

Digitized by the Internet Archive
in 2011 with funding from
Boston Library Consortium Member Libraries

<http://www.archive.org/details/unifiedtheoryofc00bate>

HB31
.M415
no. 359
1984

Dowey
S. INST. TECH.
SEP 30 1985
LIBRARIES

**working paper
department
of economics**

**A UNIFIED THEORY OF CONSISTENT ESTIMATION
FOR PARAMETRIC MODELS**

Charles Bates (Johns Hopkins)

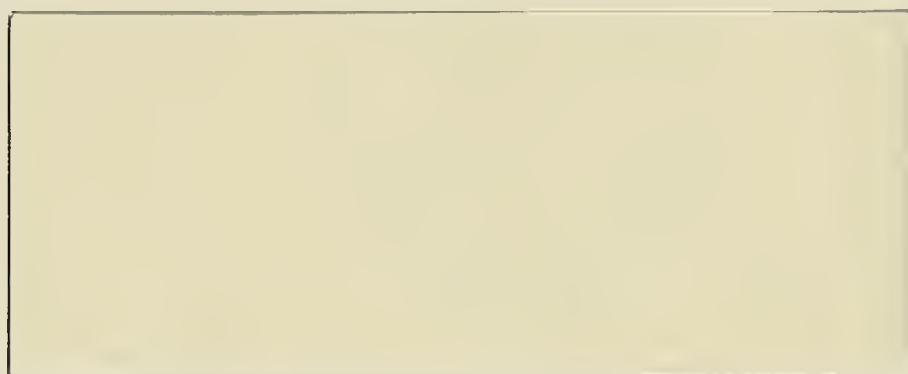
Halbert White (UCSD and MIT)

Number 359

December 1984

**massachusetts
institute of
technology**

**50 memorial drive
cambridge, mass. 02139**



A UNIFIED THEORY OF CONSISTENT ESTIMATION
FOR PARAMETRIC MODELS

Charles Bates (Johns Hopkins)

Halbert White (UCSD and MIT)

Number 359

December 1984

A Unified Theory of Consistent Estimation for
Parametric Models

by

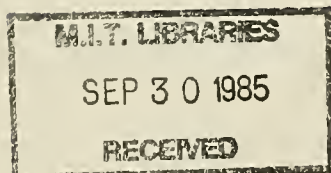
Charles Bates and Halbert White¹

Revised

November 1984

Abstract

We present a general theory of consistent estimation for possibly misspecified parametric models based on recent results of Domowitz and White [17]. This theory extends the unification of Burguete, Gallant and Souza [12] by allowing for heterogeneous, time dependent data and dynamic models. The theory is applied to yield consistency results for quasi-maximum likelihood and method of moments estimators. Of particular interest is a new generalized rank condition for identifiability.



1. Introduction

In the second section of "Misspecified Models with Dependent Observations," Domowitz and White [17] provide general results for establishing the consistency and asymptotic normality of a wide class of estimators obtained as the solution to an abstract optimization problem. Their results apply to misspecified models since the model which motivates the optimization problem need not contain the true data generating process. Because of the somewhat abstract nature of the sufficient conditions for these results, it can sometimes be difficult to determine whether a specific estimation technique falls within this class and therefore has the consequent consistency and asymptotic normality properties. Here we provide a more convenient theory of estimation specific enough to be directly applicable to particular estimation techniques applied to potentially misspecified models and yet sufficiently general so as to cover many estimation techniques of interest to econometricians. This theory, presented in Section 2, bridges the gap between the general results of Domowitz and White [17] and the myriad of single case results which characterize much of the econometrics literature. It also extends the unification achieved by Burguete, Gallant and Souza [12] for the case in which the true data generating mechanism is not subject to drift by allowing the least mean distance and method of moments estimators, which they treated separately, to be handled in a single framework; and by allowing for dynamic models and fairly arbitrary data processes which fall outside their theory. Although our focus here is on the consistency of estimators, the approach we take also allows a unified treatment of asymptotic normality and provides a very convenient foundation for addressing the issue of asymptotic efficiency for various estimators of the parameters of a correctly specified model. (These issues are addressed in Bates and White [7]).

Sections 3 and 4 present examples of specific estimation techniques covered by the results of Section 2, which represent extensions of previously published results. Section 3 deals with the consistency of the (quasi-)maximum likelihood estimator. Section 4 examines the consistency of m-estimators and of the generalized method of moments estimator of Hansen [25].

2. CONSISTENCY OF THE ABSTRACT ESTIMATOR

The available data are assumed to be generated in the following way:

ASSUMPTION A.1: The observed data are a realization of a stochastic process $\omega = \{W_t: t, 1, 2, \dots\}$ on a probability space $(\Omega, \mathbb{F}, P^*)$, where $\Omega \equiv \times_{i=1}^{\infty} R^S \equiv R_{\infty}^S$, and \mathbb{F} is the Borel σ -field generated by the measurable finite dimensional product cylinders.

The probability measure P^* provides a complete description of the stochastic relationships holding among the data both contemporaneously and over time. Usually, the "data generating mechanism" described by P^* is unknown, and our goal is to use the available data to learn about some aspect of the data generating mechanism which is of particular interest.

A common way of doing this is to specify a probability model, i.e. a family of probability measures $\mathbb{P} = \{ P_{\theta} : \theta \in \Theta \}$. P_{θ} is indexed by parameters θ taking values in some set Θ . For example, \mathbb{P} might be the family of probability measures whose finite dimensional distributions are given by

$$P_{\theta n}(E) = \int_E f_n(\omega^n, \theta) d\mu_n(\omega^n) \quad E \in \mathcal{B}(R_n^S), n=1, 2, \dots$$

where $f_n : \mathbb{R}_n^S \times \Theta \rightarrow \mathbb{R}^1$ is the likelihood function of the probability model for a sample of size n , μ_n is a σ -finite measure on $(\mathbb{R}_n^S, \mathcal{B}(\mathbb{R}_n^S))$ and $\omega^n \equiv (\omega_1, \dots, \omega_n) \in \mathbb{R}_n^S$. The notation $\mathcal{B}(\mathbb{R}_n^S)$ denotes the Borel σ -field generated by the open sets of $\mathbb{R}_n^S \equiv \times_{i=1}^n \mathbb{R}^S$. This is a probability model which is completely specified, in the sense that under general identification conditions, a specific choice for θ leads to a unique probability measure $P_\theta \in \mathcal{P}$.

However, interest only in certain features of the data generating mechanism (e.g. the conditional means) may lead to specifying a probability model \mathcal{P} such as the collection of all probability measures P_θ for which

$$E_\theta(W_t \mid \mathcal{F}_{t-1}) = W_{t-1}\theta, \quad t=1,2,\dots$$

where $E_\theta(\cdot \mid \cdot)$ is the conditional expectation implied by the probability measure P_θ and $\mathcal{F}_{t-1} \equiv \sigma(W_1, \dots, W_{t-1})$ is the Borel σ -field generated by W_1, \dots, W_{t-1} . This is the vector autoregression model of order 1 (VAR(1)). Because this model only specifies the behavior of conditional expectations under P_θ , it does not yield a "completely" specified probability model in the same way that specifying a likelihood function does since a unique probability measure is not determined for given θ . Nevertheless, if our interest lies solely in conditional expectations, then such a specification may be sufficiently complete for our purposes, and identification of the parameters of the conditional mean can be achieved under general conditions.

In most studies, it is assumed that P^* belongs to \mathcal{P} , the specified probability model, so that the model is correctly specified. Due to the complexity of economic phenomena, there is little guarantee that the probability

models obtained from the economic theory relevant to the data under consideration are correctly specified. Thus, we acknowledge from the outset that any specific probability model adopted is most plausibly treated as misspecified (so that P^* is not necessarily in \mathbb{P}) and that our model is best viewed as a way of obtaining some kind of approximation to P^* . For this reason, any theory which we develop should rely as little as possible on the suspect probability model \mathbb{P} .

Of course, there is always the hope that one's probability model is correctly specified. Given this hope, it is natural to construct estimators based on the probability model that have the property that if P^* is in \mathbb{P} (e.g. $P^* = P_{\theta^0}$ for some θ^0 in Θ) then the estimator is consistent for the "true" parameters, θ^0 . Such estimators are frequently constructed by choosing a parameter vector in Θ which implies the closest possible correspondence between the behavior exhibited by the data and that implied for the data by the probability model.

The criterion by which one measures "closeness" allows considerable latitude. For probability models with a specified likelihood function, closeness can be measured using the Kullback-Leibler [31] Information Criteria (KLIC), which provides a measure of the closeness of the specified likelihood function $f_n(\omega^n, \theta)$ to the joint density of the sample $dP_n^* / d\mu_n$ implied by the true data generating mechanism. (We define $P_n^*(E) \equiv P^*[\omega^n \in E]$, $E \in \mathbb{R}_n^S$.) This leads to the method of maximum likelihood, in which an estimator is obtained by solving the problem

$$\max_{\theta \in \Theta} \ln f_n(\omega^n, \theta).$$

For the VAR(1) model discussed above, closeness may be measured in terms of how well $W_{t-1}\theta$ approximates W_t . An estimator could be obtained by solving the

problem

$$\min_{\theta \in \Theta} \sum_{t=1}^n (W_t - W_{t-1}\theta)'(W_t - W_{t-1}\theta)$$

which gives a least squares estimator, or in the case of scalar W_t by solving

$$\min_{\theta \in \Theta} \sum_{t=1}^n |W_t - W_{t-1}\theta|.$$

This gives the least absolute deviations estimator.

The study of the consequences of applying such estimators to misspecified probability models was initiated by Huber [28] and Berk [8,9]. Work by White [42,44,45] examines the implications of model misspecification for ordinary least squares, nonlinear least squares and maximum likelihood estimation respectively under convenient assumptions less general than those of Huber, while the recent work of Burguete, Gallant and Souza [12] provides a unified framework in which estimation of misspecified models can be conveniently treated.

All of these studies address situations in which the observations are independent, and aspects of dependence and dynamic misspecification are not addressed. However, the recent work of Domowitz and White [17] does provide a framework in which the consequences of misspecification can be studied in the context of dependent observations.

The starting point for all of this work is the recognition that no matter how the data may truly be generated, the parametric probability models typically specified and the estimation criteria typically applied often lead to estimators obtained as the solution to an optimization problem, in which the optimand is a function of the observed data, ω , and the parameters θ . Formally, an estimator $\hat{\theta}_n$ is the solution to an optimization problem

$$\min_{\theta \in \Theta} Q_n(\omega, \theta).$$

The properties of $\hat{\theta}_n$ can be studied in the misspecified case by relying upon the specified probability model only to the extent that it affects the form of the optimization problem as determined by the choice of estimation technique, and then, placing as little structure on P^* as possible, using laws of large numbers and central limit theory to draw conclusions about the behavior of $\hat{\theta}_n$.

To study the consistency of $\hat{\theta}_n$ we rely on the heuristic insight that because $\hat{\theta}_n$ minimizes Q_n and because Q_n can generally be shown to tend to a real function $\bar{Q}_n : \Theta \rightarrow \mathbb{R}$ as $n \rightarrow \infty$ under mild conditions on P^* , then $\hat{\theta}_n$ should tend to θ_n^* (say) which minimizes \bar{Q}_n .

In order for this heuristic argument to work, we require two things: that Q_n converges appropriately to some function \bar{Q}_n ; and that \bar{Q}_n has a well defined (i.e. appropriately unique) minimum. For the first item, the appropriate convergence is that $Q_n(\omega, \theta) - \bar{Q}_n(\theta) \xrightarrow{a.s.} 0$ uniformly on Θ . For the second item, the appropriate uniqueness condition is supplied using the following definition of Domowitz and White [17].

DEFINITION 2.1. Let $\bar{Q}_n(\theta)$ be a real-valued continuous function on a compact metric space Θ such that $\bar{Q}_n(\theta)$ has a minimum at θ_n^* , $n=1,2,\dots$. Let $S_n(\varepsilon)$ be an open sphere centered at θ_n^* with fixed radius $\varepsilon > 0$. For each $n=1,2,\dots$, define the neighborhood $\eta_n(\varepsilon) \equiv S_n(\varepsilon) \cap \Theta$. Its complement in Θ , $\eta_n^c(\varepsilon)$, is compact. The minimizer θ_n^* is said to be identifiably unique if and only if for any $\varepsilon > 0$

$$\liminf_{n \rightarrow \infty} \left[\min_{\theta \in \eta_n^c(\varepsilon)} \bar{Q}_n(\theta) - \bar{Q}_n(\theta_n^*) \right] > 0.$$

Note that this is a global concept which essentially says that there exists a supporting hyperplane of the epigraph² of \bar{Q}_n which is uniformly bounded away from the epigraph of \bar{Q}_n outside every neighborhood of a unique supporting point θ_n^* . As such, it is easily seen to be a minimal requirement in that if it is not satisfied, a sequence $\{\tilde{\theta}_n\}$ could always be constructed such that the distance between θ_n^* and $\tilde{\theta}_n$ does not approach zero even though $|\bar{Q}_n(\theta_n^*) - \bar{Q}_n(\tilde{\theta}_n)|$ does.

For convenience, we restate the consistency result of Domowitz and White [17].

THEOREM 2.2. Given Assumption A.1, assume:

Assumption 2.2.i: For each θ in Θ , a compact subset of R^k , $Q_n(\omega, \theta)$ is a measurable function on Ω and, for each $\omega \in \Omega$, a continuous function on Θ , $n=1,2,\dots$.

Then there exists a measurable function $\hat{\theta}_n(\omega)$ such that

$$Q_n(\omega, \hat{\theta}_n(\omega)) = \inf_{\theta \in \Theta} Q_n(\omega, \theta) \text{ for all } \omega \in \Omega.$$

If, in addition,

Assumption 2.2.ii: $|Q_n(\omega, \theta) - \bar{Q}_n(\theta)| \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$, uniformly on Θ ; and

Assumption 2.2.iii: $\bar{Q}_n(\theta)$ has an identifiably unique minimizer θ_n^* , $n=1,2,\dots$;

then $\hat{\theta}_n - \theta_n^* \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$.

The form and proof of this result owe much to the antecedent work of Wald [39], Jennrich [29], Hoadley [27] and Amemiya [2] for estimation of correctly specified models.

The requirement that $\bar{Q}_n(\theta)$ have an identifiably unique minimizer allows "identification" of the parameters θ_n^* only in the sense that the objective function is not allowed to become too flat in the neighborhood of its minimum. This is quite distinct from the notion of identification which arises in the estimation of the "true" parameters of a correctly specified model. There, the knowledge of P makes it possible to discuss meaningfully the identification of true parameters without reference to the properties of the optimand defining the estimator, although lack of identification does lead to optimands with non-unique minima. When the model is correctly specified, it is typically straightforward to place convenient and plausible primitive conditions on the model to obtain identification of true parameters (as in Wald [39], Jennrich [29], Hoadley [27], and Amemiya [2]) and then verify Assumption 2.2.iii (or an appropriate analog) under appropriate primitive conditions on the estimator (as these authors also do).

In the present context, θ_n^* does not necessarily correspond to any "true" parameters, but instead is determined by the interaction of the probability model specified, the estimation technique chosen, and the behavior of the underlying stochastic process described by the unknown probability measure P^* . In particular, given a specific probability model and specific data generation mechanism, the sequence θ_n^* will generally differ under different choices of estimators, e.g. ordinary least squares vs. least absolute deviations. (In fact, such variability can be a useful indicator of model specification problems.)

Thus, it is useful and important to distinguish between the concept of the identification of the parameters of a correctly specified model, which is a property of the probability model not requiring reference to any particular estimation technique, and the concept of identifiability of the parameters of a

possibly misspecified model, which results when the probability model and estimation technique chosen interact to produce a uniquely determined parameter vector or sequence of parameter vectors to which the estimator tends.

When estimating the parameters of a correctly specified model, identification of the true parameters is a necessary condition for their identifiability, defined in this way. It is not a sufficient condition because the estimator may be constructed in such a way as not to be consistent for the true parameters, but for some other parameters which are nevertheless "identifiable". Moreover, a model may fail to be identified, but estimation of a misspecified version of the model may yield identifiable parameters. For example, suppose the correctly specified model is

$$Y_t = X_t \theta + \varepsilon_t$$

where $E(X_t' \varepsilon_t) \neq 0$, and no instrumental variables are available for X_t (e.g.

because $X_t = Y_t \gamma + u_t$ also, as in the elementary supply and demand model where

X_t is price and Y_t is quantity). In this model θ is not identified.

Nevertheless, the parameter

$$\theta^* = E(X_t' X_t)^{-1} E(X_t' Y_t)$$

is identifiable when the method of ordinary least squares is applied to i.i.d.

observations on X_t and Y_t in an attempt to estimate θ .

In the present context in which the probability model is suspect (or equivalently, in which the estimation technique chosen does not consistently estimate the true parameters of an otherwise correctly specified probability model), the concept of identification loses its immediacy, and the concept of identifiability acts as an appropriate replacement. Nevertheless, if the model

is correctly specified or incorrectly specified in some irrelevant way, it is usually quite easy to verify Assumption 2.2.iii using appealing primitive conditions on the model and estimator, as discussed by White [48, chapter 5] in a context closely related to that presented here. We provide a result in Section 3 below which provides such primitive conditions for method of moments and m-estimators not treated by White [48].

The difficulty in applying Theorem 2.2 is the need to verify Assumptions 2.2.i through 2.2.iii. Our goal here is to simplify this task by presenting convenient sufficient conditions for Assumption 2.2.i through 2.2.iii which are nevertheless sufficiently general to cover many estimators of interest to economists. The estimators considered here are defined as follows:

ASSUMPTION A.2: The estimator $\hat{\theta}_n$ is the solution to the problem

$$\min_{\theta \in \Theta} Q_n(\omega, \theta) \equiv g_n(n^{-1} \sum_{t=1}^n q_t(\omega, \theta), \hat{\pi}_n(\omega)) \quad (2.1)$$

where

(A.2.i) $\{g_n: \mathbb{R}^l \times \mathbb{R}^p \rightarrow \mathbb{R}^1\}$ is continuous uniformly in n ;

(A.2.ii) $q_t: \Omega \times \Theta \rightarrow \mathbb{R}^l$ is measurable for each $\theta \in \Theta$, a compact subset of \mathbb{R}^k , and continuous on Θ for each $\omega \in \Omega$, uniformly in $t=1,2,\dots$;

(A.2.iii) $\hat{\pi}_n: \Omega \rightarrow \mathbb{R}^p$ is measurable, $n=1,2,\dots$.

With this condition, it is straightforward to verify Assumption 2.2.i.

COROLLARY 2.3 (Existence): Given Assumptions A.1 and A.2, Assumption 2.2.i is satisfied, and there exists a measurable function $\hat{\theta}_n(\omega)$ such that

$$Q_n(\omega, \hat{\theta}_n(\omega)) = \inf_{\theta \in \Theta} Q_n(\omega, \theta) \quad \text{for all } \omega \text{ in } \Omega.$$

Proofs of all results are given in the Mathematical Appendix.

The form given in (2.1) is quite flexible. For example, letting $g_n(\psi, \pi) = -\psi$ and letting q_t be the (quasi-)log-likelihood of an observation yields the (quasi-)maximum likelihood estimator. Letting³ $\pi = \text{vech } P$ for symmetric matrices P and letting $g_n(\psi, \pi) = \psi'P\psi$ yields instrumental variables, method of moments or m-estimators.

It is typically convenient to view q_t as a composition of functions, e.g.

$$q_t(\omega, \theta) = s_t(W^t(\omega), \theta)$$

where $W^t(\omega) \equiv (W_1(\omega), \dots, W_t(\omega)) = \omega^t$ and $s_t: \mathbb{R}_t^s \times \Theta \rightarrow \mathbb{R}^l$ is appropriately measurable on \mathbb{R}_t^s and continuous on Θ . This allows us to see immediately that the least mean distance estimators of Burguete, Gallant and Souza [12] fall into the present class. In their notation, a least mean distance estimator minimizes an optimand of the form

$$Q_n(\omega, \theta) = n^{-1} \sum_{t=1}^n s(x_t, y_t, \hat{\tau}_n(\omega), \theta).$$

For many of these estimators (and in particular the feasible GLS estimators)

$$s(x_t, y_t, \hat{\tau}_n(\omega), \theta) = s_1(x_t, y_t, \theta) + s_2(\hat{\tau}_n(\omega)) + s_3(\hat{\tau}_n(\omega))' s_4(x_t, y_t, \theta)$$

where s_1 and s_2 are scalar functions and s_3 and s_4 are finite dimensional vectors. In this case

$$n^{-1} \sum_{t=1}^n s(x_t, y_t, \hat{\tau}_n(\omega), \theta) = a(\hat{\tau}_n(\omega))' n^{-1} \sum_{t=1}^n b(x_t, y_t, \theta)$$

where $a' \equiv (1, s_2, s_3')$ and $b' = (s_1, 1, s_4')$. Setting $g_n(\psi, \pi) = a(\pi)\psi$,

$q_t(\omega, \theta) = b(W_t(\omega), \theta)$, $W_t(\omega) \equiv (x_t, y_t) = \omega_t$ and $\hat{\pi}_n = \hat{\tau}_n$, we see that this is in the form (2.1).

A similar argument applies to many of the method of moments estimators discussed by Burguete, Gallant and Souza. An important exception, however, is the case of m-estimators which use a preliminary estimate of scale. These estimators are obtained as solutions to systems of equations

$$n^{-1} \sum_{t=1}^n m_{\theta}(x_t, y_t, \hat{\tau}_n(\omega), \theta) = 0$$

where $\hat{\tau}_n$ enters in a nonlinear manner and cannot be factored out in the way shown above. Nevertheless, if $\hat{\tau}_n$ is itself an m-estimator solving the equations

$$n^{-1} \sum_{t=1}^n m_{\tau}(x_t, y_t, \tau) = 0,$$

then the scaled m-estimator can be obtained as the solution to the system of equations

$$n^{-1} \sum_{t=1}^n m(x_t, y_t, \tau, \theta) = 0$$

where $m(x_t, y_t, \tau, \theta)' \equiv (m_{\tau}(x_t, y_t, \tau)', m_{\theta}(x_t, y_t, \tau, \theta)')$. Thus, the scaled m-estimator can be viewed as a particular m-estimator not involving nuisance parameters, and it therefore falls into the present framework. Similar techniques can be applied to least mean distance estimators which contain nuisance parameters τ_n entering nonlinearly. Thus, the estimators considered by Burguete, Gallant and Souza (and in particular all those they enumerate in Section 4) can be embedded in the present class.

Common feasible GLS estimators for the more general setting allowed here also fall into our class of estimators. For example, consider a system of linear regression equations

$$Y_t = X_t \theta + \varepsilon_t$$

in which the $l \times 1$ vector of errors ε_t is thought to have a VAR(1) structure

$$\varepsilon_t = R\varepsilon_{t-1} + \eta_t,$$

where η_t is assumed i.i.d., independent of ε_{t-1} and of X_τ , $\tau=1,2,\dots$ and

$$E(\eta_t \eta_t') = \Sigma.$$

For this model, one could construct a feasible GLS estimator by obtaining consistent estimates of R and Σ , say \hat{R}_n and $\hat{\Sigma}_n$, and then solving the problem

$$\begin{aligned} \min_{\theta \in \Theta} (n-1)^{-1} \sum_{t=2}^n (Y_t - X_t \theta - \hat{R}_n(Y_{t-1} - X_{t-1} \theta))' \hat{\Sigma}_n^{-1} \\ \times (Y_t - X_t \theta - \hat{R}_n(Y_t - X_{t-1} \theta)). \end{aligned}$$

This is a least mean distance estimator which has the same form as the feasible GLS estimators discussed by Burguete, Gallant and Souza, where $\hat{\tau}_n$ corresponds to \hat{R}_n , $\hat{\Sigma}_n$. Convenient choices for \hat{R}_n and $\hat{\Sigma}_n$ can be constructed using

$$\hat{\theta}_{OLS} = (X' X)^{-1} X' Y$$

where X is the $l \cdot n \times k$ matrix

$$X' \equiv [X_1', \dots, X_n']$$

and Y is the $l \cdot n \times 1$ vector

$$Y' = [Y_1', \dots, Y_n'] .$$

Letting

$$\hat{\varepsilon}' = [\hat{\varepsilon}_1', \dots, \hat{\varepsilon}_n']$$

$$\hat{\varepsilon}'_{-1} = [0, \hat{\varepsilon}_1, \dots, \hat{\varepsilon}_{n-1}],$$

where $\hat{\varepsilon}_t \equiv Y_t - X_t \hat{\theta}_{OLS}$, we form

$$\hat{R}_n \equiv \hat{\varepsilon}'_{-1} \hat{\varepsilon}_{-1} [\hat{\varepsilon}_{-1}' \hat{\varepsilon}_{-1}]^{-1}$$

and

$$\hat{\Sigma}_n = (\hat{\varepsilon} - \hat{\varepsilon}_{-1} \hat{R}'_n)' (\hat{\varepsilon} - \hat{\varepsilon}_{-1} \hat{R}'_n) / (n-1).$$

Note that this model only motivates our choice of estimator. It need not be correctly specified (e.g., ε_t could in fact be VAR(2)). Other feasible GLS estimators (e.g. those correcting for heteroskedasticity as in Chapter VII of White [46]) can be similarly treated.

To establish Assumption 2.2.ii for (2.1) we provide conditions which ensure the convergence of the arguments of g_n , i.e. $n^{-1} \sum_{t=1}^n q_t(\omega, \theta)$ and $\hat{\pi}_n$, and then verify that g_n is sufficiently well behaved to ensure the convergence of Q_n . This verification is accomplished using the following lemma, a generalization of Proposition 2.16 of White [46].

LEMMA 2.4 Let $\{g_n: \mathbb{R}^l \rightarrow \mathbb{R}^m\}$ be continuous uniformly in n . Suppose for all $n=1,2,\dots$ there exists $\psi_n: \Omega \times \Theta \rightarrow \mathbb{R}^l$ such that for each $\theta \in \Theta$, a compact subset of \mathbb{R}^k , $\psi_n(\omega, \theta)$ is measurable on Ω , and for each $\omega \in \Omega$, $\psi_n(\omega, \theta)$ is continuous on Θ . Also suppose for all $n=1,2,\dots$, there exists continuous $\bar{\psi}_n: \Theta \rightarrow \mathbb{R}^l$ such that $\psi_n(\omega, \theta) - \bar{\psi}_n(\theta) \rightarrow 0$ a.s. as $n \rightarrow \infty$, uniformly on Θ . Finally, suppose that for all $\theta \in \Theta$, $\psi_n(\theta)$ is interior to Ψ , a compact subset of \mathbb{R}^l , uniformly in n . Then $g_n(\psi_n(\omega, \theta)) - g_n(\bar{\psi}_n(\theta)) \rightarrow 0$ a.s. as $n \rightarrow \infty$, uniformly on Θ .

Given the continuity of g_n uniformly in n imposed in Assumption A.2.i, it will suffice that $n^{-1} \sum_{t=1}^n q_t(\omega, \theta)$ and $\hat{\pi}_n$ converge uniformly on Θ . Since $\hat{\pi}_n$ is independent of θ , it does so trivially. The uniform convergence of $n^{-1} \sum_{t=1}^n q_t(\omega, \theta)$ can be established using a uniform law of large numbers. A variety of such laws is available. Choice of the appropriate result depends on the behavior assumed for the stochastic process generating the data. If the process is assumed stationary and ergodic, the following generalization of Hoadley's [27] uniform law of large numbers is available.

THEOREM 2.5: Given Assumptions A.1 and A.2.ii, assume:

Assumption 2.5.i: For all t , $q_t(\omega, \theta) = q(T^t \omega, \theta)$ where $T: \Omega \rightarrow \Omega$ is measure preserving and one-to-one, and there exists d_t measurable- \mathcal{F} such that $d_t(\omega) = d(T^t \omega)$ and $|q_t(\omega, \theta)| \leq d_t(\omega)$ for all θ in Θ , where d_t is integrable;

Assumption 2.5.ii: $\{W_t\}$ is a stationary ergodic process such that $W_t(\omega) = W_1(T^{t-1} \omega)$.

Then $\{n^{-1} \sum_{t=1}^n E(q_t(\omega, \theta))\}$ is equicontinuous on Θ and $n^{-1} \sum_{t=1}^n [q_t(\omega, \theta) - E(q_t(\omega, \theta))] \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$ uniformly on Θ .

If the process is not assumed to be stationary but is instead heterogeneous, then an extension of Hoadley's [27] uniform law of large numbers due to Domowitz and White [17] is available. This result uses the following definitions.

DEFINITION 2.6: Measurable functions $d_t: \Omega \rightarrow \mathbb{R}^l$ are $r+\delta$ -integrable uniformly in t if and only if $E|d_{ti}(\omega)|^{r+\delta} \leq \Delta < \infty$ for all $t=1,2,\dots, i=1,\dots,l$.

DEFINITION 2.7: The mixing coefficients $\phi(m)$ and $\alpha(m)$ are defined as

$$\phi(m) \equiv \sup_n \left[\sup_{\{F \in \mathbb{F}_{-\infty}^n, G \in \mathbb{F}_{n+m}^\infty: P(F>0)\}} |P(G|F) - P(G)| \right]$$

$$\alpha(m) = \sup_n \left[\sup_{\{F \in \mathbb{F}_{-\infty}^n, G \in \mathbb{F}_{n+m}^\infty\}} |P(GF) - P(F)P(G)| \right]$$

where $\mathbb{F}_{-\infty}^n \equiv \sigma(\dots, W_n)$, $\mathbb{F}_{n+m}^\infty \equiv \sigma(W_{n+m}, \dots)$. If $\phi(m) = O(m^{-\lambda})$ for $\lambda > a$ then $\phi(m)$ is said to be of size a , and similarly for $\alpha(m)$.

The mixing coefficients $\phi(m)$ and $\alpha(m)$ measure the amount of dependence between events in the data generation process separated by at least m time periods. For more on mixing, see White [46]. To state the following result, we adopt the notation $\mathbb{F}_{t-\tau}^t \equiv \sigma(W_{t-\tau}, \dots, W_t)$. Theorem 2.5 of Domowitz and White [17] can now be stated.

THEOREM 2.8: Given Assumptions A.1 and A.2.ii, assume

Assumption 2.8.i: q_t is measurable- $\mathbb{F}_{t-\tau}^t$, $\tau < \infty$ and there exists d_t measurable- $\mathbb{F}_{t-\tau}^t$ such that $|q_t(\omega, \theta)| \leq d_t(\omega)$ (element by element) for all θ in Θ , and for $r \geq 1$ and some $\delta > 0$, d_t is $r+\delta$ integrable uniformly in t ;

Assumption 2.8.ii: $\{W_t\}$ is a mixing sequence such that either $\phi(m)$ is of size $r/(2r-1)$ or $\alpha(m)$ is of size $r/(r-1)$, $r > 1$.

Then $\{n^{-1} \sum_{t=1}^n E(q_t(\omega, \theta))\}$ is equicontinuous on Θ and $n^{-1} \sum_{t=1}^n [q_t(\omega, \theta) - E(q_t(\omega, \theta))] \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$ uniformly on Θ .

This result ensures a uniform law of large numbers under specific moment and memory conditions. Note that the memory conditions specify not only that $\phi(m)$ or $\alpha(m)$ tend to zero at a specific rate as $m \rightarrow \infty$, implying that the data generating process exhibits a form of asymptotic independence, but also that $q_t(\omega, \theta)$ is measurable $\mathcal{F}_{t-\tau}^t$, $\tau < \infty$, which implies that q_t depends on only a finite number of recent lagged values of W_t . This ensures that $\{q_t(\omega, \theta)\}$ is also a mixing process (see Theorem 3.49 of White [46]). The condition that $\tau < \infty$ is a great convenience, but not a necessity. For a uniform law of large numbers for functions of mixing processes with τ unrestricted, see Gallant and White [22]. The machinery required to allow $\tau \rightarrow \infty$ is rather complex; we shall avoid this complexity here.

We apply these results by imposing the following assumption.

ASSUMPTION A.3:

(A.3.i) $\{q_t\}$ and $\{W_t\}$ satisfy either Assumptions 2.5.i and 2.5.ii or Assumptions 2.8.i and 2.8.ii;

(A.3.ii) there exists a $O(1)$ sequence of non-stochastic p -vectors $\{\pi_n^*\}$ such that $\hat{\pi}_n - \pi_n^* \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$.

Finally, we need to ensure that Assumption 2.2.iii holds. For now we simply impose this directly for an appropriate choice of \bar{Q}_n . In Section 3 below we provide sufficient conditions related to the familiar rank condition for identification.

ASSUMPTION A.4: When it exists, the function $\bar{Q}_n: \Theta \rightarrow \mathbb{R}^1$ defined as

$$\bar{Q}_n(\theta) \equiv g_n \left(n^{-1} \sum_{t=1}^n E(q_t(\omega, \theta)), \pi_n^* \right)$$

has an identifiably unique minimizer θ_n^* in Θ for all n sufficiently large.

When the model is correctly specified, this condition is straightforward to verify using plausible primitive assumptions on the model and the estimator. Even when the model is not correctly specified, plausible primitive conditions are often easily available. For example, in the case of linear regression with $g_n(\psi, \pi) = \psi$, $q_t(\omega, \theta) = (Y_t - X_t\theta)^2$, it suffices that $E(X'X/n)$ is positive definite uniformly in n . The desired consistency result can now be stated.

THEOREM 2.9 (Consistency): Given Assumptions A.1-A.4, $\hat{\theta}_n - \theta_n^* \xrightarrow{a.s.} 0$.

In contrast to the consistency results of Burguete, Gallant and Souza [12], Theorem 2.9 allows for econometric models with error terms which are not i.i.d. Also, the explanatory variables of the model may be stochastic and contain both lagged dependent and explanatory variables, allowing for dynamics.

The consistency of the least mean distance and method of moments estimators are established as separate theorems in Burguete, Gallant and Souza (Theorems 2 and 5, respectively). Both types of estimators are covered by Theorem 2.9 by considering an optimand of the form (2.1). We illustrate how this is accomplished in the following two sections.

3. CONSISTENCY OF MAXIMUM LIKELIHOOD ESTIMATION

Sufficient conditions for the consistency of the maximum likelihood estimator (MLE) of Fisher [19,20] have been given for the case of a correctly

specified model by an impressive list of authors. Among the important early contributions are those of Doob [16], Cramér [14], Wald [38,39] and Le Cam [32]. Of these, Wald [38] is the first to have provided conditions ensuring the consistency of the MLE for generally dependent stochastic processes. Subsequent results for dependent processes following Wald's [38] approach have been given by Bar-Shalom [6], Weiss [40,41], Crowder [15] and Hall and Heyde [24], among others. This approach, however, studies maximum likelihood estimators obtained as solutions to the first order conditions for a maximum of the likelihood function. Such estimators are appropriately treated using methods illustrated in the following section.

In this section we treat estimators obtained directly as the solution to a maximization problem, thereby avoiding differentiability assumptions. This is the approach taken by Doob [16], Wald [39] and Le Cam [32], although none of these treat the case of dependent observations or estimation of misspecified models. Results for dependent observations have been given by Silvey [37], Bhat [10], Caines [13] and Heijmans and Magnus [26]. Results for misspecified models with independent observations have been given by Huber [28]. The results presented here allow for both dependent observations and misspecified models, and are closely related to the results of White [47]. Since we allow for the possibility of a misspecified model, we shall refer to the maximum likelihood estimator for a misspecified model as a quasi-maximum likelihood estimator (QMLE).

Given the results of the previous section, very little effort is needed to establish general conditions ensuring the consistency of the QMLE for dependent data processes. It suffices simply to choose g_n and q_t appropriately and to provide conditions ensuring that the conditions of Theorem 2.9 hold for these choices.

We maintain Assumption A.1. We replace Assumption A.2 with the following assumption.

ASSUMPTION B.1:

(B.1.i) For all n , $g_n(\psi, \pi) = -\psi$, where $\psi \in \mathbb{R}^1$.

(B.1.ii) $q_t(\omega, \theta) = \ln f_t(Y_t | Z_t, \theta)$ satisfies Assumption 2.3.ii where for each θ in Θ , f_t is a conditional density of Y_t given Z_t , where $W_t = (X_t, Y_t)$ and $Z_t(\omega)$ is measurable- $\sigma(\dots W_{t-1}, X_t)$.

Since $\hat{\pi}_n$ is irrelevant, no conditions apply to it. With this assumption, the estimation problem becomes

$$\min_{\theta \in \Theta} Q_n(\omega, \theta) = -n^{-1} \sum_{t=1}^n \ln f_t(Y_t | Z_t, \theta).$$

Note that the possibility for misspecification arises because nothing acts to guarantee that the conditional density of Y_t given Z_t implied by Assumption A.1 is given by $f_t(Y_t | Z_t, \theta^0)$ for some θ^0 in Θ . Also note that Z_t can be any subset of the vector $(W_1, \dots, W_{t-1}, X_t)$. If Z_t is measurable with respect to a sub σ -algebra of $\sigma(W_1, \dots, W_{t-1}, X_t)$ the model specified may also represent a dynamic misspecification in the sense that the true conditional density of Y_t given X_t and past W_t 's may well involve all these variables.

When W_t is an independent sequence and $Z_t = X_t$, the estimator under consideration is the conditional maximum likelihood estimator of Andersen [5]. Andersen recommends use of this estimator in situations where one is conditioning on sufficient statistics for incidental parameters. The present formulation allows a generalization of Andersen's approach to the case of dependent

observations. Note that X_t may be null, so that $Y_t = W_t$, allowing for unconditional maximum likelihood as well.

The existence of the quasi-maximum likelihood estimator can be established immediately.

COROLLARY 3.1: Given Assumptions A.1 and B.1, there exists a measurable function $\hat{\theta}_{QML}(\omega)$ such that

$$Q_n(\omega, \hat{\theta}_{QML}(\omega)) = \sup_{\theta \in \Theta} n^{-1} \sum_{t=1}^n \ln f_t(Y_t | Z_t, \theta).$$

To obtain consistency, we maintain the relevant part of Assumption A.3 (i.e. A.3.i) and maintain Assumption A.4. Without further structure, θ_n^* of Assumption A.4 must simply be interpreted as the parameter value which minimizes

$$\bar{Q}_n(\theta) = -n^{-1} \sum_{t=1}^n E(\ln f_t(Y_t | Z_t, \theta)).$$

Under additional mild conditions, White [48, chapter 4] shows that θ_n^* can be given an information theoretic interpretation in terms of the Kullback-Leibler [31] Information Criterion, in a manner analogous to that provided by Akaike [1] and White [45] for the case of independent observations. White [48, chapter 5] also provides more primitive conditions on the model which help ensure that Assumption A.4 is satisfied.

It should be pointed out that when W_t is heterogeneous and the mixing conditions are utilized, then maintaining Assumption A.3.i may force a dynamic misspecification since $f_t(Y_t | Z_t, \theta)$ is required by A.3.i to depend only on a finite number of lagged W_t 's, whereas a correct specification might require

$f_t(Y_t|Z_t, \theta)$ to depend on all past W_t 's. Thus, while correct specification for certain heterogeneous Markov processes is allowed, correct specification for heterogeneous non-Markov processes is not necessarily allowed. Stationary ergodic non-Markov processes are correctly specifiable under A.3.i, however. Results for heterogeneous non-Markov mixing processes are covered by results of Gallant and White [22].

The consistency result is

COROLLARY 3.2: Given Assumptions A.1, B.1, A.3.i and A.4, $\hat{\theta}_{QML} - \theta_n^* \xrightarrow{a.s.} 0$.

Recently, Levine [33] has shown that maximum likelihood estimators can be consistent despite the presence of dynamic misspecification. The key condition is as follows.

ASSUMPTION B.2: The conditional density of Y_t given Z_t implied by Assumption A.1 is given by $f_t(Y_t|Z_t, \theta^0)$ for some θ^0 in Θ , $t=1,2,\dots$

COROLLARY 3.3: Given Assumptions A.1, B.1, A.3.i, A.4 and B.2, $\hat{\theta}_{QML} \xrightarrow{a.s.} \theta^0$.

Note that A.4 is still required even in the presence of B.2.

To give some additional insight into the content of Corollary 3.3, we compare it with the results of Crowder [15] for consistency of the maximum likelihood estimator. The first distinction to be drawn is that the present results apply to conditional maximum likelihood estimators with the possible presence of dynamic misspecification. Crowder's results apply to correctly

specified unconditional maximum likelihood estimators. Next, no differentiability assumptions are imposed here in obtaining the present consistency results; Crowder's conditions require that the likelihood function be continuously differentiable of order two.

Crowder allows his likelihood function for a given observation to depend on all previous observations. We explicitly allow this for stationary ergodic processes and explicitly rule this out for heterogeneous mixing processes on grounds of convenience. The extension to allow dependence on all previous observations for heterogeneous mixing processes is non-trivial, and this issue is not explicitly addressed by Crowder. For this extension, see Gallant and White [22]. Crowder does consider an example involving a stationary mixing process (hence an ergodic process). The conditions imposed on the mixing coefficients here are weaker than those imposed by Crowder in his example.

The area in which Crowder's results do allow considerably greater flexibility is that Crowder does not impose domination conditions, as we do here. Thus, Crowder's results apply to such situations as ordinary least squares applied to a linear model with a time trend. This situation is not covered by our results. Such situations can be covered in a framework closely related to that adopted here by dispensing with Assumptions 2.2.ii and 2.2.iii, and replacing these with weaker conditions, as in Wooldridge [50].

Finally, Crowder provides an elegant treatment of the incidental parameters problem. We do not address that issue here.

4. CONSISTENCY OF M-ESTIMATORS AND GENERALIZED METHOD OF MOMENTS ESTIMATORS

An extremely useful estimation technique introduced by Huber [28] arises when the assumption that

$$\bar{\psi}_n(\theta_n^*) \equiv n^{-1} \sum_{t=1}^n E(q_t(\omega, \theta_n^*)) \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

for some sequence $\{\theta_n^*\}$ in Θ is used to construct an estimator $\hat{\theta}_M(\omega)$ which satisfies

$$\hat{\psi}_n(\hat{\theta}_M(\omega)) \equiv n^{-1} \sum_{t=1}^n q_t(\omega, \hat{\theta}_M(\omega)) \xrightarrow{\text{a.s.}} 0.$$

Such estimators are called *m*-estimators. (In Huber's i.i.d. setup, θ_n^* is independent of n . Here, it is natural to allow θ_n^* to depend on n .) In certain cases, such an estimator may be obtained directly as the solution to the problem

$$\hat{\psi}_n(\theta) = 0.$$

The maximum likelihood estimator occurs as the special case for which

$q_t(\omega, \theta) = \nabla_{\theta} \ln f_t(Y_t | Z_t, \theta)$, provided that the log-likelihood function is differentiable. The instrumental variables estimator of Reiserø [34] and Geary [23] occurs as the special case in which $q_t(\omega, \theta) = Z_t'(Y_t - X_t\theta)$, where now $W_t = (X_t, Y_t, Z_t)$ and Z_t is a vector of the same dimension as X_t .

In other cases, no such solution will exist (e.g., when q_t is of greater dimension than θ), but regardless of whether it does or not, an estimator can be obtained by finding the value $\hat{\theta}_M(\omega)$ which makes $\hat{\psi}_n(\theta)$ as close to zero as possible. A common measure of closeness is a metric for the space in which $\hat{\psi}_n(\theta)$ takes its values. However, a metric has properties irrelevant for present purposes. Here it suffices to measure closeness in the following way.

DEFINITION 4.1: A function $D_n: \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}^1$ is called a discrepancy from zero if and only if $D_n(\psi, 0) \geq 0$ for all ψ in \mathbb{R}^k and $D_n(\psi, 0) = 0$ if and only if $\psi = 0$.

For simplicity, D_n satisfying this definition will be called simply a

"discrepancy." In general, we could then choose $\hat{\theta}_M(\omega)$ to solve the problem

$$\min_{\theta \in \Theta} D_n(\hat{\psi}_n(\theta), 0).$$

However, in practice, one might wish to choose D_n to depend implicitly on unknown parameters π_n^* . This dependence can be made explicit by writing

$$\varepsilon_n(\hat{\psi}_n(\theta), \pi_n^*) = D_n(\hat{\psi}_n(\theta), 0).$$

We obtain an estimator as the solution to the problem

$$\min_{\theta \in \Theta} Q_n(\omega, \theta) = \varepsilon_n(\hat{\psi}_n(\theta), \hat{\pi}_n) \equiv \hat{D}_n(\hat{\psi}_n(\theta), 0),$$

where $\hat{\pi}_n$ is a consistent estimator for π_n^* . The theory of Section 2 applies immediately to such estimators. Existence follows from Corollary 2.3 and consistency from Theorem 2.9. Note that θ_n^* is defined as the parameter vector minimizing $D_n(\bar{\psi}_n(\theta), 0)$ in A.4. Because it is not required that $D_n(\bar{\psi}_n(\theta_n^*), 0)$ attain the value zero, the equations $\bar{\psi}_n(\theta) = 0$ which motivate the m-estimator are allowed to be misspecified.

For the case in which $\bar{\psi}_n(\theta_n^*) = 0$, or more generally when $\bar{\psi}_n(\theta_n^*) \rightarrow 0$ as $n \rightarrow \infty$ for some sequence $\{\theta_n^*\}$, a condition closely related to the rank condition for the identification of the parameters of systems of simultaneous equations can be shown to assist in verifying Assumption A.4. In the standard linear simultaneous equations model identification of the true parameter vector θ^0 relies on establishing that the familiar rank and order conditions are satisfied (see e.g., Fisher [18]). These assumptions ensure that distinct parameter values give rise to distinct values for the function $\bar{\psi}_n$, i.e., if this function is not one-to-one at least in a neighborhood of the true parameter vector, then the true parameter vector is not identifiable.⁴

For the linear case, a necessary condition for $\bar{\psi}_n(\theta)$ to be one-to-one is the familiar order condition. This guarantees that the range space is "large" enough for $\bar{\psi}_n(\theta)$ to be one-to-one. Clearly, this is a necessary condition for identifiability in any more general case, regardless of whether or not the probability model is correctly specified. The following definition is appropriate for the present case.

DEFINITION 4.2: Let $\{\bar{\psi}_n: \Theta \rightarrow \mathbb{R}^l\}$ be a sequence of functions. $\{\bar{\psi}_n\}$ satisfies the generalized order condition if and only if $\dim(\Theta) \equiv k \leq l$.

The rank condition of the linear case is both necessary and sufficient for identification of the parameters of a correctly specified model. When it is satisfied, the image of the parameter space under the linear function is of dimension k and there exists a one-to-one, onto mapping between the parameter space and a subset of the range space of $\bar{\psi}_n$.

To establish identifiability in the present context, a condition slightly stronger than the one-to-oneness of the individual functions $\bar{\psi}_n$ is required. This is because the one-to-oneness may vanish in the limit even though it is satisfied for each of the individual functions. This could make identifiability impossible in the limit. Sequences which satisfy the following definition will avoid this problem. Furthermore, in the standard linear simultaneous equations model, it guarantees that the standard rank and order conditions are satisfied.

Let $\|\cdot\|$ denote the standard Euclidean norm.

DEFINITION 4.3: Let $\{\bar{\psi}_n : \theta \rightarrow \mathbb{R}^l\}$ be a sequence of functions, and let $\theta_n^* \in \eta \subset \theta$
 $n = 1, 2, \dots$. $\{\bar{\psi}_n\}$ satisfies the generalized rank condition at $\{\theta_n^*\}$ on η if
 and only if for every $\epsilon > 0$ there exists $\delta > 0$ such that for any $\theta \in \eta$ for which
 $\|\theta - \theta_n^*\| > \epsilon$ it follows that $\|\bar{\psi}_n(\theta) - \bar{\psi}_n(\theta_n^*)\| > \delta$ for all n sufficiently
 large.

Although it can be shown that having the generalized rank condition hold on
 an appropriate set η is necessary for Assumption A.4 to be satisfied, it is not
 sufficient. This is because even though true parameters for a correctly
 specified model may be identified, the estimator defined by choice of D_n may fail
 to have a unique limit, and therefore fail to be identifiable.

In the present context, we control the behavior of the discrepancies D_n
 so that the limiting discrepancy from zero of a vector is bounded away from zero
 except when the vector is zero. The following definition guarantees this and
 somewhat more.

DEFINITION 4.4: Let $\{D_n\}$ be a sequence of discrepancies. $\{D_n\}$ is an
asymptotically uniform sequence of discrepancies if and only if there exists a
 monotone transformation h continuous at zero with $h(0) = 0$ and constants $\delta > 0$
 and $\Delta < \infty$ such that for any $\psi \in \mathbb{R}^l$, $\delta \|\psi\| \leq h(D_n(\psi, 0)) \leq \Delta \|\psi\|$ for all n
 sufficiently large.

The following result shows that by imposing the generalized rank condition
 and by choosing an asymptotically uniform sequence $\{D_n\}$ to define an estimator,
 one can verify the identifiable uniqueness Assumption A.4.

THEOREM 4.5: Suppose that

Assumption 4.5.i: $\{D_n(\psi, 0) \equiv g_n(\psi, \pi_n^*)\}$ is an asymptotically uniform sequence of discrepancies;

Assumption 4.5.ii: There exists a sequence $\{\theta_n^*\}$ in Θ , a compact subset of R^k , such that $\bar{\psi}_n(\theta_n^*) \rightarrow 0$ as $n \rightarrow \infty$;

Assumption 4.5.iii: There exist a constant $\rho > 0$ and integer $N_0 < \infty$ such that $\{\bar{\psi}_n\}$ satisfies the generalized rank condition at $\{\theta_n^*\}$ on $\Theta(\rho, N_0) \equiv [\theta \in \Theta : D_n(\bar{\psi}_n(\theta), 0) < \rho \text{ for some } n \geq N_0]$.

Then θ_n^* is an identifiably unique minimizer of

$$\bar{Q}_n(\theta) \equiv g_n(\bar{\psi}_n(\theta), \pi_n^*) \equiv D_n(\bar{\psi}_n(\theta), 0).$$

By showing how Assumption A.4 can be verified when the probability model is correctly specified in the sense of Assumption 4.5.ii, this result provides useful additional insight into the meaning of Assumption A.4.

We illustrate the use of this result and those of the Section 2 by providing a consistency result for the generalized method of moments (GMM) estimator of Hansen [25]. Among other things, this estimator is useful in estimating the parameters of the implicit nonlinear system of simultaneous equations

$$U_t = F_t(\omega, \theta^0), \quad t=1, \dots, n$$

when it is known that there exists

$$Z_t = G_t(\omega, \theta^0), \quad t=1, \dots, n$$

such that

$$E(U_t \otimes Z_t) = 0.$$

Letting

$$q_t(\omega, \theta) = F_t(\omega, \theta) \otimes G_t(\omega, \theta),$$

it follows that

$$\bar{\psi}_n(\theta^0) = n^{-1} \sum_{t=1}^n E(q_t(\omega, \theta^0)) = 0.$$

The GMM estimator is obtained by making the choices formalized in the following assumption.

ASSUMPTION C.1:

(C.1.i): For all n , $\varepsilon_n(\psi, \pi) = \psi' P \psi$, where $\psi \in R^l$ and P is a symmetric positive semi-definite matrix such that $\pi = \text{vech } P$;

(C.1.ii): $q_t(\omega, \theta)$ satisfies Assumption A.2.ii;

(C.1.iii): there exists $\{\hat{P}_n(\omega)\}$, a sequence of symmetric positive semi-definite matrices such that $\hat{\pi}_n(\omega) = \text{vech } \hat{P}_n(\omega)$ satisfies Assumption A.2.iii.

Given these choices, the GMM estimator solves the problem

$$\min_{\theta \in \Theta} \left(n^{-1} \sum_{t=1}^n q_t(\omega, \theta) \right)' \hat{P}_n(\omega) \left(n^{-1} \sum_{t=1}^n q_t(\omega, \theta) \right)$$

Existence of the GMM estimator follows immediately from Corollary 2.3.

COROLLARY 4.6: Given Assumptions A.1 and C.1, there exists a measurable function $\hat{\theta}_{\text{GMM}}(\omega)$ such that

$$Q_n(\omega, \hat{\theta}_{\text{GMM}}(\omega)) = \inf_{\theta \in \Theta} \left(n^{-1} \sum_{t=1}^n q_t(\omega, \theta) \right)' \hat{P}_n(\omega) \left(n^{-1} \sum_{t=1}^n q_t(\omega, \theta) \right).$$

To obtain consistency, we impose the following assumptions.

ASSUMPTION C.2:

(C.2.i): Assumption A.3.i holds for $\{q_t\}$ and $\{W_t\}$.

(C.2.ii) There exists a $O(1)$ sequence of non-stochastic symmetric positive and semi-definite matrices P_n^* such that $\hat{P}_n - P_n^* \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$.

Given Assumption C.2.ii, we can identify the discrepancy measure as

$$D_n(\psi_1, \psi_2) = (\psi_1 - \psi_2)' P_n^* (\psi_1 - \psi_2).$$

Note that $D_n(\psi_1, \psi_2)^{1/2}$ is a weighted Euclidean norm. Thus,

$$\varepsilon_n(\bar{\psi}_n(\theta), \pi_n^*) = D_n(\bar{\psi}_n(\theta), 0)$$

where $\pi_n^* = \text{vech } P_n^*$.

ASSUMPTION C.3:

(C.3.i) $\{P_n^*\}$ is uniformly positive definite;

(C.3.ii) there exists θ^0 in Θ such that $n^{-1} \sum_{t=1}^n E(q_t(\omega, \theta^0)) \rightarrow 0$ as $n \rightarrow \infty$;

(C.3.iii) there exists a constant $\rho > 0$ and an integer $N_0 < \infty$ such that

$\{\bar{\psi}_n(\theta) \equiv n^{-1} \sum_{t=1}^n E(q_t(\omega, \theta))\}$ satisfies the generalized rank condition at θ^0 on

$\Theta(\rho, N_0)$ as defined in Assumption 4.5.iii.

The consistency result for the GMM estimator is

COROLLARY 4.7: Given Assumptions A.1, C.1, C.2 and C.3, $\hat{\theta}_{GMM} \xrightarrow{a.s.} \theta^0$.

This result is broadly similar to Theorem 2.1 of Hansen [25], although it differs in certain particulars. First, Hansen requires $\{W_t\}$ to be strictly stationary and ergodic. Although we allow this possibility, we also allow for heterogeneous processes. Next, Hansen only requires (Θ, σ) to be a separable metric space with Θ locally compact. Our assumption that Θ is a compact subset of a finite dimensional Euclidean space suffices for this. Although this is a commonly encountered situation in econometrics, the main result of Theorem 2.9 is easily extended to allow (Θ, σ) to be a separable metric space.

Hansen requires the function q to be time-invariant, but we do not. Similarly, Hansen requires P_n^* to converge to a constant limit. This requirement also is not imposed here. These are natural restrictions for Hansen since his data are assumed stationary. Here we require somewhat greater flexibility.

In place of our domination conditions, Hansen imposes a super-continuity requirement (see Definition 2.2, Hansen [25]). While this condition appears somewhat weaker, it may also be somewhat more difficult to verify. Finally, Hansen directly assumes identification (hence identifiability). We provide useful sufficient conditions related to the familiar rank condition as embodied in Assumption C.3. With stationarity, it is natural to assume for identification that $E(q_t(\omega, \theta^0)) = 0$ for all $t=1,2,\dots$ as Hansen does. Here we adopt the weaker requirement that $n^{-1} \sum_{t=1}^n E(q_t(\omega, \theta^0)) \rightarrow 0$ as $n \rightarrow \infty$.

Despite these differences, the present result is clearly identical in spirit to Hansen's result. In fact, for the consistency property, it represents a useful version of the extension to the nonstationary context which Hansen [25] suggests in his concluding remarks.

As Hansen notes, his results cover the nonlinear instrumental variables estimators of Amemiya [3,4], Jorgensen and Laffont [30] and Gallant [22]. The present result also covers the nonlinear least squares estimators of White [43] and White and Domowitz [49].

5. CONCLUSION

This paper has presented a theory of consistent estimation for parametric models which bridges the gap between the general consistency result of Domowitz and White [17] and the myriad of single cases characterizing much of the econometrics literature, and which extends the unification of Burguete, Gallant and Souza to the non-i.i.d. case. Two special cases, (quasi-)maximum likelihood and generalized method of moments estimators, are discussed in some detail. These examples by no means exhaust the possible applications of the present theory. For example, minimum chi-square estimation as discussed by Rothenberg [36, p. 24] can be treated by setting $q_t(\omega, \theta) = \theta$, $\hat{\pi}_n = (\hat{\theta}_n, \text{vech } \hat{a\hat{v}ar} \hat{\theta}_n)$ and defining

$$g_n(n^{-1} \sum_{t=1}^n q_t(\omega, \theta), \hat{\pi}_n) = (\hat{\theta}_n - \theta)' [\hat{a\hat{v}ar} \hat{\theta}_n]^{-1} (\hat{\theta}_n - \theta)$$

where $\Theta = \{\theta \in \Theta^*: h(\theta) = 0\}$, and Θ^* is a compact subset of a finite dimensional Euclidean space.

The present approach also provides a convenient framework for establishing asymptotic normality results, and for selecting the most efficient estimator within any subclass of the estimators covered by this theory. These topics are addressed in a subsequent paper (Bates and White [7]).

Johns Hopkins University

University of California San Diego

and M.I.T.

MATHEMATICAL APPENDIX

PROOF OF THEOREM 2.3: As g_n is continuous on $\mathbb{R}^k \times \mathbb{R}^p$, it follows from Theorem 13.3 of Billingsley [1979] that $Q_n(\omega, \theta) = g_n(n^{-1} \sum_{t=1}^n q_t(\omega, \theta), \hat{\pi}_n(\omega))$ is measurable for each θ given the measurability of q_t and $\hat{\pi}_n$. As g_n is continuous and q_t is continuous on Θ for each ω , it follows that $Q_n(\omega, \theta)$ is continuous on Θ for each ω . By Assumption A.2.ii, Θ is a compact subset of \mathbb{R}^k . Thus Assumption 2.2.i is satisfied and the existence of $\theta_n(\omega)$ follows from Theorem 2.2, given Assumption A.1.

PROOF OF LEMMA 2.4: As $\{g_n\}$ is continuous uniformly in n , given $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$ such that if $x, y \in \Phi$, a compact set, and $\|x - y\| < \delta(\varepsilon)$, then $\|g_n(x) - g_n(y)\| < \varepsilon$.

For any $\varepsilon > 0$, define $\delta(\varepsilon)$ as immediately above. Since $\phi_n(\omega, \theta) - \bar{\phi}_n(\theta) \xrightarrow{a.s.} 0$ uniformly on Θ , there exists F in \mathcal{F} , $P(F) = 1$ such that for (any) $\delta(\varepsilon) > 0$ and each ω in F , there exists an integer $N_0(\omega, \delta(\varepsilon))$ such that for all $n > N_0(\omega, \delta(\varepsilon))$

$$\sup_{\theta \in \Theta} \|\phi_n(\omega, \theta) - \bar{\phi}_n(\theta)\| < \delta(\varepsilon).$$

Further, since $\bar{\phi}_n(\theta)$ is interior to Φ uniformly in n for all θ in Θ , for each ω in F there exists $N_1(\omega)$ such that for all $n > N_1(\omega)$ $\phi_n(\omega, \theta)$ lies interior to Φ

for all θ in Θ . Thus for each ω in F there exists $N_2(\omega, \varepsilon) = \max(N_0(\omega, \delta(\varepsilon)), N_1(\omega))$ such that for all $n > N_2(\omega, \varepsilon)$ $\psi_n(\omega, \theta)$ and $\bar{\psi}_n(\theta)$ are in Φ and $\|\psi_n(\omega, \theta) - \bar{\psi}_n(\theta)\| < \delta(\varepsilon)$ for all θ in Θ . It follows from the continuity of $\{g_n\}$ uniformly in n that for each ω in F $\|g_n(\psi_n(\omega, \theta)) - g_n(\bar{\psi}_n(\theta))\| < \varepsilon$ for all $n > N_2(\omega, \varepsilon)$ and all θ in Θ . Since ε is arbitrary and $P(F)=1$, it follows that $g_n(\psi_n(\omega, \theta)) - g_n(\bar{\psi}_n(\theta)) \rightarrow 0$ a.s. as $n \rightarrow \infty$ uniformly on Θ .

PROOF OF THEOREM 2.5: The proof is identical to that of Jennrich's [29] Theorem 2, except that the Ergodic Theorem (e.g. White [46, Theorem 3.34]) is applied in place of Komolgorov's law of large numbers for i.i.d. random variables.

PROOF OF THEOREM 2.9: We verify the conditions of Theorem 2.2. Given Assumptions A.1 and A.2, Assumption 2.2.ii is satisfied and $\hat{\theta}_n(\omega)$ exists by Corollary 2.3. Given Assumption A.3.i it follows from either Theorem 2.5 or Theorem 2.8 that $n^{-1} \sum_{t=1}^n [q_{ti}(\omega, \theta) - E(q_{ti}(\omega, \theta))] \xrightarrow{a.s.} 0$ uniformly on Θ . By Assumption A.3.ii, $\hat{\pi}_n(\omega) - \pi_n^* \xrightarrow{a.s.} 0$. Identifying $(n^{-1} \sum_{t=1}^n q_t(\omega, \theta)', \hat{\pi}_n(\omega)')$ with $\psi_n(\omega, \theta)$ and $(n^{-1} \sum_{t=1}^n E(q_{ti}(\omega, \theta))', \pi_n^*)$ with $\bar{\psi}_n(\theta)$, it follows that $\psi_n(\omega, \theta) - \bar{\psi}_n(\theta) \xrightarrow{a.s.} 0$ uniformly on Θ . The domination conditions on q_t ensure that $E(q_t(\omega, \theta))$ and therefore $n^{-1} \sum_{t=1}^n E(q_{ti}(\omega, \theta))$ are $O(1)$ uniformly in θ ; by A.3.ii, π_n^* is $O(1)$. Therefore, $\bar{\psi}_n(\theta)$ is interior to a compact subset of R^{l+p} uniformly in n . As $\{g_n\}$ is continuous uniformly in n it follows from Lemma 2.4 that $g_n(\psi_n(\omega, \theta)) - g_n(\bar{\psi}_n(\theta)) \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$ uniformly on Θ , i.e. $Q_n(\omega, \theta) - \bar{Q}_n(\theta) \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$, setting $Q_n(\omega, \theta) = g_n(\psi_n(\omega, \theta))$ and $\bar{Q}_n(\theta) = g_n(\bar{\psi}_n(\theta))$. Hence Assumption

2.2.ii holds.

Given Assumption A.4, Assumption 2.2.iii is satisfied. Thus, the conditions of Theorem 2.2 are satisfied, so that $\hat{\theta}_n(\omega) - \theta_n^* \xrightarrow{a.s.} 0$.

PROOF OF COROLLARY 3.1: The result follows by verifying that Assumption B.1 implies Assumption A.2 so that the conditions of Theorem 2.3 hold. This is trivial because $g_n(\psi, \pi) = -\psi$ is obviously continuously uniformly in n , Assumption A.2.ii holds by Assumption B.1.ii, and Assumption A.2.iii is irrelevant.

PROOF OF COROLLARY 3.2: That $\hat{\theta}_{QML} - \theta_n^* \xrightarrow{a.s.} 0$ follows immediately from Theorem 2.9 since Assumption B.1 implies A.2, and the other conditions are assumed directly.

PROOF OF COROLLARY 3.3: See White [48], Theorem 4.7, for the demonstration that $\theta_n^* = \theta^0$.

PROOF OF THEOREM 4.5: Let $\eta_n \equiv S_n(\varepsilon) \cap \Theta$ as in Definition 2.1. For some $\varepsilon > 0$ suppose that

$$\liminf_n \left[\min_{\theta \in \eta_n^c} D_n(\bar{\psi}_n(\theta), 0) - D_n(\bar{\psi}_n(\theta_n^*), 0) \right] = 0$$

so that θ_n^* is not identifiably unique. Because $D_n(\bar{\psi}_n(\theta_n^*), 0) \rightarrow 0$ given

$\bar{\psi}_n(\theta_n^*) \rightarrow 0$ (Assumption 4.5.ii) and $D_n(\bar{\psi}_n(\theta_n^*), 0) < h^{-1}(\Delta \|\bar{\psi}_n(\theta_n^*)\|)$ (Assumption

4.5.i), and because $\inf_{n \geq m} \left[\min_{\theta \in \eta_n^c} D_n(\bar{\psi}_n(\theta), 0) \right]$ is monotone increasing in m

we have

$$\inf_{n \geq m} [\min_{\theta \in \eta_n^c} D_n(\bar{\psi}_n(\theta), 0)] = 0.$$

Thus, for any $\xi > 0$ and any $M < \infty$ there exists $n > M$ such that

$$\min_{\theta \in \eta_n^c} D_n(\bar{\psi}_n(\theta), 0) < \xi. \text{ Choosing } \xi < \rho, \text{ it follows that there exists a}$$

sequence $\{\tilde{\theta}_n\}$ with $\tilde{\theta}_n$ in η_n^c (so that $\|\tilde{\theta}_n - \theta_n^*\| > \varepsilon$ for all n) such that for any $M < \infty$ there exists $n > M$ for which $D_n(\bar{\psi}_n(\tilde{\theta}_n), 0) < \xi < \rho$. This implies there exists $n > M$ such that $\tilde{\theta}_n$ is also in $\Theta(\rho, M)$ for any $M < \infty$. Since

$D_n(\bar{\psi}_n(\theta_n^*), 0) \rightarrow 0$, for any $\xi > 0$ there exists $M(\xi)$ sufficiently large so that for all $n > M(\xi)$, $D_n(\bar{\psi}_n(\theta_n^*), 0) < \xi$. Further, since $\{D_n\}$ is an asymptotically

uniform sequence of discrepancies by Assumption 4.5.i, it follows that there

exist $\delta > 0$ and $N_1 < \infty$ such that for all $n > N_1$, $\delta \|\bar{\psi}_n(\theta_n^*)\| \leq h(D_n(\bar{\psi}_n(\theta_n^*), 0))$ and $\delta \|\bar{\psi}_n(\tilde{\theta}_n)\| \leq h(D_n(\bar{\psi}_n(\tilde{\theta}_n), 0))$. Given N_0 as specified in Assumption 4.5.iii, it follows that for any ξ such that $0 < \xi < \rho$ and some $n > \max(M(\xi), N_0, N_1)$

$$\begin{aligned} \delta \|\bar{\psi}_n(\tilde{\theta}_n)\| + \delta \|\bar{\psi}_n(\theta_n^*)\| &\leq h(D_n(\bar{\psi}_n(\tilde{\theta}_n), 0)) + h(D_n(\bar{\psi}_n(\theta_n^*), 0)) \\ &< 2h(\xi). \end{aligned}$$

By the triangle inequality

$$\begin{aligned} \|\bar{\psi}_n(\tilde{\theta}_n) - \bar{\psi}_n(\theta_n^*)\| &\leq \|\bar{\psi}_n(\tilde{\theta}_n)\| + \|\bar{\psi}_n(\theta_n^*)\| \\ &< 2h(\xi)/\delta \end{aligned}$$

for some $n > \max(M(\xi), N_0, N_1)$ and any ξ such that $0 < \xi < \rho$, where

$\tilde{\theta}_n \in \Theta(\rho, N_0)$. However, since $\tilde{\theta}_n$ is also in η_n^c , and since h is continuous at

zero this violates Assumption 4.5.iii which requires the generalized rank

condition to hold on $\Theta(\rho, N_0)$, a contradiction. Hence θ_n^* must be identifiably unique.

PROOF OF COROLLARY 4.6: The result follows by verifying that Assumption C.1 implies Assumption A.2 so that the conditions of Theorem 2.3 hold. This is trivial because $g_n(\phi, \pi) = \phi'P\phi$, where $\pi = \text{vech } P$, is obviously continuous uniformly in n , Assumption A.2.ii holds by Assumption C.1.ii, and Assumption A.2.iii holds by Assumption C.1.iii.

PROOF OF COROLLARY 4.7: We verify the conditions of Theorem 2.1. Assumption A.1 is given, Assumption C.1 implies Assumption A.2 as just argued, Assumption C.2 directly implies Assumption A.3, and Assumption A.4 is satisfied for $\theta_n^* = \theta^0$ as a consequence of Theorem 4.5, given Assumption C.3. We note that Assumptions C.1.i and C.3.i ensure that

$$D_n(\phi, 0) = \phi'P_n^*\phi$$

is an asymptotically uniform sequence of discrepancies.

Footnotes

¹The authors are grateful to A. Ronald Gallant, Jeffrey Wooldridge, the editor, and especially to an anonymous referee for helpful comments and suggestions. Any remaining errors are attributable to the authors. This work was supported by NSF grant SES83-0635.

²For $\bar{Q}_n: \Theta \rightarrow \mathbb{R}$, let $A_n \equiv \{(\theta, x) \in \Theta \times \mathbb{R}^1: \bar{Q}_n(\theta) \leq x\}$. Then A_n is the epigraph of \bar{Q}_n (Roberts and Varberg [35, p. 80]).

³The operator vech maps the lower triangle of a symmetric $l \times l$ matrix into an $l(l+1)/2 \times 1$ column vector.

⁴The distinction between identification and identifiability previously discussed should be carefully borne in mind here and in what follows.

References

- [1] Akaike, H. Information Theory and an Extension of the Likelihood Principle.
In B. N. Petrov and F. Csáki (eds.), Proceedings of the Second International Symposium on Information Theory. Budapest: Akadémiai Kiado, 1973.
- [2] Amemiya, T. Regression Analysis When the Dependent Variable is Truncated Normal. Econometrica 41 (1973), 997-1016.
- [3] Amemiya, T. The Nonlinear Two-Stage Least-Squares Estimator. Journal of Econometrics 2 (1974), 105-110.
- [4] Amemiya, T. The Maximum Likelihood and Nonlinear Three Stage Least Squares Estimator in the General Nonlinear Simultaneous Equations Model. Econometrica 45 (1977), 955-968.
- [5] Andersen, E. B. Asymptotic Properties of Conditional Maximum Likelihood Estimators. Journal of the Royal Statistical Society, Series B 32 (1970) 283-301.
- [6] Bar-Shalom, Y. On the Asymptotic Properties of the Maximum Likelihood Estimate Obtained from Dependent Observations. Journal of the Royal Statistical Society, Series B 33 (1971), 72-77.
- [7] Bates, C. and H. White. Efficient Estimation of Parametric Models. Johns Hopkins Department of Political Economy Discussion Paper (1984).
- [8] Berk, H. Limiting Behavior of Posterior Distributions When the Model Is Incorrect. Annals of Mathematical Statistics 37 (1966), 51-58.
- [9] Berk, H. Consistency A Priori. Annals of Mathematical Statistics 41 (1970), 894-906.
- [10] Bhat, B. R. On the Method of Maximum Likelihood for Dependent Observations. Journal of the Royal Statistical Society, Series B 36 (1974), 48-53.

- [11] Billingsley, P. Probability and Measure. New York: Wiley, 1979.
- [12] Burguete, J., A. R. Gallant and G. Souza. On the Unification of the Asymptotic Theory of Nonlinear Econometric Models. Econometric Reviews 1 (1982), 151-212.
- [13] Caines, P. A Note on the Consistency of Maximum Likelihood Estimates for Finite Families of Stochastic Processes. The Annals of Statistics 3 (1975), 539-546.
- [14] Cramér, H. Mathematical Methods of Statistics. Princeton: Princeton University Press, 1946.
- [15] Crowder, M. Maximum Likelihood Estimation for Dependent Observations. Journal of the Royal Statistical Society, Series B 38 (1976), 45-53.
- [16] Doob, J. L. Probability and Statistics. Transactions of the American Mathematical Society 36 (1934), 759-775.
- [17] Domowitz, I. and H. White. Misspecified Models with Dependent Observations. Journal of Econometrics 20 (1982), 35-58.
- [18] Fisher, F. The Identification Problem in Econometrics. New York: McGraw-Hill, 1966.
- [19] Fisher, R. A. On the Mathematical Foundations of Theoretical Statistics. Philosophical Transactions of the Royal Society of London, Series A 222 (1922), 309-368.
- [20] Fisher, R. A. Theory of Statistical Estimation. Proceedings of the Cambridge Philosophical Society 22 (1925), 700-725.
- [21] Gallant, A. R. Three Stage Least Squares Estimation for a System of Simultaneous, Nonlinear, Implicit Equations. Journal of Econometrics 5 (1977), 71-88.

- [22] Gallant, A. R. and H. White. A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models. UCSD Department of Economics Discussion Paper (1985).
- [23] Geary, R. Determination of Linear Relations Between Systematic Parts of Variables with Errors of Observation, The Variances of Which are Unknown. Econometrica 17 (1949), 30-58.
- [24] Hall, P. and C. Heyde. Martingale Limit Theory and Its Applications. New York: Academic Press, 1980.
- [25] Hansen, L. Large Sample Properties of Generalized Method of Moments Estimators. Econometrica 50 (1982), 1029-1054.
- [26] Heijmans, R. and J. Magnus. Consistency of Maximum Likelihood Estimators when Observations are Dependent. University of Amsterdam Faculty of Actuarial Science and Econometrics Report No. AE 11/83, (1983).
- [27] Hoadley, B. Asymptotic Properties of Maximum Likelihood Estimators for the Independent Not Identically Distributed Case. Annals of Mathematical Statistics 42 (1971), 1977-1991.
- [28] Huber, P. The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions. in Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability. Berkeley: University of California Press, 1967.
- [29] Jennrich, R. I. Asymptotic Properties of Non-Linear Least Squares Estimators. Annals of Mathematical Statistics 40 (1969), 633-643.
- [30] Jorgenson, D. W. and J. Laffont. Efficient Estimation of Nonlinear Simultaneous Equations with Additive Disturbances. Annals of Economic and Social Measurement 3 (1974), 615-640.
- [31] Kullback, S. and R. A. Leibler. On Information and Sufficiency. Annals of Mathematical Statistics 22 (1951), 79-86.

- [32] Le Cam, L. On Some Asymptotic Properties of Maximum Likelihood Estimates and Related Bayes Estimates. University of California Publications in Statistics 1 (1953), 277-328.
- [33] Levine, D. A Remark on Serial Correlation in Maximum Likelihood. Journal of Econometrics 23 (1983), 337-342.
- [34] Reiersøl, O. Confluence Analysis by Means of Instrumental Sets of Variables. Arkiv for Matematik, Astronomi och Fysik 32 (1945).
- [35] Roberts, A. and Varberg, D. Convex Functions. New York: Academic Press, 1973.
- [36] Rothenberg, T. Efficient Estimation with a Priori Informaton. New Haven: Yale University Press, 1973.
- [37] Silvey, S. A Note on Maximum Likelihood in the Case of Dependent Random Variables. Journal of the Royal Statistical Society, Series B 23 (1961), 444-452.
- [38] Wald, A. Asymptotic Properties of the Maximum Likelihood Estimate of an Unknown Parameter of a Discrete Stochastic Process. Annals of Mathematical Statistics 19 (1948), 40-46.
- [39] Wald, A. Note on the Consistency of the Maximum Likelihood Estimate. Annals of Mathematical Statistics 20 (1949), 595-601.
- [40] Weiss, L. Asymptotic Properties of Maximum Likelihood in Some Nonstandard Cases: I. Journal of the American Statistical Society 66 (1971), 345-350.
- [41] Weiss, L. Asymptotic Properties of Maximum Likelihood in Some Nonstandard Cases: II. Journal of the American Statistical Society 68 (1973), 428-430.
- [42] White, H. Using Least Squares to Approximate Unknown Regression Functions. International Economic Review 21 (1980), 149-170.

- [43] White, H. Nonlinear Regression on Cross-Section Data. Econometrica 48 (1980), 721-746.
- [44] White, H. Consequences and Detection of Misspecified Nonlinear Regression Models. Journal of the American Statistical Association 76 (1981), 419-433.
- [45] White, H. Maximum Likelihood Estimation of Misspecified Models. Econometrica 50 (1982), 1-26.
- [46] White, H. Asymptotic Theory For Econometricians. New York: Academic Press, 1984.
- [47] White, H. Maximum Likelihood Estimation of Misspecified Dynamic Models. UCSD Department of Economics Discussion Papers, 1984.
- [48] White, H. Estimation, Inference and Specification Analysis. New York: Cambridge University Press, forthcoming.
- [49] White, H. and I. Domowitz. Nonlinear Regression with Dependent Observations. Econometrica 52 (1984), 143-162.
- [50] Wooldridge, J. A General Theory of Estimation and Inference for Parametric Models. UCSD Department of Economics Doctoral Dissertation, 1985.

MIT LIBRARIES



3 9080 003 062 392

Date Due

DEC 08 1988

JUL 06 1993:

Lib-26-67

bar code is on back cover

