

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125 Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1053

NOTE: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than acknowledgment that a writer had access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

Nonlinear Econometric Models
with
Deterministically Trending Variables
by
Donald W.K. Andrews and C. John McDermott

August 1993

Nonlinear Econometric Models with Deterministically Trending Variables

Donald W. K. Andrews¹
C. John McDermott

Cowles Foundation for Research in Economics
Yale University

January 1993

Revised: June 1993

Abstract

This paper considers an alternative asymptotic framework to standard sequential asymptotics for nonlinear models with deterministically trending variables. The asymptotic distributions of generalized method of moments estimators and corresponding test statistics are derived using this framework. The asymptotic distributions are shown to be the same with deterministically trending variables as with non-trending variables. That is, the distributions are normal and chi-squared respectively. The asymptotic covariance matrices of the estimators, however, are found to depend on the form of the trends. These findings provide a justification for the use of standard asymptotic approximations in nonlinear models even when the variables have deterministic trends.

JEL Classification Number: 211

Keywords: Asymptotics, deterministic trend, generalized method of moments estimator, hypothesis test, nonlinear econometric model, time trend.

1. Introduction

The purpose of this paper is to show that most of the work on nonlinear econometric models for non-trending variables done in the 1970's and 1980's can be extended to cover nonlinear models that are based on variables that have deterministic trends of quite general forms. We consider a triangular array asymptotic framework, which differs from standard sequential asymptotics. This asymptotic framework is quite tractable, since methods developed for models with non-trending variables can be applied to it.

Utilizing the triangular array asymptotics, we find that extremum estimators for nonlinear models with deterministically trending variables are approximately normal distributed and corresponding test statistics are approximately chi-square distributed under the null -- just as in models with non-trending variables. The form of the estimators' asymptotic covariance matrices, however, is found to differ when deterministic trends are present. Nevertheless, standard estimators of covariance matrices designed for the non-trending case yield consistent estimators for the deterministically trending case as well. Thus, estimation and inference for the deterministically trending case can be carried out in exactly the same fashion as in the non-trending case.

The triangular array asymptotics used here have numerous antecedents in the literature. Examples include the asymptotics used in some nonparametric regression models (e.g., see Priestley and Chao (1972)), many structural change models, some linear models with additive deterministic linear trends (see MacNeil (1978)), some linear models with additive deterministic nonlinear trends (see Phillips and Hansen (1990)), and some nonstationary models with time varying coefficients (see Dalhaus (1993)). Additional examples include the use of linear time trends and deterministically de-trended variables in nonlinear models estimated by GMM (see Eichenbaum and Hansen (1990)). In spite of these antecedents, the usefulness and generality of the method has not been fully explored nor appreciated in the nonlinear econometrics literature. For example, most review articles and treatises on nonlinear econometrics do not mention any results for models with deterministically trending variables (see Gallant (1987), Gallant and White (1988), Pötscher and Prucha (1991a, b) and Newey and McFadden (1992)). Exceptions are the review articles by Ogaki (1992) and Wooldridge (1993).

We note that Wooldridge (1986, 1993) has obtained some qualitatively similar results to those obtained here but using standard sequential asymptotics. We view the results of this paper as being complementary to those of Wooldridge. One advantage of our results is their tractability. Wooldridge's results utilize high level assumptions that can be difficult to verify. Primitive conditions for his results have been worked out for only one simple model in which a trend enters the model nonlinearly. Further discussion of the advantages and disadvantages of triangular array asymptotics relative to standard sequential asymptotics is given below.

The usefulness of the results given here depends, obviously, on the preponderance of economic variables that exhibit deterministic trends (of one form or another). It is clear that most macroeconomic variables and many financial variables exhibit trends. What is less clear is whether these trends are better viewed as being deterministic or stochastic (where a stochastically trending variable refers to a variable that is integrated). The empirical literature on this question is extensive and its conclusions are mixed (e.g., for a variety of results using a variety of methods, see Nelson and Plosser (1982), DeJong and Whiteman (1991), Phillips (1991), Kwiatkowski, Phillips, Schmidt, and Shin (1992), and Andrews and Chen (1992)). For some variables, the evidence points to the trends being deterministic and for other variables stochastic, but in only a few cases is the evidence strong. One feature of the empirical results that is relatively clear is that the evidence for deterministic trends is stronger for real variables than for nominal variables.

The empirical results referred to above consider linear deterministic trends. There is also a literature on the question of whether macroeconomic variables are better represented as having stochastic trends or deterministic trends with a break (or several breaks), e.g., see Perron (1989) and Zivot and Andrews (1992). Again the empirical findings are mixed.

The upshot of the above discussion is that some economic variables are probably better viewed as having deterministic trends while others are better viewed as having stochastic trends. Thus, there is some range of applications for the results of this paper, but it is certainly not all inclusive (since this paper does not allow for variables with stochastic trends). Examples of papers in the literature where the results are applicable include Fair's (1984) macroeconometric model which assumes trend stationary variables and includes nonlinear equations and time trend

regressors, Eichenbaum and Hansen's (1990) paper referred to above, and various papers that have appeared recently in the *American Economic Review* such as Fair and Dominguez (1991), Hamilton (1992), and Allen (1992). It appears that there are sufficient potential applications that the results of this paper should be of some interest for both theoretical and empirical reasons.

The remainder of the paper is organized as follows. Section 2 motivates and defines the triangular array asymptotics that are employed in the paper. Section 3 applies the asymptotic framework of Section 2 to GMM estimators and corresponding test statistics for nonlinear models based on deterministically trending rv 's. An Appendix provides proofs.

Throughout the paper, all limits are taken "as $T \rightarrow \infty$ " unless stated otherwise, " $\xrightarrow{p} 0$ " denotes convergence in probability, " \xrightarrow{d} " denotes convergence in distribution, "wp $\rightarrow 1$ " abbreviates "with probability that goes to one as $T \rightarrow \infty$ ", Σ_1^T denotes $\sum_{t=1}^T$, $\|\cdot\|$ denotes the Euclidean norm, $\|\cdot\|_q$ denotes the L^q norm, and A/B denotes the set of points that are in the set A but not in the set B .

2. The Asymptotic Framework

Here we consider the question of how to construct a suitable asymptotic framework for handling nonlinear econometric models when the underlying random variables (rv 's) may possess deterministic trends. The purpose of the asymptotics is to generate good approximations to the finite sample distributions of estimators and test statistics. It is this purpose that guides our construction of the asymptotic framework.

To achieve good asymptotic approximations, it is usually desirable to embed the finite sample problem of interest in a sequence of problems which mimic the finite sample problem as closely as possible. To implement this principle, one first has to decide what aspects of the problem should be closely mimicked. In the present case of models based on rv 's with deterministic trends, there are four characteristics of the underlying rv 's that we focus on. The first is the shape of the deterministic trend, the second is the magnitude of the deterministic trend relative to the stochastic component, the third is the magnitude of the deterministic trend in an absolute

sense, and the fourth is the nature of the stochastic components of the rv's. We seek an asymptotic framework in which these characteristics of the finite sample problem of interest are mimicked as closely as possible in each element of the sequence upon which the asymptotics are based.

Let the sample size of interest be denoted by T^* . The observed data consist of T^* R^k -valued rv's: W_1, \dots, W_{T^*} . These rv's can be viewed as part of an infinite sequence $W_t, t = 1, 2, \dots$. We write the rv W_t as the composition of a stochastic component Z_t and a deterministic function $d(t, \cdot)$ of Z_t :

$$(2.1) \quad W_t = d(t, Z_t) \quad \text{for } t = 1, 2, \dots,$$

where $d(\cdot, z) : (0, \infty) \rightarrow R^k$. This decomposition is not unique. For present purposes, it suffices that the decomposition is such that Z_t does not contain a deterministically trending component. For example, we want $\{Z_t : t \geq 1\}$ to satisfy a condition such as $\{Z_t\}$ are stationary or $\sup_{t \geq 1} E \|Z_t\|^\gamma < \infty$ for some $\gamma > 0$, or $\{Z_t\}$ is a tight sequence (i.e., $\forall \varepsilon > 0 \exists B_\varepsilon < \infty$ such that $\sup_{t \geq 1} P(\|Z_t\| > B_\varepsilon) < \varepsilon$).

For example, $d(t, Z_t)$ could be of the form:

$$(2.2) \quad \text{(i) } d(t) + Z_t, \quad \text{(ii) } \text{diag}\{d(t)\}Z_t, \quad \text{or (iii) } d_1(t) + \text{diag}\{d_2(t)\}Z_t,$$

where $d(\cdot)$, $d_1(\cdot)$, and $d_2(\cdot)$ are functions from $(0, \infty)$ to R^k and $\text{diag}\{d(t)\}$ is a $k \times k$ diagonal matrix with diagonal elements given by $d(t)$. Typically, the functions $d(\cdot)$, $d_1(\cdot)$, and $d_2(\cdot)$ are smooth, but they could have discontinuities. For example, $d(\cdot)$ in (2.2) could be of the form:

$$(2.3) \quad \begin{aligned} & \text{(i) } d(t) = bt + ct^2 \quad \text{for } t > 0, \quad \text{(ii) } d(t) = \exp(bt), \quad \text{or} \\ & \text{(iii) } d(t) = \begin{cases} bt & \text{for } 0 < t \leq G \\ ct & \text{for } t > G \end{cases} \end{aligned}$$

for some $0 < G < \infty$, where b and c are constant k -vectors and $\exp(\cdot)$ denotes the element by element exponentiation operator.

Form (i) of (2.2) plus (i) of (2.3) are commonly used as the form for the logarithms of "trend stationary" macroeconomic rv's (often with $c = 0$). In levels, the same rv's are of the form (ii) of (2.2) plus (ii) of (2.3). The logarithms of trend stationary macro rv's with a change in

trend are often given by (i) of (2.2) plus (iii) of (2.3). Of course, more esoteric forms of trend functions $d(t, z)$ than those exemplified in (2.2) and (2.3) are also possible.

At this point, we could consider "standard asymptotics" in which the sample of interest (W_1, \dots, W_{T^*}) is embedded in the sequence of samples $\{(W_1, \dots, W_T) : T \geq 1\}$, where $W_t = d(t, Z_t) \forall t \geq 1$. Standard asymptotics have several advantages and several disadvantages. As an advantage, they guarantee that any approximations they generate are accurate to any given degree for samples of sufficiently large size. In addition, they allow one to identify "dominating trends" (i.e., trends that are sufficiently large that they interfere with the operation of a central limit theorem (CLT)). In such cases, normal approximations for the distributions of estimators are not appropriate.

On the other hand, standard asymptotics suffer from problems of tractability in nonlinear models with deterministically trending variables. Asymptotic distributional results using standard asymptotics either rely on high level assumptions that can be difficult and/or tedious to verify (e.g., see Crowder (1976) and Wooldridge (1986, 1993)) or they apply only to special models under restrictive assumptions, e.g., see Wu (1981).

A second potential problem with standard asymptotics is that they depend on the trend function $d(t, z)$ for all values of t sufficiently large. In contrast, the finite sample distribution of (W_1, \dots, W_{T^*}) depends only on $d(t, z)$ for $t \leq T^*$; its values for $t > T^*$ are irrelevant. A problem occurs if the shape of the trend function for t large (relative to T^*) does not reflect its shape for the sample size of interest.

Third, standard asymptotics typically drive the magnitude of the deterministic trend to infinity as $T \rightarrow \infty$. For samples that exhibit small trends, this form of asymptotics does not mimic the finite sample problem of interest very well and may lead to inappropriate results. For example, for rv's $\{W_t : t \geq 1\}$ of form (ii) of (2.2) plus (ii) of (2.3), a CLT does not apply (even if the rv's $\{Z_t : t \geq 1\}$ are iid and bounded), because the Lindeberg condition fails. Yet, it may be appropriate to use a normal approximation for a normalized sum of such rv's $\{W_t : t \leq T^*\}$ when the trend coefficient b is small (since such rv's deviate little from a sample of iid rv's).

In this paper, we consider an alternative asymptotic framework to standard asymptotics. We

embed the sample of interest (W_1, \dots, W_{T^*}) in a triangular array of rv's in such a way that (i) only the deterministic trend function $d(t, z)$ for $t \leq T^*$ affects the asymptotics, (ii) the shape of the deterministic trend in the sample of interest is mimicked in the limit, (iii) the magnitude of the deterministic trend is bounded in the limit, and (iv) asymptotic distributional results for estimators and test statistics can be obtained straightforwardly using the techniques developed for nonlinear models with non-trending observations.

As with standard asymptotics, the asymptotics considered here have both advantages and disadvantages. Properties (i), (ii), and (iv) listed in the previous paragraph are advantages. Property (iii) can be an advantage or a disadvantage depending on the circumstances. It is discussed further below. A disadvantage of the triangular array asymptotics is that they do not guarantee that the approximations they generate are accurate to any given degree when the sample size T^* of interest is sufficiently large.

The triangular array asymptotics considered here are set up as follows. Let

$$(2.4) \quad W_{Tt} = d\left(\frac{T^*}{T}t, Z_t\right) \quad \forall t \leq T, \quad T \geq 1,$$

where Z_t and $d(\cdot, \cdot)$ are as above. The sequence of samples upon which the asymptotics are based is given by $\{(W_{T1}, \dots, W_{TT}) : T \geq 1\}$. Note that this framework embeds the sample of interest, since $W_{T^*t} = W_t \quad \forall t \leq T^*$. Properties (i)-(iii) stated two paragraphs above, can be verified by inspection. (Property (iii) holds provided $\sup_{t \in (0, T^*)} \|d(t, z)\| < \infty$ for all z in the supports of Z_t for $t \leq T^*$, as is usually the case.) Property (iv) holds because the boundedness of the deterministic functions $d\left(\frac{T^*}{T}t, z\right)$ over $t \leq T$ and $T \geq 1$ allows laws of large numbers (LLNs) and CLTs to be applied to the rv's $\{W_{Tt} : t \leq T, T \geq 1\}$ and to functions of them. In consequence, standard proofs of the consistency and asymptotic normality of extremum estimators and of the asymptotic chi-square null distributions of corresponding test statistics go through under almost the same conditions as when the underlying rv's exhibit no deterministic trends. One only has to replace uniform LLNs and a CLT for sequences of rv's by corresponding ones for triangular arrays.

The form of the limiting covariance matrices of extremum estimators differs when the rv's are deterministically trending as in (2.4) from when they are not, as shown in the next section.

Nevertheless, consistent estimators of the limiting covariance matrices can be defined in the same way in both cases. Thus, one can use the triangular array asymptotics of (2.4) to provide a justification for using standard asymptotic approximations (i.e., normality for extremum estimators and chi-square null distributions for the corresponding test statistics) in nonlinear models with deterministically trending rv's.

The asymptotic framework of (2.4) could be referred to as one of *bounded trend asymptotics* or *small trend asymptotics*. A disadvantage of the use of such asymptotics arises when the trend in the sample of interest is dominating. For example, suppose $W_t = \text{diag}\{\exp(bt)\}Z_t$, Z_t are iid, and the distribution of an estimator depends on the distribution of the sum $\Sigma_1^{T^*} W_t$. If the trend coefficient b is large, the distribution of $\Sigma_1^{T^*} W_t$ is determined by only the last few observations in the sample and the averaging effect of the CLT does not operate to yield approximate normality. In this case, the use of bounded trend asymptotics is inappropriate, because they generate approximations based on the CLT. On the other hand, if the trend coefficient b is small, then a large number of observations in the sum $\Sigma_1^{T^*} W_t$ determines its distribution, the CLT approximation is reasonable, and the use of bounded trend asymptotics is appropriate.

The upshot of the above discussion is that one has to be careful when using the triangular array asymptotics of (2.4). One needs to look at the rv's to which a CLT is applied (typically these rv's are functions of the rv's $\{W_t : t \leq T^*\}$) and decide whether a normal approximation is appropriate. If the variances of the rv's in the sum are increasing as polynomials in time, then the CLT can be applied using standard asymptotics or bounded trend asymptotics. In such cases, there is no problem using bounded trend asymptotics. If the variances are increasing exponentially, however, then the CLT does not apply with standard asymptotics, although it does with bounded trend asymptotics. In such cases, one has to decide whether the magnitude of the trend is sufficiently small that bounded trend asymptotics are suitable. For example, if one believes that the logarithms of the economic variables of concern are of the form $bt + Z_t$, then these variables can enter a linear model in logarithmic form without trouble, but if they enter in levels then they may cause problems if b is too large. Analogous examples for nonlinear models can be constructed.

3. Generalized Method of Moments Estimators

3.1. Introduction and Definitions

In this section, we apply the triangular array asymptotics of (2.4) to the class of GMM estimators and corresponding Wald, Lagrange multiplier (LM), and likelihood ratio-like (LR) test statistics. We consider the case where the stochastic components $\{Z_t : t \geq 1\}$ of the underlying rv's $\{W_{Tt} : t \leq T, T \geq 1\}$ are *strongly second order stationary* (SSOS) (i.e., the distribution of (Z_s, Z_t) depends only on $|s-t| \forall s, t = 1, 2, \dots$). The triangular array framework can be applied more generally, but the SSOS assumption yields relatively simple conditions and a simple expression for the asymptotic covariance matrix and should be sufficiently general to illustrate the usefulness of the approach. Asymptotically weak temporal dependence of the stochastic components $\{Z_t : t \geq 1\}$ is governed by *near epoch dependence* (NED) conditions.

Let $\theta \in \Theta \subset R^p$ be the unknown parameter to be estimated. The population orthogonality conditions that are used by the GMM estimator to estimate the true parameter θ_0 are $\frac{1}{T} \sum_1^T E m(W_{Tt}, \theta) = 0$ for a specified function $m(w, \theta)$ from $\mathcal{W} \times \Theta$ to R^v , where \mathcal{W} is a Borel subset of R^k that contains all realizations of $W_{Tt} \forall t \leq T, T \geq 1$. The underlying rv's $\{Z_t : t \geq 1\}$ take values in a Borel subset \mathcal{Z} of R^l . They are defined on a probability space (Ω, \mathcal{F}, P) . By definition, a rv is Borel measurable.

DEFINITION: A sequence of *GMM estimators* $\{\hat{\theta} : T \geq 1\}$ is any sequence of (Borel measurable) estimators in Θ that satisfies

$$(3.1) \quad \bar{m}_T(\hat{\theta})' \hat{\gamma} \bar{m}_T(\hat{\theta}) = \inf_{\theta \in \Theta} \bar{m}_T(\theta)' \hat{\gamma} \bar{m}_T(\theta) \quad \text{wp} - 1,$$

where $\bar{m}_T(\theta) = \frac{1}{T} \sum_1^T m(W_{Tt}, \theta)$ and $\hat{\gamma}$ is a random (Borel measurable) symmetric $v \times v$ matrix (which depends on T in general).

As is well-known, e.g., see Hansen (1982), the class of GMM estimators is quite broad. Among others, it includes least squares, nonlinear instrumental variables, maximum likelihood (ML), and pseudo-ML estimators.

Next, we define the NED condition. It has origins as far back (at least) as Ibragimov (1962).

The NED condition can be used to obtain LLNs, CLTs, and invariance principles for sequences and triangular arrays of temporally dependent rv's. It is one of the most general concepts of weak temporal dependence for nonlinear models that is available. See Bierens (1981), Gallant (1987), Gallant and White (1988), and Pötscher and Prucha (1991a) for examples of its application to particular econometric models. Note that the definition below is a slight variant (generalization), suggested by Pötscher and Prucha (1991a), of definitions in the literature.

DEFINITION: For $q \geq 0$, a sequence of rv's $\{Z_t : t = 1, 2, \dots\}$ is said to be L^q -NED on the strong mixing base $\{Y_t : t = \dots, 0, 1, \dots\}$ if $\{Y_t : t = \dots, 0, 1, \dots\}$ is a strong mixing (i.e., α -mixing) sequence of rv's and $\overline{\lim}_{T \rightarrow \infty} \frac{1}{T} \sum_1^T \|Z_t - E(Z_t | Y_{t-m}, \dots, Y_{t+m})\|_q \rightarrow 0$ as $m \rightarrow \infty$ when $q > 0$ or $\overline{\lim}_{T \rightarrow \infty} \frac{1}{T} \sum_1^T P(\|Z_t - E(Z_t | Y_{t-m}, \dots, Y_{t+m})\| > \varepsilon) \rightarrow 0$ as $m \rightarrow \infty \forall \varepsilon > 0$ when $q = 0$. For $q > 0$, $u > 0$, and $b > 0$, $\{Z_t : t = 1, 2, \dots\}$ is said to be L^q -NED of size $-u$ on a strong mixing base $\{Y_t : t = \dots, 0, 1, \dots\}$ of size $-b$ if $\{v_m : m \geq 1\}$ is of size $-u$ (i.e., $v_m = O(m^{-\lambda})$ for some $\lambda > u$), where $v_m = \sup_{t \geq 1} \|Z_t - E(Z_t | Y_{t-m}, \dots, Y_{t+m})\|_q$ and $\{\alpha_m : m \geq 1\}$ is of size $-b$, where $\{\alpha_m : m \geq 1\}$ are the strong mixing numbers of $\{Y_t\}$, as defined, e.g., in Gallant and White (1988, p. 23). NED triangular arrays of rv's are defined analogously.

3.2. Consistency

We now provide a set of conditions under which the GMM estimator is consistent within the context of the triangular array asymptotics of (2.4).

ASSUMPTION 1: (a) $\{Z_t : t = 1, 2, \dots\}$ is a sequence of identically distributed \mathcal{Z} -valued rv's that is L^0 -NED on a strong mixing base $\{Y_t : t = \dots, 0, 1, \dots\}$.

(b) $d(\cdot, \cdot)$ is a function from $(0, \infty) \times \mathcal{Z}$ to R^k , $d(t, z)$ is right continuous in t almost everywhere (Lebesgue) on $(0, T^*] \forall z \in \mathcal{Z}$, $d(t, z)$ is continuous in z uniformly over $t \in (0, T^*] \forall z \in \mathcal{Z}$, and $\sup_{t \in (0, T^*]} \|d(t, z)\| < \infty \forall z \in \mathcal{Z}$.

(c) $W_{Tt} = d\left(\frac{T^*}{T}t, Z_t\right) \forall t \leq T, T \geq 1$.

(d) $\hat{\gamma} \xrightarrow{P} \gamma$ for some nonsingular symmetric $v \times v$ matrix γ .

(e) Θ is a bounded subset of R^p .

(f) $m(w, \theta)$ is a function from $\mathcal{W} \times \Theta$ to R^v that is continuous in w uniformly over $\theta \in \Theta \quad \forall w \in \mathcal{W}$ and continuous in θ uniformly over $\Theta \quad \forall w \in \mathcal{W}$.

(g) $E \sup_{\theta \in \Theta} \sup_{s \in (0, T^*]} |m(d(s, Z_t), \theta)|^{1+\varepsilon} < \infty$ for some $\varepsilon > 0$.

(h) $m(\theta_0) = 0$ and for all neighborhoods Θ_0 of θ_0 , $\inf_{\theta \in \Theta/\Theta_0} m(\theta)' \gamma m(\theta) > 0$, where $m(\theta) = \lim_{T \rightarrow \infty} \frac{1}{T} \Sigma_1^T E m(W_{Tt}, \theta)$.

Assumption 1 is quite similar to consistency assumptions in the literature for nonlinear models with non-trending variables. In fact, if the trending function $d(\cdot, z)$ in part (b) is set equal to z , so that no trends appear, then Assumption 1 reduces to a typical set of such assumptions. When $d(\cdot, z)$ does not equal z , the only added features are in the definition of the underlying rv's (in parts (b) and (c)) and in a slight strengthening of a moment condition (i.e., the addition of \sup_s in part (g)).

Note that the function $d(\cdot, z)$ can be discontinuous at a countable number of points for any fixed z . This implies that the deterministic trends can exhibit (multiple) breaks.

The identification condition given in part (h) depends on the "asymptotic criterion function" $m(\theta)' \gamma m(\theta)$. Given our assumptions, we can find an explicit expression for the main component of this function, viz., $m(\theta)$.

LEMMA 1: Under Assumptions 1(a)-(c) and 1(e)-(g),

$$m(\theta) = \lim_{T \rightarrow \infty} \frac{1}{T} \Sigma_1^T E m(W_{Tt}, \theta) = \frac{1}{T^*} \int_0^{T^*} E m(d(s, Z_t), \theta) ds$$

(and the convergence as $T \rightarrow \infty$ holds uniformly over $\theta \in \Theta$).

COMMENT: The Lemma shows that $m(\theta)$ is just the average over values of the trend component of the function that arises in the case with no deterministic trends.

Consistency is established in the following theorem.

THEOREM 1: Under Assumption 1, every sequence of GMM estimators $\{\hat{\theta} : T \geq 1\}$ satisfies $\hat{\theta} \xrightarrow{P} \theta_0$.

3.3. Asymptotic Normality

This section contains the main result. We show that the GMM estimator is asymptotically normal in nonlinear models with deterministically trending random variables under the triangular array asymptotics of (2.4). First, we define the asymptotic covariance matrix V of the GMM estimator. Define

$$\begin{aligned}
 (3.2) \quad V &= (M' \gamma M)^{-1} M' \gamma S \gamma M (M' \gamma M)^{-1}, \text{ where} \\
 M &= \frac{1}{T^*} \int_0^{T^*} E \frac{\partial}{\partial \theta'} m(d(s, Z_t), \theta_0) ds \text{ and} \\
 S &= \frac{1}{T^*} \int_0^{T^*} E m(d(s, Z_t), \theta_0) m(d(s, Z_t), \theta_0)' ds \\
 &\quad + \sum_{j=1}^{\infty} \frac{1}{T^*} \int_0^{T^*} E m(d(s, Z_t), \theta_0) m(d(s, Z_{t-j}), \theta_0)' ds \\
 &\quad + \sum_{j=1}^{\infty} \frac{1}{T^*} \int_0^{T^*} E m(d(s, Z_{t-j}), \theta_0) m(d(s, Z_t), \theta_0)' ds.
 \end{aligned}$$

Note that when the trend function $d(t, z)$ equals z , so that no trends exist, the asymptotic covariance matrix V simplifies to its well-known form for models with non-trending rv's as given, e.g., in Hansen (1982). As with the function $m(\theta)$, when $d(t, z)$ does not equal z , various components of V are given by averages with respect to the trend component of the matrices that arise in the case with no time trend.

Sufficient conditions for asymptotic normality of $\hat{\theta}$ are given in the following assumption.

ASSUMPTION 2: (a) *Assumptions 1(a)-(d) and 1(f) hold.*

(b) $\{Z_t : t = 1, 2, \dots\}$ *is a strongly second-order stationary sequence of rv's.*

(c) *For some $r > 2$, $\{m(W_{Tt}, \theta_0) : t \leq T, T \geq 1\}$ is a triangular array of R^p -valued rv's that is L^2 -NED of size -1 on a strong mixing base $\{Y_t : t = \dots, 0, 1, \dots\}$ of size $-2r/(r-2)$ and $E \sup_{s \in (0, T^*]} |m(d(s, Z_t), \theta_0)|^r < \infty$.*

(d) $E m(W_{Tt}, \theta_0) = 0 \quad \forall t \leq T, T \geq 1$.

(e) $\hat{\theta} \xrightarrow{p} \theta_0 \in \Theta \subset R^p$ *and θ_0 is in the interior of Θ .*

(f) $m(w, \theta)$ *is partially differentiable in $\theta \forall \theta \in \Theta_0 \quad \forall w \in \mathcal{W}$, where Θ_0 is some neighborhood of θ_0 ,*

$\frac{\partial}{\partial \theta'} m(w, \theta)$ is continuous in (w, θ) on $\mathcal{W} \times \Theta_0$, and $E \sup_{\theta \in \Theta_0} \sup_{s \in (0, T^*)} \left\| \frac{\partial}{\partial \theta'} m(d(s, Z_t), \theta) \right\|^{1+\varepsilon} < \infty$ for some $\varepsilon > 0$.

(g) $M' \gamma M$ and S are nonsingular.

As with Assumption 1, Assumption 2 differs very little from typical NED-based conditions for the asymptotic normality of GMM estimators with non-trending rv's, e.g., see Andrews (1993). The only differences are those referred to above in Assumptions 1(b) and 1(c), plus a slight strengthening of moment conditions (by the addition of \sup_s) in Assumptions 2(c) and 2(f). We note that if $\{m(W_{Tt}, \theta_0) : t \leq T, T \geq 1\}$ is a martingale difference triangular array, then Assumption 2(c) can be replaced by sufficient conditions for a martingale difference CLT to hold for $\frac{1}{\sqrt{T}} \Sigma_1^T m(W_{Tt}, \theta_0)$, e.g., see Pötscher and Prucha (1991b, Theorem 4.1). We also note that (more primitive) sufficient conditions for Assumption 2(c), which rely on NED conditions placed on the underlying rv's $\{Z_t : t \geq 1\}$, can be obtained with some loss of generality by applying Theorem 4.2 of Gallant and White (1988) or Theorem 6.7 of Pötscher and Prucha (1991a).

The source of the matrices M and S that arise in the definition of the asymptotic covariance matrix V of $\hat{\theta}$ is made clear in the following result.

LEMMA 2: (a) Under Assumptions 2(a) and 2(f),

$$\lim_{T \rightarrow \infty} \frac{1}{T} \Sigma_1^T E \frac{\partial}{\partial \theta'} m(W_{Tt}, \theta_0) = M .$$

(b) Under Assumptions 2(a)-(d),

$$\lim_{T \rightarrow \infty} \text{Var} \left(\frac{1}{\sqrt{T}} \Sigma_1^T m(W_{Tt}, \theta_0) \right) = S .$$

Asymptotic normality of $\hat{\theta}$ is established in the following result.

THEOREM 2: Under Assumption 2, every sequence of GMM estimators $\{\hat{\theta} : T \geq 1\}$ satisfies $\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V)$.

COMMENT: Theorem 2 relies on $\{m(W_{Tt}, \theta_0) : t \leq T, T \geq 1\}$ satisfying a CLT. Under the bounded trend asymptotics of (2.4) and Assumption 2, such a CLT is guaranteed to hold. On the other hand, as discussed at the end of Section 2, the use of bounded trend asymptotics can be inappropriate if the deterministic trends are sufficiently dominating. In particular, what matters is whether $\Sigma_1^{T^*} m(W_t, \theta_0)$ can be approximated reasonably well by a normal distribution. If this sum, suitable normalized, is asymptotically normal using both bounded-trend asymptotics and standard asymptotics (i.e., if $\{m(W_{Tt}, \theta_0) : t \leq T, T \geq 1\}$ and $\{m(W_t, \theta_0) : t \geq 1\}$ both satisfy CLTs using suitable normalizations), then the normal approximation is certainly valid, at least for T^* large.

If $\{m(W_t, \theta_0) : t \geq 1\}$ does not satisfy a CLT (using standard asymptotics), however, then the normal approximation needs to be scrutinized more closely. For example, consider a linear regression model $Y_t = e^{bt}X_t'\theta_0 + U_t$ and the least squares estimator $\hat{\theta}$. Suppose $Z_t = (U_t, X_t)'$ are iid, $E\|Z_t\|^4 < \infty$, $b > 0$, $EU_tX_t = 0$, $W_t = d(t, Z_t) = (Y_t, e^{bt}X_t)$, and $m(W_t, \theta_0) = U_t e^{bt}X_t$. In this case, $\{m(W_t, \theta_0) : t \geq 1\}$ does not satisfy a CLT (because the Lindeberg condition fails), even though $\{m(W_{Tt}, \theta_0) : t \leq T, T \geq 1\}$ does. A normal approximation for $\Sigma_1^{T^*} U_t e^{bt}X_t$ is appropriate in this case only if the trend coefficient b is sufficiently small that the distribution of the sum is not dominated by a relatively small number of observations. Note that if the exponentially trending regressors $e^{bt}X_t$ are replaced by polynomially trending regressors $\sum_{j=1}^{J_1} b_j t^j + \sum_{j=1}^{J_2} c_j t^j X_t$ in this example, then there is no problem, because the CLT holds with the latter regressors under both bounded trend asymptotics and standard asymptotics.

3.4. Covariance Matrix Estimation

The covariance matrix V of $\hat{\theta}$ can be estimated by

$$(3.3) \quad \hat{V} = (\hat{M}' \hat{\gamma} \hat{M})^{-1} \hat{M}' \hat{\gamma} \hat{S} \hat{\gamma} \hat{M} (\hat{M}' \hat{\gamma} \hat{M})^{-1}, \quad \text{where}$$

$$\hat{M} = \frac{1}{T} \Sigma_1^T \frac{\partial}{\partial \theta'} m(W_{Tt}, \hat{\theta}),$$

$\hat{\gamma}$ is as in Assumption 1(d), and \hat{S} is some consistent estimator of S . A suitable choice of \hat{S} depends on the dependence properties of $\{m(W_{Tt}, \theta_0) : t \leq T, T \geq 1\}$. If the latter is uncor-

related across time, then

$$(3.4) \quad \hat{S} = \frac{1}{T} \Sigma_1^T (m(W_{Tt}, \hat{\theta}) - \bar{m}_T(\hat{\theta})) (m(W_{Tt}, \hat{\theta}) - \bar{m}_T(\hat{\theta}))'$$

is a suitable choice, where $\bar{m}_T(\theta) = \frac{1}{T} \Sigma_1^T m(W_{Tt}, \theta)$. If $\{m(W_{Tt}, \theta_0)\}$ is correlated over time, then an estimator designed for m -dependent data or a heteroskedasticity and autocorrelation consistent (HAC) estimator may be suitable. See Gallant and White (1988) for consistency results for HAC estimators under NED conditions. See Andrews (1991) and Andrews and Monahan (1992) for results concerning the choice of HAC estimator.

An important point to note is that a suitable choice of \hat{S} , just as with \hat{M} , is the same in models with deterministically trending variables as in models with non-trending variables.

Under Assumption 2, $\hat{M} \xrightarrow{P} M$ and $\hat{\gamma} \xrightarrow{P} \gamma$. (See the proof of Theorem 2 for a proof that $\hat{M} \xrightarrow{P} M$.) If, in addition, $\{m(W_{Tt}, \theta_0)\}$ are uncorrelated, then \hat{S} defined in (3.4) satisfies $\hat{S} \xrightarrow{P} S$. Otherwise, for brevity, we simply assume that \hat{S} has been chosen to be consistent.

ASSUMPTION 3: $\hat{S} \xrightarrow{P} S$.

In consequence, under Assumptions 2 and 3, \hat{V} is a consistent estimator of V .

Note that an optimal choice of asymptotic weight matrix γ is S^{-1} . Correspondingly, if $\hat{\gamma} = \hat{S}^{-1}$, then \hat{V} simplifies to $(\hat{M}' \hat{S}^{-1} \hat{M})^{-1}$.

3.5. Hypothesis Tests

Here we consider tests of nonlinear restrictions of the form $H_0 : h(\theta_0) = 0$. The upshot of the present section is that standard test statistics have asymptotic chi-square null distributions in nonlinear models with deterministically trending data using the asymptotic framework of (2.4).

The R^r -valued function $h(\cdot)$ defining the restrictions is assumed to satisfy:

ASSUMPTION 4: $h(\theta)$ is continuously differentiable in a neighborhood of θ_0 and $H = \frac{\partial}{\partial \theta'} h(\theta_0)$ has full rank r ($\leq p$).

The Wald statistic for testing H_0 is defined to be

$$(3.5) \quad W_T = Th(\hat{\theta})' (\hat{H} \hat{V} \hat{H}')^{-1} h(\hat{\theta}), \quad \text{where } \hat{H} = \frac{\partial}{\partial \theta'} h(\hat{\theta}).$$

Next, we define the LM and LR statistics. For brevity, we do so only for the case where the weight matrix $\hat{\gamma}$ is chosen optimally: $\hat{\gamma} = \hat{S}^{-1}$ and $\gamma = S^{-1}$. These test statistics depend on a *restricted GMM estimator* $\bar{\theta}$. By definition, a sequence of restricted GMM estimators $\{\bar{\theta} : T \geq 1\}$ minimizes $\bar{m}_T(\theta)' \hat{\gamma} \bar{m}_T(\theta)$ over $\{\theta \in \Theta : h(\theta) = 0\}$ wp - 1. Under Assumption 1, $\bar{\theta} \xrightarrow{P} \theta_0$ when the null hypotheses is true, by the proof of Theorem 1. In consequence, the second part of the following assumption is easily verifiable.

ASSUMPTION 5: (a) $\hat{\gamma} = \hat{S}^{-1}$ and $\gamma = S^{-1}$. (b) $\bar{\theta} \xrightarrow{P} \theta_0$.

Let \bar{V} , \bar{M} , and \bar{S} be defined as in (3.3), but with $\hat{\theta}$ replaced by $\bar{\theta}$. The LM statistic is defined by

$$(3.6) \quad LM_T = T \bar{m}_T(\bar{\theta})' \hat{\gamma} \bar{M} (\bar{M}' \bar{S}^{-1} \bar{M})^{-1} \bar{M}' \hat{\gamma} \bar{m}_T(\bar{\theta}) .$$

The LR statistic is defined by

$$(3.7) \quad LR_T = T \bar{m}_T(\bar{\theta})' \hat{\gamma} \bar{m}_T(\bar{\theta}) - T \bar{m}_T(\hat{\theta})' \hat{\gamma} \bar{m}_T(\hat{\theta}) .$$

The asymptotic null distribution of the Wald, LM and LR statistics is shown in the following theorem to be chi-squared with r degrees of freedom (χ_r^2).

THEOREM 3: Suppose θ_0 satisfies the null hypothesis $H_0 : h(\theta_0) = 0$. Then, (a) $W_T \xrightarrow{d} \chi_r^2$ under Assumptions 2-4, (b) $LM_T \xrightarrow{d} \chi_r^2$ under Assumptions 2-5 (with \bar{S} in place of \hat{S} in Assumption 3), and (c) $LR_T \xrightarrow{d} \chi_r^2$ under Assumptions 2, 4, and 5.

Given Theorem 2 and its proof, the proof of Theorem 3 is quite similar to proofs in the literature for nonlinear models with non-trending variables. In consequence, we omit its proof for brevity.

COMMENT: Theorem 3 shows that, in the case of deterministically trending rv's, one can define the Wald, LM, and LR statistics just as one would in the case of non-trending rv's and their asymptotic null distributions still are χ_r^2 using the asymptotics of (2.4).

Appendix of Proofs

PROOF OF LEMMA 1: Let $\xi_T(u, z) = d\left(\frac{T^*}{T}([Tu] + 1), z\right)$ for $u \in [0, 1]$, $\xi_T(1, z) = d(T^*, z)$, and $\xi(u, z) = d(T^*u, z)$ for $u \in [0, 1]$ and $z \in \mathcal{Z}$. By Assumption 1(b), $\xi_T(u, z) \rightarrow \xi(u, z)$ as $T \rightarrow \infty$ for (Lebesgue) almost all u in $[0, 1] \forall z \in \mathcal{Z}$. We have

$$\begin{aligned}
 & \limsup_{T \rightarrow \infty} \sup_{\theta \in \Theta} \left\| \int_{\mathcal{Z}} \left[\frac{1}{T} \sum_1^T m\left(d\left(\frac{T^*}{T}t, z\right), \theta\right) - \frac{1}{T^*} \int_0^{T^*} m(d(s, z), \theta) ds \right] dP(z) \right\| \\
 & \leq \limsup_{T \rightarrow \infty} \int_{\mathcal{Z}} \sup_{\theta \in \Theta} \left\| \frac{1}{T} \sum_1^T m\left(d\left(\frac{T^*}{T}t, z\right), \theta\right) - \int_0^1 m(\xi(u, z), \theta) du \right\| dP(z) \\
 \text{(A.1)} \quad & = \int_{\mathcal{Z}} \limsup_{T \rightarrow \infty} \sup_{\theta \in \Theta} \left\| \frac{1}{T} \sum_1^T m\left(d\left(\frac{T^*}{T}t, z\right), \theta\right) - \int_0^1 m(\xi(u, z), \theta) du \right\| dP(z) \\
 & = \int_{\mathcal{Z}} \limsup_{T \rightarrow \infty} \sup_{\theta \in \Theta} \left\| \int_0^1 [m(\xi_T(u, z), \theta) - m(\xi(u, z), \theta)] du \right\| dP(z) \\
 & \leq \int_{\mathcal{Z}} \int_0^1 \limsup_{T \rightarrow \infty} \sup_{\theta \in \Theta} \|m(\xi_T(u, z), \theta) - m(\xi(u, z), \theta)\| du dP(z) \\
 & = 0.
 \end{aligned}$$

The first equality holds using Assumption 1(g) by the dominated convergence theorem (DCT) with dominating function $2 \sup_{\theta \in \Theta} \sup_{s \in (0, T^*]} \|m(d(s, z), \theta)\|$. The second equality holds by the definition of $\xi_T(u, z)$. The second inequality utilizes the DCT with dominating functions for each fixed z given by the constant function on $[0, 1]$ with value

$2 \sup_{\theta \in \Theta} \sup_{s \in (0, T^*]} \|m(d(s, z), \theta)\|$. The last equality holds by Assumption 1(f) because $\xi_T(u, z) \rightarrow \xi(u, z)$ as $T \rightarrow \infty$ for (Lebesgue) almost all $u \in [0, 1] \forall z \in \mathcal{Z}$. \square

The proof of Theorem 1 uses the following two lemmas.

LEMMA A1: Suppose $\hat{\theta}$ minimizes a random real function $Q_T(\theta)$ over $\theta \in \Theta \subset R^p$ w.p. - 1. If (a) $\sup_{\theta \in \Theta} |Q_T(\theta) - Q(\theta)| \xrightarrow{P} 0$ for some real function Q on Θ and (b) for every neighborhood Θ_0 ($\subset \Theta$) of θ_0 , $\inf_{\theta \in \Theta/\Theta_0} Q(\theta) > Q(\theta_0)$, then $\hat{\theta} \xrightarrow{P} \theta_0$.

LEMMA A2: Suppose (a) Assumptions 1(a), 1(b), and 1(c) hold, (b) Θ_1 is a bounded subset of \mathbb{R}^s , (c) $f(w, \theta)$ is an \mathbb{R}^c -valued function on $\mathcal{W} \times \Theta_1$ that is continuous in w for all $\theta \in \Theta_1$ and continuous in θ uniformly over $\theta \in \Theta_1 \quad \forall w \in \mathcal{W}$, and (d) $E \sup_{\theta \in \Theta_1} \sup_{s \in (0, T^*]} |f(d(s, Z_t), \theta)|^{1+\varepsilon} < \infty$ for some $\varepsilon > 0$. Then, $\sup_{\theta \in \Theta_1} \left| \frac{1}{T} \Sigma_1^T [f(W_{T^*}, \theta) - Ef(W_{T^*}, \theta)] \right| \xrightarrow{P} 0$.

PROOF OF THEOREM 1: We apply Lemma A1 with $Q_T(\theta) = \left(\frac{1}{T} \Sigma_1^T m(W_{T^*}, \theta) \right)' \hat{\gamma} \left(\frac{1}{T} \Sigma_1^T m(W_{T^*}, \theta) \right)$ and $Q(\theta) = m(\theta)' \gamma m(\theta)$. Condition (b) of Lemma A1 holds by Assumption 1(h). Using Assumption 1(d), condition (a) of Lemma A-1 holds if $\sup_{\theta \in \Theta} \left\| \frac{1}{T} \Sigma_1^T m(W_{T^*}, \theta) - m(\theta) \right\| \xrightarrow{P} 0$. The latter holds if

$$(A.2) \quad \sup_{\theta \in \Theta} \left\| \frac{1}{T} \Sigma_1^T [m(W_{T^*}, \theta) - Em(W_{T^*}, \theta)] \right\| \xrightarrow{P} 0 \quad \text{and}$$

$$(A.3) \quad \sup_{\theta \in \Theta} \left\| \frac{1}{T} \Sigma_1^T Em(W_{T^*}, \theta) - m(\theta) \right\| \rightarrow 0.$$

Equation (A.2) holds by Lemma A2 under Assumptions 1(a)-(c), and 1(e)-(g). Equation (A.3) holds by Lemma 1. \square

PROOF OF LEMMA A-1: This lemma is well-known. For example, a proof of it is given in Andrews (1993, Lemma A-1). \square

PROOF OF LEMMA A2: By Theorem 4 and Lemma 4(a) of Andrews (1992) with $(Z_t, q_t(Z_t, \theta))$ set equal to $(W_{T^*}, f(W_{T^*}, \theta))$, it suffices to verify the conditions BD, P-WLLN, DM, and TSE-1B of Andrews (1992). BD holds by Assumption (b). DM holds by Assumption (d). TSE-1B holds by Assumptions (a) and (c), because $\frac{1}{T} \Sigma_1^T P(W_{T^*} \in A) - \mu(A) = \frac{1}{T^*} \int_0^{T^*} P(d(s, Z_t) \in A) ds$ by an argument analogous to that given in the proof of Lemma 1. It remains to show P-WLLN. Assumption (a) allows one to apply Theorem 6.5 of Pötscher and Prucha (1991a) to obtain that $\left\{ d\left(\frac{T^*}{T}t, Z_t\right) : t \leq T, T \geq 1 \right\}$ is L^0 -approximable by the strong mixing base $\{Y_t : t = \dots, 0, 1, \dots\}$. This result and Assumptions (a), (b), and (d) allow one to apply Theorem 6.5 of Pötscher and Prucha (1991a) to obtain that $\{f(W_{T^*}, \theta) : t \leq T, T \geq 1\}$ is L^0 -approximable by the base $\{Y_t\}$ for all $\theta \in \Theta_1$. (The assumptions that $\{Z_t : t \geq 1\}$ are identically distributed and $\sup_{s \in (0, T^*]} |d(s, z)|$

$< \infty$ guarantee that $\left\{d\left(\frac{T^*}{T}t, Z_t\right) : t \leq T, T \geq 1\right\}$ is tight, as is required for Theorem 6.5.) By Assumption (d) and Theorem 6.1 of Pötscher and Prucha (1991a), the approximators can be taken to be the conditional means $\{E(f(W_{Tt}, \theta) | Y_{t-m}, \dots, Y_{t+m}) : t \leq T, T \geq 1, m \geq 1\}$. In consequence, $\{f(W_{Tt}, \theta) : t \leq T, T \geq 1\}$ is L^0 -NED on the strong mixing base $\{Y_t\}$ for all $\theta \in \Theta_1$. The WLLN given in Lemma A-2 of Andrews (1993) with X_{Tt} equal to an element of the c -vector $f(W_{Tt}, \theta) - Ef(W_{Tt}, \theta)$ now yields P-WLLN. \square

PROOF OF LEMMA 2: Part (a) holds under Assumptions 2(a) and 2(f) by an argument analogous to that used to prove Lemma 1.

To establish part (b), we make the following definitions: Let S_J equal S with the sums $\sum_{j=1}^{\infty}$ replaced by the truncated sums $\sum_{j=1}^J$. Let $R_J = S - S_J$. Let $S_{TT} = \text{Var}(\sqrt{T} \bar{m}_T(\theta_0)) = \frac{1}{T} \sum_{s=1}^T \sum_{t=1}^T Em(W_{Ts})m(W_{Tt})'$, where $m(W_{Tt}) = m(W_{Tt}, \theta_0)$. Let S_{TJ} equal S_{TT} with the sums over s and t restricted to include only (s, t) pairs for which $|s-t| \leq J$. Let $R_{TJ} = S_{TT} - S_{TJ}$.

Below, we show that

$$(A.4) \quad \limsup_{J \rightarrow \infty} \sup_{T \geq 1} \|R_{TJ}\| = 0 ,$$

$$(A.5) \quad \lim_{J \rightarrow \infty} S_J = S , \text{ and}$$

$$(A.6) \quad \lim_{T \rightarrow \infty} S_{TJ} = S_J \quad \forall J \geq 0 .$$

Then, given any $\varepsilon > 0$, we can choose J^* sufficiently large that $\sup_{T \geq 1} \|R_{TJ^*}\| < \varepsilon/3$ and $\|S_{J^*} - S\| < \varepsilon/3$. Next, we can choose T_ε sufficiently large such that $\forall T \geq T_\varepsilon, \|S_{TT^*} - S_{J^*}\| < \varepsilon/3$. This yields the result of Lemma 2(b): $\forall T \geq T_\varepsilon$,

$$(A.7) \quad \begin{aligned} \|S_{TT} - S\| &= \|S_{TJ^*} + R_{TJ^*} - S_{J^*} + S_{J^*} - S\| \\ &\leq \|S_{TJ^*} - S_{J^*}\| + \sup_{T_1 \geq 1} \|R_{T_1 J^*}\| + \|S_{J^*} - S\| < \varepsilon . \end{aligned}$$

It remains to show (A.4)-(A.6). To show (A.4), we write

$$\begin{aligned}
\sup_{T \geq 1} |R_{TJ}| &\leq \sup_{T \geq 1} \frac{1}{T} \sum_{s=1}^T \sum_{\substack{t=1 \\ |s-t| > J}}^T |Em(W_{Ts})m(W_{Tt})'| \\
\text{(A.8)} \quad &\leq C \sup_{T \geq 1} \frac{1}{T} \sum_{s=1}^T \sum_{\substack{t=1 \\ |s-t| > J}}^T \left(\alpha_{\lfloor |s-t|/4 \rfloor}^{1/2-1/r} + v_{\lfloor |s-t|/4 \rfloor} \right) \\
&= 2C \sup_{T \geq 1} \sum_{j=J+1}^T \left(1 - \frac{j}{T} \right) \left(\alpha_{\lfloor j/4 \rfloor}^{1/2-1/r} + v_{\lfloor j/4 \rfloor} \right) \\
&\rightarrow 0 \text{ as } J \rightarrow \infty,
\end{aligned}$$

where the second inequality holds for some constant $C < \infty$ by a covariance inequality for L^r -bounded L^2 -NED rv's, see p. 110 of Gallant and White (1988), and the convergence to zero holds because $\sum_{j=1}^{\infty} \left(\alpha_j^{1/2-1/r} + v_j \right) < \infty$ by Assumption 2(c).

To show (A.5), we write

$$\begin{aligned}
|S - S_J| &\leq 2 \sum_{j=J+1}^{\infty} \frac{1}{T^*} \int_0^{T^*} |Em(d(s, Z_t)m(d(s, Z_{t-j}))'| ds \\
\text{(A.9)} \quad &\leq 2C \sum_{j=J+1}^{\infty} \left(\alpha_{\lfloor j/4 \rfloor}^{1/2-1/r} + v_{\lfloor j/4 \rfloor} \right) \\
&\rightarrow 0 \text{ as } J \rightarrow \infty,
\end{aligned}$$

where the second inequality follows by the same covariance inequality as used above.

To show (A.6), we write

$$\begin{aligned}
S_{TJ} - S_J &= \sum_{j=0}^J G_{Tj} + \sum_{j=1}^J G'_{Tj}, \text{ where} \\
\text{(A.10)} \quad G_{Tj} &= \frac{1}{T} \sum_{t=j+1}^T Em(W_{Tt})m(W_{Tt-j})' + \frac{1}{T^*} \int_0^{T^*} Em(d(s, Z_t)m(d(s, Z_{t-j}))'| ds.
\end{aligned}$$

Thus, for (A.10), it suffices to show that $\lim_{T \rightarrow \infty} |G_{Tj}| = 0 \quad \forall j = 0, 1, \dots, J$. For notational simplicity, let $m(t_1, t_2)$ abbreviate $m\left(d\left(\frac{T^*}{T}t_1, Z_{t_2}\right), \theta_0\right)$. We have

$$\begin{aligned}
A_{Tj} &= \left\| \frac{1}{T} \sum_{t=j+1}^T Em(t, t)m(t-j, t-j)' - \frac{1}{T} \sum_{t=j+1}^T Em(t, t)m(t, t-j)' \right\| \\
&= \left\| \frac{1}{T} \sum_{t=1}^{T-j} Em(t+j, 1+j)[m(t, 1)' - m(t+j, 1)'] \right\| \\
\text{(A.11)} \quad &\leq \sup_{t \leq T, T \geq 1} (E \|m(t, 1+j)\|^2)^{1/2} \cdot \frac{1}{T} \sum_{t=1}^{T-j} (E \|m(t, 1) - m(t+j, 1)\|^2)^{1/2} \\
&= C_0 \cdot \int_0^{1-j/T} \left(E \left\| m(\xi_T(u, Z_1)) - m\left(\xi_T\left(u + \frac{j}{T}, Z_1\right)\right) \right\|^2 \right)^{1/2} du,
\end{aligned}$$

where $\xi_T(u, z)$ is defined in the proof of Lemma 1 and the last inequality holds because $\xi_T\left(u + \frac{j}{T}, z\right) = d\left(\frac{T^*}{T}([Tu] + 1 + j), z\right)$. We now have

$$\begin{aligned}
0 &\leq \lim_{T \rightarrow \infty} A_{Tj} \leq C_0 \lim_{T \rightarrow \infty} \int_0^1 \left(E \left\| m(\xi_T(u, Z_1)) - m\left(\xi_T\left(u + \frac{j}{T}, Z_1\right)\right) \right\|^2 \right)^{1/2} du \\
\text{(A.12)} \quad &= C_0 \int_0^1 \left(E \lim_{T \rightarrow \infty} \left\| m(\xi_T(u, Z_1)) - m\left(\xi_T\left(u + \frac{j}{T}, Z_1\right)\right) \right\|^2 \right)^{1/2} du \\
&= 0
\end{aligned}$$

where $\xi_T(u, z) = 0$ by definition for $u > 1$, the first equality holds by the DCT applied twice, and the second equality holds because $\xi_T(u, z) \rightarrow \xi(u, z)$ and $\xi_T\left(u + \frac{j}{T}, z\right) \rightarrow \xi(u, z)$ as $T \rightarrow \infty$ for Lebesgue almost all $u \forall z \in \mathcal{Z}$. Using (A.12), we obtain

$$\begin{aligned}
\lim_{T \rightarrow \infty} G_{Tj} &= \lim_{T \rightarrow \infty} \left\| \frac{1}{T} \sum_{j=1}^{T-j} Em(t, 1+j)m(t, 1)' - \frac{1}{T^*} \int_0^{T^*} Em(d(s, Z_{1+j}))m(d(s, Z_1))' ds \right\| \\
\text{(A.13)} \quad &= \lim_{T \rightarrow \infty} \left\| \int_0^{1-\frac{j}{T}} Em(\xi_T(u, Z_{1+j}))m(\xi_T(u, Z_1))' du - \int_0^1 Em(\xi(u, Z_{1+j}))m(\xi(u, Z_1))' du \right\| \\
&\leq \int_0^1 E \lim_{T \rightarrow \infty} \left\| m(\xi_T(u, Z_{1+j}))m(\xi_T(u, Z_1))' - m(\xi(u, Z_{1+j}))m(\xi(u, Z_1))' \right\| du \\
&= 0,
\end{aligned}$$

where the inequality uses the DCT and the last equality uses the facts that $\xi_T(u, z) - \xi(u, z)$ for Lebesgue almost all $u \ \forall z \in \mathcal{Z}$ and $m(w, \theta_0)$ is continuous in w . This establishes (A.6). \square

PROOF OF THEOREM 2: Since $\hat{\theta}$ minimizes $\bar{m}_T(\theta)' \hat{\gamma} \bar{m}_T(\theta)$ over Θ and is in the interior of Θ w.p. - 1 by (Assumption) 2(d), we have

$$(A.14) \quad \left[\frac{\partial \bar{m}_T(\hat{\theta})}{\partial \theta'} \right]' \hat{\gamma} \sqrt{T} \bar{m}_T(\hat{\theta}) = o_p(1) .$$

Let $\bar{m}_{Tj}(\theta)$ denote the j -th element of $\bar{m}_T(\theta)$. A mean value expansion gives

$$(A.15) \quad \sqrt{T} \bar{m}_{Tj}(\hat{\theta}) = \sqrt{T} \bar{m}_{Tj}(\theta_0) + \frac{\partial \bar{m}_{Tj}(\theta^*)}{\partial \theta'} \sqrt{T} (\hat{\theta} - \theta_0) ,$$

where $\theta^* (= \theta_j^*)$ is a rv on the line segment joining $\hat{\theta}$ and θ_0 , $\forall j = 1, \dots, v$, and hence $\theta^* \xrightarrow{p} \theta_0$.

Below we show that (i) $\frac{\partial \bar{m}_T(\theta^*)}{\partial \theta'} \xrightarrow{p} M$ and (ii) $\sqrt{T} \bar{m}_T(\theta_0) \xrightarrow{d} N(0, S)$. Results (i) and (ii), equations (A.14) and A.15), and Assumptions 1(d) and 2(g) combine to give the desired result: $\sqrt{T}(\hat{\theta} - \theta_0) = -(M' \gamma M)^{-1} M' \gamma \sqrt{T} \bar{m}_T(\theta_0) + o_p(1) \xrightarrow{d} N(0, V)$.

To establish result (i), we write

$$(A.16) \quad \left\| \frac{\partial \bar{m}_T(\theta^*)}{\partial \theta'} - M \right\| \leq \left\| \frac{\partial \bar{m}_T(\theta^*)}{\partial \theta'} - E \frac{\partial \bar{m}_T(\theta)}{\partial \theta'} \Big|_{\theta=\theta^*} \right\| + \left\| E \frac{\partial \bar{m}_T(\theta)}{\partial \theta'} \Big|_{\theta=\theta^*} - E \frac{\partial \bar{m}_T(\theta_0)}{\partial \theta'} \right\| + \left\| E \frac{\partial \bar{m}_T(\theta_0)}{\partial \theta'} - M \right\| .$$

The first summand $\xrightarrow{p} 0$ under Assumption 2 by Lemma A-2 with $f(w, \theta) = \frac{\partial}{\partial \theta'} m(w, \theta)$. The third summand $\xrightarrow{p} 0$ by Lemma 2(a). The second summand $\xrightarrow{p} 0$, because

$$(A.17) \quad \begin{aligned} & \lim_{\theta \rightarrow \theta_0} \sup_{T \geq 1} \left\| E \frac{\partial \bar{m}_T(\theta)}{\partial \theta'} - E \frac{\partial \bar{m}_T(\theta_0)}{\partial \theta'} \right\| \\ & \leq \lim_{\theta \rightarrow \theta_0} E \sup_{s \in (0, T^*)} \left\| \frac{\partial}{\partial \theta'} m(d(s, Z_t), \theta) - \frac{\partial}{\partial \theta'} m(d(s, Z_t), \theta_0) \right\| \\ & = E \lim_{\theta \rightarrow \theta_0} \sup_{s \in (0, T^*)} \left\| \frac{\partial}{\partial \theta'} m(d(s, Z_t), \theta) - \frac{\partial}{\partial \theta'} m(d(s, Z_t), \theta_0) \right\| = 0 , \end{aligned}$$

where the first equality holds by the DCT with dominating function

2 $\sup_{\theta \in \theta_0} \sup_{s \in (0, T^*)} \left| \frac{\partial}{\partial \theta'} m(d(s, Z_t), \theta_0) \right|$ under Assumption 2(f) and the second equality holds by the continuity of $\frac{\partial}{\partial \theta'} m(w, \theta)$ in (w, θ) under Assumption 2(f).

Result (ii) holds by the Cramer-Wold device, Lemma 2(b), and a CLT of Davidson (1992, Theorem 3.6) with $X_{nt} = \alpha' m(W_{Tt}, \theta_0) / \text{Var}^{1/2}(\sqrt{T} \alpha' \bar{m}_T(\theta_0))$ and $c_{nt} = \text{Var}^{-1/2}(\sqrt{T} \alpha' \bar{m}_T(\theta_0))$ for arbitrary v -vector $\alpha \neq 0$, using Assumptions 2(b)-(d). (Davidson's CLT yields $\sqrt{T} \alpha' \bar{m}_T(\theta_0) / \text{Var}^{1/2}(\sqrt{T} \alpha' \bar{m}_T(\theta_0)) \xrightarrow{d} N(0, 1)$.) \square

Footnotes

¹We thank B. Pötscher and H. White for helpful comments. The first author gratefully acknowledges the research support of the National Science Foundation through grant number SES-9121914.

References

- Allen, S. G. (1992): "Changes in the Cyclical Sensitivity of Wages in the United States, 1891-1987," *American Economic Review*, 82, 122-140.
- Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.
- Andrews, D. W. K. (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, 241-257.
- Andrews, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, forthcoming.
- Andrews, D. W. K. and H. Y. Chen (1992): "Approximately Median-unbiased Estimation of Autoregressive Models with Application to U.S. Macroeconomic and Financial Time Series," Cowles Foundation Discussion Paper No. 1026, Yale University.
- Andrews, D. W. K. and J. C. Monahan (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953-966.
- Bierens, H. J. (1981): *Robust Methods and Asymptotic Theory*. Lecture Notes in Economics and Mathematical Systems No. 192. Berlin: Springer Verlag.
- Crowder, M. J. (1976): "Maximum Likelihood Estimation with Dependent Observations," *Journal of the Royal Statistical Society, Series B*, 38, 45-53.
- Dalhaus, R. (1993): "Fitting Time Series Models to Nonstationary Processes," unpublished manuscript, Institut für Angewandte Mathematik, Universität Heidelberg, Germany.
- Davidson, J. (1992): "A Central Limit Theorem for Globally Nonstationary Near-epoch Dependent Functions of Mixing Processes," *Econometric Theory*, 8, 313-329.
- DeJong, D. and C. H. Whiteman (1991): "Reconsidering 'Trends and Random Walks in Macroeconomic Time Series'," *Journal of Monetary Economics*, 28, 221-254.
- Eichenbaum, M. and L. P. Hansen (1990): "Estimating Models with Intertemporal Substitution Using Aggregate Time Series Data," *Journal of Business and Economic Statistics*, 8, 53-69.
- Fair, R. C. (1984): *Specification, Estimation, and Analysis of Macroeconometric Models*. Cambridge, MA: Harvard University Press.
- Fair, R. C. and K. M. Dominguez (1991): "Effects of Changing U.S. Age Distribution on Macroeconomic Equations," *American Economic Review*, 81, 1276-1294.
- Gallant, A. R. (1987): *Nonlinear Statistical Models*. New York: Wiley.
- Gallant, A. R. and H. White (1988): *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Oxford: Basil Blackwell.
- Hamilton, J. D. (1992): "Was the Inflation During the Great Depression Anticipated?" Evidence from the Commodity Futures Market," *American Economic Review*, 82, 157-182.
- Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

- Ibragimov, I. A. (1962): "Some Limit Theorems for Stationary Processes," *Theory of Probability and its Applications*, 7, 349-382.
- Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin (1992): "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root," *Journal of Econometrics*, 54, 159-178.
- MacNeill, I. B. (1978): "Properties of Sequences of Partial Sums of Polynomial Regression Residuals with Applications to Tests for Change of Regression at Unknown Times," *Annals of Statistics*, 6, 422-433.
- Nelson, C. R. and C. I. Plosser (1982): "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics*, 10, 139-162.
- Newey, W. K. and D. McFadden (1992): "Estimation in Large Samples," paper prepared for the *Handbook of Econometrics*, Vol. 4, Department of Economics, M.I.T.
- Ogaki, M. (1992): "Generalized Method of Moments: Econometric Applications," *Handbook of Statistics, Vol. 11: Econometrics*, ed. by G. S. Maddala, C. R. Rao, and H. D. Vinod. Amsterdam: North-Holland.
- Perron, P. (1989): "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Econometrica*, 57, 1361-1401.
- Phillips, P. C. B. (1991): "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends," *Journal of Applied Econometrics*, 6, 333-364.
- Phillips, P. C. B. and B. E. Hansen (1990): "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *Review of Economic Studies*, 57, 99-125.
- Pötscher, P. M. and I. R. Prucha (1991a): "Basic Structure of the Asymptotic Theory in Dynamic Nonlinear Econometric Models, Part I: Consistency and Approximation Concepts," *Econometric Reviews*, 10, 125-216.
- Pötscher, P. M. and I. R. Prucha (1991b): "Basic Structure of the Asymptotic Theory in Dynamic Nonlinear Econometric Models, Part II: Asymptotic Normality," *Econometric Reviews*, 10, 253-325.
- Priestley, M. B. and M. T. Chao (1972): "Non-parametric Function Fitting," *Journal of the Royal Statistical Society, Series B*, 34, 385-392.
- Wooldridge, J. M. (1986): *Asymptotic Properties of Econometric Estimators*. Ph.D. Thesis, University of California - San Diego.
- Wooldridge, J. M. (1993): "Estimation and Inference for Dependent Processes," prepared for the *Handbook of Econometrics*, Vol. IV. Amsterdam: North Holland, forthcoming.
- Wu, C. F. (1981): "Asymptotic Theory of Nonlinear Least Squares Estimation," *Annals of Statistics*, 9, 501-513.
- Zivot, E. and D. W. K. Andrews (1992): "Further Evidence on the Great Crash, the Oil Price Shock and the Unit Root Hypothesis," *Journal of Business and Economic Statistics*, 10, 251-270.