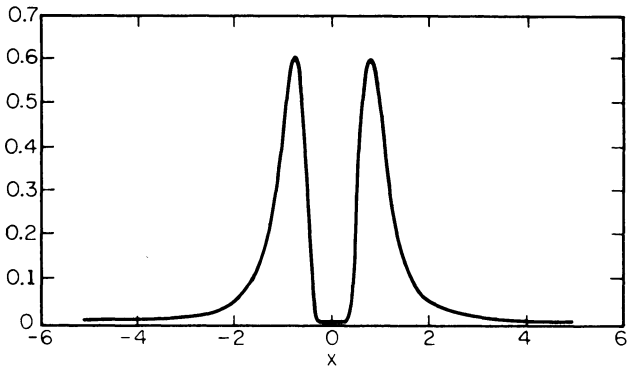

Figure 1. The Limiting Density of the Normalized Sample Median by Random Sampling From the Distribution Function (3.2).

4. CONSISTENCY WITH HARMONIC DESIGNS

When $\sum x_i^2 < \infty$, convergence of $\hat{\beta}$ typically fails, as does the consistency of the least-squares estimator. The present example shows, however, that the consistency of $\hat{\beta}$, the l_1 estimator, may be salvaged if F has sufficient mass at the median.

Consider the harmonic design $x_i = 1/i$. In this case $\sum x_i^2 \to \pi^2/6$ (Knopp 1956, p. 173), so $n^{-1}\sum x_i^2 \to 0$. As in the previous section, with $\delta > 0$,

$$\Pr[\hat{\beta}_n - \beta < \delta] = \Pr[g_n(\delta) > 0], \tag{4.1}$$

where now

$$g_n(\delta) = -\sum_{i=1}^{n} \operatorname{sgn}(u_i - x_i\delta)x_i. \tag{4.2}$$

So

$$\Pr[\hat{\beta}_n - \beta < \delta] = \Pr[h_n(\delta) > -1], \tag{4.3}$$

where

$$h_n(\delta) = (g_n(\delta) - Eg_n(\delta))/Eg_n(\delta). \tag{4.4}$$

Now

$$\operatorname{var} h_n(\delta) = \frac{Vg_n(\delta)}{(Eg_n(\delta))^2} \le \frac{\pi^2/6}{[\sum(1/i)(1 - 2F(\delta/i))]^2}, \tag{4.5}$$

which converges to zero if and only if the denominator diverges. Suppose in a neighborhood of 0, F is symmetric with

$$F(u) = 1/2 + 1/\log_2(1/u), \qquad u > 0; \tag{4.6}$$

then for some $n_0 > 0$, the denominator becomes

$$(Eg_n(\delta))^2 = [\textstyle\sum 2(i \log_2(i/\delta))^{-1}]^2, \tag{4.7}$$

which diverges. Thus by the Chebyshev inequality, $h_n(\delta)$ converges in probability to zero, and hence the left side of (4.3) converges to one. An identical argument implies that $\Pr[-(\hat{\beta}_n - \beta) < \delta] \to 1$ as well.

Thus we have an example in which the usual design condition $\sum x_i^2 \to \infty$ is violated, but the weak consistency of $\hat{\beta}$ is salvaged by making the density of u sufficiently large near the median. The least-squares estimator is inconsistent under these conditions, of course, unless F is actually degenerate at 0.

[*Received September 1983. Revised March 1984.*]

REFERENCES

BASSETT, G., and KOENKER, R. (1978), "Asymptotic Theory of Least Absolute Error Regression," *Journal of the American Statistical Association,* 73, 618–622.

CHEN, G.J., LAI, T.L., and WEI, C.Z. (1981), "Convergence Systems and Strong Consistency of Least Squares Estimates in Regression Models," *Journal of Multivariate Analysis,* 11, 319–333.

CHUNG, K.L. (1974), *A Course in Probability Theory,* New York: Academic Press.

FELLER, W. (1968), *An Introduction to Probability Theory and Its Applications* (3rd ed.), New York: John Wiley.

GALLANT, A.R., BURGUETE, J.F., and SOUZA, G. (1983), "On the Unification of the Asymptotic Theory of Nonlinear Econometric Models" (with discussion), *Econometric Reviews,* 1, 151–211.

KNOPP, K. (1956), *Infinite Sequences and Series,* New York: Dover.

LAI, T.L., ROBBINS, H., and WEI, C.Z. (1979), "Strong Consistency of Least Squares Estimates in Multiple Regression II," *Journal of Multivariate Analysis,* 9, 343–361.

OBERHOFER, W. (1982), "The Consistency of Nonlinear Regression Minimizing the l_1 Norm," *Annals of Statistics,* 10, 316–319.

WU, CHIEN-FU (1981), "Asymptotic Theory of Nonlinear Least-Squares Estimation," *Annals of Statistics,* 9, 501–513.