COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station

New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 798

ON THE PERFORMANCE OF LEAST SQUARES IN LINEAR REGRESSION WITH UNDEFINED ERROR MEANS

Donald W. K. Andrews

January 1986

Revised July 1986

ON THE PERFORMANCE OF LEAST SQUARES IN LINEAR REGRESSION WITH UNDEFINED ERROR MEANS

By Donald W. K. Andrews[1]

Cowles Foundation
Yale University
Box 2125 Yale Station
New Haven, CT 06520

HEADNOTE

This paper considers the linear regression model with multiple stochastic regressors, intercept, and errors that have undefined means. This model is of interest from a robustness perspective as a polar case. Generally, least squares estimators are inconsistent in this context. It is shown, however, that this inconsistency is restricted to the estimation of the intercept, if the regressors are highly variable. Rates of convergence of the least squares slope estimators are determined, and are shown to exceed the standard rate $n^{-1/2}$, in certain contexts. The results place no restrictions on the temporal dependence of the errors, and require an unusually weak exogeneity condition between the regressors and errors. Implications of the results for robustness theory are discussed.

Keywords: Least squares estimator; linear regression; stable distribution; fat-tails; consistency; robust.

January, 1986

Revised: July 1986

1. INTRODUCTION

This paper considers the linear regression model with intercept, multiple stochastic regressors, and errors with undefined means. This model is of interest from a robustness perspective, because it represents a polar case. The least squares (LS) estimators generally are inconsistent in this model. It is shown, however, that this inconsistency is restricted to the estimation of the intercept, if the regressors are highly variable. For example, if the regressors and errors exhibit the same degree of variability, as measured by their maximal moment exponents, then the LS slope parameter estimators are consistent. Further, in some cases, their rates of convergence to the true parameters exceed $n^{-1/2}$.

The results given below are sufficiently general to incorporate time series, cross-section, panel data, seemingly unrelated, and multivariate regression models. Somewhat surprisingly, the results are obtained without any assumption on the temporal dependence of the errors. In this regard, some interesting new results also are obtained for LS estimation of linear regression models with well-defined error means. In addition, some of the preliminary results apply to dynamic linear regression models.

The paper is organized as follows: The remainder of this section motivates consideration of the model referred to above, discusses the relevant literature, and outlines the results of the paper. Section 2 presents the basic consistency and rate of convergence results for the case of M-dependent regressors. The rate of convergence results are shown to be sharp, in some contexts. Section 3 extends the results of Section 2 to models with infinite-order moving average regressors with heterogeneous innovations and heterogeneous moving average coefficients. Section 4 considers cases where some regressors are highly variable while others are not. Section 5 provides some Monte Carlo results that illustrate that the qualitative asymptotic results carry over to finite samples, and that compare the LS estimator to a classical robust regression estimator. The final section, Section 7, draws conclusions based on the results which precede it. An Appendix provides proofs of the results given in the paper.

The statistical literature on robustness has developed from the analysis of the location model, i.e., the problem of estimating the "center" of a distribution, see Huber (1964). In the location model, there are many ways of exhibiting the lack of robustness of the least squares (LS) estimator to fat-tailed errors. A dramatic method is to consider a polar case, viz., the case of errors with undefined means. For this model, the converse to the strong law of large numbers implies that the LS estimators are not even consistent. In less fat-tailed error cases, it has been shown that the LS estimators are consistent, but are relatively inefficient asymptotically.

The location model is of limited interest in econometrics. Many techniques that were introduced and justified theoretically for the location model, however, have been extended to regression models, both linear and nonlinear, e.g., see Huber (1973), Maronna and Yohai (1981), Krasker and Welsch (1982), Koenker (1982), Bierens (1981), and Andrews (1983). Thus, the robustness literature is of relevance to econometrics. Robust regression techniques have been designed to work well when the errors have fat tails, and the evidence suggests that such phenomena do occur with some economic data (e.g., see Koenker (1982), Andrews (1983), and Judge <u>et al</u>. (1980, Ch. 7)).

For numerous reasons, a prime one being analytic tractability, theoretical justifications for various robust regression procedures are not as well developed as their analogues for location models, e.g., see Huber (1973, 1983) and Krasker and Welsch (1982). Thus, appeals to the analogy with the location model are fairly common (at least at an informal level).

This paper addresses one aspect of such analogies, viz., the question of consistency of LS when the errors have undefined means. The purpose of analyzing this question is three-fold. First, following a long tradition in mathematical analysis, we believe that in order to understand some phenomena over a wide range of circumstances, it is extremely useful to understand its behavior in polar cases. For considerations of robustness, the polar case of concern is the case in which the errors have undefined means. Second, the polar case may be of interest in and of itself. That is, in some regression applications the assumption of undefined error means may be appropriate. Third, we are interested in the extent to which analogies between the location model and the regression model are appropriate. And, if they are not appropriate in some respect, we would like to know which features of the regression model are important in explaining the differences. The analysis below seems to provide useful results for each of these three purposes.

One might expect that the properties of LS in linear regression with undefined error means would have been worked out in the robustness literature some time ago. In fact, this is not the case. Only recently have some properties been investigated, see Kanter and Steiger (1974), Chen, Lai and Wei (1981), and Cline (1983). And it is interesting to note that in none of these papers has the issue of robustness been a prime motivating factor.[2]

Chen, Lai, and Wei (1981) consider the linear regression model with fixed regressors and independent identically distributed (iid) errors. As a by-product of some general consistency results, they find that LS slope coefficient estimators are consistent even if the errors have undefined means, provided the regressors exhibit sufficient variability (see their Corollary 4).

In econometrics, the regressors in linear regression models often are random. For present purposes, it is more appropriate to analyze the regression model treating the regressors as such, than to condition on the regressors and treat them as fixed. The reason is that the consistency results depend crucially on the behavior of the regressors, and the former method forces one to relate their behavior to underlying assumptions on their joint distribution.

Cline's (1983, pp. 115-121) results for undefined error means apply to a linear regression model with a single <u>stochastic</u> regressor, no intercept, and iid, mutually independent, regressors and errors. His results include those of Kanter and Steiger (1974, Lemma 4.3). Cline shows that if the regressor and error are in the domains of attraction of stable random variables (rv's), and the regressor is sufficiently highly variable, then the LS slope estimator is consistent. In addition, he derives the asymptotic distribution of the LS estimator (for the case of univariate regression through the origin, when the regressor has smaller maximal moment exponent than the error) and finds it to be given by that of a ratio of dependent stable rv's.

In this paper, we present consistency results (including rates of convergence), for the undefined error mean case, which are considerably more general than those referred to above. We consider a linear regression model with multiple stochastic regressors and an intercept. The errors and regressors may be temporally dependent and non-identically distributed, and need not be mutually independent. Thus, our results apply in both time series and cross-sectional contexts. In fact, in the time series context, no assumption at all is placed on the extent of temporal dependence of the errors.

The consideration of an intercept reflects the almost universal use of intercepts in econometric applications. A consequence of using an intercept is that we get different rates of convergence than those found by Cline (1983). Also, by considering multiple regressors with possibly infinite second moments, we face an identification question that does not arise in the single regressor model considered by Cline. On the other hand, we do not derive limit distributions here, as is done by Cline. Our methods of proof differ considerably from those of Kanter and Steiger (1974), Chen, Lai, and Wei (1981), and Cline (1983).

The consistency results depend on the relative magnitudes of the maximal moment exponents of the regressors and the errors. For example, if the regressors and errors have equal maximal moment exponents, as might be expected if the errors contain left out variables, then the LS slope estimators are consistent. The conditions for consistency require at least the second moments of the regressors to be infinite, when the errors have undefined means.

Rates of convergence to the true parameter are given and are shown to be sharp in some contexts. These rates are found to be very fast in some cases. For example, if the regressors and errors are Cauchy, then the rate is n^{-1} . If the regressors are stable with exponent α and the errors are Cauchy, then the rate is $n^{-1/\alpha}$, for any $\alpha \in (0,1)$. Thus, the rate is arbitrarily fast, for α arbitrarily small.

This paper also considers the case where some regressor variables have infinite second moments, while others have finite second moments. Consistency of the LS slope estimators for the parameters of the more highly variable regressors is established, under certain conditions.

An interesting by-product of the results for the case of undefined error means is a new result for the case of well-defined error means. Sections 2-4 show that under no assumptions regarding the temporal dependence of the errors, the LS slope estimators are consistent, if the regressors have infinite second moments (and well-defined means). Thus, the errors may contain a time invariant random component, or, the errors may evolve in such a fashion that their future conditional distributions depend strongly on their past realizations.

Similar consistency results, that do not require any assumption on the temporal dependence of the errors, are given in Andrews (1985c), for the case of LS with integrated regressors. In these models, the regressors have finite, but exploding, variances, and the errors may have exploding means and/or variances. In contrast, the regressors and errors considered in this paper are not exploding in any sense. Stationary models are special cases of the models considered here. A distinct difference between the results of this paper, and those for models with integrated regressors, is that in the present case the "signal" from the regressors does not dominate the "noise" from the errors as the sample size increases. In fact, the consistency results given here include stationary models in which the regressors are less variable than the errors (as measured by their maximal moment exponent) for every observation.

2. CONSISTENCY AND RATE OF CONVERGENCE RESULTS

Consider the linear regression model

(1) $$Y_i = c_0 + X_i'\beta_0 + U_i \ , \quad i = 1, \ldots, n \ ,$$

where Y_i , c_0 , and U_i are real-valued, and X_i and β_0 are R^K-valued. c_0 and β_0 are unknown parameters. Our attention will focus on the estimation of β_0 .

First, we introduce some notation: The symbol $\overset{ST}{\leq}$ denotes "stochastically less than or equal to." Thus, the rv's W_1 and W_2 satisfy $W_1 \overset{ST}{\leq} W_2$ if and only if $F_{W_1}(w) \geq F_{W_2}(w)$, $\forall w \in R$, where $F_{W_1}(w)$ and $F_{W_2}(w)$ are the distribution functions (df's) of W_1 and W_2 , respectively. The *maximal moment exponent* of a rv W is defined to be $\sup\{\zeta \geq 0 : E|W|^\zeta < \infty\}$. If $E|W|^\zeta = \infty$ $(< \infty)$ for all $\zeta > 0$, the maximal moment exponent of W is defined to be 0 (∞) . Thus, the maximal moment exponent of every rv is well-defined, and lies in $[0,\infty]$. The symbol " $\wedge$ " denotes the minimum operator. That is, $a \wedge b = \min(a,b)$. The symbol $\lambda_{\min}(A)$ denotes the smallest eigenvalue of the matrix A .

Of the errors, U_i , and the regressors, X_i , we make the following assumptions (each of which is discussed below):

ASSUMPTION C1: $|U_i| \overset{ST}{\leq} U$, $\forall i$, for some rv U with maximal moment exponent $p > 0$.

ASSUMPTION C2: $|X_{ik}| \overset{ST}{\leq} X$, $\forall k = 1, \ldots, K$, $\forall i$, for some rv X with maximal moment exponent $r > 0$.

ASSUMPTION C3: $|X_{ik}U_i| \overset{ST}{\leq} G$, $\forall k = 1, \ldots, K$, $\forall i$, for some rv G with maximal moment exponent greater than or equal to $1 \wedge p \wedge r$.

ASSUMPTION C4: The sequence of rv's $\{X_i\}$ is M-dependent.

(By definition, $\{X_i\}$ is *M-dependent* if X_i and X_ℓ are independent whenever $|i - \ell| > M$. For example, M^{th}-order moving average sequences are M-dependent.)

Assumptions C1 and C2 allow heterogeneity of the errors and regressors, but prohibit explosive behavior. Since p and r may be any positive numbers, we consider cases where the errors and regressors have undefined means or variances, in addition to cases where they are well-defined. As discussed in the Introduction, the cases of primary interest, for present purposes are those with $p \leq 1$, i.e., those with undefined error means.

Note that it is always possible to construct rv's U and X that stochastically dominate $|U_i|$, $\forall i$, and $|X_{ik}|$, $\forall k$, $\forall i$ respectively. For example, take U to have df given by $F_U(u) = \inf_{i\geq 1} F_{|U_i|}(u)$, where $F_{|U_i|}(u)$ is the df of $|U_i|$. Hence, the strength of C1 and C2 lies in the assumption that p and r are not zero.

Assumption C3 controls the dependence between the regressors and the errors. For identification of β_0 , some condition such as C3 clearly is needed. It is satisfied, for example, if the regressors and errors are independent. C3 is a much weaker assumption than independence, however, since it only requires control of the dependence of the *tails* of the rv's U_i and X_i . More specifically, for any constant b , assumption C3 applied to the rv's U_i and X_{ik} is equivalent to the same condition applied to the rv's $U_i 1(|U_i| > b)$ and $X_{ik} 1(|X_{ik}| > b)$. Thus, only the tail behavior of U_i

and X_{ik} determines whether C3 is satisfied.

To illustrate the generality of C3, suppose $\{\tilde{U}_i\}$ and $\{\tilde{X}_i\}$ satisfy C1 and C2, respectively, and are mutually independent, and $\{Z_i\}$ is an identically distributed sequence of normal rv's, independent of $\{\tilde{U}_i\}$ and $\{\tilde{X}_i\}$. Let $\{U_i\} = \{\tilde{U}_i + Z_i\}$ and $\{X_i\} = \{X_i + Z_i\underset{\sim}{d}\}$, for some constant K-vector $\underset{\sim}{d}$. Then, $\{U_i\}$ and $\{X_i\}$ satisfy C1, C2, and C3, even though the regressors X_i and the errors U_i have a component in common.

To compare C3 with more standard assumptions in the literature, consider a model where $E|U_i| < \infty$ and $E|X_{ik}|^2 < \infty$, $\forall k = 1, \ldots, K$, $\forall i$. Common assumptions used to prove consistency are: $E(U_i|X_i) = 0$ a.s. and $E|U_iX_{ik}| < \infty$. In fact, the latter condition often is a consequence of the stronger assumptions $E|U_i|^2 < \infty$ and $E|X_{ik}|^2 < \infty$. The assumption C3 holds in this model if just the second condition, $E|U_iX_{ik}| < \infty$, is imposed. The assumption of conditional mean zero is not required.

Assumption C4 (i.e., M-dependence of the regressors) allows applications to time series models. Since this assumption can be somewhat restrictive, however, it is relaxed in Section 3 below. Note that no assumptions are placed on the temporal dependence of the errors.

Next we introduce an identification condition that is analogous to the usual full rank condition on the regressor matrix. To do so, it is necessary to state several definitions and introduce some notation. A non-degenerate rv W has a *stable* distribution if for all constants $a, b \in R$, there exist constants $g, h \in R$ such that

$$(2) \qquad L(aW_1 + bW_2) = L(gW + h) ,$$

where W_1 and W_2 are independent rv's with the same distribution as W, and $L(\cdot)$ denotes the distribution (or law) of a rv.[3]

A distribution F (or a rv with distribution F) is said to be in the *domain of attraction* of a stable distribution S if there exist constants $\{c_n\}$ and $\{b_n\}$ such that

$$(3) \qquad c_n^{-1}\Sigma_1^n W_i - b_n \xrightarrow{d} W_0 \quad \text{as} \quad n \to \infty ,$$

where $\{W_i\}$ are iid with $L(W_i) = F$, and $L(W_0) = S$.[4]

Let $X \in DA(\alpha)$ denote that the rv X is in the domain of attraction of some stable distribution with characteristic exponent α. (There is a family of such stable distributions, both symmetric and asymmetric, when $\alpha < 2$.) If $X \in DA(\alpha)$, for some $\alpha < 2$, then the maximal moment exponent of X is α.

Now, consider the common identification condition for iid linear models with stochastic regressors:

$$(4) \qquad EX_iX_i' \quad \text{exists and is positive definite.}$$

This condition is not appropriate for our purposes, because it requires the regressors to have two moments finite. It is equivalent, however, to

$$(5) \qquad E(\varsigma' X_i)^2 \in (0,\infty) , \quad \forall \varsigma \in R^K \quad \text{with} \quad \|\varsigma\| = 1 ,$$

and the latter holds, only if

$$(6) \qquad |\varsigma' X_i| \in DA(2) , \quad \forall \varsigma \in R^K \quad \text{with} \quad \|\varsigma\| = 1 ,$$

e.g., see Feller (1971, Corollary XVII.5.1, p. 578). (Note that the converse does not hold, i.e., (6) does not imply (5). Thus, (6) actually is a weaker condition than (5).)

For the case of regressors whose maximal moment exponent is less than two, the natural analogue of (6) is the condition:

$$(7) \qquad |\varsigma' X_i| \in DA(\alpha) \ , \quad \forall \varsigma \in R^K \ \text{ with } \ \|\varsigma\| = 1 \ ,$$

for some $\alpha \in (0,2)$. This is the condition we adopt, although we generalize it to allow the elements of the regressor vector to be in different domains of attraction and to allow heterogeneity over $i = 1, 2, \ldots$:

ASSUMPTION C5: $\{X_i\}$ satisfies: $\forall \varsigma \in R^K$ with $\|\varsigma\| = 1$,

$$|\varsigma' X_i| \overset{ST}{\geq} X_\varsigma \ , \quad \forall i \ ,$$

for some real-valued rv $X_\varsigma \in DA(\alpha)$, where $0 < \alpha \leq 2$.

For purposes of illustration, we give two examples of regressors that satisfy C5. First, suppose the regressor vectors are identically distributed and the regressor elements X_{ik} are mutually independent, for $k = 1, \ldots, K$, and $X_{ik} \in DA(\alpha_k)$, for some $\alpha_k \in (0,2]$, $\forall k = 1, \ldots, K$. Then, C5 holds with $\alpha = \max\{\alpha_k : k = 1, \ldots, K\}$. Analogous non-identically distributed examples are easy to construct.

Alternatively, suppose the regressors are identically distributed, and the distribution of X_i is such that each element $X_{ik} \in DA(\alpha_k)$, for some $\alpha_k \in (0,2]$, $\forall k = 1, \ldots, K$, and each sub-vector of X_i , comprised of elements that are in the domains of attraction of distributions with the same characteristic exponent α_k , is in the domain of attraction of a full multivariate stable distribution with characteristic exponent α_k .[5] In this case, C5 also is satisfied with $\alpha = \max\{\alpha_k : k = 1, \ldots, K\}$.

The LS slope estimators $\hat{\beta}$ of β_0 are defined by

$$\hat{\beta} = \left(\Sigma_1^n X_i X_i' - \tfrac{1}{n}\Sigma_1^n X_i \Sigma_1^n X_i'\right)^{-}\left(\Sigma_1^n X_i Y_i - \tfrac{1}{n}\Sigma_1^n X_i \Sigma_1^n Y_i\right) ,$$

where $(\cdot)^{-}$ denotes some generalized inverse.

Our first result for the LS estimators is the following:

THEOREM 1: *Under assumptions* C1-C5, *the* LS *slope estimators* $\hat{\beta}$ *of* β_0 *in model* (1) *satisfy:*

(a) $\hat{\beta} \to \beta_0$ *as* $n \to \infty$ a.s., *if*

$$\alpha < \begin{cases} 2p & \text{when } p \le 1 \text{ and } r \ge 1 \\ 2p/(1+p(1/r - 1)) & \text{when } p \le 1 \text{ and } r \le 1 , \\ 2 & \text{when } p \ge 1 \end{cases}$$

provided $r > 2\alpha/(2+\alpha)$,

(b) *if* $r = \alpha$ *(as occurs, for example, if the regressors* $\{X_i\}$ *are identically distributed and each element* $X_{ik} \in DA(\alpha)$), *then* $\hat{\beta} \to \beta_0$ *as* $n \to \infty$ a.s., *provided*

$$\alpha < \begin{cases} p/(1-p) & \text{when } p \le 1/2 \\ 2p & \text{when } 1/2 \le p \le 1 , \\ 2 & \text{when } p \ge 1 \end{cases}$$

(c) $n^{\xi}(\hat{\beta} - \beta_0) \to \underset{\sim}{0}$ *as* $n \to \infty$ a.s., *for all* $\xi \geq 0$ *such that*

$$\xi < \begin{cases} 2/\alpha - 1/p & \textit{when} \quad p \leq 1 \quad \textit{and} \quad r \geq 1 \\ 2/\alpha - 1/p + 1 - 1/r & \textit{when} \quad p \leq 1 \quad \textit{and} \quad r \leq 1 \\ 2/\alpha - 1 & \textit{when} \quad p \geq 1 \quad \textit{and} \quad r \geq 1 \\ 2/\alpha - 1/r & \textit{when} \quad p \geq 1 \quad \textit{and} \quad r \leq 1 \end{cases},$$

provided $r > 2\alpha/(2+\alpha)$.

The proofs of Theorem 1 and all of the results that follow are given in the Appendix.

COMMENTS: 1. The Theorem yields consistency and rates of convergence for the LS estimators of slope parameters for cases where the errors have undefined $(p \leq 1)$ and well-defined $(p > 1)$ means. The crucial factors are the relative variabilities of the regressors and the errors, as measured by r and α , and p , respectively. Parts (a) and (b) and Figure 1 show that consistency may hold even if the errors have undefined means, and are more highly variable than the regressors, i.e., $p \leq 1$ and $p < r \leq \alpha$. Part (c) and Figure 2 show that very fast rates of convergence occur in certain contexts. In particular, no matter how small the maximal moment exponent p of the errors, the LS estimators may have a faster rate of convergence to β_0 than the standard rate $n^{-1/2}$.

2. For the case of well-defined error means $(p \geq 1)$, the consistency results generalize existing results for models without trending regressors by avoiding assumptions on the temporal dependence of the errors, and by avoiding the assumption that the errors have conditional mean zero given the regressors.

3. The length of the interval $[r,\alpha]$ indicates the variation in the maximal moment exponents of the different elements of the regressor vector X_i. If each element has the same moment exponent, then we often have $r = \alpha$. (This is necessarily so in the case of identically distributed regressor vectors.) The condition $r > 2\alpha/(2+\alpha)$ in parts (a) and (c) is a restriction on the length of $[r,\alpha]$. Figure 3 indicates the admissible $[r,\alpha]$ ranges for the results of the Theorem.

4. Bounded influence estimators (see Maronna and Yohai (1981) and Krasker and Welsch (1982)) are, at best, $n^{1/2}$-consistent in the above context. Hence, the LS estimators may exhibit faster rates of convergence than bounded influence estimators, even when the errors have undefined means. On the other hand, if the regressors are not highly variable, then LS will be inconsistent whereas bounded influence estimators still will be consistent. (In addition, bounded influence estimators may be more stable than LS (see Andrews (1986b)), and may perform better than LS under different forms of model failure.)

5. The proof of Theorem 1 makes extensive use of a result of Loeve (1955, Theorem 29.E.1, p. 387), see Lemma A-1 of the Appendix.

The next result shows that even though the assumptions of Theorem 1 are quite weak, the bounds on the rates of convergence, and on the values of α, are best possible when the errors have undefined means and $r = \alpha$:

THEOREM 2: *In Theorem 1 above, and Theorems 3 and 4 below, the upper bounds on* α *and* ξ *are sharp for the case* $r = \alpha \neq 1$ *and* $p < 1$ *. That is, in these cases, the replacement of the strict inequality* " < " *by a weak inequality* " ≤ " *in any of the expressions bounding the values of* α *and* ξ *in parts (a), (b), or (c) renders the corresponding result false.*

COMMENTS: 1. The proof of the Theorem proceeds by producing counterexamples to the Theorem when " < " is replaced by " ≤ ". These counterexamples are not pathological. Rather, they consist of quite standard models. In particular, they assume mutually independent, iid regressors and errors that are in the domains of attraction of stable laws.

2. The proof uses some results of Cline (1983) concerning the domain of attraction of the product of two rv's.

3. INFINITE ORDER MOVING AVERAGE REGRESSORS

In Section 2 we considered the case of M-dependent regressors. Unfortunately, the assumption of M-dependence can be restrictive. For example, it precludes the case of autoregressive or ARMA regressors. In this section we relax this assumption by considering regressor vectors that are infinite order vector moving averages with heterogeneous innovations and heterogeneous moving average (MA) coefficients. That is, we suppose the regressors are of the form $X_i = \sum_{j=0}^{\infty} A_{ij}\varepsilon_{i-j}$, where the A_{ij} are $K \times D$ non-random coefficient matrices, and $\{\varepsilon_i : i = \ldots -1, 0, 1, \ldots\}$ is a sequence of random innovation D-vectors. Under the assumptions introduced below, the infinite sum converges absolutely with probability one, so X_i is well-defined. Infinite order moving averages allow for a rich array of temporal dependence of the regressors. This is illustrated in the stationary finite variance case, by the general results for Wold decompositions.

Of the innovations, $\{\varepsilon_i\}$, and the moving average coefficients, $\{A_{ij}\}$, we make the following assumptions (which are discussed below):

ASSUMPTION D1: $\{\varepsilon_i\}$ is a sequence of independent random vectors.

ASSUMPTION D2: $\|\varepsilon_i\| \overset{ST}{\leq} \Xi$, $\forall i$, for some rv Ξ with maximal moment exponent $r > 0$ (where $\|\cdot\|$ denotes the Euclidean norm).

ASSUMPTION D3: $\{\varepsilon_i\}$ satisfies: $\forall \zeta \in R^D$ with $\|\zeta\| = 1$,

$$|\zeta' \varepsilon_i| \overset{ST}{\geq} \Xi_\zeta , \quad \forall i ,$$

for some real-valued rv $\Xi_\zeta \in DA(\alpha)$, where $0 < \alpha < 2 \wedge 2r$.

ASSUMPTION D4: $\{A_{ij}\}$ satisfies: $\|A_{ij}\| \leq M_j$, $\forall i$, for some constants M_j , $j = 0, 1, \ldots$, that satisfy $\sum_{j=0}^{\infty} (M_j)^b < \infty$, for some $b < r \wedge 1$.

ASSUMPTION D5: $\{A_{ij}\}$ satisfies: For some j and some $\delta > 0$, $\lambda_{min}(A_{ij}A'_{ij}) \geq \delta$, $\forall i$.

ASSUMPTION D6: $|\varepsilon_{i-j,\ell} U_i| \overset{ST}{\leq} \tilde{G}$, $\forall \ell = 1, \ldots, D$; $\forall j = 0, 1, 2, \ldots$, $\forall i = 1, 2, \ldots$, for some rv $\tilde{G}$ with maximal moment exponent greater than or equal to $1 \wedge p \wedge r$, and D4 holds with some $b < 1 \wedge p \wedge r$.

(Note that $\varepsilon_{i-j,\ell}$ denotes the ℓ^{th} element of ε_{i-j} .)

Assumptions D2 and D4 allow the regressors to be non-identically distributed. The condition on the decay rate of the dominating constants M_j of the MA coefficients is stronger than the standard assumption

$\sum_{j=0}^{\infty} (M_j)^2 < \infty$. The stronger condition is used to establish almost sure convergence of $\sum_{j=0}^{\infty} A_{ij}\varepsilon_{i-j}$ (and is used elsewhere in the proof of Theorem 3 below).

Assumptions D3 and D5 are identification conditions that combine to imply C5. These assumptions could be replaced by C5 in the results that follow, but it is preferable to impose conditions that are as primitive as possible. Thus, all assumptions are placed on the underlying innovations and MA coefficients rather than on the regressors themselves. D3 can be motivated in the same manner as C5 above.

To see the content of assumption D5, consider the case where A_{ij} does not depend on i . For example, this occurs with stationary regressors. In this case, D5 holds if and only if A_jA_j' is positive definite for some j . This condition is used to ensure sufficient independent variability of each of the elements of the regressor vector X_i .

Assumption D6, like assumption C3, is used to control the tail dependence between the regressors and the errors. Clearly, it holds if the errors and the regressor innovations are independent. It is a much weaker assumption than independence, however, as the discussion following C3 indicates.

To see that $\sum_{j=0}^{\infty} \|A_{ij}\|\cdot\|\varepsilon_{i-j}\| < \infty$ a.s., and hence, that X_i is well-defined, take b as in D4, i.e., $b < r \wedge 1$. Then,

$$
\begin{aligned}
E\left(\sum_{j=0}^{\infty}\|A_{ij}\|\cdot\|\varepsilon_{i-j}\|\right)^b &\le E\sum_{j=0}^{\infty}(M_j)^b\cdot\|\varepsilon_{i-j}\|^b = \sum_{j=0}^{\infty}(M_j)^b\cdot E\|\varepsilon_{i-j}\|^b \\
&\le \sum_{j=0}^{\infty}(M_j)^b\cdot E\Xi^b < \infty ,
\end{aligned}
\tag{8}
$$

using D2, D4, and the inequality $(v+w)^b \leq v^b + w^b$ for $v \geq 0$, $w \geq 0$, and $0 \leq b \leq 1$. This implies immediately that $\sum_{j=0}^{\infty} \|A_{ij}\| \cdot \|\varepsilon_{i-j}\| < \infty$ a.s.

With the above assumptions replacing assumptions C2-C5 of Theorem 1, we have:

THEOREM 3: *Suppose the regressors are of the form* $\sum_{j=0}^{\infty} A_{ij}\varepsilon_{i-j}$ *in model* (1). *Under assumptions* C1 *and* D1-D6, *the* LS *slope estimators* $\hat{\beta}$ *satisfy the convergence results of parts (a), (b), and (c) of Theorem* 1.

COMMENT: This result applies in a wide variety of time series contexts. No assumptions are placed on the temporal dependence of the errors, and the regressors can be infinite order moving averages with heterogeneous innovations and MA coefficients.

4. REGRESSION WITH SOME REGRESSORS HIGHLY VARIABLE AND OTHERS LESS VARIABLE

This section considers regression models in which different regressors exhibit differing degrees of variability. Such situations arise quite frequently in economics. An obvious example is the case of regression with random regressors as well as a seasonal dummy. For the case of errors with undefined means, we show that the LS slope parameter estimators of highly variable regressors are consistent, even in the presence of less variable regressors.

Consider the model

$$(9) \qquad Y_i = X_i'\beta_0 + Z_i'\emptyset_0 + U_i \ , \ i = 1, \ldots, n \ ,$$

where $Y_i \in R$, $X_i \in R^K$, $Z_i \in R^L$, and $U_i \in R$. The regressors $\{X_i, Z_i\}$ are split into two groups depending upon whether their maximal moment exponent is less than two, or greater than or equal to two. The regressors $\{X_i\}$ constitute the highly variable regressors, and are taken to satisfy the same assumptions C2-C5 as in Section 2 above, or D1-D6 as in Section 3 above. The regressors $\{Z_i\}$ are less variable. They may contain non-random variables, such as an intercept or a seasonal dummy, as well as random variables. Assume the regressors $\{Z_i\}$ satisfy:

ASSUMPTION C6: $\|Z_i\| \overset{ST}{\lesssim} Z$, $\forall i$, for some rv $Z \in L^2$.

ASSUMPTION C7: $\{Z_iZ_i' : i = 1, 2, \ldots\}$ satisfy a strong law of large numbers. That is, $\frac{1}{n}\Sigma_1^n(Z_iZ_i' - EZ_iZ_i') \xrightarrow{n\to\infty} \underset{\sim}{0}$ a.s.

ASSUMPTION C8: For some $\varepsilon > 0$, $\lambda_{\min}\left(\frac{1}{n}\Sigma_1^n EZ_iZ_i'\right) \geq \varepsilon > 0$, for n large.

ASSUMPTION C9: $|Z_{i\ell}U_i| \overset{ST}{\lesssim} G_1$, $\forall \ell = 1, \ldots, L$, $\forall i$, and $|Z_{i\ell}X_{ik}| \overset{ST}{\lesssim} G_2$, $\forall \ell = 1, \ldots, L$, $\forall k = 1, \ldots, K$, $\forall i$, where G_1 and G_2 are rv's with maximal moment exponents greater than or equal to $1 \wedge p$ and $1 \wedge r$, respectively.

Assumption C6 is the basic condition that classifies a regressor as being "less variable." For random regressors, assumption C7 requires some form of asymptotic weak dependence of $\{Z_i\}$, such as ergodicity, strong mixing, or infinite-order moving average form. For fixed regressors, C7 is satisfied trivially, since $Z_iZ_i' = EZ_iZ_i'$, $\forall i$.

Assumption C8 is a standard identification condition for the parameter vector ϕ_0. It is possible to relax this assumption to allow for lack of identification of ϕ_0. This relaxation, however, lengthens the proof of

Theorem 4 below considerably. Hence, for brevity, we adopt the assumption as stated. (Note that C8 cannot be dispensed with entirely without changing the rate of convergence results. In particular, for the results given below, C8 can be relaxed to the extent that it holds with the vector Z_i replaced by some one of its elements.)

Assumption C9 is analogous to C3. It limits the degree of tail dependence between the less variable regressors, and (i) the errors and (ii) the highly variable regressors.

The LS estimators of β_0 in model (9) can be written as

$$(10)\quad \tilde{\beta} = \left(\Sigma_1^n X_i X_i' - \Sigma_1^n X_i Z_i'(\Sigma_1^n Z_i Z_i')^{-}\Sigma_1^n Z_i X_i'\right)^{-}\left(\Sigma_1^n X_i Y_i - \Sigma_1^n X_i Z_i'(\Sigma_1^n Z_i Z_i')^{-}\Sigma_1^n Z_i Y_i\right) ,$$

where $(\cdot)^{-}$ denotes some generalized inverse. These estimators satisfy the following consistency and rate of convergence results:

THEOREM 4: *Under assumptions* C1-C9, or assumptions C1, D1-D6, *and* C6-C9, *the* LS *slope estimators* $\tilde{\beta}$ *of* β_0 *for the model* (9) *satisfy the results of Theorem* 1 *parts (a), (b), and (c) with* $\tilde{\beta}$ *in place of* $\hat{\beta}$.

As stated in Theorem 2 above, the rates of convergence given in this Theorem are sharp for the case of undefined error means $(p < 1)$ and $r = \alpha$. When the error means are undefined, the LS estimators of the parameter vector of the less variable regressors, $\emptyset_0$, generally are inconsistent.

5. MONTE CARLO RESULTS

This section presents some Monte Carlo results for the LS estimator and a well-known robust estimator, viz., the Huber M-estimator (see Huber (1973, p. 815)). The results indicate that the general qualitative results for LS obtained above in an asymptotic framework carry over, more or less, to finite samples. The Monte Carlo results also provide comparisons of LS with a robust procedure. These comparisons show that, although LS may perform quite well in estimating a slope coefficient when both the regressors and errors are thick-tailed, as compared to a thinner-tailed situation, LS may perform much worse than other available procedures.

We consider the simple linear regression model

$$Y_i = c_0 + X_i\beta_0 + \tilde{U}_i \ , \quad i = 1, \ldots, n \ , \tag{11}$$

where c_0 and β_0 are scalar parameters, and $\{X_i\}$ and $\{U_i\}$ are sequences of mutually independent iid scalar rv's. Since LS and the Huber M-estimator are equivariant, we can take $c_0 = \beta_0 = 0$ without loss of generality (e.g., see Andrews (1986a)). The distribution of the regressor X_i and the error U_i is taken to be symmetric stable with characteristic exponent α equal to 2.0 (normal), 1.5 (Holtsmark), 1.0 (Cauchy), or 0.5. (The same distribution is taken for X_i and U_i .) The sample sizes investigated are 25, 50, 100, and 200. Fifty thousand repetitions are used.

The Huber estimator considered here is the classical Huber M-estimator which solves for (c,β,σ) the equations:

$$\Sigma_1^n \psi_\kappa((Y_i - c - X_i\beta)/\sigma)\begin{pmatrix}1\\X_i\end{pmatrix} = \underset{\sim}{0}$$

(12)

$$\Sigma_1^n \psi_\kappa^2((Y_i - c - X_i\beta)/\sigma)/2 = g \ ,$$

where σ is a scale parameter, $\psi_\kappa(u) = \min(\kappa, \max(-\kappa,u))$, $\kappa = 1.35$, $g = (n-2)\cdot\int\psi_\kappa^2(u)\frac{1}{2}\phi(u)du$, and $\phi(u)$ is the standard normal distribution. This estimator does not downweight outlying regressor values, and hence, is not a bounded influence estimator (cf., Krasker and Welsch (1982)).

Our interest is focused on the estimation of the slope coefficient β_0 . We describe the distributions of the LS and Huber M-estimator of β_0 using their 75^{th}, 85^{th}, and 95^{th} percentiles.[6] Since both estimators are symmetric about the true parameter value zero in this case (see Andrews (1986a)), the negatives of these percentiles give the 25^{th}, 15^{th}, and 5^{th} percentiles, respectively. The [25%, 75%] interval is the interquartile range--a common measure of dispersion. The [5%, 95%] interval is a measure of thickness of tails. The [15%, 85%] interval is intermediate between the two.

The first set of Monte Carlo results concerns the LS estimator of β_0 , see Table I. These results illustrate the following points:

1. The LS estimator has the bulk of its probability mass (as measured by its 75^{th} or 85^{th} percentile) more concentrated about the true value, the smaller is the value of α (i.e., the greater is the variability of the regressor and error). This is true for all sample sizes considered. The relative increase in concentration with declining α increases with the sample size. This is to be expected, given the faster rate of convergence to zero as $n \to \infty$ for smaller α values.

2. The LS estimator has thicker tails (as measured by its 95^{th} percentile), the smaller is α. This holds for all sample sizes. In addition, the relative increase in tail thickness with declining α decreases with the sample size. Again, this can be attributed to the faster rate of convergence to zero as $n \to \infty$ for smaller values of α.

3. The LS percentiles decline as the sample size increases at *roughly* the rate suggested by Theorem 1, viz., n^{-1} for $\alpha = 0.5$ and $\alpha = 1.0$, $n^{-2/3}$ for $\alpha = 1.5$, and $n^{-1/2}$ for $\alpha = 2.0$.[7] For $\alpha = 1.0$ and $\alpha = 1.5$ this decline is somewhat slower than n^{-1} and $n^{-2/3}$, respectively, whereas for $\alpha = 0.5$ it is somewhat faster than n^{-1}.

Table I shows that depending upon the shape of one's loss function, one might prefer the behavior of LS with $\alpha = 1.0$ (Cauchy) over $\alpha = 2.0$ (normal) for samples as small as twenty-five. A tradeoff exists between the concentration of the bulk of the density about the true value, and the thickness of its tails. For larger sample sizes, this tradeoff moves increasingly in favor of the performance of LS with smaller α values.

To conclude, the results show that not only does the LS estimator not perform disastrously in some cases in which the errors have undefined means, but it actually may exhibit superior performance in comparison with cases in which the errors are normally distributed. (Of course, if the distribution of the regressor is held fixed for these comparisons, the results are quite different. When the error is comprised of left out variables, however, it often may be reasonable to assume that the regressors and errors display similar properties, such as similar tail behavior, as is crucial here.)

The second set of Monte Carlo results compares the LS estimator of β_0 with a robust regression estimator, see Table II. This Table illustrates quite clearly the desirable features of the Huber M-estimator in the present context. It is much more concentrated about the true value, and has thinner tails, than the LS estimator (as indicated by the numbers in parentheses) when $\alpha = 1.5$, 1.0, or 0.5. Only in the case of normal regressors and errors is the LS estimator preferred over the Huber M-estimator, and in this case the difference between the two is quite small.

From the above results, we conclude that the LS estimator of the slope coefficient does not perform poorly with thick-tailed regressors and errors, in comparison to its performance with thinner-tailed variables. Other procedures, however, perform very much better with thick-tailed variables than with thinner-tailed ones. In consequence, the performance of LS relative to other available procedures is poor when the errors and regressors are thick-tailed.

6. CONCLUSIONS

The main conclusions of this paper are summarized as follows:

1. When assessing the performance of estimation procedures, analogies between the regression model and the location model can be misleading. In the location model, least squares is inconsistent if the errors have undefined means. In the regression model, however, least squares may be consistent for some of the parameters, even when the errors have undefined means.

2. In the linear regression model, the performance of least squares depends crucially on the relative variability of the regressors and the errors --regardless of how fat are the tails of the error distributions. Under classical assumptions on the errors, the importance of this relative variability has been recognized widely. It follows immediately from the

asymptotic covariance matrix of the LS estimators, viz., $\lim_{n\to\infty}\left(\frac{1}{n}\Sigma_1^n EX_iX_i'\right)^{-1}\sigma^2$. The present results show that its importance carries over to the case of undefined error means.

3. The performance of least squares in models with fat-tailed errors _may_ not be nearly as poor in regression models as in location models. This does not imply that LS is necessarily a desirable procedure in the case of highly variable errors and regressors. Other procedures may perform much better (as the Monte Carlo results indicate). Nevertheless, the use of LS in this context is far from disastrous--in contrast to the location case. On the other hand, if the errors have undefined means and the regressors are not highly variable, then LS can perform just as poorly in the linear regression model as in the location model.

4. When the regressors are highly variable, the justification for the use of bounded influence estimators cannot rely solely on the existence of fat-tailed errors. It needs to be supplemented by some failure of the model, such as measurement errors in the regressor variables or incorrect functional form. This follows because LS estimation may be preferable to bounded influence estimation, even when the errors have undefined means, due to the faster rate of convergence of LS for (sufficiently) highly variable regressors.

5. Regardless of the temporal dependence in the errors, the LS slope parameter estimators are consistent, if the regressors are highly variable. This holds even if the errors are highly variable.

Cowles Foundation for Research in Economics, Yale University

APPENDIX

This Appendix provides proofs of Theorems 1-4 given in Sections 2-4 above. First, we consider the proof of Theorem 1. The proof is broken into several parts for ease of reading, and to allow the use of these parts in the proofs of Theorems 3 and 4 below.

We introduce an assumption that is shown below (in Theorem A-2) to be weaker than assumptions C4 and C5:

ASSUMPTION C45: $\underline{\lim}_{n\to\infty} n^{\gamma}\lambda_{\min}(\Sigma_1^n X_i X_i') > 0$ a.s., for some $\gamma < \gamma(p,r)$,

$$\text{where} \quad \gamma(p,r) = \begin{cases} -1 & \\ -1/p & \\ 1 - 1/p - 1/r & \\ 1 - 2/r & \end{cases} \text{when} \quad \begin{matrix} r \geq 1, & p \geq 1 \\ r \geq 1, & p \leq 1 \\ r \leq 1, & p \leq r \\ r \leq 1, & p \geq r \end{matrix} .$$

Assumption C45 requires the regressors to exhibit a certain degree of variability. Conditions of this sort, on the eigenvalues of the sample second moment matrices of the regressors, are common in recent analyses of consistency of LS in dynamic and non-dynamic regression models, e.g., see Anderson and Taylor (1979), Christopeit and Helmes (1980), and Lai and Wei (1982a,b). Such assumptions are satisfactory when the regressors are fixed. When the regressors are stochastic, however, it is highly desirable to replace them by more primitive assumptions on the underlying distribution of the regressors. Whence our use of the primitive assumptions C4 and C5 in the text above, rather than C45.

With assumption C45, we have the following consistency result:

THEOREM A-1: *Under assumptions* C1-C3 *and* C45, *the* LS *slope estimators* $\hat{\beta}$ *satisfy*

(a) $\hat{\beta} \to \beta_0$ *as* $n \to \infty$ a.s., *and*

(b) $n^{\xi}(\hat{\beta} - \beta_0) \to 0$ *as* $n \to \infty$ a.s., *for all* $\xi < \nu(p,r) - \gamma$,

$$\text{where} \quad \nu(p,r) = \begin{cases} -1 & r \geq 1 \ , \quad p \geq 1 \\ -1/p & r \geq 1 \ , \quad p \leq 1 \\ 1 - 1/p - 1/r & \text{when} \quad r \leq 1 \ , \quad p \leq 1 \\ -1/r & r \leq 1 \ , \quad p \geq 1 \end{cases} .$$

The proof of Theorem A-1 follows that of Theorem 1.

COMMENT: Theorem A-1 holds without any assumption on the temporal dependence of the errors or the regressors. (Thus, even the zero-one result for the LS estimator, see Andrews (1985a), does not apply in this context.) All that is needed is that the regressors display sufficient variability to satisfy C45. In consequence, this result incorporates dynamic models, i.e., models with lagged values of Y_i entering as elements of X_i . One just needs to verify assumption C45 (although this may require considerable effort). Since our focus in this paper is on the standard linear regression model, we do not verify this condition, nor replace it by more primitive conditions, for dynamic regression models, but only for standard regression models. (See Andrews (1985b) for some related results for the dynamic regression model.)

Next we show that assumptions C4 and C5 imply that C45 holds with a particular value of γ :

THEOREM A-2: *Suppose* $\{X_i\}$ *satisfies assumptions C4 and C5. Then,*

$$n^{\eta}\lambda_{min}(\Sigma_1^n X_i X_i') \xrightarrow{n\to\infty} \infty \quad \text{a.s.,} \quad \forall\eta > -2/\alpha \ .$$

COMMENT: The proof of Theorem A-2 makes use of results of Derman and Robbins (1955) and Miller (1967).

Theorems A-1 and A-2 combine to establish Theorem 1:

PROOF OF THEOREM 1: By Theorem A-2, C45 holds with any $\gamma \in (-2/\alpha,\ \gamma(p,r))$. Hence, C45 holds if $\gamma(p,r) > -2/\alpha$. By definition of $\gamma(p,r)$ and $\nu(p,r)$, this occurs iff $\nu(p,r) > -2/\alpha$ and $1 - 2/r > -2/\alpha$. Since the upper bound on ξ in part (c) equals $\nu(p,r) + 2/\alpha$, the statement of part (c) only has content if $\nu(p,r) > -2/\alpha$, as desired. Also, the condition $1 - 2/r > -2/\alpha$ is equivalent to $r > 2\alpha/(2+\alpha)$, which is assumed. Thus, using Theorem A-1, part (c) holds.

Part (a) follows from part (c) by determining conditions on α and p such that the upper bound on ξ is positive. Part (b) follows from part (a) by noting that (i) $\alpha > 2\alpha/(2+\alpha)$ always holds, (ii) $\alpha < 2p/(1 + p(1/\alpha - 1))$ iff $\alpha < p/(1-p)$, (iii) $p/(1-p) \leq 1$ iff $p \leq 1/2$, (iv) $2p > 1$ iff $p > 1/2$, and (v) $p/(1-p) < 2p$ if $p < 1/2$. *Q.E.D.*

We now turn to the proof of Theorem A-1. Four lemmas are used in its proof. The first lemma is due to Loeve (1955, Theorem 29.E.1, p. 387). It is used extensively.

LEMMA A-1 (Loeve): *Let* $\{B_i : i \geq 1\}$ *and* B *be rv's such that* $|B_i| \overset{ST}{\leq} B$, $\forall i$, *and* $B \in L^d$, *for some* $d \in (0,1)$, *then*

$$n^{-1/d}\, \Sigma_1^n B_i \xrightarrow{n\to\infty} 0 \quad \text{a.s.}$$

The proof of this Lemma uses a classical truncation argument, an L^2 martingale convergence theorem, and a moment inequality (see Loeve (1955) or Chatterji (1969)).

Let $M_n = \Sigma_1^n X_i U_i - \frac{1}{n}\Sigma_1^n X_i \Sigma_1^n U_i$, and $Q_n = \lambda_{min}(\Sigma_1^n X_i X_i' - \frac{1}{n}\Sigma_1^n X_i \Sigma_1^n X_i')$.

LEMMA A-2: *Suppose* (i) $\lim_{n\to\infty} n^{\nu} M_n = 0$ a.s., *for some* ν , *and* (ii) $\underline{\lim}_{n\to\infty} n^{\tau} Q_n > 0$ a.s., *for some* τ . *Then, the* LS *slope estimators* $\hat{\beta}$ *satisfy*

$$n^{\nu-\tau}(\hat{\beta} - \beta_0) \xrightarrow{n\to\infty} 0 \quad \text{a.s.}$$

LEMMA A-3: *Suppose* C2 *and* C45 *hold, then*

$$\underline{\lim}_{n\to\infty} n^{\tau} Q_n > 0 \quad \text{a.s.,} \quad \forall \tau \geq \gamma \text{ , } \textit{for } \gamma \textit{ as in } \text{C45.}$$

LEMMA A-4: *Suppose* C1-C3 *hold, then*

$$n^{\nu} M_n \xrightarrow{n\to\infty} 0 \quad \text{a.s.,} \quad \forall \nu < \nu(p,r) \text{ .}$$

The proofs of these three lemmas follow that of Theorem A-1.

PROOF OF THEOREM A-1: Lemmas A-2, A-3, and A-4 give

$$\text{(A.1)}\qquad n^{\nu-\tau}(\hat{\beta}-\beta_0) \xrightarrow{n\to\infty} 0 \quad \text{a.s., for all} \quad \nu < \nu(p,r) \quad \text{and} \quad \tau \ge \gamma ,$$

i.e., for $\nu-\tau < \nu(p,r) - \gamma$, which establishes part (b). By C45, $\nu(p,r) - \gamma \ge \gamma(p,r) - \gamma > 0$, so part (a) follows. *Q.E.D.*

PROOF OF LEMMA A-2: The LS slope estimators $\hat{\beta}$ solve

$$\text{(A.2)}\qquad \left[\Sigma_1^n X_i X_i' - \tfrac{1}{n}\Sigma_1^n X_i \Sigma_1^n X_i'\right](\hat{\beta}-\beta_0) = M_n , \quad \text{or}$$

$$\text{(A.3)}\qquad \|\hat{\beta}-\beta_0\|\tilde{Q}_n = (\hat{\beta}-\beta_0)'M_n/\|\hat{\beta}-\beta_0\| ,$$

where $\tilde{Q}_n = (\hat{\beta}-\beta_0)'\left[\Sigma_1^n X_i X_i' - \frac{1}{n}\Sigma_1^n X_i \Sigma_1^n X_i'\right](\hat{\beta}-\beta_0)/\|\hat{\beta}-\beta_0\|^2$. Note that $\tilde{Q}_n \ge \inf_{\varsigma:\|\varsigma\|=1} \varsigma'\left[\Sigma_1^n X_i X_i' - \frac{1}{n}\Sigma_1^n X_i \Sigma_1^n X_i'\right]\varsigma = Q_n$. Hence, using assumption (i), we get

$$\text{(A.4)}\qquad 0 \le n^{\nu-\tau}\|\hat{\beta}-\beta_0\|\cdot n^{\tau}Q_n \le n^{\nu}\|\hat{\beta}-\beta_0\|\tilde{Q}_n = n^{\nu}(\hat{\beta}-\beta_0)'M_n/\|\hat{\beta}-\beta_0\| \xrightarrow{n\to\infty} 0 \text{ a.s.}$$

In view of assumption (ii), this yields $n^{\nu-\tau}\|\hat{\beta}-\beta_0\| \xrightarrow{n\to\infty} 0$ a.s., as desired. *Q.E.D.*

PROOF OF LEMMA A-3: It suffices to consider the case where $\tau = \gamma$. We have

$$\text{(A.5)}\qquad \begin{aligned} n^{\gamma}Q_n &\ge n^{\gamma}\inf_{\varsigma:\|\varsigma\|=1}\varsigma'\left[\Sigma_1^n X_i X_i'\right]\varsigma - n^{\gamma}\sup_{\varsigma:\|\varsigma\|=1}\varsigma'\left[\tfrac{1}{n}\Sigma_1^n X_i \Sigma_1^n X_i'\right]\varsigma \\ &\ge n^{\gamma}\lambda_{min}(\Sigma_1^n X_i X_i') - n^{\gamma}\underset{\sim}{1}'\left|\tfrac{1}{n}\Sigma_1^n X_i \Sigma_1^n X_i'\right|\underset{\sim}{1} , \end{aligned}$$

where $\underset{\sim}{1}$ is a K-vector of ones, and $|A|$ denotes the matrix of absolute values of the elements of A . Using C45, we get $\underline{\lim}_{n\to\infty} n^{\gamma} Q_n > 0$ a.s., if we can show

(A.6) $$n^{\gamma/2-1/2}\Sigma_1^n X_i \xrightarrow{n\to\infty} 0 \text{ a.s., for } \gamma \text{ as in C45.}$$

Let $s = (r \wedge 1) - \varepsilon$, for $\varepsilon > 0$ arbitrarily small. By C2, $|X_{ik}| \overset{ST}{\lesssim} X \in L^s$. Lemma A-1 now gives

(A.7) $$n^{-1/s}\Sigma_1^n X_i \xrightarrow{n\to\infty} 0 \text{ a.s.}$$

Thus, (A.6) holds if $-1/s \geq \gamma/2 - 1/2$. For the case $r \geq 1$, we have $s = 1-\varepsilon$, so $-1/s$ is less than, but arbitrarily close to, -1. Thus, $-1/s \geq \gamma/2 - 1/2$ holds, if $-1 > \gamma/2 - 1/2$, or $\gamma < -1$. The latter follows from C45 by definition of γ . For the case $r < 1$, $\gamma/2 - 1/2 < (1-2/r)/2 - 1/2 = -1/r$ implies $\gamma/2 - 1/2 \leq -1/(r-\varepsilon) = -1/s$, for ε sufficiently small. Hence, (A.6) holds and the proof is complete.

Q.E.D.

PROOF OF LEMMA A-4: By assumption C3, $|X_{ik}U_i| \overset{ST}{\lesssim} G \in L^{(p^*\wedge r)-\varepsilon}$, for all $\varepsilon > 0$, $\forall k$, $\forall i$, where $p^* = p \wedge 1$. By Lemma A-1, this implies

(A.8) $$n^s\Sigma_1^n X_i U_i \xrightarrow{n\to\infty} 0 \text{ a.s., } \forall s < (-1/p^*) \wedge (-1/r) .$$

A fortiori, we have $n^{\nu}\Sigma_1^n X_i U_i \xrightarrow{n\to\infty} 0$, $\forall \nu < \nu(p,r)$, because $\nu(p,r) \leq (-1/p^*) \wedge (-1/r)$.

Since $p^* \leq 1$, Lemma A-1 gives

(A.9) $$n^{-1/p^*-\varepsilon}\Sigma_1^n U_i \xrightarrow{n\to\infty} 0 \text{ a.s., } \forall \varepsilon > 0 .$$

If $r < 1$, we also get $n^{-1/r-\varepsilon}\Sigma_1^n X_i \xrightarrow{n\to\infty} 0$ a.s., $\forall\varepsilon > 0$. Thus, for $r < 1$,

$$(A.10) \qquad n^{-1/p^*-1/r-2\varepsilon}\Sigma_1^n X_i \Sigma_1^n U_i \xrightarrow{n\to\infty} 0 \quad \text{a.s.}, \quad \forall\varepsilon > 0 .$$

Since $\nu < 1 - 1/p^* - 1/r$, (A.10) gives $n^{\nu}\frac{1}{n}\Sigma_1^n X_i \Sigma_1^n U_i \xrightarrow{n\to\infty} 0$ a.s., as desired, by taking ε sufficiently small.

If $r \geq 1$, then for all $\varepsilon > 0$, condition C2, Lemma A-1 and (A.9) yield

$$(A.11) \qquad n^{-(1+\varepsilon)}\Sigma_1^n X_i \xrightarrow{n\to\infty} 0 \quad \text{a.s., and} \quad n^{-1-1/p^*-2\varepsilon}\Sigma_1^n X_i \Sigma_1^n U_i \xrightarrow{n\to\infty} 0 \quad \text{a.s.}$$

Now, $\nu < -1/p^*$ implies $\nu-1 < -1 - 1/p^* - 2\varepsilon$, for some ε sufficiently small. Hence, $n^{\nu}\frac{1}{n}\Sigma_1^n X_i \Sigma_1^n U_i \xrightarrow{n\to\infty} 0$ a.s., which completes the proof. *Q.E.D.*

We now prove Theorem A-2. Its proof uses the following lemma:

LEMMA A-5: *Suppose* $\{V_i\}$ *are non-negative, M-dependent rv's and* $V_i \overset{ST}{\geq} V$ *for all* i, *for some* rv $V \in DA(\tilde{\alpha})$, *where* $0 < \tilde{\alpha} \leq 2$. *Then*

$$n^{\kappa}\Sigma_1^n V_i \xrightarrow{n\to\infty} \infty \quad \text{a.s.}, \quad \forall\kappa > -1/\tilde{\alpha} .$$

PROOF OF THEOREM A-2: Using C4 and C5, apply Lemma A-5 with $V_i = (\varsigma' X_i)^2$ *to get:* $\forall\varsigma \in R^K$ *with* $\|\varsigma\| = 1$ *and* $\kappa > -2/\alpha$,

$$(A.12) \qquad \varsigma'(n^{\kappa}\Sigma_1^n X_i X_i')\varsigma \xrightarrow{n\to\infty} \infty \quad \text{a.s.}$$

This implies $\left[n^{\kappa}\Sigma_1^n X_i X_i'\right]^{-1} \xrightarrow{n\to\infty} 0$ a.s. (e.g., see Wu (1980, Theorem 1)). Thus,

$$\text{(A.13)} \quad \lambda_{max}\left((n^{\kappa}\Sigma_1^n X_i X_i')^{-1}\right) \xrightarrow{n\to\infty} 0 \quad \text{a.s., and} \quad \lambda_{min}(n^{\kappa}\Sigma_1^n X_i X_i') \xrightarrow{n\to\infty} \infty \quad \text{a.s.,}$$

as desired. *Q.E.D.*

PROOF OF LEMMA A-5: It suffices to consider the case where $\kappa < 0$. The summands of $\Sigma_1^n V_i$ are non-negative, so it suffices to prove the result with $\Sigma_1^n V_i$ replaced by a sum that only contains every $(M+1)^{st}$ term of the sequence $\{V_i : i = 1, 2, \ldots\}$. That is, we can assume $\{V_i\}$ is an independent sequence without loss of generality.

The proof now proceeds in a somewhat similar fashion to that of Derman and Robbins (1955, Theorem 1), also see Miller (1967). For any $c > 0$,

$$\text{(A.14)} \quad \begin{aligned} P(n^{\kappa}\Sigma_1^n V_i \le c) &\le P(\max_{i\le n} V_i \le n^{-\kappa}c) = \prod_{i=1}^{n} P(V_i \le n^{-\kappa}c) \\ &\le \left[P(V \le n^{-\kappa}c)\right]^n \le \exp[-n(1 - P(V \le n^{-\kappa}c))] = \exp[-n^{1+\kappa\tilde{\alpha}}c^{-\tilde{\alpha}}L(n^{-\kappa}c)] \ , \end{aligned}$$

since $V \in DA(\tilde{\alpha})$ implies that $P(V > x) = x^{-\tilde{\alpha}}L(x)$, where $L(\cdot)$ is a slowly varying function, e.g., see Feller (1971, Corollary XVII.5.2, p. 578). The last inequality follows by the result: If $z \in [0,1]$, then $z^n \le e^{-n(1-z)}$, see Galambos (1978, Lemma 1.3.1). Thus, we have

$$\text{(A.15)} \quad \sum_{n=1}^{\infty} P(n^{\kappa}\Sigma_1^n V_i \le c) \le \sum_{n=1}^{\infty} \exp[-n^{1+\kappa\tilde{\alpha}}c^{-\tilde{\alpha}}L(n^{-\kappa}c)] < \infty \ ,$$

since $1 + \kappa\tilde{\alpha} > 0$, $L(x) \ge x^{-\varepsilon}$ for x large and any $\varepsilon > 0$, and for present purposes $\kappa < 0$. The first Borel-Cantelli Lemma now gives: $\forall c > 0$, (A.16) $\quad P(n^{\kappa}\Sigma_1^n V_i \le c \text{ for infinitely many } n) = 0$.

Thus, $P(\lim_{n\to\infty} n^{\kappa}\Sigma_1^n V_i = \infty) = 1$, as desired. *Q.E.D.*

This concludes the proofs of Theorem 1 and all the results used in its proof. Next we prove Theorem 2:

PROOF OF THEOREM 2: We consider only Theorem 1 because the same counterexamples given below apply to Theorems 3 and 4. Also, parts (a) and (b) of Theorem 1 are derived from part (c), so it suffices to show that, under assumptions C1-C5, part (c) of Theorem 1 does not necessarily hold if the strict inequality, $<$, is replaced by a weak inequality, $\le$, for the case $r = \alpha \ne 1$ and $p < 1$.

Toward this end, suppose (i) $\{X_i\}$ and $\{U_i\}$ are independent, iid sequences of scalar rv's, (ii) X_i is in the domain of normal attraction of a strictly stable rv with exponent $\alpha \in (0,2)$, denoted $X_i \in DNA(\alpha)$, and (iii) U_i is a strictly stable rv with exponent $p \in (0,1)$. (By definition, a stable rv is *strictly stable* if the constant h of equation (2) can be taken to be zero. Also, a rv is in the *domain of normal attraction* of a stable rv with characteristic exponent α if it is in the domain of attraction of the rv and the constants c_n of equation (3) equal $n^{-1/\alpha}$, for all n .)

For these rv's, conditions C3 and C4 clearly hold. By (ii) and (iii), the maximal moment exponent r of X_i is α , and that of U_i is p (e.g., see Feller (1971, Lemma XVII.5.1, p. 578)). Thus, conditions C1 and C2 hold. Condition C5 holds with $X_\zeta = |X_i|$, because the conditions on the tail probabilities of a rv that determine whether a distribution is in the domain of attraction of some rv imply: If $X_i \in DA(\alpha)$, then $|X_i| \in DA(\alpha)$ (e.g., see Feller (1971, Theorem XVII.5.2, p. 577)). Thus, $\{X_i\}$ and $\{U_i\}$ satisfy all the conditions of Theorem 1, and it suffices to show:

$$(A.17) \quad P\left[n^{\xi}(\hat{\beta}-\beta_0) \xrightarrow{n\to\infty} 0\right] < 1 \ , \text{ for } \ \xi = \begin{cases} 2/\alpha-1/p & \text{when } \alpha > 1 \text{ and } p < 1 \\ 1/\alpha-1/p+1 & \text{when } \alpha < 1 \text{ and } p < 1 \ . \end{cases}$$

First, we establish several preliminary results. By assumption (ii),

$$(A.18) \qquad n^{-1/\alpha}\Sigma_1^n X_i \xrightarrow{d} W_1 \quad \text{as} \quad n \to \infty \ ,$$

where W_1 is a stable rv with exponent α . Also, using the conditions on tail probabilities for a rv to be in the domain of normal attraction of some rv (see Gnedenko and Kolmogorov (1954, Theorem 7.35.5, p. 181)), it is straightforward to show that $X_i \in DNA(\alpha)$ implies $X_i^2 \in DNA(\alpha/2)$. Thus,

$$(A.19) \qquad n^{-2/\alpha}\Sigma_1^n X_i^2 \xrightarrow{d} W_2 \quad \text{as} \quad n \to \infty \ ,$$

for some non-negative stable rv W_2 . (Using the fact that $\alpha/2 < 1$, we have taken the centering constants to be zero.) Note that the rv W_2 is positive with probability one, because stable rv's have continuous distribution functions.

Combining these results, and using the continuous mapping theorem, yields

$$(A.20) \qquad n^{2/\alpha}\left(\Sigma_1^n X_i^2 - \tfrac{1}{n}\Sigma_1^n X_i \Sigma_1^n X_i'\right)^{-1} \xrightarrow{d} 1/W_2 \quad \text{as} \quad n \to \infty \ ,$$

where $1/W_2 > 0$ with probability one. By the definition of $\hat{\beta}$, this implies that $n^{\xi}(\hat{\beta} - \beta_0)$ converges in probability to zero only if

$$(A.21) \qquad n^{\xi-2/\alpha}\left(\Sigma_1^n X_i U_i - \tfrac{1}{n}\Sigma_1^n X_i \Sigma_1^n U_i\right) \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty .$$

We will show that (A.21) does not hold, and hence, $n^{\xi}(\hat{\beta} - \beta_0)$ does not converge to 0 in probability, or *a fortiori*, almost surely.

By assumption (iii),

$$(A.22) \qquad n^{-1/p}\Sigma_1^n U_i \xrightarrow{d} W_3 \quad \text{as} \quad n \to \infty ,$$

where W_3 is a stable rv with exponent p.

Under the assumptions, $X_i U_i \in DA(\alpha \wedge p)$, by Cline (1983, Corollary 3.2, Part III, p. 80). Hence, there exists a sequence of constants $\{c_n : n = 1, 2, \ldots\}$ such that

$$(A.23) \qquad c_n^{-1}\Sigma_1^n X_i U_i \xrightarrow{d} W_4 \quad \text{as} \quad n \to \infty ,$$

where W_4 is a stable rv with exponent $\alpha \wedge p$ (< 1).

Consider the case $\alpha < 1$. Then, $\xi - 2/\alpha = 1 - 1/\alpha - 1/p$. By (A.18) and (A.22),

$$(A.24) \qquad n^{1-1/\alpha-1/p}\, \tfrac{1}{n}\Sigma_1^n X_i \Sigma_1^n U_i \xrightarrow{d} W_1 \cdot W_3 \quad \text{as} \quad n \to \infty ,$$

where W_1 and W_3 are independent. And, by Lemma A-1, $\forall s < \alpha \wedge p$, $n^{-1/s}\Sigma_1^n X_i U_i \xrightarrow{n\to\infty} 0$ a.s. When $\alpha < 1$ and $p < 1$, we get $1 - 1/\alpha - 1/p < -1/s$, for s sufficiently close to $\alpha \wedge p$. Thus,

$$(A.25) \qquad n^{1-1/\alpha-1/p}\, \Sigma_1^n X_i U_i \xrightarrow{n\to\infty} 0 \quad \text{a.s.}$$

Combining (A.24) and (A.25) shows that (A.21) does not hold, and so, (A.17) holds for $\alpha < 1$.

Next, consider the case $\alpha > 1$. Then, $\xi - 2/\alpha = -1/p$. By (A.18) and (A.22),

$$\text{(A.26)} \qquad n^{-1/p} \tfrac{1}{n}\Sigma_1^n X_i \Sigma_1^n U_i \xrightarrow{P} 0 \quad \text{as} \quad n \to \infty .$$

Combining (A.23) and (A.26), we find that if

$$\text{(A.27)} \qquad \overline{\lim_{n\to\infty}}\, c_n/n^{1/p} > 0 ,$$

then (A.21) does not hold, and hence, the desired result (A.17) does hold.

To show (A.27), we use an argument similar to that of Cline (1983, pp. 119-120). The norming constants for $U_i \in \mathrm{DNA}(p)$ and $X_iU_i \in \mathrm{DA}(p \wedge \alpha)$ are $n^{1/p}$ and c_n , respectively. By DeHaan (1970, p. 22) , $c_n/n^{1/p} \xrightarrow{n\to\infty} 0$ iff $C(t)/B(t) \xrightarrow{t\to\infty} 0$, where $C(t) = \frac{1}{t^2}\int_0^t u^2 dF_{|XU|}(u)$, $B(t) = \frac{1}{t^2}\int_0^t u^2 dF_U(u)$, and $F_{|XU|}$ and F_U denote the distribution functions of $|XU|$ and U , respectively. By Cline (1983, p. 120), $\lim_{t\to\infty}(1 - F_{|XU|}(t))/C(t) = (2-p)/p$, provided $\alpha > p$. And by Cline (1983, Lemma 2.1, Part III, p. 71), $\lim_{t\to\infty}(1 - F_{|XU|}(t))/P(|U| > t) = E|X_i|^p$. Thus,

$$\text{(A.28)} \quad \overline{\lim_{t\to\infty}} \frac{C(t)}{B(t)} = \frac{pE|X_i|^p}{2-p} \overline{\lim_{t\to\infty}} \frac{P(|U| > t)}{B(t)} = \frac{pE|X_i|^p}{2-p} \overline{\lim_{t\to\infty}} \frac{t^2 P(|U| > t)}{\int_0^t u^2 dF_U(u)} .$$

The right-hand-side of (A.28) equaling zero is a necessary and sufficient condition for the weak limit of $c_n^{-1}\Sigma_1^n U_i$ to be normal, see Gnedenko and

Kolmogorov (1954, p. 172). Since $U_i \epsilon$ DNA(p) , $p \neq 2$, we have $\lim_{t\to\infty} C(t)/B(t) \neq 0$, which establishes (A.27). *Q.E.D.*

We now prove Theorem 3. The proof is analogous to that of Theorem 1, except that we first show that assumptions D1-D5, rather than C4 and C5, imply C45. This is done in the following theorem:

THEOREM A-3: *Suppose the regressors* $\{X_i\}$ *are of the form* $X_i = \sum_{j=0}^{\infty} A_{ij}\epsilon_{i-j}$. *Under assumptions* D1-D5, *we have* $n^{\eta}\lambda_{\min}(\Sigma_1^n X_i X_i') \to \infty$ *as* $n \to \infty$ a.s., $\forall \eta > -2/\alpha$.

COMMENT: The proof of Theorem A-3 makes use of that of Theorem A-2.

PROOF OF THEOREM 3: To prove Theorem 3, we need to verify the conditions of Theorem A-1, viz., C1-C3 and C45. Then the proof of Theorem 1 establishes the desired rates of convergence for part (c).

By Theorem A-3, condition C45 holds. To get C2, let $\{\epsilon_i^* : i = \ldots 0, -1, 0, \ldots\}$ be an iid sequence of real-valued rv's such that ϵ_i^* has the same distribution as Ξ (specified in D2), for all i . Note that

$$(A.29) \qquad |X_{ik}| \leq \|X_i\| = \left\|\sum_{j=0}^{\infty} A_{ij}\epsilon_{i-j}\right\| \leq \sum_{j=0}^{\infty}\|A_{ij}\|\cdot\|\epsilon_{i-j}\| \overset{ST}{\leq} \sum_{j=0}^{\infty} M_j\cdot\epsilon_{i-j}^* .$$

Suppose $r \leq 1$. Then, for all s in $(0,r)$, the right-hand-side of (A.29) is in L^s , since for $s \epsilon (b,r)$ (where b is as in D4) we have

$$(A.30) \qquad E\left[\sum_{j=0}^{\infty} M_j\epsilon_{i-j}^*\right]^s \leq \sum_{j=0}^{\infty}(M_j)^s E(\epsilon_{i-j}^*)^s = \sum_{j=0}^{\infty}(M_j)^s\cdot E\Xi^s < \infty ,$$

which establishes C2 (since it is clear that under D2 the maximal moment exponent of $\sum_{j=0}^{\infty} M_j \varepsilon^*_{i-j}$ cannot exceed r). Now suppose $r > 1$. Let $\|\cdot\|_s$ denote the L^s norm of a rv. By Minkowski's inequality and the monotone convergence theorem, we get

$$\text{(A.31)} \quad \left\| \sum_{j=0}^{\infty} M_j \varepsilon^*_{i-j} \right\|_s = \left[E\left(\sum_{j=0}^{\infty} M_j \varepsilon^*_{i-j} \right)^s \right]^{1/s} \leq \sum_{j=0}^{\infty} \|M_j \varepsilon^*_{i-j}\|_s = \left(\sum_{j=0}^{\infty} M_j \right) \cdot \|\varepsilon^*_i\|_s < \infty,$$

for all s in $(1,r)$, using D2 and D4. Hence, C2 holds.

To show C3, take d less than, but arbitrarily close to, $1 \wedge p \wedge r$. We have

$$\text{(A.32)} \qquad E\|X_{ik}U_i\|^d \leq E \sum_{j=0}^{\infty} \|A_{ij}\varepsilon_{i-j}U_i\|^d \leq \sum_{j=0}^{\infty} (M_j)^d E\|\varepsilon_{i-j}U_i\|^d < \infty ,$$

by the monotone convergence theorem, and assumptions D4 and D6. *Q.E.D.*

PROOF OF THEOREM A-3: It suffices to consider the case $\eta = -2/\alpha + \bar{\delta}$, for some $\bar{\delta} > 0$ arbitrarily small. Write

$$\text{(A.33)} \quad \begin{aligned} n^{\eta}\Sigma_1^n X_i X_i' &= n^{\eta} \sum_{i=1}^{n} \left[\sum_{\substack{j=0 \\ j\neq v}}^{\infty} \sum_{v=0}^{\infty} A_{ij}\varepsilon_{i-j}\varepsilon'_{i-v}A'_{iv} \right] + \sum_{j=0}^{\infty} \left[n^{\eta} \sum_{i=1}^{n} A_{ij}\varepsilon_{i-j}\varepsilon'_{i-j}A'_{ij} \right] \\ &= n^{\eta} \sum_{i=1}^{n} C_{1i} + \sum_{j=0}^{\infty} C_{2jn} . \end{aligned}$$

We want to show (i) $\left\| n^{\eta} \sum_{i=1}^{n} C_{1i} \right\| \xrightarrow{n\to\infty} 0$ a.s., and (ii) $\underline{\lim}_{n\to\infty} \lambda_{\min}(C_{2jn}) > 0$ a.s. for some j . This proves the Theorem, since $C_{2jn} \geq 0$, $\forall j$, $\forall n$.

To show (ii), let $W_{ij} = A_{ij}\varepsilon_{i-j}$. We claim that for the index j of

D5, $\{W_{ij} : i = 1, 2, \ldots\}$ satisfies the conditions C4 and C5 on $\{X_i\}$ for Theorem A-2. Thus, Theorem A-2 gives

$$\text{(A.34)} \qquad \underline{\lim}_{n\to\infty} \lambda_{\min}(C_{2jn}) = \underline{\lim}_{n\to\infty} n^{\eta}\lambda_{\min}\left(\sum_{i=1}^{n} W_{ij}W_{ij}'\right) > 0 \quad \text{a.s.},$$

as desired. Condition C4 follows immediately from D1. To show C5, consider any $\zeta \in R^K$ with $\|\zeta\| = 1$. Let $\eta_{ij} = A_{ij}'\zeta \in R^D$ and $\tilde{\zeta} = \eta_{ij}/\|\eta_{ij}\|$. We have

$$\text{(A.35)} \qquad \begin{aligned} |\zeta' W_{ij}| &= |\zeta' A_{ij}\varepsilon_{i-j}| = \|\eta_{ij}\|\cdot|(\eta_{ij}/\|\eta_{ij}\|)'\varepsilon_{i-j}| \\ &\geq \delta^{1/2}\cdot|\tilde{\zeta}'\varepsilon_{i-j}| \geq \delta^{1/2}\Xi_{\tilde{\zeta}} \in DA(\alpha) \ , \end{aligned}$$

using D3, since $\|\eta_{ij}\|^2 = \zeta' A_{ij}A_{ij}'\zeta \geq \lambda_{\min}(A_{ij}A_{ij}') \geq \delta > 0$, by D5. This establishes C5.

To show (i), consider a single element of the $K \times K$ matrix C_{1i}. For notational simplicity we omit the extra subscripts. We have

$$\text{(A.36)} \qquad |C_{1i}| \leq \sum_{j=0}^{\infty} \sum_{\substack{v=0 \\ j\neq v}}^{\infty} \|A_{ij}\|\cdot\|\varepsilon_{i-j}\|\cdot\|\varepsilon_{i-v}\|\cdot\|A_{iv}\| \overset{ST}{\leq} \sum_{j=0}^{\infty} \sum_{\substack{v=0 \\ j\neq v}}^{\infty} M_j M_v \varepsilon_{i-j}^* \varepsilon_{i-v}^* \ ,$$

where $\{\varepsilon_i^*\}$ is as in the proof of Theorem 3, and where the stochastic inequality follows easily by a succession of conditioning arguments, using the independence of $\{\varepsilon_i^*\}$.

The rv on the right-hand-side of (A.36) is identically distributed for $i = 1, 2, \ldots$. Also, if $r \leq 1$, then it is in L^q for all $q < r$, since for $q \in (b,r)$ (where b is as in D4) we have

$$\text{(A.37)}\quad \begin{aligned} E\left[\sum_{j=0}^{\infty}\sum_{\substack{v=0\\ j\neq v}}^{\infty} M_j M_v \varepsilon^*_{i-j}\varepsilon^*_{i-v}\right]^q &\le \sum_{j=0}^{\infty}\sum_{\substack{v=0\\ j\neq v}}^{\infty}(M_j)^q(M_v)^q E(\varepsilon^*_{i-j})^q E(\varepsilon^*_{i-v})^q \\ &\le \left[\sum_{j=0}^{\infty}(M_j)^q\right]^2 [E\Xi^q]^2 < \infty , \end{aligned}$$

by D2 and D4. Then, by taking q arbitrarily close to r, Lemma A-1 and the assumption $\alpha < 2r$ give the desired result $n^{\eta}\sum_{i=1}^{\infty} C_{1i} \xrightarrow{n\to\infty} 0$ a.s., for $\eta = -2/\alpha + \tilde{\delta}$, where $\tilde{\delta} > 0$ is arbitrarily small (for the case $r \le 1$).

Next, suppose $r \ge 1$. By D4 and the assumption $\alpha < 2$ in D3, there exists a constant $d < 1$ such that $\sum_{j=0}^{\infty}(M_j)^d < \infty$ and $d > \alpha/2$. For $r \ge 1$, the rv on the right-hand side of (A.36) is in L^d, since

$$\text{(A.38)}\quad \begin{aligned} E\left[\sum_{j=0}^{\infty}\sum_{\substack{v=0\\ j\neq v}}^{\infty} M_j M_v \varepsilon^*_{i-j}\varepsilon^*_{i-v}\right]^d &\le \sum_{j=0}^{\infty}\sum_{\substack{v=0\\ j\neq v}}^{\infty}(M_j)^d(M_v)^d E(\varepsilon^*_{i-j})^d E(\varepsilon^*_{i-v})^d \\ &\le \left[\sum_{j=0}^{\infty}(M_j)^d\right]^2 [E\Xi^d]^2 < \infty . \end{aligned}$$

Hence, Lemma A-1 gives the result $n^{\eta}\sum_{i=1}^{\infty} C_{1i} \xrightarrow{n\to\infty} 0$ a.s. (since $d > \alpha/2$ implies $-1/d > -2/\alpha$, and so, $-1/d > \eta = -2/\alpha + \tilde{\delta}$, for $\tilde{\delta} > 0$ arbitrarily small). Thus (i) is established, and the proof is complete. *Q.E.D.*

This completes the proof of Theorem 3 and its related result Theorem A-3. We now prove Theorem 4. To do this, we first give an analogue of Theorem A-1:

THEOREM A-4: *Under assumptions* C1-C3, C45, *and* C6-C9, *the* LS *slope estimators* $\tilde{\beta}$ *of* β_0 *for the model* (9) *satisfy*

$$n^{\xi}(\tilde{\beta} - \beta_0) \xrightarrow{n\to\infty} \underset{\sim}{0} \quad \text{a.s.}, \quad \forall \xi < \nu(p,r) - \gamma .$$

PROOF OF THEOREM 4: Theorems A-2 and A-4, and the proof of Theorem 1 combine to give the desired results under assumptions C1-C9. Theorems A-3 and A-4, and the proof of Theorem 3 do likewise under assumptions C1, D1-D6, and C6-C9. *Q.E.D.*

PROOF OF THEOREM A-4: Define $M_n^* = \Sigma_1^n X_i U_i - \Sigma_1^n X_i Z_i'(\Sigma_1^n Z_i Z_i')^{-}\Sigma_1^n Z_i U_i$, and $Q_n^* = \Sigma_1^n X_i X_i' - \Sigma_1^n X_i Z_i'(\Sigma_1^n Z_i Z_i')^{-}\Sigma_1^n Z_i X_i'$. It suffices to show that Lemmas A-2, A-3, and A-4 hold with M_n and Q_n replaced by M_n^* and Q_n^* .

The proof of Lemma A-2 goes through with M_n^* and Q_n^* in exactly the same way as with M_n and Q_n .

To show Lemma A-3, it suffices to consider the case where $\tau = \gamma$. By an argument analogous to that of (A.5), it suffices to show

$$\text{(A.39)} \qquad H_n = n^{\gamma-1}\Sigma_1^n X_i Z_i'\left(\frac{1}{n}\Sigma_1^n Z_i Z_i'\right)^{-}\Sigma_1^n Z_i X_i' \xrightarrow{n\to\infty} 0 \quad \text{a.s.}$$

Using C9, Lemma A-1 gives $n^{g}\Sigma_1^n X_i Z_i' \xrightarrow{n\to\infty} 0$ a.s., $\forall g < (-1/r) \wedge (-1)$. And by C7 and C8, $\frac{1}{n}\Sigma_1^n Z_i Z_i'$ has smallest eigenvalue bounded away from zero for n sufficiently large a.s. Thus, $\forall \delta > 0$, $n^{-\delta}\left(\frac{1}{n}\Sigma_1^n Z_i Z_i'\right)^{-} \xrightarrow{n\to\infty} 0$ a.s. Hence, $H_n \xrightarrow{n\to\infty} 0$ a.s., if $\gamma - 1 - 2g + \delta \le 0$. This holds for δ sufficiently small and g sufficiently close to $(-1/r) \wedge (-1)$, since

$$\gamma \le \gamma(p,r) - 2\delta \le [(1-2/r)\wedge(-1)] - 2\delta = 1+2[(-1/r)\wedge(-1)] - 2\delta \le 1 + 2g - \delta .$$

Next we establish Lemma A-4 with M_n^* in place of M_n . By the same argument as in the proof of Lemma A-4, $n^{\nu}\Sigma_1^n X_i U_i \xrightarrow{n\to\infty} 0$ a.s. By Lemma A-1,

(A.40) $$n^w \Sigma_1^n Z_i U_i \xrightarrow{n\to\infty} \underset{\sim}{0} \text{ a.s.}, \quad \forall w < -1/p^* , \quad \text{where} \quad p^* = p \wedge 1 .$$

Hence, $n^{\nu} M_n^* \xrightarrow{n\to\infty} \underset{\sim}{0}$ a.s., provided $\nu - 1 - g - w + \delta \le 0$. The latter holds for δ sufficiently small, g sufficiently close to $(-1/r) \wedge (-1)$, and w sufficiently close to $-1/p$, because

$$\begin{aligned}\nu &\le \nu(p,r) - 2\delta \le [(1 - 1/p^* - 1/r) \wedge (-1/p^*)] - 2\delta \\ &= 1 + [(-1/r) \wedge (-1)] - 1/p^* - 2\delta \le 1 + g + w - \delta .\end{aligned}$$

Q.E.D.

FOOTNOTES

1. I would like to thank Chris Monahan for excellent research assistance, and Roger Koenker, Stephan Morganthaler, Charles Manski, and Jim Powell for helpful comments. I am indebted to the Sloan Foundation and the National Science Foundation for research support through a Research Fellowship and grant number SES-8419789, respectively.

2. In addition, the results of Kanter and Steiger (1974), Chen, Lai, and Wei (1981), and Cline (1983) have not been recognized in the robustness literature. There are no robustness papers in the literature, that I am aware of, that make reference to them.

3. Stable distributions are the only possible limit distributions of normalized sums of iid rv's. The class of stable distributions is a four parameter family with parameters for location, scale, skewness, and kurtosis. Expressions for stable characteristic functions are well-known, e.g., see Feller (1971, p. 570). Their densities exist in closed form only for special cases. The <u>characteristic exponent</u> of a stable rv, usually denoted α, is the parameter that indicates kurtosis. It lies in the interval $(0,2]$. If the characteristic exponent is less than two, then it equals the maximal moment exponent of the stable rv. Examples of symmetric stable distributions are the normal $(\alpha = 2)$ and the Cauchy $(\alpha = 1)$. See Feller (1971, pp. 173-176) for additional examples. The normal distribution is the only stable distribution with finite variance.

4. Necessary and sufficient conditions for a distribution F to be in the domain of attraction of a stable distribution with characteristic exponent α are well-known, e.g., see Feller (1971, Theorem IX.8.1a, p. 313, or Corollary XVII.5.2, p. 578). For the case $\alpha \in (0,2)$, we have $X \in DA(\alpha)$ iff the tail sum, $1 - F(x) + F(-x)$, is regularly varying with exponent $-\alpha$ (where $F = \mathcal{L}(X)$), and the tails are balanced. That is,

$$1 - F(x) + F(-x) = x^{-\alpha}L(x) \ , \quad \forall x \in R \ ,$$

where $L(x)$ is some function that is slowly varying at infinity, and

$$\frac{1 - F(x)}{F(-x)} \xrightarrow{x\to\infty} Q \ , \quad \text{for some constant } Q \in [0,\infty] \ .$$

A function $L(x)$ is <u>slowly varying at infinity</u> if

$$L(sx)/L(s) \xrightarrow{s\to\infty} 1 \ , \quad \forall x > 0 \ .$$

Examples of slowly varying functions include: functions that converge to a positive limit, and all powers of $\log x$. If $L(x)$ is slowly varying, then for all $\varepsilon > 0$ and all x sufficiently large,

$$x^{-\varepsilon} < L(x) < x^{\varepsilon} \ .$$

5. For present purposes we adopt the definition of Gine and Hahn (1983), and say that a distribution F on R^K is <u>multivariate stable</u> if: For all $a, b \in R$, there exists $g \in R$ and $h \in R^K$ such that

$$(*) \qquad L(a\underset{\sim}{W}_1 + b\underset{\sim}{W}_2) = L(gW + h) ,$$

where W_1, W_2, and W are independent random K-vectors with distribution F. A multivariate stable distribution has a characteristic exponent α, and each of its marginal distributions is stable with characteristic exponent α. (Conversely, if a multivariate distribution has all marginal distributions are infinitely divisible, then it is multivariate stable with characteristic exponent α, see Gine and Hahn (1983, Theorem 1).)

A distribution F on R^K (or a random K-vector) is said to be in the <u>domain of attraction</u> of a multivariate stable law S if there exist real-valued constants $\{c_n\}$ and R^K-valued constants $\{b_n\}$ such that

$$c_n^{-1} \sum_{i=1}^{n} W_i - b_n \xrightarrow{d} W_0 \quad \text{as} \quad n \to \infty ,$$

where W_i are iid with $L(W_i) = F$, and $L(W_0) = S$.

A distribution of R^K is <u>full</u> if its characteristic function equals one only when evaluated at the origin. Thus, a full distribution is not concentrated on any lower dimensional linear subspace than the dimension of the distribution.

The above definition of multivariate stable distributions can be generalized considerably. One can consider multivariate distributions called G-<u>stable</u> distributions, see Schmidt (1975) and Jurek (1982); also see Resnick and Greenwood (1979), and DeHaan, Omey, and Resnick (1984). Such distributions must satisfy (*) with a, b, and g given by $K \times K$ matrices in some group of matrices G. By considering G-stable distributions and their domains of attraction, one can generate endless additional examples of regressor sequences that satisfy C5.

6. The variance of the estimators is not a very informative measure in this context, because the LS estimator has infinite variance when $\alpha = 1.5$, 1.0, or 0.5, due to thick tails, even though its distribution may be tightly concentrated about the true value.

7. For $\alpha = 1.5$, Theorem 1 does not yield a sharp result, and hence, does not provide a rate of decline. For the present simple iid set-up with independent regressor and error, however, the supremum rate of convergence of the LS estimator can be shown to be $n^{-2/3}$.

REFERENCES

[1] Anderson, T. W. and J. B. Taylor: "Strong Consistency of Least Squares Estimates in Dynamic Models," *Annals of Statistics*, 7 (1979), 484-489.

[2] Andrews, D. W. K.: "Robust and Efficient Estimation of Nonlinear Regression Models with Dependent Errors," manuscript, Cowles Foundation, Yale University, 1983.

[3] __________: "A Zero-one Result for the Least Squares Estimator," *Econometric Theory*, 1 (1985a), 85-96.

[4] __________: "Consistency of Least Squares Estimators in Dynamic Regression with Infinite Variances," manuscript, Cowles Foundation, Yale University, 1985b.

[5] __________: "Time Series Regression with Integrated Regressors and Explosive, Non-Mixing Errors," manuscript, Cowles Foundation, Yale University, 1985c.

[6] __________: "A Note on the Unbiasedness of Feasible GLS, Quasi-Maximum Likelihood, Robust, Adaptive, and Spectral Estimators of the Linear Model," *Econometrica*, 54 (1986a), forthcoming.

[7] __________: "Stability Comparisons of Estimators," *Econometrica*, 54 (1986b), forthcoming.

[8] Beasley, J. D. and S. G. Springer: "The Percentage Points of the Normal Distribution," *Applied Statistics (Journal of the Royal Statistical Society C)*, 26 (1977), 118-121.

[9] Bierens, H. J.: *Robust Methods and Asymptotic Theory in Nonlinear Econometrics*. Lecture Notes in Economics and Mathematical Systems No. 192. New York: Springer-Verlag, 1981.

[10] Chatterji, S. D.: "An L^p-Convergence Theorem," *Annals of Mathematical Statistics*, 40 (1969), 1068-1070.

[11] Chen, G. J., T. L. Lai, and C. Z. Wei: "Convergence Systems and Strong Consistency of Least Squares Estimators in Regression Models," *Journal of Multivariate Analysis*, 11 (1981), 319-333.

[12] Christopeit, N. and K. Helms: "Strong Consistency of Least Squares Estimators in Linear Regression Models," *Annals of Statistics*, 8 (1980), 778-788.

[13] Cline, D. B. H.: "Estimation and Linear Prediction for Regression, Autoregression, and ARMA with Infinite Variances," Ph.D Thesis, Department of Statistics, Colorado State University, Boulder, Colorado, 1983.

[14] DeHaan, L.: "On Regular Variation and its Applications to the Weak Convergence of Sample Extremes," *Math. Centre Tracts*, 32. Amsterdam: Mathematisch Centrum, 1970.

[15] DeHaan, L., E. Omey, and S. Resnick: "Domains of Attraction and Regular Variation in R^d," *Journal of Multivariate Analysis*, 14 (1984), 17-33.

[16] Derman, C. and H. Robbins: "The Strong Law of Large Numbers when the First Moment Does Not Exist," *Proceedings of the National Academy of Sciences USA*, 41 (1955), 586-587.

[17] Feller, W.: *An Introduction to Probability Theory and its Applications*, Volume II. New York: Wiley, 1971.

[18] Galambos, J.: *The Asymptotic Theory of Extreme Order Statistics*. New York: Wiley, 1978.

[19] Gine, E. and M. G. Hahn: "On Stability of Probability Laws with Univariate Stable Marginals," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 64 (1983), 157-165.

[20] Gnedenko, B. V. and A. N. Kolmogorov: *Limit Distributions for Sums of Independent Random Variables*. Cambridge, MA: Addison-Wesley, 1954.

[21] Huber, P. J.: "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics*, 35 (1964), 73-101.

[22] __________: "Robust Regression: Asymptotics, Conjectures, and Monte Carlo," *Annals of Statistics*, 1 (1973), 799-821.

[23] __________: "Minimax Aspects of Bounded-Influence Regression," *Journal of the American Statistical Association*, 78 (1983), 66-72.

[24] Judge, G. G., W. E. Griffiths, R. C. Hill, and T.-C. Lee: *The Theory and Practice of Econometrics*. New York: Wiley, 1980.

[25] Jurek, Z. J.: "Domains of Normal Attraction for *G*-Stable Measures on R^d," *Theory of Probability and its Applications*, 27 (1982), 419-424.

[26] Kanter, M. and W. L. Steiger: "Regression and Autoregression with Infinite Variance," *Advances in Applied Probability*, 6 (1974), 768-783.

[27] Koenker, R.: "Robust Methods in Econometrics," *Econometric Reviews*, 1 (1982), 213-255.

[28] Krasker, W. S. and R. E. Welsch: "Efficient Bounded-Influence Regression Estimation," *Journal of the American Statistical Association*, 77 (1982), 595-604.

[29] Lai, T. L. and C. Z. Wei: "Asymptotic Properties of Projections with Applications to Stochastic Regression Problems," *Journal of Multivariate Analysis*, 12 (1982a), 346-370.

[30] __________: "Least Squares Estimates in Stochastic Regression Models with Applications to Identification and Control of Dynamic Systems," *Annals of Statistics*, 10 (1982b), 154-166.

[31] Loeve, M.: *Probability Theory*, 1st Edition. Princeton, NJ: Van Nostrand, 1955.

[32] Maronna, R. A. and V. J. Yohai: "Asymptotic Behaviour of General M-estimates for Regression and Scale with Random Carriers," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 58 (1981), 7-20.

[33] Miller, H. D.: "A Note on Sums of Independent Random Variables with Infinite First Moment," *Annals of Mathematical Statistics*, 38 (1967), 751-758.

[34] Resnick, S. I. and P. Greenwood: "A Bivariate Stable Characterization and Domains of Attraction," *Journal of Multivariate Analysis*, 9 (1979), 206-221.

[35] Schmidt, K.: "Stable Probability Measures on R^{ν}," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 33 (1975), 19-31.

[36] Wichmann, B. A. and I. D. Hill: "An Efficient and Portable Pseudo-random Number Generator," *Applied Statistics (Journal of the Royal Statistical Society C)*, 31 (1982), 188-190.

[37] Wu, C.-F.: "Characterizing the Consistent Directions of Least Squares Estimates," *Annals of Statistics*, 8 (1980), 789-801.

TABLE I

SELECTED PERCENTILES OF THE LEAST SQUARES ESTIMATOR OF β_0 IN THE SIMPLE LINEAR REGRESSION MODEL WITH SYMMETRIC STABLE REGRESSORS AND ERRORS BOTH WITH CHARACTERISTIC EXPONENT α

Sample Size	Characteristic Exponent	Percentiles of Least Squares Estimator[a,b]					
n	α	75%		85%		95%	
25	2.0	.14	(1.00)	.22	(1.00)	.35	(1.00)
	1.5	.12	(.86)	.22	(1.00)	.47	(1.34)
	1.0	.083	(.59)	.21	(.95)	.81	(2.31)
	0.5	.048	(.34)	.28	(1.27)	4.71	(13.46)
50	2.0	.096	(1.00)	.15	(1.00)	.24	(1.00)
	1.5	.080	(.83)	.15	(1.00)	.32	(1.33)
	1.0	.046	(.48)	.12	(.80)	.46	(1.92)
	0.5	.022	(.23)	.13	(.87)	1.93	(8.04)
100	2.0	.068	(1.00)	.10	(1.00)	.17	(1.00)
	1.5	.054	(.79)	.097	(.97)	.21	(1.24)
	1.0	.026	(.38)	.068	(.68)	.27	(1.59)
	0.5	.010	(.15)	.058	(.58)	.95	(5.59)
200	2.0	.048	(1.00)	.072	(1.00)	.12	(1.00)
	1.5	.036	(.75)	.065	(.90)	.14	(1.17)
	1.0	.015	(.31)	.039	(.54)	.16	(1.33)
	0.5	.0047	(.10)	.028	(.39)	.44	(3.67)

[a]The numbers in parentheses are the ratios of the percentile for the α value in question to the percentile for $\alpha = 2$ (i.e., the normal distribution).

[b]These are Monte Carlo results based on 50,000 repetitions using the stable random number algorithm of Kanter and Steiger (1974, pp. 769-770), the uniform [0,1] generator of Wichmann and Hill (1982), the inverse normal algorithm of Beasley and Springer (1977), and the ROBETH regression package.

TABLE II

SELECTED PERCENTILES OF THE HUBER M-ESTIMATOR OF β_0 IN THE SIMPLE LINEAR REGRESSION MODEL WITH SYMMETRIC STABLE REGRESSORS AND ERRORS BOTH WITH CHARACTERISTIC EXPONENT α

Sample Size	Characteristic Exponent	Percentiles of Huber M-Estimator[a,b]					
n	α	75%		85%		95%	
25	2.0	.14	(1.00)	.22	(1.00)	.36	(1.03)
	1.5	.10	(.83)	.18	(.82)	.35	(.74)
	1.0	.059	(.71)	.14	(.67)	.46	(.57)
	0.5	.022	(.46)	.13	(.46)	1.87	(.40)
50	2.0	.10	(1.04)	.15	(1.00)	.24	(1.00)
	1.5	.063	(.79)	.11	(.73)	.21	(.66)
	1.0	.030	(.65)	.072	(.60)	.23	(.50)
	0.5	.0078	(.35)	.044	(.34)	.58	(.30)
100	2.0	.070	(1.03)	.11	(1.10)	.17	(1.00)
	1.5	.040	(.74)	.069	(.71)	.13	(.62)
	1.0	.016	(.62)	.038	(.56)	.12	(.44)
	0.5	.0027	(.27)	.015	(.26)	.21	(.22)
200	2.0	.049	(1.02)	.075	(1.04)	.12	(1.00)
	1.5	.026	(.72)	.044	(.68)	.083	(.59)
	1.0	.0083	(.55)	.021	(.54)	.064	(.40)
	0.5	.00096	(.20)	.0055	(.20)	.075	(.17)

[a] The numbers in parentheses are the ratios of the Huber estimator percentiles to the least squares estimator percentiles, both with the same α value.

[b] Same as footnote b of Table I.

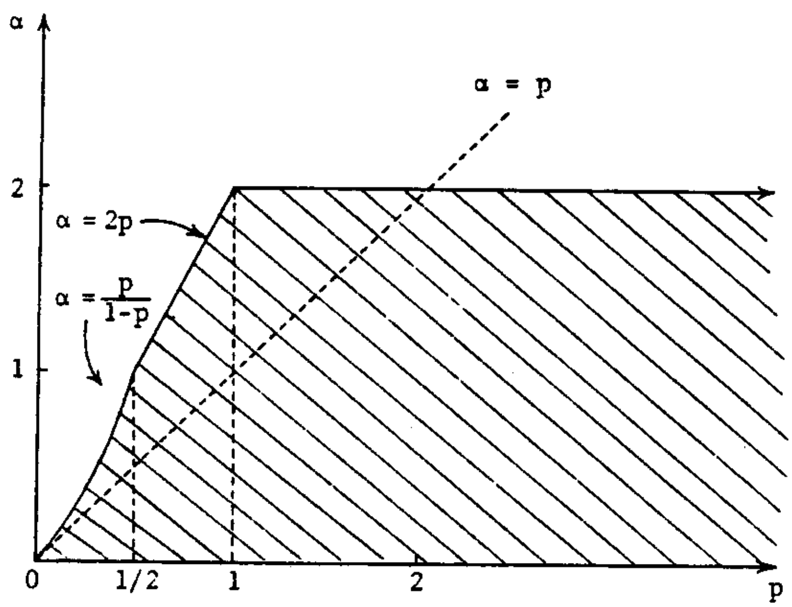

FIGURE 1. Region of Consistency of Least Squares for Case $r = \alpha$ in Theorems 1, 3, and 4.

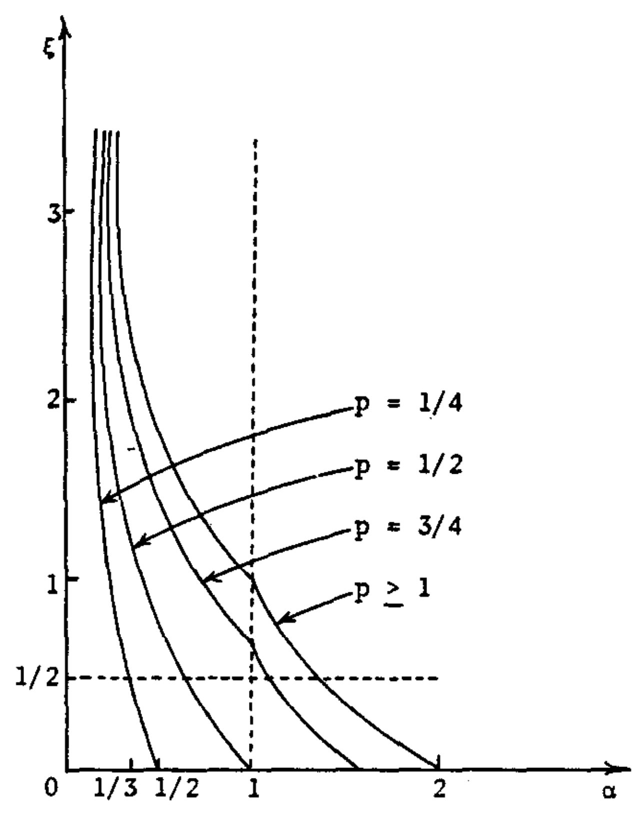

FIGURE 2. Rates of Convergence $n^{-\xi}$ of Least Squares for Case $r = \alpha$ in Theorems 1, 3, and 4.

FIG3

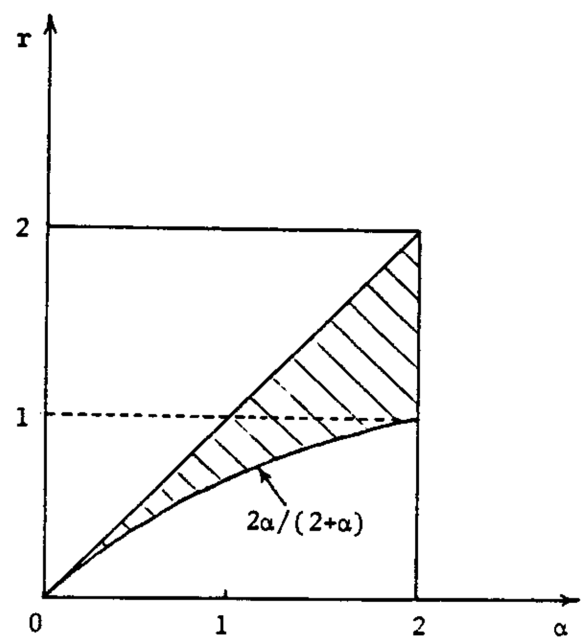

FIGURE 3. Admissable [r,α] Region for Theorems 1, 3, and 4.