ESTIMATING LINEAR RELATIONSHIPS FOR MODELS BASED ON RANDOM VARIABLES WITH INFINITE VARIANCE

BY: Marek Kanter
William L. Steiger

Technical Report # 31

Department of Computer Science
Rutgers, The State University of New Jersey
New Brunswick, New Jersey 08903

Department of Computer Science Technical Report # 31
December, 1974

Classification: Probability / Statistics

Title: Estimating Linear Relationships for Models Based on Random Variables with Infinite Variance

Authors: Marek Kanter[1]
Department of Mathematics
Sir George Williams University
Montreal, Qué.

William L. Steiger[2]
Department of Computer Science
Rutgers University
The State University of New Jersey
New Brunswick, N. J.

Key words: Stable Process, Autoregressive Process, Moving Average Process, Regression

Running Title: Estimating Linear Relationships

AMS (MOS) Classification Numbers: Primary 62J05, 60G10, 62G05

Secondary 62P20, 93E10

1. Research partially supported by Grant #A8753, National Research Council of Canada, and Action concertée, Ministère de l'Education du Québec.

2. Research partially supported by Action concertée, Ministère de l'Education du Québec.

ABSTRACT

We sketch the proof of some theorems that show how to estimate the parameters in linear regressions, finite moving averages, and in finite order, stationary autoregressions. Some of these estimates have not been studied yet, but the chief novelty is that existing theory is extended to include processes with infinite variance. A main result is that ordinary least squares estimates are consistent for both finite moving average processes and finite order autoregressions. The sampling properties of some of these estimates are indicated.

1. Introduction

This paper is a summary of results on estimation in linear regressions, autoregressions, and moving average processes, based on random variables with infinite variance. Some of the work appears in previous papers [7], [8], and [9], and some is new.

On a probability space (Ω, A, P), E denotes Lebesgue integration with respect to P. As usual, L_q denotes the random variables X on Ω with $E(|X|^q) < \infty$, $q > 0$, and $E(\cdot|X)$ is a version of the conditional expectation operator that maps L_1 into the subclass of functions measurable with respect to the sigma field generated by X. For $A \in A$, I_A denotes the function $\Omega \to R$ that is 1 on A and 0 otherwise. Finally, $L(X)$ denotes the distribution of the random variable X; $\xrightarrow{L}$, $\xrightarrow{P}$, $\xrightarrow{a.s.}$ denote, as usual, convergence in law, convergence in probability, and almost sure convergence, respectively.

We shall be dealing with stable random variables (see Feller [3]). A random variable X_0 is stable of index $\alpha \in (0,2]$ if and only if, for each $n > 0$, there is a real γ_n such that $L(X_0) = L((X_1+\ldots+X_n)/ n^{1/\alpha} - \gamma_n)$, where the X_i are independent and $L(X_i) = L(X_0)$, $i \geq 1$. If X is stable of index $\alpha < 2$ and $P(X=0) < 1$, X is asymptotically Pareto of index α; i.e., $t^{\alpha}P(|X|>t) \to k > 0$ as $t \to \infty$. This implies that for $\alpha < 2$ $E(|X|^{\beta})$ is infinite for $\beta \geq \alpha$ and finite for $\beta<\alpha$. In particular a stable X of index $\alpha < 2$ has infinite variance.

In the next section we consider the problem of estimating the slope, λ, of the linear regression $E(X|Z) = \lambda Z$ a.s., where X and Z may have infinite variance. Several estimates are presented, including the L_1 analogue of least squares, and their properties are discussed. In Section 3, we treat the problem of estimating coefficients in finite order autoregressive processes. The interesting fact that emerges here is that whereas least squares estimates are not consistent for infinite variance regressions, they are consistent for autoregressions. Finally in Section 4 the estimation problem for stationary, finite order, infinite variance moving average processes is discussed, along with the consistency of least squares estimates.

2. Regressions

Let X and Z be L_1 random variables that satisfy the linear regression

1) $$E(X|Z) = \lambda Z \quad \text{a.s.}$$

The problem is to estimate λ from a sample of size n; i.e., a stationary and ergodic sequence $\{(X_i, Z_i)\}$ with $L(X_i,Z_i) = L(X,Z)$, all i. A familiar estimate of λ is the least squares one

2) $$C_n = \sum_{i=1}^{n} X_i Z_i \Big/ \sum_{i=1}^{n} Z_i^2$$

which, if $E(Z^2) < \infty$, is strongly consistent; a simple application of the ergodic theorem shows that $C_n \xrightarrow{a.s.} \lambda$ as $n \to \infty$. This fails to hold in infinite variance regressions. Take X=U, Z=U+V, U and V independent, symmetric stable random variables of index $\alpha \in (1,2)$. The structure forces a linear regression because
$E(X|Z) = E(U|U+V) = E(V|U+V) = E(U+V|U+V)/2 = (U+V)/2 = Z/2.$
Nevertheless it is easy to show that $C_n \xrightarrow{L} S/(S+T)$, where S and T are independent, positive stable random variables of index $\alpha/2$ [7].

In seeking consistent estimates of λ, it is tempting to consider the L_1 analogue of C_n. The minimum absolute deviation (MAD) estimate, M_n, based on a sample of size n, is defined by

$$2) \qquad M_n = \min\,(u\colon f(u) \le f(v) \text{ for all } v),$$

where $f(u) = \sum_{i=1}^{n} |X_i - uZ_i|$. The MAD estimate converges to a constant almost surely, as follows.

Theorem 1: Given $X, Z, Z \in L_1$ with $P\{Z=0\} = 0$, form the distribution function $F(u) = E(|Z| I_{\{X/Z \le u\}}) / E(|Z|)$. If F is increasing at the point $y = \min(u\colon F(u) \ge 1/2)$, then

$$M_n \xrightarrow{a.s.} y$$

Proof: Define the random distribution function F_n by $F_n(u) = \sum_{i=1}^{n} (|Z_i| I_{\{X_i/Z_i \le u\}}) / \sum_{i=1}^{n} |Z_i|$. By [10], $M_n = \min(u\colon F_n(u) \ge 1/2)$. By the ergodic theorem, $F_n(u) \xrightarrow{a.s.} F(u)$, uniformly in u. The conclusion follows by the monotonicity of F_n and the fact that F is increasing at y.$\square$

Certain linear regression structures have joint distributions that satisfy the above conditions. For them, M_n is useful because it is strongly consistent.

Corollary 1: Let $X = \lambda Z+W$, $Z \epsilon L_1$ and $P\{Z=0\} = 0$. If 0 is the unique median of W given Z, $M_n \xrightarrow{a.s.} \lambda$. If also $W \epsilon L_1$ and $E(W \mid Z) = 0$ a.s., $E(X|Z)=\lambda Z$, so M_n is strongly consistent.

Proof: $F(\lambda) = E(|Z| I_{\{(\lambda Z+W)/Z\leq\lambda\}})/E(|Z|) = E(|Z| I_{\{W/Z\leq 0\}})/E(|Z|) = 1/2$

because 0 is a median of the conditional distribution of W/Z. Since it is unique, F increases at λ.□

Remark 1: There are interesting consequences of the corollary. Taking W and Z independent and $\lambda=0$ shows MAD to be a strongly consistent estimate of 0 in the trivial regression $E(X|Z)=0$ between independent random variables X and Z. Next, if X and Z are jointly normal with zero mean, they may be represented as in Corollary 1 for some λ and W, so MAD is a strongly consistent estimate of the slope of the linear regression relating such normal random variables.

Finally, MAD is a strongly consistent estimate in a class of infinite variance regressions that contains the example given earlier, in which $C_n \xrightarrow{L}$ to a non-degenerate limit. This appears in the following statement whose proof is omitted.

Corollary 2: Let $U_1,\ldots,U_n$ be i.i.d. random variables in L_1 and put $X=cU_1$, $Z=U_1+\ldots+U_n$. Then

$$M_n \xrightarrow{a.s.} c/n ,$$

the slope of the linear regression between X and Z.

We now consider two other estimates of λ in (1). Let $X, Z \in L_1$ satisfy (1). For numbers $0 < d_1 < d_2 \leq \infty$, and $0 < d_3 < \infty$ put $B = \{|Z| \in (d_1, d_2)\}$ and $B' = \{|Z| < d_3\}$. For $P(B) > 0$ the screened ratio estimate, R, of λ is defined by

$$R = XZ^{-1}I_B \,/\, P(B). \tag{3}$$

It is unbiased as is shown in the next theorem, along with another fact.

Theorem 2: $E(R) = \lambda$ and $E(XZI_{B'}) = \lambda E(Z^2 I_{B'})$

Proof: Note that R and $XZI_{B'}$ are in L_1 and that $E(Xf(Z)) = E(E(X|Z)f(Z)) = \lambda E(Zf(Z))$ for any measurable function f. □

Remark 2: Take a sample of size n and put $R_n \equiv \sum_{i=1}^{n} X_i Z_i^{-1} I_{B_i} \,/\, \sum_{i=1}^{n} I_{B_i}$, where $B_i = \{|Z_i| \in (d_1, d_2)\}$. Obviously $R_n \xrightarrow{a.s.} \lambda$. Similarly, the screened least squares estimate $S_n \equiv \sum_{i=1}^{n} X_i Z_i I_{B_i'} \,/\, \sum_{i=1}^{n} Z_i^2 I_{B_i'}$ is strongly consistent, where $B_i' = \{|Z_i| < d_3\}$. For an application, take U_i i.i.d. symmetric stable random variables of index $\alpha \in (1,2]$ and define $X = a_1U_1 + \ldots + a_kU_k$, $Z = b_1U_1 + \ldots + b_kU_k$. Then (1) holds with $\lambda = (a_1b_1^{\alpha-1} + \ldots + a_kb_k^{\alpha-1})/(|b_1|^\alpha + \ldots + |b_k|^\alpha)$, where x^y means $\mathrm{sign}(x)|x|^y$, by [6], and by the above remarks, $R_n \xrightarrow{a.s.} \lambda$

and $S_n \xrightarrow{a.s.} \lambda$. In fact, $R_n \xrightarrow{a.s.} \lambda$ for $\alpha \leq 1$ as long as $d_2 < \infty$ and $b_i = 0 \Rightarrow a_i = 0$ since one can still write $E(XI_B \mid ZI_B) = \lambda ZI_B$ a.s. Monte-Carlo studies indicate that $R_n \to \lambda$ rather slowly while $S_n \to \lambda$ more rapidly.

3. Autoregressions

Let $\ldots U_{-1}, U_0, U_1, \ldots$ be an i.i.d. sequence of random variables and $a_1, \ldots, a_k$ numbers such that $a_1x + \ldots + a_kx^k \neq 1$ for any complex x with $|x| \leq 1$. If $U_i \in L_2$, it is familiar that there is a unique stationary solution, $\{ X_n \}$, to the k^{th} order autoregressive scheme

4) $$X_n = a_1X_{n-1} + \cdots + a_kX_{n-k} + U_n$$

with X_i and U_{i+j} independent for $j > 0$. Then, from the Yule-Walker equations

5) $$\rho_i = \sum_{j=1}^{k} a_j \rho_{i-j} \qquad i=1,\ldots,k,$$

where $\rho_i = E(X_nX_{n-i})/E(X_n^2)$, the a_i's may be estimated by solving the corresponding sample equations

$$\hat{\rho}_i(n) = \sum_{j=1}^{k} \hat{a}_j(n)\hat{\rho}_{i-j}(n), \quad i=1,\ldots,k, \tag{6}$$

for the $\hat{a}_i(n)$; here, $\hat{\rho}_i(n) = \sum_{j=i}^{n} X_j X_{j+i} / \sum_{j=i}^{n} X_j^2$. Using the fact that

$$\hat{\rho}_i(n) \xrightarrow{a.s.} \rho_i \quad , \quad i=1,\ldots,n \tag{7}$$

by the ergodic theorem,

$$\hat{a}_i(n) \xrightarrow{a.s.} a_i \quad , \quad i=1\ldots,n, \tag{8}$$

so the least squares estimates (6) are strongly consistent.

This remains more or less true in the infinite variance context. Let the U_i be in the domain of attraction of a stable law of index α. There is a.s. a unique stationary solution to (4) because it may be written uniquely as $X_n = \sum_{j=0}^{\infty} c_j U_{n-j}$ by inverting (4); $a_1x+\ldots+a_kx^k \neq 1$, $|x| \leq 1$, guarantees that $c_j \leq k/t^j$ for some constants $k > 0$, $t > 1$, so $\sum_{j=0}^{\infty} |c_j|^\alpha < \infty$, which ensures the a.s. convergence of $\sum_{j=0}^{\infty} c_j U_{n-j}$ (see [7]). Next, although autocorrelations in (5) don't exist if $\alpha < 2$, (6) may still be used to estimate the a_i's. That this produces consistent estimates is our main result.

<u>Theorem 3:</u> Let $\{\ldots,U_{-1},U_0,U_1\ldots\}$ be an i.i.d. sequence of random variables in the domain of attraction of a stable law of index $\alpha \in (0, 2)$ and let $\{X_i\}$ be a stationary solution of (4), with X_n and U_{n+j} independent, $j > 0$. Then for $1 \leq i \leq k$,

9) $$n^{\delta} \mid \hat{\rho}_i(n) - \rho_i \mid \overset{P}{\rightarrow} 0, \quad n \to \infty,$$

for all $\delta < 2/\alpha - 1/\min(1,\alpha)$, where $\hat{\rho}_i(n) = \sum_{j=1}^{n} X_j X_{i+j} / \sum_{j=1}^{n} X_j^2$ and the ρ_i are the unique solution of (5).

<u>Proof:</u> Use (4) in the definition of $\hat{\rho}(n)$ to see that $\hat{\rho}_i(n) = \sum_{j=1}^{n} X_j(a_1X_{i+j-1}+\ldots+a_kX_{i+j-k}+U_{i+j})/\sum_{j=1}^{n} X_j^2$. Thus

$$\hat{\rho}_i(n) = a_1\hat{\rho}_{i-1}(n)+\ldots+a_i+a_{i+1}(\hat{\rho}_1(n)-X_nX_{n+i}/\sum_{j=1}^{n} X_j^2) +$$

10) $$\ldots+a_k(\hat{\rho}_{k-i}(n) - (X_{n+1}X_{n+i-k+1}+\ldots+X_{n+k-i}X_n)/\sum_{j=1}^{n} X_j^2) + \sum_{j=1}^{n} X_jU_{i+j}/\sum_{j=1}^{n} X_j^2,$$

where we have taken $X_j = 0$, $j \leq 0$.

To prove the consistency of the $\hat{\rho}_i(n)$ we need only show that $n^{\delta} \sum_{j=i}^{n} X_jU_{i+j}/\sum_{j=i}^{n} X_j^2$ and $n^{\delta}(X_{n+1}X_{n-m+1}+\ldots+X_{n+m}X_n)/\sum_{j=1}^{n} X_j^2$, $1 \leq m \leq k-i$, each $\overset{P}{\rightarrow} 0$ for all $\delta < 2/\alpha - 1/\min(1,\alpha)$.

To prove the first assertion, choose β, γ, $\alpha/2 < \beta < \gamma < \min(1,\alpha)$, so close to $\min(1,\alpha)$, that $\delta < 2/\alpha - 1/\beta < 2/\alpha - 1/\delta < 2/\alpha - 1/\min(1,\alpha)$, and multiply numerator and denominator of $n^{\delta} \sum_{j=i}^{n} X_jU_{i+j}/\sum_{j=i}^{n} X_j^2$ by $n^{-1/\beta}$ to obtain

(11) $$n^{-1/\beta} \sum_{j=1}^{n} X_j U_{i+j} / (n^{-\delta} n^{-1/\beta} \sum_{j=1}^{n} X_j^2);$$

we will show that the numerator $\xrightarrow{a.s.} 0$ and that the denominator $\xrightarrow{P} \infty$. First, note that $|n^{-1/\beta} \sum_{j=1}^{n} X_j U_{i+j}|^{\gamma} \leq n^{-(\gamma-\beta)/\beta}(n^{-1} \sum_{j=1}^{n} |X_j|^{\gamma}|U_{j+i}|^{\gamma})$ because $\gamma < 1$. The right hand side $\xrightarrow{a.s.} 0$ because $\gamma > \beta$ and, by the ergodic theorem, the expression in parentheses $\xrightarrow{a.s.}$ to $E(|X_1|^{\gamma}|U_{1+i}|^{\gamma}) < \infty$, so the numerator of (11) $\to 0$, as asserted.

Next note that the denominator of (11) satisfies

$$n^{-\delta} n^{-1/\beta} \sum_{j=1}^{n} X_j^2 = n^{2/\alpha - \delta - 1/\beta}(n^{-2/\alpha} \sum_{j=1}^{n} (a_1 X_{j-1} + \dots + a_k X_{j-k} + U_j)^2)$$

$$\geq n^{2/\alpha - \delta - 1/\beta}(n^{-2/\alpha} \sum_{j=1}^{n} U_j^2)$$

$$+ 2n^{2/\alpha - \delta - 1/\beta}(n^{-2/\alpha} \sum_{j=1}^{n} U_j(a_1 X_{j-1} + \dots + a_k X_{j-k})).$$

Because $\alpha/2 < \beta$, the second term $\xrightarrow{a.s.} 0$ by the foregoing. The denominator of (11) $\xrightarrow{P} \infty$ then, because $2/\alpha - \delta - 1/\beta > 0$ and because $n^{-2/\alpha} \sum_{j=1}^{n} U_j^2 \xrightarrow{L}$ to a positive stable random variable of index $\alpha/2$ (see [3]).

Finally, since we now know that $n^{-\delta} \sum_{j=1}^{n} X_j^2 \xrightarrow{P} \infty$, and because

$(X_{n+1} X_{n-m+1} + \dots + X_{n+m} X_n)$, $1 \leq m \leq k-i$ has a fixed distribution, independent of n, $n^{\delta}(X_{n+1} X_{n-m+1} + \dots + X_{n+m} X_n) / \sum_{j=1}^{n} X_j^2 \xrightarrow{P} 0$. The proof is complete. □

Remark 3: In view of the behavior of least squares estimates in infinite variance regressions, Theorem 3 is somewhat of a surprise So is the rate result which has the convergence $o(n^{1/\gamma})$: thus, for example, least squares in autoregressions with normal errors converge slowest; with Cauchy errors, the estimates converge faster, like $1/n$. The theorem is false if $\xrightarrow{p}$ is replaced by $\xrightarrow{a.s.}$. However X_n need not be stationary. Finally, the least squares estimates $\hat{\rho}_i(n)$ are a special case of a family of least beta estimates defined by $\hat{b}_i(n) = \sum_{j=1}^{n} X_j X_{j+i}{}^{1/(\beta-1)} / \sum_{j-1}^{n} |X_j|^{\beta/(\beta-1)}$, $1 < \alpha < \beta \leq 2$ and Theorem 3 is a special case of a corresponding statement for $\hat{b}_i(n)$. However the convergence rate is greatest when $\beta = 2$ in which case $\hat{b}_i(n) = \hat{\rho}_i(n)$, so the least betas are of no real interest.

4. Moving Averages

Consider the finite moving average process

$$X_n = \sum_{j=-\infty}^{\infty} a_j U_{n+j} \tag{12}$$

where $a_j = 0$ for $|j| > N > 0$, and the U_i are an i.i.d. sequence of random variables. When $E(U_i^2) < \infty$, it is familiar that the least squares estimate $\hat{\rho}_i(n) \xrightarrow{a.s.}$ to constants. A corresponding result holds in the infinite variance case.

Theorem 4. Let U_i be as in Theorem 3 and consider the stationary process in (12). The least squares estimates obey

(13) $$\hat{\rho}_i(n) \equiv \sum_{j=1}^{n} X_j X_{i+j} / \sum_{j=1}^{n} X_j^2 \overset{P}{\to} \alpha_i/\alpha_o,$$

where $\alpha_i = \sum_{j=-\infty}^{\infty} a_j a_{i+j}$.

Proof: Take $i > 0$. For n large we can write

(14) $$\hat{\rho}_i(n) = (\sum_{j=N}^{n-2N} \alpha_i U_j^2 + D_i(n))/(\sum_{j=N}^{n-2N} \alpha_o U_j^2 + D_o(n))$$

where $D_i(n) = \Sigma c_{jk}(i) U_j U_k + \Sigma b_j(i) U_j^2$; the first sum is taken over the set $\{j \neq k, |j - k| \leq 2N+2i+1, -N \leq j, k \leq n+N+i\}$, while the second is over the set $\{-N \leq j \leq N\} \cup \{n-2N \leq j \leq n+N+i\}$. The coefficients $c_{jk}(i)$ and $b_j(i)$ may be computed from the a_i's and are bounded by a constant times α_o.

Multiply the numerator and denominator of (14) by $n^{-1/\beta}$, $\alpha/2 < \beta < \min(1,\alpha)$. The claim is that $n^{-1/\beta} D_i(n) \overset{P}{\to} 0$ as $n \to \infty$, for each i, from which the result will follow via (14). First note that $n^{-1/\beta} \Sigma b_j(i) U_j^2 \overset{P}{\to} 0$ as $n \to \infty$ because the sum is always over a finite number ($< 3N$) of terms. Next, $n^{-1/\beta} \Sigma c_{jk}(i) U_j U_k \xrightarrow{a.s.} 0$ because, for $\beta < \gamma < \min(1,\alpha)$, $|n^{-1/\beta} \Sigma c_{jk}(i) U_j U_k|^\gamma < n^{-(\gamma-\beta)/\beta} (n^{-1} \Sigma |c_{jk}(i) U_j U_k|^\gamma$; since $j \neq k$ in the sum, the term in parentheses is a.s. $< KE(|U_1|^\gamma) E(|U_2|^\gamma)$ for all n, by the ergodic theorem and the boundedness of the $c_{jk}(i)$. □

Remark 4: We conjecture that if $X_n = \sum_{j=-\infty}^{\infty} c_j U_{n+j}$ is an infinite order stationary moving average process with $\sum_{j=-\infty}^{\infty} |c_j|^{\alpha} < \infty$, then if the U_i are i.i.d. and in the domain of attraction of a stable law of index α ϵ (or 2) the least squares estimates will continue to be consistent; i.e., that

$$\rho_i(n) \overset{P}{\to} \sum_{j=-\infty}^{\infty} c_j c_{i+j} / \sum_{j=-\infty}^{\infty} c_j^2 \equiv \zeta_i$$

This is already true in case the process arises as an inverted, stationary autoregressive scheme, (4), by Theorem 3. Incidentally it has been shown in [5] that the distribution of the process X_n determines the coefficients c_j for the symmetric stable case when $0 < \alpha < 2$.

REFERENCES

[1] Anderson, T. W. (1971) *The Statistical Analysis of Time Series* (New York, John Wiley).

[2] Blattberg, R. and Sargent, T. (1971) "Regression with Non-Gaussian Stable Disturbances : Some Sampling Results", *Econometrika* 39, 501-510.

[3] Feller, W. (1971) *Introduction to Probability Theory and its Application, Vol. II.* (second edition, New York, John Wiley).

[4] Granger, C. and Orr, D. (1972) "Infinite Variance and Research Strategy in Time Series", *J. Amer. Statist. Assoc. 67*, 275-285.

[5] Kanter, Marek (1973) "On the L_p Norm of Sums of Translates of a Function", *Trans. Amer. Math. Soc. 179*, 35-47.

[6] Kanter, Marek (1972) "Linear Sample Spaces and Stable Processes", *J. Funct. Anal. 9*, 441-459.

[7] Kanter, Marek and Steiger, W. L. (1974) "Regression and Autoregression with Infinite Variance", to appear, *Adv. Applied Prob.*

[8] Kanter, Marek and Steiger, W. L. (1974) "Minimum Absolute Deviation Estimates in Linear Regressions", Preprint #353, Centre de Recherches Mathématiques, U. de Montreal.

[9] Kanter, Marek and Steiger, W. L. (1974) "Sampling Properties of Some Estimates of Parameters for Regressive and Autoregressive Models Based on Stable Random Variables", Preprint #405, Centre de Recherches Mathématiques, U. de Montreal.

[10] Singleton, J. (1940) "A Method for Minimizing the Sum of Absolute Deviations", *Ann. Math. Statist. 11*, 301-310.