COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station

New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 710-R

(Revision no. 1)

STABILITY COMPARISONS OF ESTIMATORS

Donald W.K. Andrews

Revised: May 1985

STABILITY COMPARISONS OF ESTIMATORS

by Donald W. K. Andrews[1]

HEADNOTE

This paper investigates a property of estimators called stability. The stability exponent of an estimator is defined to be a measure of the effect of any single observation in the sample on the realized value of the estimator. High stability is often desirable for robustness against misspecification and against highly variable observations.

Stability exponents are determined and compared for a wide variety of estimators and econometric models. They are found to depend on the maximal moment exponent (i.e., the number of finite moments) of the estimator's influence curve. Since it is possible often to construct estimators with specified influence curves, estimators with different stability exponents can be constructed.

1. INTRODUCTION AND CONCLUSION

This paper investigates a property of estimators called <u>stability</u>. The stability exponent of an estimator is a measure of the magnitude of the effect of any single observation in the sample on the realized value of the estimator. A number of reasons related to robustness suggest that often it is desirable for an estimator to be relatively insensitive to any particular observation in the sample, i.e., to have high stability. In addition, it is useful for diagnostic purposes to have knowledge of the stability exponents of different estimators, in order to know which estimators are likely to rely more heavily on some single observation.

The paper is organized as follows: Section 1 introduces the basic idea contained in the paper, attempts to motivate it, and summarizes the results in an informal manner. Section 2 presents definitions, assumptions, and the general results. For purposes of illustration, the linear regression model with the least squares estimator is used as a running example throughout this section. Section 3 discusses numerous additional applications of the general results. An Appendix contains proofs of the results given in Section 2.

In words, the stability exponent of an estimator is the greatest normalization factor such that the normalized deviation of the estimator, due to the deletion of a single observation, converges to zero with probability one as the sample size goes to infinity, for any sequence of deletions. More specifically, we make the following definition.

1

DEFINITION: The _stability exponent_ of an estimator $\hat{\theta} \equiv \{\hat{\theta}_n : n = 1, 2, \ldots\}$ of some R^J-valued parameter θ is defined to be

$$(1.1) \quad \Lambda(\hat{\theta}, P) = \sup\{\xi \in R : n^{\xi}(\hat{\theta}_n - \hat{\theta}_{n,k_n}) \xrightarrow{n \to \infty} \underset{\sim}{0} \text{ a.s. } [P], \forall\{k_n\}\},$$

where $\hat{\theta}_{n,k_n}$ is the estimator applied to the sample of size n with the k_n^{th} observation deleted, $\underset{\sim}{0}$ is a J-vector of zeros, a.s. abbreviates almost surely, P is the underlying probability distribution generating the data, and $\{k_n\} \equiv \{k_n : k_n \leq n, n = 1, 2, \ldots\}$ is any fixed sequence of indices of deleted observations, one for each sample size.

Thus, the stability exponent of an estimator is an asymptotic measure of the sensitivity of the estimator to observations actually in the sample[2] (rather than to non-random hypothetical observations, as is measured by the influence curve, see Hampel [25]). Under fairly general conditions, stability exponents lie between zero and one, with the extreme values being attained by certain estimators. The results of this paper concern the determination of the stability exponents of estimators in a fairly broad class, and for an extensive array of different econometric models. Among others, models for which the results apply include: linear and nonlinear regression (with fixed or random regressors), linear and nonlinear simultaneous equations, panel data, and limited dependent variable (such as logit, probit, truncated and censored regression, and self-selection).

The class of estimators considered in this paper is defined to include all estimators that can be written as solutions (for θ) to a system of equations:

$$(1.2) \qquad \sum_{i=1}^{n} r_i(Z_i, \theta) = \underline{0} \, ,$$

where $r_i(\cdot, \cdot)$ is a specified function that defines the estimator, and Z_i is a random vector of observed variables comprising the ith observation (see Huber [30]). Note that Z_i may include variables in Z_ℓ for $\ell < i$. For example, in time series regression and simultaneous equations models, Z_i may include lagged variables. The number of estimators that can be written in the form (1.2) is quite large. For example, the following estimators are included: least squares, maximum likelihood (including full-information (FIML) and limited information (LIML) estimators of simultaneous equations models), instrumental variables, M-, and various multi-stage estimators such as Zellner's [58] seemingly unrelated regressions estimator, Heckman's [27] estimator of censored regression and self-selection models, two stage least squares (2SLS), and three stage least squares (3SLS). These examples are discussed below in Section 3.

Under suitable regularity conditions (outlined below), it is possible to write estimators in the class defined above in a linearized form:

$$(1.3) \qquad \hat{\theta}_n = \theta_0 - I_n \frac{1}{n} \sum_{i=1}^{n} A^{-1} r_i(Z_i, \theta_0) \, ,$$

where θ_0 is the estimand, I_n is a $J \times J$ random matrix equal to the identity matrix plus a matrix of small order one as $n \to \infty$ a.s., and A is a $J \times J$ non-random non-singular matrix. If $r_i(\cdot, \cdot)$ is independent of i for i sufficiently large, then $A^{-1} r(z, \theta_0)$ is the influence curve of $\hat{\theta}_n$ evaluated at z, as defined by Hampel [25].

It is shown that the stability exponent of $\hat{\theta}_n$ is directly related to the maximal moment exponent (i.e., the number of finite moments) of

$r_i(Z_i, \theta_0)$, $i = 1, 2, \ldots$. In particular, if r_U and r_L are sto-chastically greater than or equal to, and less than or equal to, $|r_i(Z_i, \theta_0)|$ for all $i = 1, 2, \ldots$, respectively, then the stability exponent of $\hat{\theta}_n$ lies in the interval $[1 - 1/p, 1 - 1/q]$, where r_U and r_L have maximal moment exponents equal to p and q , respectively. If p equals q , the stability exponent of $\hat{\theta}_n$ is established. Otherwise, the stability exponent of $\hat{\theta}_n$ is given by a more complicated expression involving the tail probabilities of the random vectors $r_i(Z_i, \theta_0)$, $i = 1, 2, \ldots$.

Thus, the qualitative result is obtained that the stability exponent of an estimator depends on the maximal moment exponent of its linearized form (or influence curve)--the greater the maximal moment exponent the greater the stability exponent. Further, there is no upper bound beyond which additional moments no longer increase the stability exponent of the estimator. Since $r_i(\cdot,\cdot)$ is chosen by the investigator, it is often straightforward to obtain estimators with a specified linearized form (e.g., see Krasker and Welsch [37, 38] and Stefanski, Ruppert, and Carroll [48]). Hence, estimators with different stability exponents can be constructed.

It should be noted that stability results depend on the maximal moment exponent of the linearized estimator, not on the maximal moment exponent of the estimator itself. The latter has received considerable attention in the econometrics literature, e.g., see Kinal [35], since common estimators of simultaneous equations models have fewer than all moments finite even with normal errors. These results have no clear implications for stability since they deal with moments of the estimator rather than moments of the linearized form.

The examples of Section 3 provide a variety of models, estimators, and stability characteristics of these estimators. We briefly summarize

the results here: In the linear regression model with fixed regressors, the least squares (LS) estimator has stability exponent that depends on the maximal moment exponent of the errors. On the other hand, Huber [31] M-estimators have the maximum stability exponent of one in this model, regardless of the distribution of the errors. In the linear regression model with random regressors, the LS estimator has stability exponent that depends on the maximal moment exponent of the errors and the regressors, whichever is smaller. In contrast, Krasker and Welsch's [37] bounded influence regression estimator has stability exponent equal to one for all error and regressor distributions. Results for the LS estimator and M-estimators in the nonlinear regression model parallel those in the linear model, except the dependence on the maximal moment exponent of the regressors, when applicable, is replaced by that of the derivative of the regression function (with respect to the parameter vector) evaluated at the true parameter.

The instrumental variables (IV) estimator of a single equation from a system of linear equations has stability exponent that depends on the maximal moment exponent of the errors and the instruments. In comparison, Krasker and Welsch's [38] weighted instrumental variables (WIV) estimator for this model has a bounded influence function, and hence, has stability exponent equal to one--the maximum--regardless of the distribution of the errors and instruments.

The stability exponents of maximum likelihood (ML) and pseudo-ML estimators depend on the maximal moment exponents of their score functions. In logit and probit models, this corresponds to the maximal moment exponent of the regressors. In the censored regression model, it corresponds to the maximal moment exponent of the errors and regressors. Heckman's [27]

two-stage estimator of this model has the same stability properties as the ML estimator. Similarly, the ML estimator and Zellner's [58] feasible Aitken estimator for the seemingly unrelated nonlinear regressions model have the same stability properties. Their stability exponents depend on the maximal moment exponent of the errors and the derivatives of the regression functions (with respect to the parameter vector) evaluated at the true parameter. Following the examples of Section 3, the calculation of stability exponents of other estimators for other models is straightforward.

Clearly, if $r_i(Z_i, \theta_0)$, $i = 1, 2, \ldots$ are uniformly bounded, then all of their moments exist and the maximum possible value for the stability exponent is attained regardless of the true distribution of the data. Bounded influence estimators, referred to above, are characterized by this property. In contrast, other estimators have stability exponents that depend on the true underlying probability distribution, since the true distribution determines the maximal moment exponent of $r_i(Z_i, \theta_0)$, $i = 1, 2, \ldots$. This is illustrated by the examples of Section 3.

For reasons discussed below, high stability often is a desirable property of estimators. Hence, it may be of interest to determine whether an estimator has high stability for a given problem. Two factors are pertinent here. First, the stability exponent of an estimator generally depends on the true distribution of the data, and second, in practice this true distribution (or a parametric family containing it) is never known precisely. Thus, one can be certain that an estimator has high stability in the context at hand, only if it has a high stability exponent for all distributions close to the postulated true distribution or true parametric family of distributions. This leads to the following definition:

DEFINITION: An estimator is <u>stability-robust</u> at a distribution P (with respect to some given topology) if the infimum of its stability exponent over some neighborhood of P is positive. Further, an estimator is <u>strongly stability-robust</u> at P if its stability exponent equals one, the maximum, for all distributions in some neighborhood of P .

Clearly, bounded influence estimators are strongly stability-robust. Conversely, in i.i.d. (independent identically distributed) models it is not hard to show that for any estimator of the form (1.2) (i.e., any M-estimator) and any distribution P , there is a distribution P' arbitrarily close to P (in terms of the weak topology on the marginal distributions) for which the maximal moment exponent of $r_i(Z_i, \theta_0)$ is less than or equal to one. (See Hampel [24] for a justification of the choice of the weak topology.) That is, the estimator has stability exponent equal to zero at P' . Thus, an estimator is <u>not</u> stability-robust at any distribution, if it has an unbounded influence function.

To extend this converse result to the case of models with i.n.i.d. (independent non-identically distributed) observations is not too difficult. But for models with dependent observations one needs to find an appropriate topology or measure of closeness of distributions of the whole sequence of observations, Z_i , i = 1, 2, ... This is much less straightforward than the i.i.d. or i.n.i.d. cases, but several possibilities have been explored in the robustness literature (see Andrews [6, 7] and Papantoni-Kazakos and Grey [45]). For the neighborhoods considered by Andrews, the same converse result as above holds.

The above conditions for stability-robustness of an estimator can be compared to Hampel's [24] classical <u>qualitative robustness</u> concept.

In i.i.d. location and linear regression models, estimators of form (1.2) are qualitatively robust if and only if their influence function is bounded (and their estimating equations have a unique solution in the limit). That is, the conditions for qualitative robustness and stability-robustness are equivalent. Thus, we see that for M-estimators the stability property considered in this paper is closely related to the classical robustness properties of qualitative robustness and bounded influence.

As mentioned above, several reasons related to robustness suggest that high stability is often a desirable property for estimators. We now discuss these reasons. First, economic _data_ are rarely so "clean" that it is prudent to put great weight on a single observation. For example, the imprecisions of economic data are manifested by the continual revisions made to macroeconomic time series, and the subjective nature of some microeconomic survey data.

Several factors contribute to this imprecision: There is pure measurement error at the data collection stage. The correspondence between observed or "constructed" variables and the variables that are relevant from the perspective of economic theory is usually imperfect, and sometimes considerably so. The precise definitions of variables may be problematic even from a theoretical perspective, as exemplified by the money supply and market shares (in a nebulous market). Finally, recording errors made in stages of data collection, transmission, and analysis are inevitable. Such errors are often beyond the control of the econometrician who might not have any input into the collection and transmission stages. In fact, the econometrician might have only scant knowledge of the degree of imprecision of the data. In such cases, it is unwise to let any single observation have great weight in determining an estimator's value.

The imprecision of econometric models also adds to the desirability of high stability. Economic theory cannot yield complete model specifications, so even in the presence of a simple true model, a specified model is likely to be just an approximation. Moreover, the existence of a simple true model is usually questionable. In most cases, econometric models are approximations, at best, of much more complicated socio-economic phenomena. In this context, an observation that appears to be highly informative, may be so, only because of a spuriously precise specification of the model. For example, in a linear regression model an observation that is an outlier in the space of regressor variables can be highly informative. That is, it can greatly reduce estimator variances. If it is recognized, however, that the extension of the regression function to the outlying observation may be nonlinear with unknown functional form, then the informative content of the observation is drastically reduced. In such a case, the effect of the observation on the computed variance of an estimator with low stability exponent is spurious and deceptive. Such an observation also can cause a significant bias for an estimator with a low stability exponent. An estimator with a higher stability exponent is more robust to such specification difficulties because no single observation is given excessive weight.

A third reason for interest in high stability is that, in some models, estimators that are highly sensitive to a single observation perform quite poorly even if the model is specified correctly and the variables are measured without error. This phenomena may occur if the observations are highly variable. In this case, any single observation is potentially a randomly generated outlier with little informative content, and hence, should not be given disproportionate weight. For example, in

a regression model or simultaneous equations model with fat-tailed errors, the least squares (LS) estimator has a low stability exponent, because an outlying error realization can dramatically alter the value of the estimator. As expected, the relative efficiency of the LS estimator is quite poor in this situation. On the other hand, various robust procedures have high stability, and consequently, perform quite well even with highly variable observations. The statistical literature on robustness has analyzed problems of this sort in some detail, see Huber [32].

The above arguments for high stability are not always applicable, of course, and so, estimators with high stability are not always preferable. For diagnostic purposes, however, it still may be useful to know which estimation procedures are more likely to weight some single observation heavily. Hence, even in this case, estimator stability is of interest.

Note that stability comparisons can be made between different estimators for the same model or between estimators of different models. If an econometrician is more familiar with one model than another, stability comparisons of the latter sort may yield useful qualitative information about the second estimator's sensitivity to single observations in the sample, based on knowledge of the first estimator's sensitivity.

The stability exponent of an estimator is based on the deviations $\hat{\theta}_n - \hat{\theta}_{n,k}$, $k = 1, \ldots, n$. In the literature these deviations have been found useful for other related purposes. In analyzing the behavior of the least squares estimator in the linear regression model, Cook [16, 17] and Belesley, Kuh, and Welsch [10] use these deviations to help detect influential observations. Also, these deviations are proportional to the deviations of an estimator from its jackknifed pseudo-values. Tukey [50] has suggested a nonparametric estimator of the variance of the original

estimator, $\hat{\theta}_n$, based on the latter derivations (see also Miller [44]).
The relationship between the stability exponent and the influence curve, a
very important tool of robust statistics, has been discussed above. A finite
sample analogue of the influence curve suggested by Tukey [51], viz., the
sensitivity curve, is also related to stability. If we denote the sensi-
tivity curve of $\hat{\theta}_n$ formed using all n observations except the kth,
by $SC_{n,k}(z)$, then $SC_{n,k}(z)$ evaluated at the deleted observation Z_k
is proportional to the deviation $\hat{\theta}_n - \hat{\theta}_{n,k}$. That is, Tukey's finite
sample sensitivity curve (constructed with an observation deleted) eval-
uated at points in the actual sample is the basis of the stability exponent.

2. GENERAL RESULTS

2.1. Asymptotic Framework and Estimator Assumptions

The general asymptotic framework considered in this paper consists
of an infinite sequence $\{Z_i\} \equiv \{Z_i : i = 1, 2, \ldots\}$ of random vectors of
arbitrary dimensions. A sample of size n corresponds to the observation
of the first n terms in this sequence. For increased generality, the
i^{th} term Z_i is allowed to include elements of the random vectors Z_ℓ ,
for $\ell < i$. Thus, Z_i may include lagged variables. The distribution
of the sequence $\{Z_i\}$ is denoted P . All probabilistic statements below
are made for $\{Z_i\}$ distributed according to P . Thus, "almost surely"
means "almost surely under P ."

The sequence $\{Z_i\}$ is assumed to be weakly dependent over time.
That is, the dependence between random vectors dies out as the difference
in subscripts of the variables becomes infinitely large. (For the case
of cross-sectional data, the observations are often independent and this
requirement is satisfied.) More precisely, $\{Z_i\}$ is assumed to be strong

mixing. This is a realistic assumption for many economic time-series (and cross-section) situations. It is considerably weaker than other assumptions, such as independence, m-dependence, or auto-regressive moving average (ARMA) structure (see Withers [55], but cf. Andrews [5, 8]), that often are used in econometric models. Moreover, strong mixing does not imply stationarity or any assumption related to identical distributions.

Strong mixing is defined as follows: Let $\{Q_i : i = 1, 2, \ldots \}$ be a sequence of random vectors. Let $\mathcal{B}_{i\ell}$ denote the σ-field generated by $Q_i, Q_{i+1}, \ldots, Q_\ell$ for $1 \leq i < \ell \leq \infty$. That is, $\mathcal{B}_{i\ell}$ is the collection of all events determined by $Q_i, Q_{i+1}, \ldots, Q_\ell$. $\{Q_i\}$ is $\underline{\text{strong mixing}}$ if $\alpha(s) \to 0$ as $n \to \infty$, where $\alpha(n)$ are the strong mixing numbers of $\{Q_i\}$ defined by

$$(2.1) \qquad \alpha(s) \equiv \sup_{\ell \geq 1} \ \sup_{A \in \mathcal{B}_{1,\ell}, B \in \mathcal{B}_{\ell+s,\infty}} \left| P(A \cap B) - P(A)P(B) \right| .$$

Note, if $\{Q_i\}$ are independent, then $\alpha(s) = 0$, $\forall s \geq 1$; if $\{Q_i\}$ are m-dependent, then $\alpha(s) = 0$, $\forall s > m$; and if $\{Q_i\}$ have ARMA structure with absolutely continuous innovations, then $\alpha(s)$ declines to zero at an exponential rate as $s \to \infty$ (see Withers [55]). We assume:

A1) $\{Z_i\}$ is strong mixing with strong mixing numbers $\alpha(s)$ that satisfy $\alpha(s) = o(s^{-\alpha/(\alpha-1)})$ as $s \to \infty$, for some $\alpha \geq 1$ (where $\alpha = 1$ requires $\alpha(s) \equiv 0$ for s sufficiently large).

We consider the case where the investigator postulates a model that purports to describe some feature of the true distribution P of the data. This model is assumed to depend upon an unknown parameter vector θ. An estimator $\hat{\theta}_n$ is used by the investigator to estimate θ. It may be the case that the parametric model is correctly specified, i.e., it correctly corresponds to some aspect of the true distribution P. In this case,

a "true" parameter vector θ_0 is unambiguously defined. Alternatively, the parametric model may be misspecified. Depending upon the type of mis- specification, a "true" parameter vector θ_0 may or may not be well- defined. Fortunately, the possible difficulties in defining a true param- eter vector can be disregarded in the present analysis, provided the esti- mator considered converges to some fixed point (which might depend upon the estimation procedure itself).[3] Thus, the results allow one to deter- mine the affect of different forms of misspecification on the stability exponent of the estimator.

Once an estimator has been chosen, the parametric model specified by the investigator has no impact on the analysis of the stability exponent. Hence, the assumptions imposed below are stated in terms of the stochastic behavior of the estimating equations under the true distribution P , and make no mention of the parametric model. The assumptions we use are not the most primitive possible. That is, we do not place separate assumptions on P and on the estimating equations, but rather, on their interaction. Although the use of primitive assumptions is desirable in many contexts, their use in the present context would detract from the main point of the paper and weaken its focus. More primitive assumptions than those given can be deduced in given examples either from the existing literature or from first principles.

We now turn to two simple examples that we carry through this section to illustrate the more general framework and results. Section 3 discusses other applications of the results of this section. The first example con- sidered here is the classical linear regression (CLR) model,

(2.2) $y_i = x_i'\theta_0 + u_i$, $i = 1, 2, \ldots, n$,

where y_i is the observed dependent variable, x_i is the observed R^j-vector of fixed regressors, u_i is an independent, identically distributed (i.i.d), mean zero, unobserved error, and θ_0 is an R^J-valued unknown parameter vector. In this case, $Z_i = (y_i, x_i')'$. We suppose that the regressors are uniformly bounded, and that the $J \times J$ matrix $H = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} x_i x_i'$ exists and is non-singular. Note that this is only a partial description of the true distribution P , since the exact sequence of regressors and the error distribution are not specified. The parametric model specified by the investigator may or may not correspond to the true distribution P described above.

The second example we consider is the random regressor linear regression (RRLR) model. This model is identical to the CLR model except the regressors are assumed to be random, not fixed. We assume the regressors are i.i.d. and independent of the errors, and $H = E x_i x_i'$ is non-singular. Clearly, Al is satisfied in both of these models with $\alpha = 1$. Note that the rather restrictive assumptions placed on these models are for purposes of exposition; the general results given below allow them to be relaxed considerably.

The class of estimators considered for the general model includes all estimators that can be written as (measurable) solutions for θ to a system of equations of the form

(2.3) $$\sum_{i=1}^{n} r_i(Z_i, \theta) = \underset{\sim}{0} ,$$

for some R^J-valued (measurable) functions $r_i(\cdot, \cdot)$, $i = 1, 2, \ldots$, that are defined on some neighborhood of the true parameter θ_0 . For notational

convenience we abbreviate $r_i(Z_i, \theta)$ by $r_i(\theta)$. The jth element of $r_i(\theta)$ is denoted $r_{ij}(\theta)$. Section 3 shows that many well-known estimators of econometric models can be written as such.

For the two models used as examples in this section, we consider the least squares (LS) estimator. For this estimator,

$$(2.4) \qquad r_i(Z_i, \theta) = (y_i - x_i'\theta)x_i \equiv r_i^{LS}(\theta) \ .$$

Results concerning the stability of an estimator $\hat{\theta}_n$ are of interest only if the estimator satisfies certain minimal conditions regarding its performance. One such condition is the following:

B1a) $\{r_i(\theta)\}$ is sufficiently well-defined that a (measurable) solution $\hat{\theta}_n$ to (2.3) exists (though is not necessarily unique) for n sufficiently large a.s., and $\hat{\theta}_n \xrightarrow{n\to\infty} \theta_0$ a.s., for some θ_0 .

b) Further, $\hat{\theta}_{n,k_n} \xrightarrow{n\to\infty} \theta_0$ a.s., for any fixed sequence of positive integers $\{k_n\}$ with $k_n \leq n$, $\forall n$.

Conditions that imply almost sure convergence of the estimator $\hat{\theta}_n$ usually also imply almost sure convergence of $\hat{\theta}_{n,k_n}$, the estimator that ignores the $(k_n)^{th}$ observation. Most estimators considered in econometrics satisfy these conditions under fairly broad assumptions on the underlying model. Such assumptions can be found in the literature. In particular, the LS estimator for the CLR and RRLR models satisfies B1, see [3, 39, 53].

We now state several definitions used below.

DEFINITION: The <u>maximal moment exponent of a random variable</u> (rv) X is given by

$$(2.5) \qquad g = \sup\{\delta \geq 0 : E|X|^{\delta} < \infty\} .$$

If $E|X|^{\delta} = \infty$ $(<\infty)$ for all $\delta > 0$, the <u>maximal moment exponent</u> of X is defined to be 0 (∞) .

Thus, every rv has a unique maximal moment exponent g , and $g \in [0,\infty]$. For examples, a normal rv has a maximal moment exponent equal to ∞ , and a t rv with d degrees of freedom has a maximal moment exponent equal to d .

DEFINITION: The <u>maximal moment exponent of a random vector or matrix</u> is defined to be the smallest maximal moment exponent of any of its elements.

For a random vector or matrix X , let $|X|$ denote X with all of its elements replaced by their absolute values, and $\|X\|$ denote the Euclidean norm of X .

DEFINITION: A rv X is said to be <u>stochastically less (greater) than or equal to</u> a rv Y , and we write $X \overset{ST}{\leq} Y$ $(X \overset{ST}{\geq} Y)$, if $F_X(x) \geq F_Y(x)$ $(F_X(x) \leq F_Y(x))$, $\forall x \in R$, where F_X and F_Y are the distribution functions (df's) of X and Y , respectively. The same term is applied to random vectors and matrices if the above condition is satisfied element by element.

Next we construct a random vector, r_U , that is stochastically greater than or equal to $|r_i(\theta_0)|$ for all i . Let $F_U(w)$ be a J-vector with j^{th} element given by $\inf_{i \geq 1} P(|r_{ij}(\theta_0)| \leq w)$, for $j = 1, \ldots, J$ and $w \in R$. Let r_U be a random J-vector whose elements have univariate df's given by the vector $F_U(w)$.

DEFINITION: The maximal moment exponent of r_U is denoted by p .

The maximal moment exponent p turns out to be the key determinant of the stability exponent $\Lambda(\hat{\theta}, P)$ in many situations. It also arises in several regularity assumptions that are used in deriving the results.

One of the more primitive assumptions usually needed for assumption B1a to hold, i.e., for convergence of $\hat{\theta}_n$ to θ_0 , is that the expectation of the defining equations evaluated at θ_0 is zero, or approaches zero, as the sample size increases. We need to make this assumption explicit:

B2) $n^{\nu-1} \sum_{i=1}^{n} E r_i(\theta_0) \xrightarrow{n \to \infty} \underset{\sim}{0}$, $\forall \nu < 1 - 1/(2 \wedge (p/\alpha))$,

where " $\wedge$ " is the minimum operator. In the CLR and RRLS models $E r_i^{LS}(\theta_0) \equiv \underset{\sim}{0}$, so B2 is satisfied. Crowder [19] has shown that if $\frac{1}{n} \sum_{i=1}^{n} E r_i(\theta_0)$ is uniformly bounded away from zero, and the strong law of large numbers (SLLN) applies to $\{r_i(\theta_0)\}$, then $\hat{\theta}_n \neq \theta_0$ a.s. Thus, in the presence of B3 below (which guarantees that the SLLN applies), B2 is almost an implication of B1a.

The next assumption requires that $|r_i(\theta_0)|$ for $i = 1, 2, \ldots$ are stochastically dominated by an $L^{2\alpha-1}$ random vector (where α is a measure of the dependence of the sequence $\{Z_i\}$, see A1):

B3) $E|r_U|^{2\alpha-1} < \underset{\sim}{\infty}$, where $\underset{\sim}{\infty}$ is a J-vector of infinities. Equivalently, $p > 2\alpha-1$.

(Note that B3 rules out the case where some element of r_U is point mass at infinity.) In general, if B3 does not hold, then either $\hat{\theta}_n$ is strongly consistent, but it is somewhat more difficult than usual to prove (e.g., see

[26]), or $\hat{\theta}_n$ is not strongly consistent (as exemplified by the LS estimator when the errors in the CLR or RRLR model have undefined mean). In consequence, B3, or conditions that imply B3, is a common assumption in the literature (e.g., see assumptions 3 and 5 of Burguete, Gallant, and Souza [14, pp. 162 and 167]).

The LS estimator in the CLR and RRLR models satisfies B3 since $r_U^{LS} \stackrel{ST}{\leq} |u_1| \cdot \sup_{i \geq 1} |x_i|$ and $E|u_1| < \infty$ in the CLR model, and $r_U^{LS} \stackrel{ST}{\leq} |u_1 \cdot x_1|$ and $E|u_1 \cdot x_1| < \infty$ in the RRLR model. Note, since p is not necessarily greater than or equal to 2, $\hat{\theta}_n$ is not necessarily asymptotically normal.

We now construct a random matrix, Dr , that is stochastically greater than or equal to $\left|\frac{\partial}{\partial\theta} r_i(\theta_0)\right|$ for all i . Let $F_{Dr}(w)$, $w \in R$, be a $J \times J$ matrix with $(\ell,j)^{th}$ element $\inf_{i \geq 1} P\left(\left|\frac{\partial}{\partial\theta_\ell} r_{ij}(\theta_0)\right| \leq w\right)$, for $\ell, j = 1, \ldots, J$. Let Dr be a $J \times J$ random matrix whose elements have univariate df's given by the matrix $F_{Dr}(w)$. Dr is used to state a uniform smoothness condition on $r_i(\theta)$ at θ_0 . We assume:

B4a) $A \equiv \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E\frac{\partial}{\partial\theta} r_i(\theta_0)$ exists and is non-singular.

b) $E\|Dr\|^\eta < \infty$, for some η satisfying $\eta \geq 2$ and $\eta > \alpha$.

(Note that the assumption $\eta \geq 2$ can be relaxed in the results that follow.) Assumption B4a is common in the literature (e.g., see assumption 6 of [14, p. 169]) because it is necessary for asymptotic normality (with a non-singular covariance matrix) using the standard $\sqrt{n}$ normalization factor. The estimators considered here are not necessarily asymptotically normal, but this particular assumption is still used. It does restrict the form of heterogeneity of the observations somewhat. For the LS estimator in

the CLR and RRLR models, B4a corresponds to the assumptions above that

$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} x_i x_i'$ and $Ex_i x_i'$ exist and are non-singular, respectively. B4b

holds in the CLR model since the x_i are uniformly bounded, and in the

RRLR model if $E(x_1'x_1)^2 < \infty$.

The result $\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} r_i(\hat{\theta}_n) \xrightarrow{n \to \infty} A$ a.s. is commonly used in the liter-

ature when showing asymptotic normality of an estimator $\hat{\theta}_n$. We also use

this result, and impose the following additional smoothness condition on

$r_i(\theta)$ to ensure that it holds:[4]

B5a) $\sup_{i \geq 1} EW_{ij}^{\alpha + \delta} < \infty$, for some $\delta > 0$, for $j = 1, \ldots, J$, where

$W_{ij} \equiv \sup_{\tilde{\theta} \in \theta_0} \left\| \frac{\partial^2}{\partial \theta \partial \theta'} (r_{ij}(\tilde{\theta}) - r_{ij}(\theta_0)) \right\|$, θ_0 is some neighbourhood of

θ_0 ; and

b) $\frac{1}{n} \sum_{i=1}^{n} \left\| \frac{\partial^2}{\partial \theta \partial \theta'} r_{ij}(\theta_0) \right\| = O(1)$ as $n \to \infty$, a.s., $\forall j = 1, \ldots, J$.

For the LS estimator in the CLR and RRLR models B5 is automatically satis-

fied, since $\frac{\partial^2}{\partial \theta \partial \theta'} r_i(\theta) \equiv \underset{\sim}{0}$, where $\underset{\sim}{0}$ is a matrix of zeros.

2.2. Stability Results

First, we present a result that gives a linearized form of the estimator

$\hat{\theta}_n$. It also shows that the smoothness conditions on $r_i(\theta)$ are sufficient

to yield almost sure convergence of $\hat{\theta}_n$ and $\hat{\theta}_{n,k_n}$ to θ_0 at a faster

rate of convergence than n^0 .

THEOREM 1: *Under assumptions* A1 *and* B1-B5,

(a) $\hat{\theta}_n = \theta_0 - I_n \frac{1}{n} \sum_{i=1}^{n} A^{-1} r_i(\theta_0)$, *where* I_n *is a* $J \times J$ *random matrix*

equal to the identity matrix plus a matrix of small order one as

$n \to \infty$ a.s., *and*

(b) *for all sequences of positive integers* $\{k_n\}$ *with* $k_n \leq n$,

$\lim_{n \to \infty} n^{\nu}(\hat{\theta}_{n,k_n} - \theta_0) = \underset{\sim}{0}$ a.s., $\forall \nu < 1 - 1/(2 \wedge (p/\alpha))$.

COMMENTS: 1. The linearized form of $\hat{\theta}_n$, viz., $\theta_0 - \frac{1}{n} \sum_{i=1}^{n} A^{-1} r_i(\theta_0)$,

highlights the importance of the rv's $r_i(\theta_0)$, $i = 1, \ldots, n$, in de-

termining the stochastic properties of the estimator $\hat{\theta}_n$. In particular,

the linearized form suggests that the stability of $\hat{\theta}_n$ may be related to

the tail behavior of $r_i(\theta_0)$, $i = 1, \ldots, n$. It is shown below that

this is the case.

2. If $r_i(\cdot, \cdot)$ is independent of i for i sufficiently large,

as is often the case, then the influence curve of $\hat{\theta}_n$ is $A^{-1} r(z, \theta_0)$.

Thus the linearized form of $\hat{\theta}_n$ is determined by its influence curve.

3. Part a is a rather trivial consequence of A1 and B1-B3. It is

stated explicitly only because of the importance of the linearized form

for understanding the properties of the estimator.

4. Part b of the Theorem shows that the rate of convergence of $\hat{\theta}_n$

to θ_0 depends on the number of finite moments of $r_i(\theta_0)$, $i = 1, \ldots, n$

(as measured by the maximal moment exponent p of the stochastically domi-

nating random vector r_U). In addition, there is a tradeoff between the

maximal moment exponent of r_U and the degree of dependence over time (as

indexed by α , see A1). Note that the dependence of the rate of converg-

ence, ν , on the maximal moment exponent p of r_U and the degree of

dependence, ν , only exists below a cut off point. If $p \geq 2\alpha$, then

the maximal rate of convergence is obtained, and additional moments are of no consequence. This contrasts with the results obtained below for the stability exponent of $\hat{\theta}_n$. In the latter case no such cut off point exists.

5. In the CLR and RRLR models, the linearized form of the LS estimator is $\theta_0 - \frac{1}{n} \sum_{i=1}^{n} H^{-1} u_i x_i$, α equals one, and p equals the maximal moment exponent of u_i and $u_i x_i$, respectively. In both models, if u_i has two or more moments, the maximal rate of convergence is obtained, i.e., the upper bound on ν is one-half.

6. The proof (see Appendix) makes use of McLeish's [43] three series theorem for strong mixing rv's, and a result of Loeve [40].

We now establish two lower bounds on the stability exponent of an estimator $\hat{\theta} \equiv \{\theta_n : n = 1, 2, \ldots\}$:

THEOREM 2: *Let* A1 *and* B1-B5 *hold. Then*

(a) $\Lambda(\hat{\theta}, P) \geq 1 - 1/p$, *and*

(b) $\Lambda(\hat{\theta}, P) \geq \sup\{\xi \in R : \sum_{n=1}^{\infty} [1 - F_{nj}^*(n^{1-\xi})] < \infty, \forall j = 1, \ldots, J\}$ $(\geq 1 - 1/p)$,

where $F_{nj}^*(x) \equiv \min_{i \leq n} F_{ij}(x)$, *and* $F_{ij}(\cdot)$ *is the df of* $r_{ij}(\theta_0)$.

COMMENTS: 1. The lower bound of part a is more readily interpretable than that of part b, but part b is a stronger result. That is, the lower bound of part b is greater than or equal to that of part a.

2. The lower bound of part a is a linear function of the reciprocal of the maximal moment exponent p of r_{ij} . The lower bound increases strictly and continuously from 0 to 1 as p increases from 1 to ∞ . This result differs from rate of convergence results for almost sure convergence (see Theorem 1). The latter exhibit a cut off point beyond which additional moments do not increase the rate of convergence.

3. For the LS estimator in the CLR model, p equals the maximal moment exponent of the error u_1. For example, if u_1 has t-distribution with d degrees of freedom, then the lower bound given by part a is $1 - 1/d$, and it ranges continuously from 0 for the Cauchy $(d = 1)$ to 1 for the normal $(d = \infty)$. With regard to part b of Theorem 2, $F_{nj}^*(n^{1-\xi}) = F_{u_1}(n^{1-\xi}/(\max_{i \leq n}|x_{ij}|))$ in this case, where $F_{u_1}(\cdot)$ is the df of u_1. Note, $\sum_{n=1}^{\infty} [1 - F_{u_1}(n^{1-\xi}/(\max_{i \leq n}|x_{ij}|))] < \infty$, $\forall j$, if and only if $\sum_{i=1}^{n} [1 - F_{u_1}(n^{1-\xi})] < \infty$. And,

$$(2.6) \quad \sum_{n=1}^{\infty} [1 - F_{u_1}(n^{1-\xi})] = \sum_{n=1}^{\infty} P(|u_1|^{1/(1-\xi)} > n) \in [E|u_1|^{1/(1-\xi)}, \ E|u_1|^{1/(1-\xi)} + 1],$$

using Loeve's [40, p. 242] moments inequality. Thus, in this case, the lower bound of part b reduces to $1 - 1/p$, as in part a.

4. In the RRLR model, p equals the maximal moment exponent of $u_1 \cdot x_1$. If x_1 has as many or more moments than u_1, then the situation is exactly as above in the CLR model. If x_i has fewer moments than u_1, however, then the variability of the regressors determines the value of p and the lower bound $1 - 1/p$ is less than in the CLR model (with the same error distribution). For the RRLR model, $F_{nj}^*(n^{1-\xi}) = F_{nj}(n^{1-\xi})$, $\forall j$, and an argument similar to that of comment 3 shows that the lower bound of part b reduces to $1 - 1/p$.

5. The condition $\eta \geq 2$ in assumption B4b can be relaxed in this Theorem. Specifically, (a) under the assumptions of Theorem 2 except that of $\eta \geq 2$, for any $\tilde{p} \in (2\alpha-1, p]$, if $\eta \geq 2 \wedge (\tilde{p}/\alpha)$, then $\Lambda(\hat{\theta}, P) \geq 1 - 1/\tilde{p}$, and (b) under the assumptions of Theorem 2 except those of $\eta \geq 2$ and $p > 2\alpha-1$ (see B3), for any $\tilde{p} > 0$, if $\eta \geq 2 \wedge (\tilde{p}/\alpha)$ and $p > 1$, then $\Lambda(\hat{\theta}, P) \geq \sup\{\xi \in R : \sum_{n=1}^{\infty} [1 - F_{nj}^*(n^{1-\xi})] < \infty$ and

$\xi < 2\{1 - 1/(2 \wedge (\hat{p}/\alpha))\}$.

6. The proof makes use of a Taylor expansion of $\sum_{i=1}^{n} r_i(\hat{\theta}_n)$, the first Borel-Cantelli Lemma, a moment inequality of Loéve [40], and Theorem 1b (to show various terms are $o(1)$ as $n \to \infty$ a.s.).

The next result provides an upper bound on the stability exponent of an estimator $\hat{\theta}_n$ in terms related to the maximal moment exponent of $r_i(\theta_0)$, $i = 1, 2, \ldots$. Further, it shows that the stability exponent of $\hat{\theta}_n$ actually equals the lower bound of Theorem 2 part b. This result requires a stronger condition on the asymptotic weak dependence of the process $\{Z_i\}$ than strong mixing, because the second Borel-Cantelli (2BC) Lemma is used in its proof. The 2BC Lemma usually is stated for independent sequences, but it also holds for some strong mixing processes (see Lemma 4 in the Appendix). One might think that the 2BC Lemma holds for all strong mixing processes, since strong mixing processes satisfy a related result, viz., Kolmogorov's zero-one law, see Andrews [9]. It is shown in Lemma 4, however, that this is not the case. Hence, we need to strengthen the assumption regarding asymptotic weak dependence.

A sequence of random vectors $\{Q_i\}$ is φ-mixing if $\varphi(s) \to 0$ as $s \to \infty$, where $\varphi(s)$ are the φ-mixing numbers of $\{Q_i\}$ defined by

(2.7) $\varphi(s) \equiv \sup_{\ell \geq 1} \sup_{A \in \mathcal{B}_{1,\ell}:P(A)>0; B \in \mathcal{B}_{\ell+s,\infty}} |P(A \cap B) - P(A)P(B)|/P(A)$

$= \sup_{\ell \geq 1} \sup_{A \in \mathcal{B}_{1,\ell}:P(A)>0; B \in \mathcal{B}_{\ell+s,\infty}} |P(B|A) - P(B)|$,

where $\mathcal{B}_{i,\ell}$ is the σ-field generated by $\{Q_i, Q_{i+1}, \ldots, Q_\ell\}$. Note that $\varphi(s) \leq 1$, for all s . Sequences of independent and m-dependent rv's clearly are φ-mixing. Billingsley [13] provides additional examples. The

φ-mixing condition, however, is considerably stronger than the strong mixing condition. For example, stationary Gaussian sequences of rv's are φ-mixing if and only if they are m-dependent, see [33], whereas they are strong mixing under the weak condition that they possess a continuous, positive spectral density, see [36]. Thus, the φ-mixing assumption may be stronger than is reasonable for some economic applications.

For present purposes, we do not require the full strength of the φ-mixing assumption (i.e., $\varphi(s) \to 0$ as $s \to \infty$) for $\{Z_i\}$. We only require strong mixing and the additional assumption that $\varphi(s) < 1$, for some $s = 1, 2, \ldots$. This condition is intermediate between strong mixing and φ-mixing. Precisely how much more general it is than φ-mixing is undetermined as yet.

For the next result we assume:

(A1') $\{Z_i\}$ are strong mixing with strong mixing numbers as in A1, and

φ-mixing number $\varphi(s) < 1$, for some $s = 1, 2, \ldots$.

For the upper bound on the stability exponent of $\hat{\theta}_n$ given below, we need to construct a random vector, r_L , that is stochastically less than or equal to $|r_i(\theta_0)|$ for all i . Let $F_L(w)$ be the J-vector whose j^{th} element is $\sup_{i \geq 1} P(|r_{ij}(\theta_0)| \leq w)$ for $j = 1, \ldots, J$ and $w \in R$, and let r_L be a random J-vector whose elements have univariate df's given by the vector $F_L(w)$.

DEFINITION: The maximal moment exponent of r_L is denoted by q .

Note that the maximal moment exponent p of r_U is necessarily less than or equal to q .

THEOREM 3: *Let* A1' *and* B1-B5 *hold. Then,*

(a) $\Lambda(\hat{\theta},P) \leq 1 - 1/q$, *provided* $p > 2\alpha q/(q+1)$, *and*

(b) $\Lambda(\hat{\theta},P) = \inf\{\xi \in R : \sum_{n=1}^{\infty} [1 - F_{nj}^{*}(n^{1-\xi})] = \infty$, *for some integer j in* $\{1,\ldots,J\}\}$

$$= \sup\{\xi \in R : \sum_{n=1}^{\infty} [1 - F_{nj}^{*}(n^{1-\xi})] < \infty, \forall j = 1, \ldots, J\} ,$$

where F_{nj}^{*} *is as in Theorem 2.*

COMMENTS: 1. Part a holds for all $q < \infty$. If $q = \infty$, part a is shown

to hold (see Section 4) provided r_L is not identically $\underset{\sim}{0}$. In consequence,

the right-hand-side in part b is less than or equal to one provided r_L

is not identically $\underset{\sim}{0}$.

2. In some cases (e.g., when the observations are identically dis-

tributed), p equals q , and the stability exponent of an estimator is

given by the maximal moment exponent of the linearized form of the estimator

--the more moments, the greater the stability. In particular, there is no

cut off beyond which the existence of more moments is of no consequence.

If p is less than q , then the stability exponent of $\hat{\theta}_n$ lies in an

interval determined by p and q , and its exact value is given by the

somewhat complicated expression of part b.

3. For the LS estimator in the RRLR model, $r_L \overset{ST}{=} |u_1 \cdot x_1| \overset{ST}{=} r_U$,

and so, q = p and the stability exponent of $\hat{\theta}_n$ is $1 - 1/p$. In the CLR

model, $r_L \overset{ST}{\geq} |u_1| \cdot \underset{i \geq 1}{\min}|x_i|$. If the regression function has a constant term,

for example, then q is less than or equal to the maximal moment exponent

of $|u_1|$, which is p . Hence, $q \doteq p$ and the stability exponent of

$\hat{\theta}_n$ is $1 - 1/p$. For example, if the errors have t-distribution with d

degrees of freedom, then the stability exponent of $\hat{\theta}_n$ is $1 - 1/d$ in the

CLR model. Further, the stability exponent of the usual estimator

$\hat{\sigma}^2 \equiv \frac{1}{n-J} \sum_{i=1}^{n} (y_i - x_i'\hat{\theta}_{LS})^2$ of the error variance, σ^2 , is $1 - 2/p$ in the

CLR model. Thus, the variance estimator is less stable than the LS estimator of the regression parameters. This corroborates results found in the literature comparing the robustness of these two estimators.

4. The condition $\eta \geq 2$ of assumption B4b can be relaxed in this Theorem. Specifically, (a) the assumptions $\eta \geq 2$ and $p > 2\alpha q/(q+1)$ can be replaced in Theorem 3 part a by $\eta \geq 2\wedge(\tilde{p}/\alpha)$ for some $\tilde{p} \in (2\alpha q/(q+1), p]$, and (b) under the assumptions of Theorem 3 except those of $\eta \geq 2$ and $p > 2\alpha - 1$, if $\eta \geq 2\wedge(\tilde{p}/\alpha)$ for some $\tilde{p} > 0$, then $\Lambda(\hat{\theta}_n, P_{\theta_0}) \leq \inf C$, where $C \equiv \{\xi \in R: \sum_{n=1}^{\infty} [1 - F_{nj}^*(n^{1-\xi})] = \infty \text{ for some } j , \text{ and}$ $\xi < 2(1 - 1/(2\wedge(\tilde{p}/\alpha)))\}$. (Note that the infimum of a null set is defined to be infinity.)

3. EXAMPLES

This section contains a number of examples where the general results of Section 2 apply. The models and estimators are described as briefly as possible. In consequence, sufficient conditions for strong consistency (assumption B1) are not always given in their entirety. Such conditions can be found in the references cited. In all cases, the defining functions of the estimators, viz., $r_i(\theta)$, $i = 1, \ldots, n$, are assumed to be chosen to satisfy the conditions B2-B5.

It is possible to include sbme two and three stage estimators in the class considered in Section 2, e.g., Heckman's [27] two stage estimator of the Tobit model--Example 3.7 below, Zellner's [58] feasible Aitken estimator for the seemingly unrelated nonlinear regression model--Example 3.8, 2SLS, and 3SLS.

To see that many multi-stage estimators can be written in the form of (2.3), proceed as follows: Suppose part of the parameter vector θ_0, call the part λ_0, is estimated in a first stage via the solution to $\sum_{i=1}^{n} r_{1i}(\lambda) = \underset{\sim}{0}$, and (a not necessarily disjoint) part, call it β_0, is estimated in a second stage via the solution to $\sum_{i=1}^{n} r_{2i}(\hat{\lambda}_n, \beta) = \underset{\sim}{0}$, where $\hat{\lambda}_n$ is the first stage estimator. In place of θ_0 consider an alternative parameter vector $\tilde{\theta}_0 = (\lambda_0', \beta_0')'$. Now, a single stage estimator, $\underset{\sim}{\tilde{\theta}}_n$, of the desired form can be defined by taking

$$(3.1) \qquad r_i(\tilde{\theta}) = \begin{pmatrix} r_{1i}(\lambda) \\ r_{2i}(\lambda,\beta) \end{pmatrix}, \quad \text{for} \quad \tilde{\theta} = \begin{pmatrix} \lambda \\ \beta \end{pmatrix}.$$

This estimator satisfies A1 or A1', and B1-B5, if the separate stage estimators do. (The matrix $E\frac{\partial}{\partial\theta}r_i(\tilde{\theta}_0)$ in B4a is triangular, and hence, is non-singular if the diagonal blocks are non-singular.) Thus, the results of Section 2 apply. The extension for three stage estimators is straightforward.

In the examples that follow we assume independence of the observations, because this is the usual assumption made in the references cited. In most cases, this assumption can be relaxed by replacing it with an assumption of strong mixing. Strong consistency is proved, then, using the strong law of large numbers for strong mixing rv's (see McLeish [43]).

3.1. <u>Classical linear regression (CLR) model--Classical M-estimators</u> (References: Huber [31], Yohai and Maronna [57]). The model is the CLR model described in Section 2. We adopt slightly different notation from that of Section 2:

(3.2) $y_i = x_i'\beta_0 + u_i$, $i = 1, \ldots, n$, $Z_i \equiv (y_i, x_i')'$, $\theta_0 = (\beta_0', \sigma_0)'$.

The estimator $\hat{\theta}_n$ is defined by

(3.3) $r_i(\theta) = \begin{pmatrix} \psi((y_i - x_i'\beta)/\sigma) x_i \\ \psi^2(|y_i - x_i'\beta|/\sigma) - c \end{pmatrix}$, for $\theta = \begin{pmatrix} \beta \\ \sigma \end{pmatrix}$,

where c is a given constant, ψ is a bounded, smooth, odd function, and
the true parameter σ_0 solves $E\psi^2(|u_i|/\sigma_0) = c$. The estimator $\hat{\theta}_n$ has
the maximum stability exponent, viz., one, whether or not the errors u_i
have any moments. This contrasts sharply with the LS estimator, see Sec-
tion 2.

3.2. Random regressor linear regression (RRLR) model--General M-estimators

(References: Krasker and Welsch [37], Maronna and Yohai [42]). The model
is the RRLR model described in Section 2 with the notation of Example 3.1.
The estimator $\hat{\theta}_n$ is defined by

(3.4) $r_i(\theta) = \begin{pmatrix} \tilde{\psi}(x_i, (y_i - x_i'\beta)/\sigma) x_i \\ \chi(|y_i - x_i'\beta|/\sigma) \end{pmatrix}$,

where, for each x_i , $\tilde{\psi}(x_i, \cdot)$ is bounded, odd, and non-negative on R^+ ,
$\chi(\cdot)$ is nondecreasing and bounded, $E|x_i| \sup_u |\tilde{\psi}(x_i, u)| < \infty$, and the true
parameter σ_0 solves $E\chi(|u_i|/\sigma_0) = 0$. The stability exponent of $\hat{\theta}_n$
depends upon the maximal moment exponent of $\tilde{\psi}(x_i, u_i/\sigma_0) x_i$. If $\tilde{\psi}$ is
taken such that this is bounded uniformly for x_i and u_i (as in [37],
for example), then the general M-estimator is a bounded influence estimator,
and has stability exponent equal to one--the maximum.

3.3. <u>Linear, limited information simultaneous equations model--Instrumental variables (IV) estimator</u> (References: Heiler [28], Sargan [47]). The model is the same as the RRLR model but the regressors and errors are not necessarily independent:

(3.5) $y_i = x_i'\theta_0 + u_i$, $i = 1, \ldots, n$, $Z_i = (y_i, x_i', w_i')'$,

where w_i is a random vector of instrumental variables that is independent of the error u_i but not of the regressor x_i . The estimator $\hat{\theta}_n$ is defined by

(3.6) $r_i(\theta) = (y_i - x_i'\theta)w_i$.

The stability results for the IV estimator $\hat{\theta}_n$ are the same as for the LS estimator in the RRLR model with the maximal moment exponent of the instruments replacing that of the regressors. In particular, if the instruments or the errors have fewer than all moments finite, the IV estimator has stability exponent less than one.

3.4. <u>Linear, limited information simultaneous equations model--Weighted instrumental variables (WIV) estimator</u> (References: Krasker and Welsch [38]). The model is as in Example 3.3 with a slight change in notation:

(3.7) $y_i = x_i'\beta_0 + u_i$, $i = 1, \ldots, n$, $Z_i \equiv (y_i, x_i', w_i')$, $\theta_0 = (\beta_0', \alpha_0')'$.

The estimator $\hat{\theta}_n$ is defined by

(3.8) $r_i(\theta) = \begin{pmatrix} \min\{1, \ c/[|(y_i - x_i'\beta)/\sigma| \cdot (w_i'B^{-1}w_i)^{1/2}]\} \cdot (y_i - x_i'\beta)w_i \\ S \ vec[\gamma(c^2/w_i'B^{-1}w_i) \cdot w_iw_i' - B] \end{pmatrix}$,   for $\theta = \begin{pmatrix} \beta \\ \alpha \end{pmatrix}$,

where c and σ given constants, the parameter vector $\alpha \equiv S \text{ vec } B$, S

is a known $[J(J+1)/2] \times J^2$ selection matrix such that S vec B is the vec-

tor obtained by vectorizing the lower triangle of the symmetric $J \times J$ matrix

B , $\gamma(t) \equiv E \min(\eta^2, t)$ for $\eta \sim N(0,1)$, and the true parameter vector

$\alpha_0 \equiv S \text{ vec } B_0$ solves $B_0 = E\gamma(c^2/w_i'B_0^{-1}w_i)w_iw_i'$. (Note that σ can be esti-

mated by adding it to the parameter vector θ and adding an element to

$r_i(\theta)$.) As defined, $r_i(\theta)$ does not satisfy our conditions for smooth-

ness in θ . A version of $r_i(\theta)$ that is smoothed at the corners, however,

does satisfy our conditions and differs very little from $r_i(\theta)$.

It can be seen that $r_i(\theta_0)$ is a bounded random vector. Hence,

$\hat{\theta}_n$ and the WIV estimator of β_0 , given by the sub-vector $\hat{\beta}_n$, have

stability exponent equal to one.

3.5. <u>Nonlinear regression model--Least squares estimator</u> (References: [12,

20, 34, 41, 56]). The model is

(3.9) $y_i = f(x_i, \beta_0) + u_i$, $i = 1, \ldots, n$, $Z_i = (y_i, x_i')'$, $\theta_0 \equiv \beta_0$,

where the errors u_i are strong mixing, mean zero rv's, the regressors

x_i may be fixed or random but are independent of u and must satisfy con-

ditions for "proper" behavior as $n \to \infty$ (see references), and the regres-

sion function $f(\cdot, \cdot)$ is smooth. The LS estimator $\hat{\theta}_n$ is defined by the

function

(3.10) $r_i(\theta) = (y_i - f(x_i, \theta))\frac{\partial}{\partial\theta}f(x_i, \theta)$.

The stability exponent of $\hat{\theta}_n$ depends on the random vectors

$\left| u_i \frac{\partial}{\partial\theta}f(x_i, \theta_0) \right|$ in the manner described in Theorem 5. For examples, if

the regressors are i.i.d. random vectors or are fixed and uniformly bounded,

then its stability exponent is $1 - 1/p$, where p is the maximal moment
exponent of $\left| u_1 \frac{\partial}{\partial\theta} f(x_1, \theta_0) \right|$.

3.6. Nonlinear regression model--Classical M-estimators (References: [4,
12, 14]). The model is as in Example 3.5, except $\theta_0 \equiv (\beta_0', \sigma_0)'$ and the
assumption of mean zero errors is replaced by the assumption that
$E\psi(u_i/\sigma_0) = 0$, for ψ and σ_0 given below.[5] The estimator $\hat{\theta}_n = (\hat{\beta}_n', \hat{\sigma}_n)'$
is defined by the function

$$(3.11) \qquad r_i(\theta) = \begin{pmatrix} \psi((y_i - f(x_i, \beta))/\sigma) \cdot \frac{\partial}{\partial\theta} f(x_i, \beta) \\ \psi^2((y_i - f(x_i, \beta))/\sigma) - \gamma \end{pmatrix} , \quad \text{for } \theta \equiv (\beta', \sigma)' ,$$

where γ is a (known) constant given by $\gamma = \int \psi^2(s) d\Phi(s)$ for $\Phi(\cdot)$ the
standard normal df,[6] σ_0 is an unknown scale parameter defined by
$E\psi^2(u_i/\sigma_0) = \gamma$, and ψ is a bounded function as in Example 3.1.

Since ψ is bounded, the stability exponents of $\hat{\theta}_n$ and $\hat{\beta}_n$ depend
on the vectors $\left| \frac{\partial}{\partial\theta} f(x_i, \theta_0) \right|$. If the regressors are i.i.d. random vectors,
their stability exponents are $1 - 1/p$, where p is the number of finite
moments of $\left| \frac{\partial}{\partial\theta} f(x_1, \beta_0) \right|$. If the regressors are fixed and uniformly bounded,
their stability exponents equal one.

3.7. Censored regression (or Tobit) model--Heckman's two stage estimator
(Reference: Heckman [27]). The model

$$(3.12) \qquad y_i = (x_i'\beta_0 + u_i) \vee 0 , \quad i = 1, \ldots, n , \quad Z_i = (y_i, x_i')' ,$$

where " $\vee$ " is the maximum operator, the regressors x_i are i.i.d. random
vectors, the errors u_i are independent, normal$(0, \sigma_0^2)$ rv's, and
$\theta_0 \equiv (\beta_0', \sigma_0)'$. Heckman's two stage procedure uses an estimator of the

form (2.3) at each stage:

1st stage: The estimator $\hat{\lambda}_n$ is a maximum likelihood (ML) probit estimator of $\lambda_0 \equiv \beta_0/\sigma_0$. Its defining function is

$$(3.13) \qquad r_{1i}(\lambda) = \frac{1_{[y_i>0]} - \Phi(x_i'\lambda)}{\Phi(x_i'\lambda)(1 - \Phi(x_i'\lambda))}\phi(x_i'\lambda)x_i \; ,$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and distribution function, respectively, and $1_{[\cdot]}$ denotes the indicator function.

2nd stage: The estimator $(\hat{\beta}_n', \hat{\sigma}_n)'$ is the LS estimator of $(\beta_0', \sigma_0)'$ given $\hat{\lambda}_n$, using only the uncensored observations. Its defining function is

$$(3.14) \qquad r_{2i}(\hat{\lambda}_n, \beta, \sigma) = (y_i - x_i'\beta - (\hat{\phi}_i/\hat{\Phi}_i)\sigma)\begin{pmatrix} x_i \\ \hat{\phi}_i/\hat{\Phi}_i \end{pmatrix} \cdot 1_{[y_i>0]} \; ,$$

where $\hat{\phi}_i = \phi_i(x_i'\hat{\lambda}_n)$ and $\hat{\Phi}_i = \Phi_i(x_i'\hat{\lambda}_n)$.

This two stage estimator can be written as a single stage estimator of form (2.3) by considering the estimator $\hat{\tilde{\theta}}_n = (\hat{\lambda}_n', \hat{\beta}_n', \hat{\sigma}_n)'$ of $\tilde{\theta}_0 \equiv (\beta_0'/\sigma_0, \beta_0', \sigma_0)'$ defined by the function

$$(3.15) \qquad r_i(\tilde{\theta}) = \begin{pmatrix} r_{1i}(\lambda) \\ r_{2i}(\lambda,\beta,\sigma) \end{pmatrix} .$$

Since the errors have all moments finite in this example, the stability exponent of $\hat{\tilde{\theta}}_n$ under P_{θ_0} depends on the maximal moment exponent p of the regressors x_i . In particular, the stability exponent of $\hat{\tilde{\theta}}_n$ and of $(\hat{\beta}_n', \hat{\sigma}_n)'$ is $1 - 1/p$.

3.8. Seemingly unrelated nonlinear regression--Zellner's feasible Aitken

estimator (References: Gallant [21], Zellner [58]). The model consists

of M equations:

(3.16) $\quad y_{im} = f_m(x_{im}, \beta_{0m}) + u_{im}$, $\quad m = 1, \ldots, M$, $\quad i = 1, \ldots, n$;

$$Z_i = (y_{i1}, \ldots, y_{iM}, x'_{i1}, \ldots, x'_{iM})' \ .$$

Under Gallant's [21] assumptions, the error vectors $u_i \equiv (u_{i1}, \ldots, u_{iM})'$

satisfy $Eu_i = \underset{\sim}{0}$, $Eu_i u'_i = \Sigma_0$, and $Eu_i u'_\ell = \underset{\approx}{0}$ for $i \neq \ell$; the variables

x_{im} are fixed; and the regression functions $f_m(x_{im}, \beta_{0m})$ are smooth, have

bounded first derivatives, and behave like i.i.d. rv's for n large. Let

$\theta_0 \equiv (\beta'_0, \alpha'_0)'$, where $\beta_0 \equiv (\beta'_{01}, \ldots, \beta'_{0M})'$, $\alpha_0 \equiv S \text{ vec } \Sigma_0$, and S is

the known $M(M+1)/2 \times M^2$ selection matrix such that α_0 is the vector ob-

tained by vectorizing the lower triangle of the symmetric matrix Σ_0 .

The feasible Aitken estimator has three stages, each of which yields

an estimator that is the solution to a system of equations.

1st stage: The estimator $\hat{\lambda}_n$ is an equation by equation LS estimator of

β_0 . Its defining function is

$$(3.17) \qquad r_{1i}(\beta) = \begin{pmatrix} (y_{i1} - f_1(x_{i1}, \beta_1)) \frac{\partial}{\partial \beta_1} f_1(x_{i1}, \beta_1) \\ \vdots \qquad\qquad \vdots \\ (y_{iM} - f_M(x_{iM}, \beta_M)) \frac{\partial}{\partial \beta_M} f_M(x_{iM}, \beta_M) \end{pmatrix} , \text{ for } \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_M \end{pmatrix} .$$

2nd stage: The estimator $\hat{\alpha}_n$ of α_0 is based on the first-stage residuals.

Its defining function is

$$(3.18) \qquad r_{2i}(\hat{\lambda}_n, \alpha) = S \text{ vec}(\Sigma - \hat{u}_i \hat{u}'_i) \ ,$$

where $\hat{u}_i \equiv (y_{i1} - f_1(x_{i1}, \hat{\lambda}_{1n}), \ldots, y_{iM} - f_M(x_{iM}, \hat{\lambda}_{Mn}))'$, Σ is an $M \times M$ matrix defined by $\text{vec } \Sigma = D\alpha$, and D is the known $M^2 \times M(M+1)/2$ duplication matrix defined such that Σ is symmetric and α is the vectorization of the lower triangle of Σ .

3rd stage: The estimator $\hat{\beta}_n$ is a multi-equation weighted LS estimator of β_0 . Its defining function is

$$(3.19) \qquad r_{3i}(\hat{\alpha}_n, \beta) = (y_i - f_i(\beta))' \hat{\Sigma}_n^{-1} \frac{\partial}{\partial \beta} f_i(\beta) ,$$

where $\text{vec } \hat{\Sigma}_n \equiv D\hat{\alpha}_n$, $y_i \equiv (y_{i1}, \ldots, y_{iM})'$, and $f_i(\beta) \equiv (f_i(x_{i1}, \beta_1),$ $\ldots, f_M(x_{iM}, \beta_M))'$.

We write this multi-stage estimator in the single stage form of (2.3) by taking

$$(3.20) \qquad r_i(\tilde{\theta}) = (r_{1i}(\lambda)', r_{2i}(\lambda, \alpha)', r_{3i}(\alpha, \beta)')' ,$$

to yield an estimator $\hat{\tilde{\theta}}_n \equiv (\hat{\lambda}_n', \hat{\alpha}_n', \hat{\beta}_n')'$ of the parameter vector $\tilde{\theta}_0 \equiv (\beta_0', \alpha_0', \beta_0')'$.

In this example, the stability exponents of the estimators $\hat{\tilde{\theta}}_n$ and $(\hat{\beta}_n, \hat{\alpha}_n)$ equal $1 - 2/p$, where p is the maximal moment exponent of the errors u_i . Since $\hat{\alpha}_n$ is basically a LS variance estimator, it is not surprising that we get the same stability exponent here, as we got with the LS variance estimator in the linear-regression model (see Comment 3 following Theorem 3).

3.9. Maximum likelihood (ML) and pseudo-maximum likelihood estimators (References: [1, 18, 29, 30, 54]). ML and pseudo-ML estimators (both defined as solutions to likelihood equations) can be written in the form (2.3) for virtually all econometric models, provided the log-likelihood (or pseudo-

log-likelihood) function is differentiable in its parameter θ. (See Crowder [18] for a treatment of maximum likelihood estimation with dependent observations.) In the case of independent observations ML estimators are defined by the score function

(3.21) $r_i(\theta) = \frac{\partial}{\partial\theta}\log p(z_i, \theta)$,

where $p(z_i, \theta)$ is the density of Z_i with respect to some measure μ. Pseudo-ML estimators are defined identically, except $p(z_i, \theta)$ is some specified density that is not necessarily assumed to be the true density of Z_i. In addition, an estimator defined by (3.21) is called a pseudo-ML estimator if the observations are not independent, since in this case $\sum_{i=1}^{n} \log p(z_i, \theta)$ is not the log-likelihood of the sample.

For the results of Section 2 to hold, all that is needed is that the observations are strong mixing (with φ-mixing number $\varphi(s) < 1$ for some s) and that the score function satisfies the conditions B1-B5 on $r_i(\theta)$. Under quite general conditions, ML and pseudo-ML estimators have been shown to be strongly consistent, so B1 is not a problem. Further, assumptions B2-B5 are easy to verify and are satisfied in most econometric models.

The stability exponent of ML and pseudo-ML estimators depends on the maximal moment exponent (and perhaps tail behavior) of their score functions as established in Section 2. Examples include:

(i) Binary logit model: The rv y_i takes values 0 or 1. The probability that y_i equals 1 is $P_i(\theta) \equiv \exp(x_i'\theta)/(1 + \exp(x_i'\theta))$, where x_i is a fixed or random explanatory variable. The ML estimator is defined by

$$(3.22) \qquad r_i(\theta) = [y_i - \exp(x_i'\theta)/(1 + \exp(x_i'\theta))]x_i \ .$$

The first multiplicand of $r_i(\theta)$ lies in $(-1,1)$, so the stability exponent of $\hat{\theta}_n$ depends on the explanatory variables x_i , $i = 1, \ldots, n$. If the x_i are fixed and uniformly bounded, the stability exponent of $\hat{\theta}_n$ is one. If the x_i are i.i.d. with maximal moment exponent p , the stability exponent is $1 - 1/p$. (Note that the extension of the multinomial logit model is straightforward.)

(ii) <u>Binary probit model</u>: The model is the same as the logit model, except $P_i(\theta) \equiv \Phi(x_i'\theta)$, where $\Phi(\cdot)$ is the standard normal df. The ML probit estimator $\hat{\theta}_n$ is defined by

$$(3.23) \qquad r_i(\theta) = \frac{y_i - \Phi(x_i'\theta)}{\Phi(x_i'\theta)[1 - \Phi(x_i'\theta)]}\phi(x_i'\theta)x_i \ .$$

It is easy to see that the stability properties of the ML probit estimator are the same as those of the ML logit estimator.

(iii) <u>Censored regression (Tobit) model</u> (see Amemiya [2]): The model is the same as in Example 3.7. The ML estimator $\hat{\theta}_n = (\hat{\beta}_n', \hat{\sigma}_n)'$ is defined by

$$(3.24) \quad r_i(\theta) = \begin{pmatrix} \dfrac{\phi_i(\theta)x_i}{1 - \Phi_i(\theta)}1_{[y_i=0]} + \dfrac{1}{\sigma^2}(y_i - x_i'\beta)x_i 1_{[y_i>0]} \\[3ex] \dfrac{\phi_i(\theta)x_i'\beta}{\sigma^2(1 - \Phi_i(\theta))}1_{[y_i=0]} + \dfrac{1}{\sigma}\left[\left(\dfrac{y_i - x_i'\beta}{\sigma}\right)^2 - 1\right]1_{[y_i>0]} \end{pmatrix} \ ,$$

where $\phi_i(\theta) \equiv \phi(x_i'\beta/\sigma)$ and $\Phi_i(\theta) \equiv \Phi(x_i'\beta/\sigma)$. The form of $r_i(\theta)$ shows that the ML estimator has the same stability properties as Heckman's two stage estimator (see Example 3.7).

(iv) <u>Seemingly unrelated nonlinear regressions model</u>: The model is the same as in Example 3.8 where $\theta_0 \equiv (\beta_0', \alpha_0')'$. The pseudo-ML estimator of θ_0 formed using the multivariate normal $(\underset{\sim}{0}, \Sigma)$ distribution for the errors $u_i \equiv (u_{i1}, \ldots, u_{iM})'$ is defined by

$$(3.25) \quad r_i(\theta) = \begin{pmatrix} (y_i - f_i(\beta))' \Sigma^{-1} \frac{\partial}{\partial \beta} f_i(\beta) \\ S \ vec[\Sigma - (y_i - f_i(\beta))(y_i - f_i(\beta))'] \end{pmatrix} ., \quad \text{for} \quad \theta = (\beta', \alpha')' ,$$

where $\alpha = S \ vec \ \Sigma$ and S is defined in Example 3.8. The pseudo-ML estimator is very similar to the feasible Aitken estimator of Example 3.8. They both have the same stability properties.

For brevity we have not included the 2SLS, 3SLS, LIML, and FIML estimators of linear simultaneous equations models in the examples given above. 2SLS and 3SLS can be written in the form (2.3) via the method of Examples 3.7 and 3.8 (using Theil's [49] interpretation of 2SLS). LIML can be so written using its interpretation as the FIML estimator of an incomplete system of equations (see [23; 46, pp. 276, 351]). Finally, FIML is trivially of the form (2.3) under the assumption of independent errors.

Cowles Foundation, Department of Economics, Yale University

APPENDIX

The proofs of Theorem 1 and other results below use the following Lemma:

LEMMA 1: *Let* $\{Y_i\}$ *be a sequence of mean zero, strong mixing rv's with strong mixing numbers that satisfy* A1. *Assume* $\sup\limits_{i>1} E|Y_i|^{c-\delta} < \infty$, *for some* $c > 1$ *and all* δ *arbitrarily small and positive. Then, for any sequence of positive integers* $\{k_n\}$ *with* $k_n \leq n$,

$$n^\zeta \overline{Y}_{n,k_n} \xrightarrow{n \to \infty} 0 \quad a.s., \quad \forall \zeta < 1 - 1/(2 \wedge (c/\alpha)) ,$$

where $\overline{Y}_{n,k_n} \equiv \dfrac{1}{n-1} \sum\limits_{i=1,i\neq k_n}^{n} Y_i$.

PROOF OF THEOREM 1: Let $\overset{\sim}{\underset{i}{\sum}}$ denote a summation over i from 1 to n with $i \neq k_n$. Using $\overset{\sim}{\underset{i}{\sum}} r_{ij}(\hat{\theta}_{n,k_n}) = 0$, a Taylor expansion of $n^{\nu-1} \overset{\sim}{\underset{i}{\sum}} r_{ij}(\hat{\theta}_{n,k_n})$ about θ_0 yields

$$(4.1) \quad 0 = n^{\nu-1}\overset{\sim}{\underset{i}{\sum}} r_{ij}(\theta_0) + \frac{1}{n}\overset{\sim}{\underset{i}{\sum}}\left[\frac{\partial}{\partial\theta'}r_{ij}(\theta_0) + (\hat{\theta}_{n,k_n} - \theta_0)'\frac{\partial^2}{\partial\theta\partial\theta'}r_{ij}(\theta^*_{n,k_n})\right]n^\nu(\hat{\theta}_{n,k_n} - \theta_0) ,$$

and so,

$$(4.2) \quad 0 = o(1) + (a_j + o(1))n^\nu(\hat{\theta}_{n,k_n} - \theta_0) \quad a.s., \quad \forall \nu < 1 - 1/(2 \wedge (p/\alpha)) ,$$

for $j = 1, \ldots, J$, where θ^*_{n,k_n} is a random vector on the line segment joining $\hat{\theta}_{n,k_n}$ and θ_0, a_j is the jth row of A, and $o(1)$ is a random vector of appropriate dimension which is of small order one as $n \to \infty$ a.s. (4.2) follows from (4.1) using (i) Lemma 1 and B3 to show

$$n^{\nu-1} \sum_i^{\sim} [r_{ij}(\theta_0) - Er_{ij}(\theta_0)] = o(1) \quad \text{as} \quad n \to \infty \text{ a.s.,}$$ (ii) the assumption B2 that

$$n^{\nu-1} \sum_i^{\sim} Er_{ij}(\theta_0) = o(1) \quad \text{as} \quad n \to \infty,$$ (iii) Lemma 1 and B4 to show

$$\frac{1}{n} \sum_i^{\sim} \left[\frac{\partial}{\partial\theta} r_{ij}(\theta_0) - E\frac{\partial}{\partial\theta} r_{ij}(\theta_0) \right] = o(1) \quad \text{as} \quad n \to \infty \text{ a.s.,}$$ and (iv) equation (4.5) below and the strong consistency of $\hat{\theta}_{n,k_n}$ to give $\frac{1}{n} \sum_i^{\sim} (\hat{\theta}_{n,k_n} - \theta_0)' \frac{\partial^2}{\partial\theta\partial\theta'} r_{ij}(\theta^*_{n,k_n})$

$= o(1)$ as $n \to \infty$ a.s. Stacking equations (4.2) for $j = 1, \ldots, J$ to form a system of equations yields part b of the Theorem, since A is non-singular. Part a follows in a similar fashion from (4.1) and (4.2) taking $\nu = 0$ and $k_n = n-1$.

It remains to show (4.5). By Lemma 1 and B5a,

(4.3) $\dfrac{1}{n} \sum_i^{\sim} (W_{ij} - EW_{ij}) \xrightarrow{n\to\infty} 0$ a.s., and $\dfrac{1}{n} \sum_i^{\sim} EW_{ij} = O(1)$ as $n \to \infty$, $\forall j = 1, \ldots, J$.

Now, for any sequence of rv's $\{\tilde{\theta}_n\}$ such that $\tilde{\theta}_n \xrightarrow{n\to\infty} \theta_0$ a.s., $\tilde{\theta}_n$ is in Θ_0 for n sufficiently large a.s. (where Θ_0 is some neighborhood of θ_0, see B5), and so, for n sufficiently large,

(4.4) $\dfrac{1}{n} \sum_i^{\sim} \left\| \dfrac{\partial^2}{\partial\theta\partial\theta'} r_{ij}(\tilde{\theta}_n) - \dfrac{\partial^2}{\partial\theta\partial\theta'} r_{ij}(\theta_0) \right\| \leq \dfrac{1}{n} \sum_i^{\sim} W_{ij} = O(1)$ as $n \to \infty$ a.s.,

by (4.3). Hence, using B5b,

(4.5) $\dfrac{1}{n} \sum_i^{\sim} \left\| \dfrac{\partial^2}{\partial\theta\partial\theta'} r_{ij}(\tilde{\theta}_n) \right\| = O(1)$ as $n \to \infty$, a.s., $\forall j = 1, \ldots, J$. *Q.E.D.*

PROOF OF LEMMA 1: First we show that under the assumptions of the Lemma

$$(4.6) \qquad n^{\zeta}\overline{Y}_n \xrightarrow{n \to \infty} 0 \quad \text{a.s.,} \quad \forall \zeta < 1 - 1/(2 \wedge (c/\alpha)) \ .$$

We apply McLeish's [43] Lemma 2.9 to the rv's $X_n \equiv Y_n/n^{1-\zeta}$, where using

his notation we set $d_n = 1$, $\forall n$, $g_n(x) = |x|^{s(\delta)}$ for $s(\delta) \equiv (c-\delta) \wedge 2\alpha$,

and $\overline{X}_n \equiv X_n 1_{[|X_n| \leq d_n]}$. Since $\sum_{n=1}^{\infty} E^{1/\alpha} g_n(|X_n|) < \infty$, provided $\zeta < 1 - \alpha/s(\delta)$,

we have $\sum_{n=1}^{\infty} (X_n - E\overline{X}_n)$ converges a.s. by his Lemma 2.9. Now, by the proof

of Loeve's [40] Theorem 16.4.A (p. 241),

$$(4.7) \qquad \sum_{n=1}^{\infty} (EX_n - E\overline{X}_n) < \infty \ , \quad \text{provided} \quad |s(\delta)| \geq 1 \ .$$

Thus, provided $(c-\delta) \wedge 2\alpha \geq 1$ (which requires $c > 1$ and δ arbitrarily

small and positive), we have $\sum_{n=1}^{\infty} X_n$ converges a.s. Applying Kronecker's

Lemma gives (4.6) for $\zeta < 1 - \alpha/s(\delta)$. Since δ is arbitrarily small,

(4.6) holds for all $\zeta < 1 - \alpha/s(0)$, as desired.

Now, simple algebra gives $\overline{Y}_n - \overline{Y}_{n,k_n} = \frac{1}{n}\overline{Y}_{k_n} - \frac{1}{n}\overline{Y}_{n,k_n}$, and so,

$$(4.8) \qquad n^{\zeta}\overline{Y}_{n,k_n}\left(1 - \frac{1}{n}\right) = n^{\zeta}\overline{Y}_n - n^{\zeta-1}Y_{k_n} \ .$$

Using (4.6) we have

$$(4.9) \qquad n^{\zeta-1}Y_n = n^{\zeta}\overline{Y}_n - n^{\zeta}\left(\frac{n-1}{n}\right)\overline{Y}_{n-1} \xrightarrow{n \to \infty} 0 \quad \text{a.s.}$$

Thus, for any subsequence $\{k_n\}$ of $\{n\}$ with $k_n \leq n$, $\forall n$,

$$(4.10) \qquad n^{\zeta-1}|Y_{k_n}| \leq k_n^{\zeta-1}|Y_{k_n}| \xrightarrow{n\to\infty} 0 \quad \text{a.s.,} \quad \forall \zeta < 1 - 1/(2 \wedge (c/\alpha)) \ .$$

Combining (4.6), (4.8), and (4.10) gives the desired result. $\qquad$ Q.E.D.

The proof of Theorem 2 uses the following Lemma:

LEMMA 2: *Let* $\{Y_i\}$ *be as in Lemma 1 and assume* $|Y_i| \overset{ST}{\leq} Y$, $\forall i$, *for some rv* Y *that satisfies* $E|Y|^{c-\delta} < \infty$ *for some* $c \geq \alpha$ *and all* δ *arbitrarily small and positive. If*

$$(4.11) \qquad \sum_{n=1}^{\infty} [1 - G_n^*(n^{1-\tau})] < \infty \ , \quad \text{for some} \ \tau < 1 \ ,$$

where $G_n^*(x) \equiv \min_{1 \leq i \leq n} G_i(x)$ *and* $G_i(x)$ *is the df of* Y_i , *then for all sequences of positive integers* $\{k_n\}$ *with* $k_n \leq n$,

$$\lim_{n\to\infty} n^{\zeta}(\overline{Y}_n - \overline{Y}_{n,k_n}) = 0 \quad \text{a.s.,} \quad \forall \zeta < \tau \ ,$$

where $\overline{Y}_{n,k_n}$ *is as in Lemma 1.*

PROOF OF THEOREM 2: We prove the results of Comment 5 following Theorem 2. These results imply those of the Theorem. We prove Comment 5 part b first. It suffices to show: if

$$(4.12) \qquad \sum_{n=1}^{\infty} [1 - F_{nj}^*(n^{1-\xi})] < \infty \ , \quad \forall j = 1, \ldots, J \ , \quad \text{and} \quad \xi < 2(1 - 1/(2 \wedge (\widetilde{p}/\alpha))) \ ,$$

then for all sequences of positive integers $\{k_n\}$ with $k_n \leq n$, we have

$$(4.13) \qquad \lim_{n\to\infty} n^{\zeta}|\hat{\theta}_n - \hat{\theta}_{n,k_n}| = 0 \quad \text{a.s.,} \quad \forall \zeta < \xi \ .$$

Let $\bar{r}_n = \frac{1}{n}\sum_{i=1}^{n} r_i(\theta_0)$, $\bar{r}_{n,k_n} = \frac{1}{n-1}\sum_{i}^{\sim} r_i(\theta_0)$, $A_n = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}r_i(\theta_0)$, and

$A_{n,k_n} = \frac{1}{n-1}\sum_{i}^{\sim}\frac{\partial}{\partial\theta}r_i(\theta_0)$. Using (4.1), (4.2), and Theorem 1 part b, we have

(4.14) $\quad \underset{\sim}{0} = \bar{r}_n - \bar{r}_{n,k_n} + A_n(\hat{\theta}_n - \theta_0) - A_{n,k_n}(\hat{\theta}_{n,k_n} - \theta_0) + o(n^{-2\nu})$ a.s.,

for $\nu < 1 - 1/(2\wedge(\widetilde{p}/\alpha))$. By definition of ν and ζ we can take

ν such that $2\nu \geq \zeta$. This, plus manipulation of (4.14), gives

(4.15) $\quad -n^{\zeta}(\hat{\theta}_n - \hat{\theta}_{n,k_n}) = n^{\zeta}A_n^{-1}(\bar{r}_n - \bar{r}_{n,k_n}) + n^{\zeta}A_n^{-1}(A_n - A_{n,k_n})(\hat{\theta}_{n,k_n} - \theta_0) + o(1)$ a.s.,

where A_n^{-1} exists for n sufficiently large a.s., since A is non-singular

and $A_n \xrightarrow{n\to\infty} A$ a.s. by Lemma 1 and B4.

For all $\ell, j = 1, 2, \ldots, J$, and $\tau \equiv 1 - 1/n$,

(4.16) $\quad \sum_{n=1}^{\infty} P\left(\left|\frac{\partial}{\partial\theta_{\ell}}r_{nj}(\theta_0)\right| > n^{1-\tau}\right) \leq \sum_{n=1}^{\infty} P([Dr]_{\ell,j}^{1/(1-\tau)} > n) \leq E[Dr]_{\ell,j}^{1/(1-\tau)} + 1 < \infty$,

where the first inequality uses the definition of Dr , the second inequal-
ity follows by Loeve [40], Moments Inequality, p. 242], and the third in-
equality follows by B4b for all η satisfying $\eta \geq 2\wedge(\widetilde{p}/\alpha)$ and $\eta > \alpha$.
Lemma 2 applied element by element now gives

(4.17) $\quad n^{\delta}(A_n - A_{n,k_n}) \xrightarrow{n\to\infty} \underset{\sim}{0}$ a.s., $\forall\delta < \tau$ ($\equiv 1 - 1/n$) ,

where $\underset{\sim}{0}$ is a $J \times J$ matrix of zeros. Thus, using Theorem 1 part b,

(4.18) $\quad n^{\zeta}A_n^{-1}(A_n - A_{n,k_n})(\hat{\theta}_{n,k_n} - \theta_0) = o(1)$ as $n \to \infty$ a.s.,

provided $\zeta - \delta - \nu \leq 0$. Algebraic manipulation verifies this inequality.

For $\zeta < \xi$ where $\sum_{n=1}^{\infty} [1 - F_{nj}^*(n^{1-\xi})] < \infty$, $\forall j = 1, \ldots, J$, Lemma 2 gives

(4.19) $n^{\zeta}(\bar{r}_n - \bar{r}_{n,k_n}) \xrightarrow{n \to \infty} \underset{\sim}{0}$ a.s.

Equations (4.15), (4.18), and (4.19), and the result $A_n^{-1} \xrightarrow{n \to \infty} A^{-1}$ a.s. yield (4.13), as desired.

Now we show that Comment 5 part b implies Comment 5 part a. For all $j = 1, \ldots, J$, and all $\xi < 1 - 1/p$,

(4.20) $\sum_{n=1}^{\infty} [1 - F_{nj}^*(n^{1-\xi})] \leq \sum_{n=1}^{\infty} P((r_{Uj})^{1/(1-\xi)} > n) \leq E(r_{Uj})^{1/(1-\xi)} + 1 < \infty$,

where the third inequality holds for all $\xi < 1 - 1/p$ since r_U has maximal moment exponent equal to p , the second inequality follows by Loeve [40, Moments Inequality, p. 242], and the first inequality holds by definition of r_U . For $\tilde{p}$ $(2\alpha-1, p]$ (which requires $\tilde{p} > 2\alpha-1$) and $\xi < 1 - 1/\tilde{p}$, we can show $\xi < 2(1 - 1/(2 \wedge (\tilde{p}/\alpha)))$. This and (4.20) give part a of Comment 5. *Q.E.D.*

PROOF OF LEMMA 2: Simple algebra gives,

(4.21) $n^{\zeta}(\bar{Y}_n - \bar{Y}_{n,k_n}) = n^{\zeta-1} Y_{k_n} - n^{\zeta-1} \bar{Y}_{n,k_n}$.

By assumption,

(4.22) $\sum_{n=1}^{\infty} P(n^{\tau-1}|Y_{k_n}| \geq 1) \leq \sum_{n=1}^{\infty} [1 - G_n^*(n^{1-\tau})] < \infty$,

so the first Borel-Cantelli Lemma gives $P(n^{\tau-1}|Y_{k_n}| \geq 1 \text{ i.o.}) = 0$, where i.o. abbreviates "infinitely often." Thus, $\forall \zeta < \tau$,

(4.23) $n^{\zeta-1}|Y_{k_n}| \xrightarrow{n\to\infty} 0$ a.s.

Lemma 1 and the assumption $c \geq \alpha$ give

(4.24) $n^{\zeta-1}|\overline{Y}_{n,k_n}| \xrightarrow{n\to\infty} 0$ a.s.,

since $\zeta-1 < 0$. Equations (4.21), (4.23), and (4.24) combine to give the desired result. $Q.E.D.$

The proof of Theorem 3 uses the following Lemmas:

LEMMA 3: *Let* $\{Y_i\}$ *be as in Lemma 2 with the further assumption that* $\{Y_i\}$ *has φ-mixing numbers* $\{\varphi(s)\}$ *with* $\varphi(s) < 1$ *for some* $s = 1, 2, \ldots$. *If*

$$\sum_{n=1}^{\infty} [1 - G_n^*(n^{1-\tau})] = \infty , \quad \text{for some } \tau < 1 ,$$

where $G_n^*(\cdot)$ *is as in Lemma 2, then for some sequence of positive integers* $\{k_n\}$ *with* $k_n \leq n$,

$$\limsup_{n\to\infty} n^{\zeta}|\overline{Y}_n - \overline{Y}_{n,k_n}| = \infty \quad a.s., \quad \forall \zeta > \tau ,$$

where $\overline{Y}_{n,k_n}$ *is as in Lemma 1.*

LEMMA 4: (a) *Let* $\{X_n\}$ *be a strong mixing sequence of random vectors with φ-mixing number* $\varphi(s) < 1$ *for some* $s = 1, 2, \ldots,$ *and let* $\{D_n\}$ *be a sequence of events such that* $D_n \in \mathcal{B}_n$, $\forall n$, *where* $\mathcal{B}_n$ *is the σ-field generated by* X_n . *If* $\sum_{n=1}^{\infty} P(D_n) = \infty$, *then* $P(D_n \text{ i.o.}) = 1$ *(where i.o. abbreviates "infinitely often").*

(b) *The assumption on the φ-mixing numbers in part* (a) *is not redundant.*

COMMENT: Lemma 4(a) is a generalization of the second Borel-Cantelli Lemma.
It generalized Cohn's [15] Theorem 1.2. Lemma 4(b) shows that Theorem 3
does not hold with the weaker assumption A1 replacing A1'.

PROOF OF THEOREM 3: We prove the results of Comment 4 following Theorem 3.
These results (and Theorem 2) imply those of Theorem 3. Consider Comment
4 part b first. The result is trivial if C is null, so assume C is
non-empty. It suffices to show that for $\xi \in C$, any ζ larger than but
arbitrarily close to ξ , and some sequence $\{k_n\}$ with $k_n \leq n$,

$$(4.25) \qquad \lim_{n \to \infty} n^\zeta |\hat{\theta}_n - \hat{\theta}_{n,k_n}| = \underset{\sim}{0} \quad \text{a.s.}$$

does not hold. For ξ , ζ , and $\{k_n\}$ as above, (4.15) and (4.18) yield

$$(4.26) \qquad n^\zeta |\hat{\theta}_n - \hat{\theta}_{n,k_n}| = n^\zeta |A_n^{-1}(\bar{r}_n - \bar{r}_{n,k_n})| + o(1) \quad \text{a.s.,}$$

since $\zeta < 2(1 - 1/(2 \wedge (\widetilde{p}/\alpha)))$, and provided $\zeta - \delta - \nu \leq 0$ for some $\delta \leq 1 - 1/\eta$,
where A_n^{-1} exists for n sufficiently large a.s. Given the former condi-
tion on ζ , the latter condition holds if $\eta \geq 2 \wedge (\widetilde{p}/\alpha)$, as is assumed.

Using a proof by contradiction we show that for some sequence $\{k_n\}$

$$(4.27) \qquad \limsup_{n \to \infty} n^\zeta |A_n^{-1}(\bar{r}_n - \bar{r}_{n,k_n})| = \underset{\sim}{\infty}^1 \quad \text{a.s.,}$$

where $\underset{\sim}{\infty}^1$ denotes a J-vector with at least one element equal to ∞ . Let
ω denote a realization of the process $\{Z_i\}$. If (4.27) does not hold,
then for all ω in a set with positive probability we have

$$(4.28) \qquad n^\zeta |(A_n^\omega)^{-1}(\bar{r}_n^\omega - \bar{r}_{n,k_n}^\omega)| \leq M^\omega \cdot \underset{\sim}{e} , \quad \forall n = 1, 2, \ldots ,$$

for some scalar $M^\omega < \infty$, where $\underset{\sim}{e}$ is a vector of ones and the superscript ω indicates the particular realization ω . For such ω and n sufficiently large,

$$(4.29) \quad n^\zeta |\bar{r}_n - \bar{r}_{n,k_n}| = n^\zeta |A_n A_n^{-1} (\bar{r}_n - \bar{r}_{n,k_n})| \leq M \cdot |A_n| \cdot \underset{\sim}{e} \leq M \cdot |A + \varepsilon \underset{\sim}{ee}'| \cdot \underset{\sim}{e} < \underset{\sim}{\infty}^2$$

where the superscript ω has been omitted in (4.29) for notational convenience, $\underset{\sim}{\infty}^2$ denotes a J-vector of infinities, the first inequality holds by simple algebra, and the second inequality holds for n sufficiently large given $\varepsilon > 0$ since $A_n \xrightarrow{n \to \infty} A$ a.s. But, Lemma 3 implies $\limsup\limits_{n \to \infty} n^\zeta |\bar{r}_n - \bar{r}_{n,k_n}| = \underset{\sim}{\infty}^1$ a.s. for some sequence $\{k_n\}$. This contradicts (4.29) and implies (4.27) is true. (4.25) and (4.27) combine to give the result of Comment 4 part b.

Next we show Comment 4 part b implies Comment 4 part a. For $q < \infty$, it suffices to show $\xi \equiv 1 - 1/(q+\varepsilon)$ is in C for ε arbitrarily small and positive. For this ξ , $E|r_{Lj}|^{q+\varepsilon} = \infty$ for some integer j in $\{1, \ldots, J\}$. Thus,

$$(4.30) \quad \sum_{n=1}^{\infty} [1 - F_{nj}^*(n^{1-\xi})] \geq \sum_{n=1}^{\infty} [1 - F_{Lj}(n^{1/(q+\varepsilon)})] \geq E|r_{Lj}|^{q+\varepsilon} - 1 = \infty ,$$

where the second inequality follows by Loeve [40, Moments Inequality, p. 242]. In addition, algebraic manipulation shows that $\tilde{p} \in (2\alpha q/(q+1), p]$ and $\eta \geq 2 \wedge (\tilde{p}/\alpha)$ implies $\xi < 2(1 - 1/(2 \wedge (\tilde{p}/\alpha)))$, for ε sufficiently small. Hence, $\xi \in C$.

For the case $q = \infty$, part a says $\Lambda(\hat{\theta}, P) \leq 1$. The latter is true whether or not $q = \infty$, if $r_L \neq 0$. To see this, consider $\xi = 1 + \varepsilon$ for ε arbitrarily small and positive. For this ξ ,

$$(4.31) \qquad \sum_{n=1}^{\infty} [1 - F_{nj}^{*}(n^{1-\xi})] \geq \sum_{n=1}^{\infty} [1 - F_{Lj}(n^{-\varepsilon})] = \infty \ ,$$

unless $F_{Lj}(0) = 1$. Since $r_{Lj} \geq 0$, $\forall j$, (4.31) holds for all $\varepsilon > 0$ and all j unless $r_L \equiv \underline{0}$. Thus, part a holds for $q = \infty$, and more generally, $\Lambda(\hat{\theta}, P) \leq 1$, provided $r_L \neq \underline{0}$. $\qquad Q.E.D.$

PROOF OF LEMMA 3: It suffices to show the result for $\zeta \in (\tau, 1]$. Using (4.21), we have

$$(4.32) \qquad n^{\zeta} |\overline{Y}_n - \overline{Y}_{n,k_n}| \geq n^{\zeta-1} |Y_{k_n}| - n^{\zeta-1} |\overline{Y}_{n,k_n}| \ .$$

Let $\{k_n\}$ be a sequence such that $G_{k_n}(n^{1-\tau}) = G_n^{*}(n^{1-\tau})$, $\forall n$. Then,

$$(4.33) \qquad \sum_{n=1}^{\infty} P(n^{\tau-1} |Y_{k_n}| > 1) = \sum_{n=1}^{\infty} [1 - G_n^{*}(n^{1-\tau})] = \infty \ .$$

Lemma 4 now gives $P(n^{\tau-1} |Y_{k_n}| > 1 \text{ i.o.}) = 1$. Thus,

$$(4.34) \qquad \limsup_{n \to \infty} n^{\zeta-1} |Y_{k_n}| = \infty \ \text{a.s.}, \ \forall \zeta \in (\tau, 1] \ .$$

Also, since $\zeta \leq 1$, $n^{\zeta-1} |\overline{Y}_{n,k_n}|$ converges to zero as $n \to \infty$ a.s. by Lemma 1 and the assumption $c \succeq \alpha$. Thus, (4.34) and (4.32) combine to give the desired result. $\qquad Q.E.D.$

PROOF OF LEMMA 4: To prove part (a), note that Cohn's [15] Lemma 1.1 shows that $P(A_n \text{ i.o.}) > 0$. The Kolmogorov 0-1 Law for strong mixing random vectors (see Andrews [9]) now implies that $P(A_n \text{ i.o.}) = 1$, since $\{A_n \text{ i.o.}\}$ is a tail event.

To prove part (b), we present a counter-example. Let $\{\varepsilon_n\}$ be a sequence of independent Bernoulli random variables with probability $q_n \equiv n/(n+1)$ that ε_n equals zero. Let $\{X_n\}$ be a Markov chain defined by $X_1 = \varepsilon_1$, and for all $n > 1$, $X_n = X_{n-1}$ if $X_{n-1} = 1$, and $X_n = \varepsilon_n$, if $X_{n-1} = 0$. The state 1 is an absorbing state for $\{X_n\}$. Set $D_n = \{X_n = 0\}$. We have

$$P(D_n) = P(\varepsilon_m = 0, \forall m = 1, 2, \ldots, n) = \prod_{m=1}^{n} m/(m+1) = 1/(n+1) ,$$

so that $\sum\limits_{n=1}^{\infty} P(D_n) = \infty$. Further,

$$P(D_n \text{ i.o.}) = P(X_n = 0, \forall n = 1, 2, \ldots) = \lim_{N \to \infty} \prod_{n=1}^{N} q_n = 0 .$$

Thus, $\sum\limits_{n=1}^{\infty} P(D_n) = \infty$ does not imply $P(D_n \text{ i.o.}) = 1$. It remains to show that $\{X_n\}$ is strong mixing.

Let $\mathcal{B}_{ij}$ denote the σ-field generated by $X_i, \ldots, X_j$. Take A^0 to be any set in $\mathcal{B}_{1,n}$, and B^0 to be any set in $\mathcal{B}_{n+s,\infty}$. Let A^1 and B^1 denote the complements of A^0 and B^0 , respectively. Simple algebra shows that

(4.35) $\left| P(A^0 \cap B^0) - P(A^0)P(B^0) \right| = \left| P(A^j \cap B^\ell) - P(A^j)P(B^\ell) \right|$, $\forall j = 0, 1; \forall \ell = 0, 1$.

Define $\widetilde{A}$ and $\widetilde{B}$ by

$$\widetilde{A} = \begin{cases} A^0 & \text{if } A^0 \subset D_n^c \\ A^1 & \text{otherwise,} \end{cases} \quad \text{and} \quad \widetilde{B} = \begin{cases} B^0 & \text{if } B^0 \subset D_{n+s} \\ B^1 & \text{otherwise,} \end{cases}$$

where the superscript c denotes the complement of a set. Note that $\widetilde{A} \subset D_n^c$ and $\widetilde{B} \subset D_{n+s}$. (This follows because $\mathcal{B}_{1,n}$ ($\mathcal{B}_{n+s,\infty}$) does not contain any proper subsets of D_n (D_{n+s}^c).) Since D_n^c and D_{n+s} are disjoint, so are $\widetilde{A}$ and $\widetilde{B}$. Hence, using (4.35),

$$|P(A^0 \cap B^0) - P(A^0)P(B^0)| = |P(\widetilde{A} \cap \widetilde{B}) - P(\widetilde{A})P(\widetilde{B})| = P(\widetilde{A})P(\widetilde{B})$$

$$\leq P(D_{n+s}) \leq P(D_s) = 1/(s+1) .$$

Since $1/(s+1) \to 0$ as $s \to \infty$, $\{X_n\}$ is strong mixing. $\qquad Q.E.D.$

REFERENCES

[1] Akaike, H.: "Information Theory and an Extension of the Likelihood Principle," in *Proceedings of the Second International Symposium on Information Theory*, ed. by B. N. Petrov and F. Csáki. Budapest: Akadémai Kiado, 1973.

[2] Amemiya, T.: "Regression Analysis When the Dependent Variable Is Truncated Normal," *Econometrica*, 41 (1973), 997-1016.

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

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

[5] Andrews, D. W. K.: "Non-strong Mixing Autoregressive Processes," *Journal of Applied Probability*, 21 (1984), 930-934.

[6] Andrews, D. W. K.: "Robust and Asymptotically Efficient Estimation of Location in a Stationary Strong Mixing Gaussian Parametric Model," Cowles Foundation Discussion Paper No. 659, Yale University, 1984.

[7] Andrews, D. W. K.: "Robust Estimation of Location in a Stationary Strong Mixing Gaussian Parametric Model: II," Cowles Foundation Discussion Paper No. 697, Yale University, 1984.

[8] Andrews, D. W. K.: "A Nearly Independent, but Non-strong Mixing, Triangular Array," forthcoming in *Journal of Applied Probability* (1985).

[9] Andrews, D. W. K.: "A Zero-one Result for the Least Squares Estimator," *Econometric Theory*, 1 (1985), 85-96.

[10] Belsley, D. A., E. Kuh, and R. E. Welsch: *Regression Diagnostics*. New York: John Wiley & Sons, 1980.

[11] Bickel, P. J., and E. L. Lehmann: "Descriptive Statistics for Nonparametric Models. I. Introduction," *Annals of Statistics*, 3 (1975), 1038-1044.

[12] 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.

[13] Billingsley, P.: *Convergence of Probability Measures*. New York: John Wiley & Sons, 1968.

[14] Burguete, J. F., A. R. Gallant, and G. Souza: "On the Unification of the Asymptotic Theory of Nonlinear Econometric Models," *Economic Reviews*, 1 (1982), 151-190.

[15] Cohn, H.: "On a Class of Dependent Random Variables," *Revue Roumania de Mathematique Pures et Appliquees*, 10 (1965), 1593-1606.

[16] Cook, R. D.: "Detection of Influential Observations in Linear Regression," *Technometrics*, 19 (1977), 15-18.

[17] Cook, R. D.: "Influential Observations in Linear Regression," *Journal of the American Statistical Association*, 74 (1979), 169-174.

[18] Crowder, M. J.: "Maximum Likelihood Estimation for Dependent Observations," *Journal of the Royal Statistical Society, Series B*, 38 (1976), 45-53.

[19] Crowder, M. J.: "On Consistency and Non-consistency of Estimating Equations," manuscript, Department of Mathematics, Sussey University, Guilford, U.K., 1985.

[20] Domowitz, I., and H. White: "Misspecified Models with Dependent Observations," *Journal of Econometrics*, 20 (1982), 35-58.

[21] Gallant, A. R.: "Seemingly Unrelated Nonlinear Regressions," *Journal of Econometrics*, 3 (1975), 35-50.

[22] Giltinan, D. M., R. J. Carroll, and D. Ruppert: "Some New Estimation Methods for Weighted Regression When There Are Possible Outliers," manuscript, University of North Carolina, Chapel Hill, 1984.

[23] Godfrey, L. G., and M. R. Wickens: "The Estimation of Incomplete Models Using Subsystem LIML," Essex University Discussion Paper No. 99, 1978.

[24] Hampel, F. R.: "A General Qualitative Definition of Robustness," *Annals of Mathematical Statistics*, 42 (1971), 1887-1896.

[25] Hampel, F. R.: "The Influence Curve and Its Role in Robust Estimation," *Journal of the American Statistical Association*, 69 (1974), 383-393.

[26] Hannan, E. J., and M. Kanter: "Autoregressive Processes with Infinite Variance," *Journal of Applied Probability*, 14 (1977), 411-415.

[27] Heckman, J. J.: "Sample Selection Bias as a Specification Error," *Econometrica*, 47 (1979), 153-161.

[28] Heiler, S.: "Strong and Weak Consistency of Instrumental Variables Estimates and Application to Dynamic Models," manuscript, Dortmund University, 1979.

[29] Hoadley, B.: "Asymptotic Properties of Maximum Likelihood Estimators for the Independent Not Identically Distributed Case," *Annals of Mathematical Statistics*, 42 (1971), 1977-1991.

[30] Huber, P. J.: "The Behaviour of Maximum Likelihood Estimates under Nonstandard Conditions," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press, 1967, pp. 221-233.

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

[32] Huber, P. J.: *Robust Statistics*. New York: John Wiley & Sons, 1981.

[33] Ibragimov, I. A., and Yu. V. Linnik: *Independent and Stationary Sequences of Random Variables*. The Netherlands: Wolters-Noordhoff Publishing Co., 1971.

[34] Jennrich, R. I.: "Asymptotic Properties of Non-linear Least Squares Estimators," *Annals of Mathematical Statistics*, 40 (1969), 633-643.

[35] Kinal, T. W.: "The Existence of Moments of k-Class Estimators," *Econometrica*, 42 (1980), 517-527.

[36] Kolmogorov, A. N., and Y. A. Rozonov: "On Strong Mixing Conditions for Stationary Guassian Processes," *Theory of Probability and Its Applications*, 5 (1960), 204-208.

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

[38] Krasker, W. S., and R. E. Welsch: "Resistant Estimation for Simultaneous-equations Models Using Weighted Instrumental Variables," forthcoming in Econometrica, 47(1985).

[39] Lai, T. L., H. Robbins, and C. Z. Wei: "Strong Consistency of Least Squares Estimates in Multiple Regression: II," *Journal of Multivariate Analysis*, 9 (1979), 343-361.

[40] Loeve, M.: *Probability Theory*. New York: D. Van Nostrand, 1955.

[41] Malinvaud, E.: "The Consistency of Least Squares," *Annals of Mathematical Statistics*, 41 (1970), 956-969.

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

[43] McLeish, D. L.: "A Maximal Inequality and Dependent Strong Laws," *Annals of Probability*, 3 (1975), 829-839.

[44] Miller, R. G.: "The Jackknife--A Review," *Biometrika*, 61 (1974), 1-15.

[45] Papantoni-Kazakos, P. and R. M. Gray: "Robustness of Estimators on Stationary Observations," *Annals of Probability*, 7 (1979), 989-1002.

[46] Phillips, P. C. B., and M. R. Wickens: *Exercises in Econometrics, Volume II.* Cambridge, MA: Ballinger Publishing Co., 1978.

[47] Sargan, D.: "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26 (1958), 393-415.

[48] Stefanski, L.A., R.J. Carroll, and D. Ruppert: "Optimally Bounded Score Functions for Generalized Linear Models with Applications to Logistic Regression," Discussion Paper, Cornell University, 1985.

[49] Theil, H.: "Repeated Least-squares Applied to Complete Equation Systems," manuscript, Central Planning Bureau, The Netherlands, 1953.

[50] Tukey, J.: "Bias and Confidence in Not-quite Large Samples," Abstract, *Annals of Mathematical Statistics*, 29 (1958), 614.

[51] Tukey, J.: *Exploratory Data Analysis.* Preliminary edition. Reading, MA: Addison-Wesley, 1970.

[52] Wald, A.: "Note on the Consistency of the Maximum Likelihood Estimate," *Annals of Mathematical Statistics*, 20 (1949), 595-601.

[53] White, H.: "Using Least Squares to Approximate Unknown Regression Functions," *International Economic Review*, 21 (1980), 149-170.

[54] White, H.: "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50 (1982), 1-25.

[55] Withers, C. S.: "Conditions for Linear Processes to be Strong-Mixing," *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 57 (1981), 477-480.

[56] Wu, C. F.: "Asymptotic Theory of Nonlinear Least Squares Estimation," *Annals of Statistics*, 9 (1981), 501-513.

[57] Yohai, V. J., and R. A. Maronna: "Asymptotic Behaviour of M-Estimators for the Linear Model," *Annals of Statistics*, 7 (1979), 258-268.

[58] Zellner, A.: "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association*, 57 (1962), 348-368.

FOOTNOTES

[1] For their helpful comments and suggestions, I would like to thank Peter C. B. Phillips, two referees, the participants of the econometrics workshops at the California Institute of Technology, Queen's University, University of Toronto, University of California at San Diego, and University of Western Ontario, and the participants of the statistics seminar at M.I.T.

[2] The asymptotics used here are analogous to the standard asymptotics based on weak convergence (i.e., convergence in distribution). For example, suppose one has two estimators that can be normalized to have standard normal asymptotic distributions. The finite sample variabilities of these two estimators are usually compared by comparing the relative magnitudes of the two normalization factors. Similarly, the finite sample stability properties of two estimators are compared by comparing the normalization factors that are, by definition, their stability exponents.

[3] One solution to the problem of defining the estimand in a misspecified model is to take the estimand θ_0 to be the a.s. limit of the estimator under P (e.g., see [11, 31, 42, 53, 54]). That is, θ_0 may be defined as the unique solution to $\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E_P r_i(\theta) = 0$. Depending upon the circumstance, this solution may be more or less satisfactory.

[4] Assumption B5 requires that $r_i(\theta)$ is twice differentiable in some neighborhood of θ_0. This is not needed for asymptotic normality in general, but is needed for the stability results (see equation (4.14) of the proof of Theorem 2).

[5] This altering of the assumption of mean zero errors only affects the definition of the constant term, and hence, is relatively innocuous.

[6] The normal df arises in the definition of the constant γ, because with this definition if the true error distribution is normal, then the scale parameter σ equals the standard deviation of the errors. Such a definition is not essential, but often is reasonable.