DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

INFERENCE IN ECONOMETRIC MODELS WITH STRUCTURAL CHANGE

Donald W. K. Andrews
California Institute of Technology
and Cowles Foundation, Yale University

Ray C. Fair
Cowles Foundation, Yale University

SOCIAL SCIENCE WORKING PAPER 636

February 1987

INFERENCE IN ECONOMETRIC MODELS

WITH STRUCTURAL CHANGE

Donald W. K. Andrews and Ray C. Fair
Cowles Foundation
Yale University
February, 1987

ABSTRACT AND HEADNOTE

This paper extends the classical Chow (1960) test for structural change in linear regression models to a wide variety of nonlinear models, estimated by a variety of different procedures. Wald, Lagrange multiplier-like, and likelihood ratio-like test statistics are introduced. The results allow for heterogeneity and temporal dependence of the observations.

In the process of developing the above tests, the paper also provides a compact presentation of general unifying results for estimation and testing in nonlinear parametric econometric models.

JEL Classification Numbers: 211, 212

Keywords: Chow Test; Dynamic Model; Econometric Model; Lagrange Multiplier Test; Likelihood Ratio Test; Structural Change; Wald Test.

Address for Editorial Communication: Donald W. K. Andrews

Until June 5, 1987: Division of Humanities and Social Sciences
California Institute of Technology
Pasadena, CA, 91120, USA

From then until August 18, 1987: 3404 Point Grey Road
Vancouver, B.C., V6R1A5, CANADA

Thereafter: Cowles Foundation
Box 2125, Yale Station
New Haven, CT, 06520, USA

1. INTRODUCTION

This paper is concerned with testing for structural change in nonlinear models. For the classical linear regression model the Chow (1960) test commonly is used, and for the linear simultaneous equations model the Lo and Newey (1985) extension of the Chow test can be used. Somewhat surprisingly, however, more general cases have not been considered in the literature. We consider fairly wide classes of models, estimators, and test statistics in this paper. We also cover the case where the structural change is only partial, i.e., it pertains to only a subset of the coefficients in the model. Some of the test statistics we present can be computed using the output from standard software packages.

The models we consider may be dynamic, simultaneous, and nonlinear and may include limited dependent variables. The error terms may show a very general form of temporal dependence and heteroskedasticity. The estimators include nonlinear least squares (LS), two stage least squares (2SLS), three stage least squares (3SLS), maximum likelihood (ML), and M-estimators. The tests covered are the Wald (W) test, a Lagrange multiplier-like (LM) test, and a likelihood ratio-like (LR) test. Under certain conditions, we show that the test statistics are asymptotically chi-square under the null hypothesis of no structural change and asymptotically noncentral chi-square under sequences of local alternatives.

The paper is organized as follows. The general case is considered in Section 2, with proofs in the Appendix. Three special cases then are considered in Sections 3, 4, and 5. The three cases are (1) the single equation nonlinear regression model, (2) the nonlinear simultaneous equations

model, and (3) any model estimated by maximum likelihood. For those primarily interested in the application of the tests, Section 2 can be skipped and Sections 3, 4, and 5 can be skimmed up to the point where the formulae for the test statistics are presented and the computational requirements are discussed.

The general results of Section 2 have the added feature that in several respects they provide the most general unifying results in the econometrics literature for estimation and testing in dynamic and nondynamic, nonlinear, finite dimensional parametric models. Also, they do so in a much more economical fashion than is available elsewhere, such as in Gallant (1987) or Gallant and White (1987).[2] In contrast to Gallant (1987, Chs. 3 and 7), least mean distance and method of moment estimators are treated simultaneously. In contrast to Gallant and White (1987), a more complete treatment of multi-step procedures is given.[3]

The approach taken in Section 2 is a variant of that of Gallant (1987, Ch. 7). In contrast to Gallant (1987), however, the results are stated such that they can be applied with any uniform law of large numbers and any central limit theorem. This allows developments in these areas--especially with respect to temporal dependence-- to be adopted readily.

2. GENERAL RESULTS

This section gives general results for estimation and testing in models with structural change. The basic approach that we adopt is one that has evolved in a long series of papers on inference in nonlinear models. Such papers include those of Wald (1949), Huber (1967), Jennrich (1969), Burguete, Gallant, and Souza (1982) (denoted BGS (1982)), Domowitz and White

(1982), Bates and White (1985), Gallant (1987), and Gallant and White (1987). The present approach most closely follows that of BGS (1982) and Gallant (1987) and our notation is chosen to be as compatible as possible with them.

This section is outlined as follows: We first consider a class of extremum estimators for models where structural change may or may not occur. Consistency and asymptotic normality of these estimators are established. Consistent estimators of their asymptotic covariance matrices are provided. We then consider tests of general nonlinear restrictions. Wald, Lagrange multiplier-like, and likelihood ratio-like tests are shown to be asymptotically chi-square under the null hypothesis and asymptotically non-central chi-square under local alternatives under certain conditions.

2.1. *Consistency of Estimators*

The data are given by a doubly infinite sequence of random vectors (rv's) $\{W_t\} = \{W_t : t = \ldots, -2, -1, 1, 2, \ldots\}$ defined on some probability space (Ω, F, P). Probability statements made below refer to probabilities calculated under P. The observed sample of size $T = T_1 + T_2$ is $\{W_t : t = -T_1, \ldots, -1, 1, \ldots, T_2\}$. The point $t = 0$ is the point of structural change, if such change occurs. (For notational convenience, the sequence $\{W_t\}$ is indexed such that no W_0 rv exists.) In most cases, the asymptotics used below correspond to situations where

$$\pi_{1T} = T_1/T \to \pi_1 \in (0,1) \text{ and } \pi_{2T} = T_2/T \to \pi_2 \in (0,1) \quad \text{as} \quad T \to \infty. \qquad (2.1)$$

Extremum estimators are defined as follows.

DEFINITION: *A sequence of extremum estimators* $\{\hat{\theta}\} = \{\hat{\theta} : T = 1, 2, \ldots\}$ *is any sequence of* rv's *such that*

$$d(\bar{m}_T(\hat{\theta}, \hat{\tau}), \hat{\tau}) = \inf_{\theta \in \Theta} d(\bar{m}_T(\theta, \hat{\tau}), \hat{\tau}) \qquad (2.2)$$

with probability that goes to one as $T \to \infty$, *where* $\bar{m}_T(\theta, \tau) = \frac{1}{T} \sum_{t=-T_1}^{T_2} m_t(\theta, \tau)$, $m_t(\theta, \tau) = m_t(W_t, \theta, \tau)$, and $m_t(\cdot,\cdot,\cdot) : R^{k_t} \times \Theta \times T_1 \to R^v$ *where* $T_1 \subset R^u$, $m_0(\theta, \tau) \equiv 0$, $\hat{\tau}$ *is a random* u-*vector (which depends on* T *in general), and* $d(\cdot,\cdot)$ *is a non-random real-valued function (which does not depend on* T).

For notational simplicity, we let $\bar{m}_T(\theta)$ abbreviate $\bar{m}_T(\theta, \hat{\tau})$ and we let Σ_a^b denote $\sum_{t=a}^{b}$ for arbitrary integers $a \le b$.

In the case of <u>pure structural change</u>, the parameter vector θ can be partitioned into two sub-vectors $(\theta_1', \theta_2')'$ such that $m_t(\theta, \tau)$ does not depend on θ_1 for $t > 0$ or on θ_2 for $t < 0$. In the case of <u>partial structural change</u>, the parameter vector θ can be partitioned as $(\theta_1', \theta_2', \theta_3')'$, where θ_1 and θ_2 are as above and θ_3 is unrestricted.

We now describe briefly several common estimators in terms of the above framework; more details are given in Sections 3-5. Consider the following nonlinear regression model with partial structural change: $Y_t = f_t(X_t, \theta_j, \theta_3) + U_t$, $t = -T_1, \ldots, -1, 1, \ldots, T_2$, where $j = 1$ for $t < 0$ and $j = 2$ for $t > 0$. Let $W_t = (Y_t, X_t')'$. The nonlinear least squares estimator of $\theta = (\theta_1', \theta_2', \theta_3')'$ can be defined either as one that minimizes the sum of squared residuals or one that solves the first order conditions of this minimization problem. Thus, we can take either $m_t(\theta, \tau) = (Y_t - f_t(X_t, \theta_j, \theta_3))^2$ and $d(m, \tau) = m$ or $m_t(\theta, \tau) = (Y_t - f_t(X_t, \theta_j, \theta_3))\frac{\partial}{\partial\theta} f_t(X_t, \theta_j, \theta_3)$ and $d(m, \tau) = m'm/2$.

For the LS estimator, no nuisance parameter τ appears in the functions $m_t(\theta, \tau)$ and $d(m, \tau)$. If an M-estimator is used, however, then $m_t(\theta, \tau)$ is set equal to $\bar{\rho}((Y_t - f_t(X_t, \theta_j, \theta_3))/\tau)$ or

$\psi((Y_t - f_t(X_t, \theta_j, \theta_3))/\tau)\frac{\partial}{\partial\theta}f_t(X_t, \theta_j, \theta_3)$, where the nuisance parameter τ is a scale parameter, $\psi(x) = \frac{d}{dx}\tilde{\rho}(x)$, and $d(\cdot,\cdot)$ is as above. Huber (1981) discusses different choices for the function $\tilde{\rho}(\cdot)$.

Next, consider two stage least squares (2SLS) estimation of a single, nonlinear, simultaneous equation with pure structural change. The model is: $f_t(Y_t, X_t, \theta_j) = U_t$, for $t = -T_1, \ldots, T_2$, where $j = 1$ for $t < 0$ and $j = 2$ for $t > 0$, Y_t is a vector of endogenous variables, and X_t is a vector of predetermined variables. Let Z_t be a vector of instrumental variables that can be partitioned as $Z_t = (Z_{1t}', Z_{2t}')'$, where $Z_{1t} = \underset{\sim}{0}$ for $t > 0$ and $Z_{2t} = \underset{\sim}{0}$ for $t < 0$. Let $W_t = (Y_t', X_t', Z_t')'$. The 2SLS estimator of θ is defined by taking $m_t(\theta,\tau) = f_t(Y_t, X_t, \theta_j)Z_t$ and $d(m,\tau) = m'D(\tau)m/2$, where τ equals the non-redundant elements of $D(\tau) = \left(\lim_{T\to\infty} \frac{1}{T} \Sigma_{T_1}^{T_2} EZ_tZ_t'\right)^{-1}$ and $D(\hat{\tau}) = \left(\frac{1}{T} \Sigma_{T_1}^{T_2} Z_tZ_t'\right)^{-}$.

We now return to the general case. In what follows we avoid imposing conditions that are used just to ensure measurability of $\hat{\theta}$ by stating results that hold for any sequence of <u>rv's</u> $\{\hat{\theta}\}$. Such results have content only if such a sequence exists. Clearly, sequences $\{\hat{\theta}\}$ that satisfy (2.2), but are not necessarily measurable, always exist, since Θ is assumed below to be compact. Further, we note that one set of sufficient conditions for the existence of a measurable sequence $\{\hat{\theta}\}$ is that $d(\bar{m}_T(\theta), \hat{\tau})$, viewed as a function from $\Omega \times \Theta$ to R, is continuous in θ for each $\omega \in \Omega$ and is measurable for each fixed $\theta \in \Theta$, and Θ is a compact subset of some Euclidean space (see Jennrich (1969), Lemma 2).

For consistency we assume the following.

ASSUMPTION 1: (a) *Θ is compact.*

(b) *$\hat{\tau}$ is a rv and $\hat{\tau} \xrightarrow{p} \tau_0$ as $T \to \infty$ for some $\tau_0 \in T_1 \subset R^u$.*

(c) *There exists a Borel measurable function $m(\cdot,\cdot) : \Theta \times T \to R^v$ such that $\bar{m}_T(\theta,\tau) \xrightarrow{p} m(\theta,\tau)$ uniformly over $(\theta,\tau) \in \Theta \times T$ as $T \to \infty$, where $T \subset T_1$ is some compact neighborhood of τ_0.*

(d) *$d(\cdot,\cdot)$ is uniformly continuous on $m(\Theta,T) \times T$ and $m(\theta_0, \tau)$ is continuous in τ at τ_0.*

(e) *$\underline{\lim}_{\theta_*\to\theta,\tau_*\to\tau_0} d(m(\theta_*, \tau_*), \tau_*) \geq d(m(\theta_0, \tau_0), \tau_0)$ with equality iff $\theta = \theta_0$.*

For notational simplicity, we often denote $m(\theta, \tau_0)$ by $m(\theta)$.

Assumption 1(a) is standard in the nonlinear econometrics literature. Assumption 1(b) can be verified straightforwardly by application of a weak law of large numbers (WLLN) in some cases and by the application of Theorem 1 below to get consistency of $\hat{\tau}$ rather than $\hat{\theta}$ in other cases. The function $m(\theta,\tau)$ of assumption 1(c) generally is given by $\lim_{T\to\infty} \frac{1}{T} \Sigma_{T_1}^{T_2} Em_t(\theta,\tau)$. Thus, assumption 1(c) holds if these limits exist and if $\left\{\frac{1}{T_1} \Sigma_{T_1}^{-1} m_t(\theta,\tau)\right\}$ and $\left\{\frac{1}{T_2} \Sigma_1^{T_2} m_t(\theta,\tau)\right\}$ satisfy uniform WLLNs over $\Theta \times T$. The latter hold under conditions that allow considerable heterogeneity and temporal dependence. It is sufficient that $\{m_t(\theta,\tau)\}$ satisfy a smoothness condition in (θ,τ), a moment condition, and a condition of asymptotically weak temporal dependence --see Andrews (1987b), Gallant (1987, Ch. 7, Thm. 1), Potscher and Prucha (1986), or Bierens (1984, Lemma 2). Assumption 1(d) holds trivially in most applications, since $d(m,\tau)$ usually is continuous on $R^v \times R^u$ and $m(\Theta,T) \times T$ is contained in a compact subset of $R^v \times R^u$. Assumption 1(e) is the uniqueness assumption that ensures that $\{\hat{\theta}\}$ converges to a point θ_0 rather than to a multi-element subset Θ_0 of Θ. Assumption 1(e) is satisfied if $m(\cdot,\cdot)$ is continuous on $\Theta \times T$ and θ_0 uniquely minimizes $d(m(\theta, \tau_0), \tau_0)$ over Θ.[4]

THEOREM 1: *Under assumption* 1, *every sequence of extremum estimators* $\{\hat{\theta}\}$ *satisfies* $\hat{\theta} \xrightarrow{P} \theta_0$ *as* $T \to \infty$ *under* P.

The proofs of Theorem 1 and other results below are given in the Appendix.

2.2. *Asymptotic Normality of Estimators*

We now establish the asymptotic normality of sequences of extremum estimators $\{\hat{\theta}\}$ for models that may exhibit structural change. Their asymptotic covariance matrix V is defined as follows. Let

$S = \lim_{T\to\infty} Var_P(\sqrt{T}\, \bar{m}_T(\theta_0, \tau_0))$, $M = \lim_{T\to\infty} \frac{1}{T} \Sigma_{-T_1}^{T_2} E\frac{\partial}{\partial\theta'}m_t(\theta_0, \tau_0)$,

$D = \frac{\partial^2}{\partial m \partial m'}d(m(\theta_0, \tau_0), \tau_0)$, $J = M'DM$, $I = M'DSDM$, and $V = J^{-1}IJ^{-1}$, where $m_t(\cdot,\cdot)$ and $d(\cdot,\cdot)$ need not be defined as in assumption 1 (see footnote 4). For the LS estimator, M-estimators, and ML estimators, $m_t(\cdot,\cdot)$ and $d(\cdot,\cdot)$ must be chosen in this sub-section and the next to correspond to their first order conditions definition. For the 2SLS estimator, $m_t(\cdot,\cdot)$ and $d(\cdot,\cdot)$ are defined as in Section 2.1.

Let $\|\cdot\|$ denote the Euclidean norm and let $\frac{\partial}{\partial m}d(\cdot,\cdot)$ denote the derivative of $d(\cdot,\cdot)$ with respect to its first argument. We assume:

ASSUMPTION 2: (a) $\hat{\theta} \xrightarrow{P} \theta_0 \in R^P$ *as* $T \to \infty$.

(b)i. $\sqrt{T}(\hat{\tau} - \tau_0) = O_P(1)$ *as* $T \to \infty$ *for some* $\tau_0 \in T_1$,

ii. $\frac{\partial}{\partial m}d(E\bar{m}_T(\theta_0, \tau_0), \tau_0) = \underset{\sim}{0}$ $\forall T$ *large, and* iii. $\frac{\partial^2}{\partial \tau \partial m'}d(m(\theta_0), \tau_0) = \underset{\sim}{0}$.

(c) $\{m_t(\theta_0, \tau_0)\}$ *satisfy a central limit theorem* (CLT) *with covariance matrix* S. *That is,* $\sqrt{T}(\bar{m}_T(\theta_0, \tau_0) - E\bar{m}_T(\theta_0, \tau_0)) \xrightarrow{d} N(\underset{\sim}{0}, S)$ *as* $T \to \infty$.

(d) $\Theta \subset R^P$ *and* Θ *contains a convex neighborhood* Θ_c *of* θ_0.

(e) $\frac{\partial}{\partial m}d(m,\tau)$, $\frac{\partial^2}{\partial m \partial m'}d(m,\tau)$ *and* $\frac{\partial^2}{\partial \tau \partial m'}d(m,\tau)$ *exist and are continuous for* $(m,\tau) \in \mathcal{M} \times T$, *where* $\mathcal{M}$ *is some neighborhood of* $m(\theta_0, \tau_0)$.

(f) $m_t(\theta,\tau)$ *is once and twice continuously differentiable in* τ *and* θ, *respectively, on* $\Theta_c \times T$, $\forall t$, $\forall \omega \in \Omega$. $\{m_t(\theta,\tau)\}$, $\left\{\frac{\partial}{\partial\theta}m_t(\theta,\tau)\right\}$, $\left\{\frac{\partial}{\partial\tau}m_t(\theta,\tau)\right\}$, *and* $\left\{\sup_{\substack{(\theta^*,\tau^*)\in\Theta_c\times T,\\ a=1,\ldots,p}} \left\|\frac{\partial^2}{\partial\theta_a \partial\theta'}m_t(\theta^*, \tau^*)\right\|\right\}$ *are sequences of* $\mathcal{F}$\Borel-*measurable* rv's *that satisfy uniform* WLLNs *over* $(\theta,\tau) \in \Theta_c \times T$.

$m(\theta,\tau) = \lim_{T\to\infty} \frac{1}{T} \Sigma_{-T_1}^{T_2} Em_t(\theta,\tau)$, $M(\theta,\tau) = \lim_{T\to\infty} \frac{1}{T} \Sigma_{-T_1}^{T_2} E\frac{\partial}{\partial\theta'}m_t(\theta,\tau)$, *and*

$dm(\theta,\tau) = \lim_{T\to\infty} \frac{1}{T} \Sigma_{-T_1}^{T_2} E\frac{\partial}{\partial\tau}m_t(\theta,\tau)$ *exist uniformly for* $(\theta,\tau) \in \Theta_c \times T$[5] *and are continuous and* $dm(\theta_0, \tau_0) = \underset{\sim}{0}$.

(g) M'DM *is nonsingular.*

Assumption 2(a) can be established by Theorem 1 or some other consistency proof. Assumption 2(b) can be verified by applying a CLT to $\hat{\tau}$ in some cases and by applying the result of Theorem 2 below to $\hat{\tau}$ rather than $\hat{\theta}$ in other cases. Assumption 2(c) can be verified by defining $m_{Tt} = m_{t+T_1+1}(\theta_0, \tau_0)$ for $t = -T_1, \ldots, T_2$ to get a triangular array $\{m_{Tt} : t = 1, \ldots, T+1; T = 1, 2, \ldots\}$ to which any of a number of CLTs apply. Thus, assumption 2(c) holds under conditions that allow considerable heterogeneity and temporal dependence. It is sufficient that $Em_t(\theta_0, \tau_0) = \underset{\sim}{0}$, $\forall t$, and that $\{m_t(\theta_0, \tau_0)\}$ satisfy standard moment conditions and a condition of asymptotically weak temporal dependence--see Gallant (1987, Ch. 7, Thm. 2), McLeish (1975b, Thms 2.6, 3.8, and 4.2; 1977, Thm 2.4 and Cor. 2.11), Herrndorf (1984, Thm. and Cor. 1-4), or Withers (1981, Thms. 2.1-2.3).

Assumption 2(d) is standard. Assumption 2(e) often is satisfied trivially, since $d(m,\tau)$ often equals m or $m'D(\tau)m$, where $D(\tau)$ is a square matrix comprised of the elements of τ. Assumption 2(f) is a standard requirement of smoothness of $m_t(\theta,\tau)$ in θ and τ, the existence of certain limiting averages of expectations, and non-explosive non-trending behavior of the summands $\{m_t(\theta,\tau)\}$ and their first two derivatives. The smoothness conditions are stronger than necessary (cf., Huber (1967) and Pollard (1985)), but are satisfied in a large fraction of the cases encountered in practice. Assumption 2(g) is standard. For example, it reduces to nonsingularity of the information matrix in iid ML contexts.

THEOREM 2: *For any sequence of* extremum estimators $\{\hat{\theta}\}$ *that satisfies assumption* 2,

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(\underset{\sim}{0}, V) \quad as \quad T \to \infty .$$

Next we consider estimation of the covariance matrix V. Let $\hat{M} = \frac{1}{T} \Sigma_{-T_1}^{T_2} \frac{\partial}{\partial\theta'} m_t(\hat{\theta},\hat{\tau})$, $\hat{D} = \frac{\partial^2}{\partial m \partial m'} d(\bar{m}_T(\hat{\theta},\hat{\tau}), \hat{\tau})$, and $\hat{J} = \hat{M}'\hat{D}\hat{M}$. Let $\hat{S}$ be an estimator of S. If $\{m_t(\theta_0, \tau_0)\}$ is a sequence of independent rv's, then we can take $\hat{S} = \frac{1}{T} \Sigma_{-T_1}^{T_2} m_t(\hat{\theta},\hat{\tau}) m_t(\hat{\theta},\hat{\tau})'$. If $\{m_t(\theta_0, \tau_0)\}$ is a sequence of temporally dependent rv's, however, a more complicated estimator is required. The following choice is analogous to estimators suggested by Gallant (1987, pp. 551, 556) and Newey and West (1987). Let

$$\hat{S} = \hat{S}(\hat{\theta}), \text{ where } \hat{S}(\theta) = \pi_{1T}\hat{S}_1(\theta) + \pi_{2T}\hat{S}_2(\theta),$$

$$\hat{S}_1(\theta) = \frac{1}{T_1} \Sigma_{-T_1}^{-1} m_t(\theta)m_t(\theta)' + \sum_{v=1}^{\ell(T_1)} w\left(\frac{v}{\ell(T_1)}\right) \frac{1}{T_1} \Sigma_{-T_1+v}^{-1} [m_t(\theta)m_{t-v}(\theta)' + m_{t-v}(\theta)m_t(\theta)'], \tag{2.3}$$

$$\hat{S}_2(\theta) = \frac{1}{T_2} \Sigma_1^{T_2} m_t(\theta)m_t(\theta)' + \sum_{v=1}^{\ell(T_2)} w\left(\frac{v}{\ell(T_2)}\right) \frac{1}{T_2} \Sigma_{1+v}^{T_2} [m_t(\theta)m_{t-v}(\theta)' + m_{t-v}(\theta)m_t(\theta)'],$$

$m_t(\theta) = m_t(\theta,\hat{\tau})$, $\ell(T_1) = [T_1^{1/5}]$, $\ell(T_2) = [T_2^{1/5}]$, $[t]$ denotes the largest integer less than or equal to t, and $w(\cdot)$ yields the Parzen weights, i.e.,

$$w(x) = \begin{cases} 1 - 6x^2 + 6x^3 & \text{for } 0 \le x \le 1/2 \\ 2(1-x)^3 & \text{for } 1/2 \le x \le 1 \end{cases},$$ or the Bartlett weights, i.e., $w(x) = 1-x$ for $0 \le x \le 1$.

Conditions under which this estimator is consistent can be found in the references above. These conditions require $\{m_t(\theta_0, \tau_0)\}$ to have more moments finite than are required for $\{m_t(\theta_0, \tau_0)\}$ to satisfy an LLN or a CLT. See Gallant (1987) for conditions using near epoch dependence and Newey and West (1987) for conditions using strong mixing. Given the availability of such conditions, it is straightforward to verify the following assumption.

ASSUMPTION 3: $\hat{S} \xrightarrow{p} S$ *as* $T \to \infty$ *(where S is as in assumption 2).*

Let $\hat{I} = \hat{M}'\hat{D}\hat{S}\hat{D}\hat{M}$ and $\hat{V} = \hat{J}^{-}\hat{I}\hat{J}^{-}$, where $(\cdot)^{-}$ denotes some reflexive g-inverse (such as the Moore-Penrose inverse).

THEOREM 3: *Under assumptions* 2 *and* 3, $\hat{M} \xrightarrow{p} M$, $\hat{D} \xrightarrow{p} D$, *and* $\hat{V} \xrightarrow{p} V$ *as* $T \to \infty$.

Comment: When V simplifies, as occurs in many applications, then $\hat{V}$ simplifies or simpler estimators than $\hat{V}$ can be constructed.

2.3. *Tests of Hypotheses Concerning Structural Change*

We now consider tests of null hypotheses of the form $H_0 : h(\theta) = \underset{\sim}{0}$. Of particular interest are tests of pure and partial structural change. For testing pure structural change, the null hypothesis is $H_0 : \theta_1 = \theta_2$, where $\theta = (\theta_1', \theta_2')'$ and θ_1 and θ_2 are parameters associated only with the observations indexed by $t < 0$ and $t > 0$, respectively. In the case of partial structural change, the null hypothesis is $H_0 : \theta_1 = \theta_2$ where $\theta = (\theta_1', \theta_2', \theta_3')'$, θ_1 and θ_2 are as above, and θ_3 is a parameter that may be associated with the observations from all time periods. A third class of hypotheses of interest are joint null hypotheses of no structural change (pure or partial) plus certain nonlinear restrictions. In this case, the null hypothesis is $H_0 : \theta_1 = \theta_2$ and $h^*(\theta_1) = \underset{\sim}{0}$ when $\theta = (\theta_1', \theta_2')'$ or $H_0 : \theta_1 = \theta_2$ and $h^*(\theta_1, \theta_3) = \underset{\sim}{0}$ when $\theta = (\theta_1', \theta_2', \theta_3')'$. The present framework also includes tests of nonlinear restrictions that do not involve testing for structural change. Results for such hypotheses, however, already are available in the literature--see Gallant (1987, Ch. 7) and Gallant and White (1987, Ch. 7).

The function $h(\cdot)$ defining the restrictions is assumed to satisfy:

ASSUMPTION 4: (a) *$h(\theta)$ is continuously differentiable in a neighborhood of θ_0 and $H = \frac{\partial}{\partial\theta'}h(\theta_0)$ has full rank r ($\le p$).*
(b) *V is nonsingular (where V is as in assumption 2).*

The Wald statistic is defined as

$$W_T = Th(\hat{\theta})'(\hat{H}\hat{V}\hat{H}')^{-}h(\hat{\theta}) \ , \tag{2.4}$$

where $\hat{H} = \frac{\partial}{\partial\theta'}h(\hat{\theta})$. Since $\hat{H}\hat{V}\hat{H}' \xrightarrow{p} HVH'$ as $T \to \infty$ and HVH' is nonsingular under assumption 4, the g-inverse $(\cdot)^{-}$ equals the usual inverse $(\cdot)^{-1}$ with probability that goes to one as $T \to \infty$.[6]

In the case of testing for pure structural change, W_T is given by

$$W_T = T(\hat{\theta}_1 - \hat{\theta}_2)'(\hat{V}_1/\pi_{1T} + \hat{V}_2/\pi_{2T})^{-}(\hat{\theta}_1 - \hat{\theta}_2) \ , \tag{2.5}$$

where $\hat{V}_1$ and $\hat{V}_2$ are the estimators of the asymptotic covariance matrices of $\hat{\theta}_1$ and $\hat{\theta}_2$, which are analogous to the estimator $\hat{V}$ of V and which use the

observations indexed by $t = -T_1, \ldots, -1$ and $t = 1, \ldots, T_2$, respectively. This formula holds in the standard case where $\hat{D}$ is block diagonal with two blocks (for some ordering of its rows and columns) and $m_t(\hat{\theta}, \hat{\tau})$ has elements corresponding to the first block of $\hat{D}$ that are non-zero only if $t < 0$ and other elements that are non-zero only if $t > 0$.

The LM and LR statistics defined below make use of a restricted estimator of θ_0:

DEFINITION: *A sequence of restricted extremum estimators $\{\tilde{\theta}\} = \{\tilde{\theta} : T = 1, 2, \ldots\}$ is any sequence of rv's such that*

$$d(\bar{m}_T(\tilde{\theta}), \hat{\tau}) = \inf\{d(\bar{m}_T(\theta), \hat{\tau}) : \theta \in \Theta, \ h(\theta) = \underset{\sim}{0}\} \tag{2.6}$$

with probability that goes to one as $T \to \infty$.

Suppose the null hypothesis is true and $h(\cdot)$ is continuous on Θ. If assumption 1 holds for the parameter space Θ it also holds for the parameter space $\Theta_0 = \{\theta \in \Theta : h(\theta) = \underset{\sim}{0}\}$, since Θ_0 is compact and $\theta_0 \in \Theta_0$. Thus, assumption 1, Theorem 1, and continuity of $h(\cdot)$ over Θ imply that $\tilde{\theta} \xrightarrow{p} \theta_0$ as $T \to \infty$ under the null hypothesis. In consequence, the following assumption is straightforward to verify:

ASSUMPTION 5: *$\tilde{\theta} \xrightarrow{p} \theta_0$ as $T \to \infty$ under the null hypothesis.*

The LM statistic uses an estimator of V that is constructed with the restricted estimator $\tilde{\theta}$ in place of $\hat{\theta}$. Let $\tilde{M} = \frac{1}{T}\Sigma_{-T_1}^{T_2}\frac{\partial}{\partial\theta'}m_t(\tilde{\theta}, \hat{\tau})$, $\tilde{D} = \frac{\partial^2}{\partial m \partial m'}d(\bar{m}_T(\tilde{\theta}, \hat{\tau}), \hat{\tau})$, $\tilde{J} = \tilde{M}'\tilde{D}\tilde{M}$, $\hat{S} = \hat{S}(\tilde{\theta})$, and $\tilde{H} = \frac{\partial}{\partial\theta'}h(\tilde{\theta})$ (where $m_t(\cdot,\cdot)$ and $d(\cdot,\cdot)$ are as in assumption 2). Note that the estimator of the nuisance parameter τ_0 still is denoted $\hat{\tau}$, even though it may be a restricted estimator of τ_0. The same is true of the estimator $\hat{S}$ of S. With this notation, we do not need to adjust assumptions 2(b) or 3 when a restricted estimator of τ_0 is used. Let $\tilde{I} = \tilde{M}'\hat{D}\tilde{S}\tilde{D}\tilde{M}$ and $\tilde{V} = \tilde{J}^{-}\tilde{I}\tilde{J}^{-}$. As above with $\hat{V}$, the estimator $\tilde{V}$ can be simplified when V simplifies, as often occurs in applications of interest.

The LM statistic is defined as

$$LM_T = T\frac{\partial}{\partial\theta'}d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau})\tilde{J}^{-}\tilde{H}'(\tilde{H}\tilde{V}\tilde{H}')^{-}\tilde{H}\tilde{J}^{-}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau}) \tag{2.7}$$

(where $m_t(\cdot,\cdot)$ and $d(\cdot,\cdot)$ are as in assumption 2). As shown below, this statistic often simplifies considerably.

The LR statistic (defined below) has the desired asymptotic chi-square distribution under the null in two particular contexts contained within the general framework considered thus far. Outside of these contexts, the LR statistic generally is not asymptotically chi-square under the null. The first context is defined by the following assumption.

ASSUMPTION 6a: *Under the null hypothesis*, $I = bJ$ *for some scalar constant* $b \neq 0$ *and* $\hat{b} \xrightarrow{p} b$ or $\tilde{b} \xrightarrow{p} b$ *as* $T \to \infty$ *for some sequence of non-zero* rv's $\{\hat{b}\}$ *or* $\{\tilde{b}\}$ *(where* $m_t(\cdot,\cdot)$ *and* $d(\cdot,\cdot)$ *are as in assumption 2).*

Assumption 6a is satisfied by 2SLS and 3SLS estimators of nonlinear simultaneous equations models under certain assumptions regarding the heterogeneity and temporal dependence of the equation errors--see Section 4 below.

The second context is defined by the following assumption.

ASSUMPTION 6b: *Let* $m_t(\cdot,\cdot)$ *and* $d(\cdot,\cdot)$ *be as in assumption* 2.

(i) $d(m,\tau) = m'm/2$. *There exist functions* $\rho_t(W_t,\ \theta,\ \tau)$ *such that* $m_t(W_t,\ \theta,\ \tau) = \frac{\partial}{\partial\theta}\rho_t(W_t,\ \theta,\ \tau)$, $\forall t$. *With probability that goes to one as* $T \to \infty$, $\hat{\theta}$ *solves* $\bar{\rho}_T(\hat{\theta},\hat{\tau}) = \inf\{\bar{\rho}_T(\theta,\hat{\tau}) : \theta \in \Theta\}$ *and* $\tilde{\theta}$ *solves* $\bar{\rho}_T(\tilde{\theta},\hat{\tau}) = \inf\{\bar{\rho}_T(\theta,\hat{\tau}) : \theta \in \Theta,\ h(\theta) = \underset{\sim}{0}\}$, *where* $\bar{\rho}_T(\theta,\hat{\tau}) = \frac{1}{T}\sum_{T_1}^{T_2}\rho_t(W_t,\ \theta,\ \hat{\tau})$.

(ii) *Under the null hypothesis*, $S = cM$ *for some scalar* $c \neq 0$ *and* $\hat{c} \xrightarrow{p} c$ *or* $\tilde{c} \xrightarrow{p} c$ *as* $T \to \infty$ *for some sequence of non-zero* rv's $\{\hat{c}\}$ *or* $\{\tilde{c}\}$.

Assumption 6b is satisfied by ML estimators for general parametric models. Assumption 6b(i) is satisfied by the LS estimator and many M-estimators for the nonlinear regression model. Assumption 6b(ii) is satisfied with these estimators only under special conditions on the heterogeneity and temporal dependence of the errors--see Section 3 below.

Note that assumption 6b(i) is compatible with the definitions of $\hat{\theta}$ and $\tilde{\theta}$, because an estimator $\hat{\theta}$ $(\tilde{\theta})$ that minimizes $\bar{\rho}_T(\theta,\hat{\tau})$ (subject to $h(\theta) = \underset{\sim}{0}$) is in the interior of Θ with probability that goes to one as $T \to \infty$ under assumption 2, and hence, also minimizes $d(\bar{m}_T(\theta),\ \hat{\tau})$ (subject to $h(\theta) = \underset{\sim}{0}$) with probability that goes to one as $T \to \infty$.

The LR statistic is defined as

$$LR_T = \begin{cases} 2T(d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau}) - d(\bar{m}_T(\hat{\theta}),\ \hat{\tau}))/\hat{b} & \text{when 6a holds} \\ 2T(\bar{\rho}_T(\tilde{\theta},\hat{\tau}) - \bar{\rho}_T(\hat{\theta},\hat{\tau}))/\hat{c} & \text{when 6b holds.} \end{cases} \tag{2.8}$$

where $m_t(\cdot,\cdot)$ and $d(\cdot,\cdot)$ are as in assumption 2.[7] The nuisance parameter estimator $\hat{\tau}$ may be a restricted or an unrestricted estimator of τ_0. It must be the same in both criterion functions used to calculate LR_T, however, and it must be such that both $\hat{\theta}$ and $\tilde{\theta}$ are consistent under the null hypothesis. Otherwise, the LR statistic generally does not have the desired asymptotic distribution. That is, for use of the LR statistic, $\hat{\theta}$ and $\tilde{\theta}$ must be rv's that minimize the same criterion function subject to no restrictions and to the restrictions $h(\theta) = \underset{\sim}{0}$, respectively.

THEOREM 4: *Suppose assumptions 2-4 hold under the null hypothesis*, $h(\theta_0) = \underset{\sim}{0}$, *and the null hypothesis is true. Then the following results hold:*

(a) $W_T \xrightarrow{p} \chi_r^2$ *as* $T \to \infty$, *where* r *is the number of restrictions,*

(b) $LM_T \xrightarrow{p} \chi_r^2$ *as* $T \to \infty$ *provided assumption* 5 *also holds, and*

(c) $LR_T \xrightarrow{p} \chi_r^2$ *as* $T \to \infty$ *provided assumption* 5 *holds and either* 6a *or* 6b *holds in place of assumption* 3, *where* χ_r^2 *denotes the chi-square distribution with* r *degrees of freedom.*

Comments: 1. When assumption 6a holds, as occurs with 2SLS and 3SLS estimators in nonlinear simultaneous equations models with independent identically distributed (iid) errors (see Section 4 below), then we usually have $\hat{I} = \hat{b}\hat{J}$ for some scalar rv $\hat{b} \neq 0$. In the latter case, $\hat{V}$ and W_T simplify. We get $\hat{V} = \hat{b}\hat{J}^-$ and $W_T = Th(\hat{\theta})'(\hat{H}\hat{J}^-\hat{H}')^-h(\hat{\theta})/\hat{b}$.

Similarly, if $\tilde{I} = \tilde{b}\tilde{J}$ for some scalar rv $\tilde{b} \neq 0$, then $\tilde{V}$ and LM_T simplify. We get $\tilde{V} = \tilde{b}\tilde{J}^-$ and $LM_T \doteq T\frac{\partial}{\partial\theta'}d(\bar{m}_T(\tilde{\theta}), \hat{\tau})\tilde{J}^-\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}), \hat{\tau})/\tilde{b}$ (where $\doteq$ denotes equality that holds with probability that goes to one as $T \to \infty$), since $\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}), \hat{\tau}) \doteq \tilde{H}'\tilde{\lambda}$ for some vector $\tilde{\lambda}$ of Lagrange multipliers.

2. When assumption 6b(i) holds, both W_T and LM_T simplify. In this case, $D = I_p$, $M = \lim_{T\to\infty} \frac{1}{T}\Sigma_{-T_1}^{T_2} E\frac{\partial^2}{\partial\theta\partial\theta'}\rho_t(W_t, \theta, \tau)$, $J = M^2$, $I = MSM$, $\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}), \hat{\tau}) = \tilde{M}\bar{m}_T(\tilde{\theta})$, and by 2(g), M is nonsingular. We get $W_T \doteq Th(\hat{\theta})'(\hat{H}\hat{M}^-\hat{S}\hat{M}^-\hat{H}')^-h(\hat{\theta})$ and $LM_T \doteq T\bar{m}_T(\tilde{\theta})'\tilde{M}^-\tilde{H}'(\tilde{H}\tilde{M}^-\hat{S}\tilde{M}^-\tilde{H}')^-\tilde{H}\tilde{M}^-\bar{m}_T(\tilde{\theta})$.

If, in addition, $\hat{S} = \hat{c}\hat{M}$ or $\hat{S} = \tilde{c}\tilde{M}$ for some scalar rv's $\hat{c}$, $\tilde{c} \neq 0$ (as usually occurs when assumption 6b(ii) holds), then W_T and LM_T simplify to $W_T \doteq Th(\hat{\theta})'(\hat{H}\hat{M}^-\hat{H}')^-h(\hat{\theta})/\hat{c}$ and $LM_T \doteq T\bar{m}_T(\tilde{\theta})'\tilde{M}^-\bar{m}_T(\tilde{\theta})/\tilde{c}$, respectively. The latter holds because $\tilde{M}\bar{m}_T(\tilde{\theta}) \doteq \tilde{H}'\tilde{\eta}$ for some vector of Lagrange multipliers $\tilde{\eta}$ under assumption 6b(i).

3. One would expect the small sample properties of W_T, LM_T, and LR_T to be improved by replacing the devisors T, T_1, and T_2 that arise in various sample averages by their counterparts with the estimated number of parameters subtracted off. The relevant number of estimated parameters to subtract off may or may not include the elements of $\hat{\tau}$ and may or may not include all of the elements of $\hat{\theta}$, depending upon the context.

Next, we present asymptotic local power results for the three tests considered above. These results can be used to approximate the power functions of the tests. We assume:

ASSUMPTION 7: *There exists a sequence of distributions* (P_T) *on* (Ω, F) *such that assumption* 2 *holds under* (P_T) *with* θ_0 *replaced by* $\theta_T = \theta_0 + \eta/\sqrt{T}$ *in parts* 2(b)ii *and* 2(c) *for some* $\eta \in R^p$.

The distributions (P_T) usually are determined quite easily in applications. For example, in the nonlinear regression model, the sequence of models is $Y_{Tt} = f_t(\theta_T) + U_t$, $t = -T_1, \ldots, T_2$, for $T = 1, 2, \ldots$, and P_T is just the distribution $((Y_{Tt}, X_t, U_t) : t = \ldots, -1, 1, \ldots)$ for $T = 1, 2, \ldots$.

Verification that assumption 2(a) holds under (P_T) can be made by showing that assumption 1 holds under (P_T).

We define the following analogues of assumptions 3, 5, 6a, and 6b:

ASSUMPTION 8: *Assumption* 3 *holds under* (P_T).

ASSUMPTION 9: $\tilde{\theta} \xrightarrow{p} \theta_0$ *under* (P_T) *as* $T \to \infty$.

ASSUMPTION 10a: *Assumption* 6a *holds and* $\hat{b} \xrightarrow{p} b$ *or* $\tilde{b} \xrightarrow{p} b$ *under* (P_T) *as* $T \to \infty$.

ASSUMPTION 10b: *Assumption* 6b *holds and* $\hat{c} \xrightarrow{p} b$ *or* $\tilde{c} \xrightarrow{p} b$ *under* (P_T) *as* $T \to \infty$.

Note that assumption 9 holds if Θ_0 is compact and assumption 1 holds under (P_T).

THEOREM 5: *Under assumptions* 4, 7, *and* 8,

(a) $W_T \xrightarrow{d} \chi_r^2(\delta^2)$, *where* $\delta^2 = \eta' H'(HVH')^{-1}H\eta$,

(b) $LM_T \xrightarrow{d} \chi_r^2(\delta^2)$ *provided assumption* 9 *also holds, and*

(c) $LR_T \xrightarrow{d} \chi_r^2(\delta^2)$ *provided assumption* 9 *holds and either* 10a *or* 10b *holds in place of assumption* 8, *where* $\chi_r^2(\delta^2)$ *denotes the noncentral chi-square distribution with noncentrality parameter* δ^2 *and* r *degrees of freedom.*

COMMENTS: 1. Since $\sqrt{T}h(\theta_T) \to H\eta$ as $T \to \infty$, power approximations can be based on a $\chi_r^2(\delta_T^2)$ distribution, where $\delta_T^2 = Th(\theta_T)'(HVH')^{-1}h(\theta_T)$. In particular, to approximate the power of a test against an alternative θ when the sample size is T, we set $\theta = \theta_T$ and take $\delta_T^2 = Th(\theta)'(HVH')^{-1}h(\theta)$.

2. Due to the local nature of the alternatives in Theorem 5, the approximations described in Comment 1 usually are more accurate for close alternatives to the null hypothesis than for distant alternatives.

3. NONLINEAR REGRESSION

Here we consider structural change in the nonlinear regression model

$$Y_t = f_t(X_t, \theta_0) + U_t, \quad t = -T_1, \ldots, -1, 1, \ldots, T_2, \tag{3.1}$$

where $Y_t \in R^1$ and $X_t \in R^K$ are observed, $U_t \in R^1$ is unobserved, $f_t(\cdot,\cdot) \in R^1$ is a known function, and $\theta_0 \in \Theta \subset R^p$ is unknown. The vector X_t may include lagged values of Y_t. For brevity, we only consider the LS estimator and tests based on it in this section. Using the results of Section 2, one can treat the more general class of M-estimators analogously.

3.1 *Least Squares Estimation*

A sequence of <u>least squares estimators</u> of θ_0 for T = 1, 2, ... is defined to be any sequence of rv's $\{\hat{\theta}\}$ such that

$$\frac{1}{T}\sum_{-T_1}^{T_2}(Y_t - f_t(\hat{\theta}))^2 = \inf_{\theta\in\Theta}\frac{1}{T}\sum_{-T_1}^{T_2}(Y_t - f_t(\theta))^2 \tag{3.2}$$

with probability that goes to one as $T \to \infty$.

The following assumption R1 guarantees the existence of a sequence of LS estimators $\{\hat{\theta}\}$. Also, it implies assumption 1 of Section 2 with $m_t(\theta,\tau) = (Y_t - f_t(\theta))^2 - U_t^2$ and $d(m,\tau) = m$ for $m \in R^1$ (see the Appendix). Hence, using Theorem 1, assumption R1 guarantees the consistency of any such sequence. We note that each variable and vector that appears in this assumption and the others below is assumed (implicitly) to be $\mathcal{F}\backslash$Borel-measurable.

ASSUMPTION R1: (a) Θ *is a compact subset of* R^p.

(b) $EU_t f_t(X_t, \theta) = 0$, $\forall\theta \in \Theta$, $\forall t$.

(c) $\lim_{T\to\infty}\frac{1}{T}\sum_{-T_1}^{T_2}E(f_t(\theta) - f_t(\theta_0))^2$ *exists uniformly for* $\theta \in \Theta$ *and is positive unless* $\theta = \theta_0$.

(d) $\{(X_t, U_t)\}$ *is strong mixing with strong mixing numbers* $\{\alpha(s)\}$ *that satisfy* $\alpha(s) = o(s^{-\alpha/(\alpha-1)})$ *for some* $\alpha > 1$.[8]

(e) $\sup_t E\sup_{\theta\in\Theta}[|f_t(\theta) - f_t(\theta_0)|^{2\xi} + |U_t(f_t(\theta) - f_t(\theta_0))|^{\xi}] < \infty$ *for some* $\xi > \alpha$.

(f) $f_t(\theta)$ *is defined and differentiable in* θ, $\forall t$, *for all realizations of* X_t, $\forall\theta \in \Theta^*$, *where* Θ^* *is some convex or open set that contains* Θ *and*

$$\lim_{T\to\infty}\frac{1}{T}\sum_{-T_1}^{T_2}E\sup_{\theta\in\Theta^*}\left(\left\|(f_t(\theta) - f_t(\theta_0))\frac{\partial}{\partial\theta}f_t(\theta)\right\| + \left\|U_t\frac{\partial}{\partial\theta}f_t(\theta)\right\|\right) < \infty.$$

Assumptions R1(a), (b), and (c) are standard compactness, orthogonality, and identification assumptions, respectively. The strong mixing assumption

R1(d) is used to ensure that a law of large numbers (LLN) holds for certain rv's. This condition is quite convenient and fairly general, but is not all encompassing (see Andrews (1984, 1985)). For cases where this assumption fails, one can substitute an alternative condition of asymptotic weak dependence (see references in Section 2) and use the results of Section 2 to establish consistency and asymptotic normality.

Note that assumption R1(e) does not require the errors to have finite variances. Assumption R1(f) is used to convert an LLN into a uniform LLN. It could be replaced by a weaker continuity or Lipschitz condition (see Andrews (1987b, Cor. 2 and 3)) and assumption 1 still would follow. This assumption is convenient, however, since differentiability of $f_t(\theta)$ is used below for asymptotic normality anyway.

For the case of a model with no structural change, assumption R1 is quite similar to the consistency assumptions of White and Domowitz (1984).[9]

Next, we introduce an assumption R2 such that assumptions R1 and R2 imply assumption 2 of Section 2 with $m_t(\theta,\tau) = (Y_t - f_t(\theta))\frac{\partial}{\partial\theta}f_t(\theta)$ and $d(m,\tau) = m'm/2$ (see the Appendix). Hence, by Theorem 2, under assumptions R1 and R2, $\sqrt{T}(\hat\theta - \theta_0)$ has an asymptotic normal distribution as $T\to\infty$ with mean vector $\underset{\sim}{0}$ and covariance matrix $V = M^{-1}SM^{-1}$, where

$M = M(\theta_0) = -\lim_{T\to\infty} \frac{1}{T}\sum_{T_1}^{T_2} E\frac{\partial}{\partial\theta}f_t(\theta_0)\frac{\partial}{\partial\theta'}f_t(\theta_0)$ and S is as in R2(c).

ASSUMPTION R2: (a) Θ *contains a convex compact neighborhood* Θ_c *of* θ_0.

(b) $m(\theta) = \lim_{T\to\infty}\frac{1}{T}\sum_{T_1}^{T_2}E(Y_t - f_t(\theta))\frac{\partial}{\partial\theta}f_t(\theta)$ *and*

$M(\theta) = \lim_{T\to\infty}\frac{1}{T}\sum_{T_1}^{T_2}E\frac{\partial}{\partial\theta'}[(Y_t - f_t(\theta))\frac{\partial}{\partial\theta}f_t(\theta)]$ *exist uniformly for* $\theta\in\Theta_c$ *and are continuous on* Θ_c *and* $M(\theta_0)$ *is nonsingular.*

(c) $S = \lim_{T\to\infty}\mathrm{Var}\left(\frac{1}{\sqrt{T}}\sum_{T_1}^{T_2}U_t\frac{\partial}{\partial\theta}f_t(\theta_0)\right)$ *exists.*

(d) $f_t(\theta)$ *is three times continuously differentiable with respect to* θ *on* Θ_c *for all realizations of* X_t, $\forall t$, *and*

$$\sup_t E\sup_{\theta\in\Theta_c}\left[\|m_t(\theta_0)\|^{2\xi} + \|m_t(\theta)\|^{\xi} + \left\|\frac{\partial}{\partial\theta}m_t(\theta)\right\|^{\xi} + \left\|\frac{\partial^2}{\partial\theta_a\partial\theta}m_t(\theta)\right\|^{\xi}\right] < \infty,\ \textit{for}$$

some $\xi > \alpha$, $\forall a = 1, \ldots, p$, *where* $m_t(\theta) = (Y_t - f_t(\theta))\frac{\partial}{\partial\theta}f_t(\theta)$.[10]

We now define a consistent estimator of the covariance matrix V. Let

$$\hat V = \hat M^{-}\hat S\hat M^{-}, \tag{3.3}$$

where $\hat M = \hat M(\hat\theta)$, $\hat M(\theta) = \frac{1}{T}\sum_{T_1}^{T_2}\frac{\partial}{\partial\theta}f_t(\theta)\frac{\partial}{\partial\theta'}f_t(\theta)$, and $\hat S = \hat S(\hat\theta)$ for $\hat S(\theta)$ as defined in equation (2.3) with $m_t(\theta,\tau) = (Y_t - f_t(\theta))\frac{\partial}{\partial\theta}f_t(\theta)$ and $w(\cdot)$ corresponding to the Parzen or Bartlett weights. (Gallant (1987, p. 533) recommends the Parzen weights.)

For consistency of $\hat S$, we use the additional assumption:

ASSUMPTION R3: $\sup_t E\|U_t\frac{\partial}{\partial\theta}f_t(\theta_0)\|^{4\xi} < \infty$ *for some* $\xi > \alpha$.

The estimator $\hat S$ can be replaced by a simpler estimator in certain cases. If $EU_tU_s\frac{\partial}{\partial\theta}f_t(\theta_0)\frac{\partial}{\partial\theta'}f_s(\theta_0) = \underset{\sim}{0}$, $\forall t\neq s$, one can take $\ell(T_1) = \ell(T_2) = 0$ in the definition of $\hat S$. This yields

$$\hat S = \frac{1}{T}\sum_{T_1}^{T_2}(Y_t - f_t(\hat\theta))^2\frac{\partial}{\partial\theta}f_t(\hat\theta)\frac{\partial}{\partial\theta'}f_t(\hat\theta)\ . \tag{3.4}$$

If, in addition, $E(U_t^2|X_t) = \sigma^2$ a.s., $\forall t$, then $EU_t^2\frac{\partial}{\partial\theta}f_t(\theta_0)\frac{\partial}{\partial\theta'}f_t(\theta_0) = \sigma^2E\frac{\partial}{\partial\theta}f_t(\theta_0)\frac{\partial}{\partial\theta'}f_t(\theta_0)$, $\forall t$, and we can take

$$\hat S = \hat\sigma^2\hat M\ , \quad \text{where} \quad \hat\sigma^2 = \frac{1}{T}\sum_{T_1}^{T_2}(Y_t - f_t(\hat\theta))^2\ . \tag{3.5}$$

The following consistency results for $\hat{S}$, $\hat{M}$, and $\hat{V}$ make use of Theorem 3 and a result of Newey and West (1987):

THEOREM 6: (a) *Under assumptions* R1-R3, $\hat{S} \xrightarrow{p} S$, $\hat{M} \xrightarrow{p} M$, *and* $\hat{V} \xrightarrow{p} V$ *as* $T \to \infty$ *for* $\hat{S}$ *as defined in equation* (3.3). (b) *Under assumptions* R1 *and* R2, $\hat{S} \xrightarrow{p} S$, $\hat{M} \xrightarrow{p} M$, *and* $\hat{V} \xrightarrow{p} V$ *as* $T \to \infty$ *for* $\hat{S}$ *as defined in equation* (3.4) *or* (3.5), *provided the additional conditions outlined above* (3.4) *or* (3.5) *are satisfied, respectively.*

3.2. *Tests of Structural Change*

We now consider tests of $H_0 : h(\theta) = \underset{\sim}{0}$, where $h(\cdot)$ satisfies assumption 4 of Section 2. The LM and LR test statistics make use of a restricted LS estimator $\tilde{\theta}$. By definition, a sequence of <u>restricted LS estimators</u> of θ_0 is any sequence of rv's $\{\tilde{\theta}\} = \{\tilde{\theta} : T = 1, 2, \ldots\}$ such that equation (3.2) holds (with probability that goes to one as $T \to \infty$) with $\tilde{\theta}$ in place of $\hat{\theta}$, where the infimum is taken over $\Theta_0 = \{\theta \in \Theta : h(\theta) = \underset{\sim}{0}\}$. We assume:

ASSUMPTION R5: Θ_0 *is compact.*

Assumption R5 holds if $h(\cdot)$ is continuous on Θ, as is usually the case.

Assumptions R1 and R5 guarantee that a sequence of restricted LS estimators exists and that any such sequence is consistent for θ_0 when θ_0 satisfies the null hypothesis $h(\theta) = \underset{\sim}{0}$ (see the Appendix).

When the restrictions $h(\theta) = \underset{\sim}{0}$ correspond to a test of pure or partial structural change, $\tilde{\theta}$ generally is easy to compute. It equals $(\tilde{\theta}_1', \tilde{\theta}_1')'$ or $(\tilde{\theta}_1', \tilde{\theta}_1', \tilde{\theta}_3')'$, where $\tilde{\theta}_1$ or $(\tilde{\theta}_1', \tilde{\theta}_3')'$ are just the estimators obtained from the whole sample under the assumption of no structural change.

The LM test statistic defined in equation (2.7) uses a consistent estimator of the covariance matrix V that is based on the restricted estimator $\tilde{\theta}$. In the present context, we take

$$\tilde{V} = \tilde{M}^{-}\tilde{S}\,\tilde{M}^{-}, \text{ where } \tilde{M} = \hat{M}(\tilde{\theta}) \text{ and } \tilde{S} = \hat{S}(\tilde{\theta}) \tag{3.6}$$

and $\hat{M}(\theta)$ and $\hat{S}(\theta)$ are given in equation (3.3). As in equations (3.4) and (3.5), $\tilde{S}$ can be replaced by the simpler estimator

$$\tilde{S} = \frac{1}{T}\Sigma_{T_1}^{T_2}(Y_t - f_t(\tilde{\theta}))^2 \frac{\partial}{\partial\theta}f_t(\tilde{\theta})\frac{\partial}{\partial\theta'}f_t(\tilde{\theta}) \quad \text{or} \quad \tilde{S} = \tilde{\sigma}^2\tilde{M}\,, \tag{3.7}$$

where $\tilde{\sigma}^2 = \frac{1}{T}\Sigma_{T_1}^{T_2}(Y_t - f_t(\tilde{\theta}))^2$, when the conditions outlined above (3.4) or or (3.5) hold, respectively. By the proof of Theorem 6, $\tilde{S}$, $\tilde{M}$, and $\tilde{V}$ are consistent for S, M, and V, respectively, when the null hypothesis is true under the conditions of Theorem 6 and assumption R5.

The following assumption R6b implies assumption 6b of Section 2, which is used to ensure that the LR statistic has an asymptotic chi-square null distribution. This assumption is not needed for the W and LM test statistics.

ASSUMPTION R6b: *Under the null hypothesis,*

$$EU_tU_s\,\frac{\partial}{\partial\theta}f_t(\theta_0)\frac{\partial}{\partial\theta'}f_s(\theta_0) = \begin{cases} \sigma^2 E\frac{\partial}{\partial\theta}f_t(\theta_0)\frac{\partial}{\partial\theta'}f_t(\theta_0) & \textit{when } t = s \\ \underset{\sim}{0} & \textit{when } t \neq s \end{cases}$$

for all t, *where* $\sigma^2 = EU_t^2$ *for all* t.

Assumption R6b holds if U_t and U_s are independent conditional on X_t and X_s a.s., $\forall t \neq s$, and U_t has homoskedastic variance σ^2 conditional on X_t a.s. $\forall t$. These conditions restrict the temporal dependence and heterogeneity of the errors considerably. It often is possible, however, to transform a model with temporally dependent or heteroskedastic errors into a model with iid errors. A prime example is when the original model has stationary, autoregressive errors (e.g., see Fair (1970) and Gallant and Goebel (1976)).

We now have assumptions R1-R3, 4, R5, and R6b for LS estimation of the nonlinear regression model that imply assumptions 1-5 and 6b of Section 2. In consequence, Theorem 4 holds and the W, LM, and LR statistics defined in equations (2.4), (2.7), and (2.8) (with $m_t(\theta,\tau) = (Y_t - f_t(\theta))\frac{\partial}{\partial\theta}f_t(\theta)$) are asymptotically chi-square with r degrees of freedom under the null hypothesis.

Next, we introduce assumptions that guarantee that the W, LM, and LR test statistics have noncentral chi-square distributions under sequences of local alternatives (using Theorem 5 of Section 2).

ASSUMPTION R7: *Given* $\eta \in R^p$, *let* $\theta_T = \theta_0 + \eta/\sqrt{T}$ *and* $Y_{Tt} = f_t(\theta_T) + U_t$. *Let* P_T *denote the distribution of* $\{(Y_{Tt}, X_t, U_t)\}$ *for* T = 1, 2, ... *Suppose assumption* R1 *holds with* θ_0 *replaced by* θ_T *in* R1(e), R1(f), *and the first time it appears in* R1(c). *Suppose assumption* R2 *holds with* Y_t *given by* Y_{Tt} *in* R2(b), *with* θ_0 *replaced by* θ_T *in* R2(c), *with* $\sup_t$ *replaced by* $\sup_{t\le T, T=1,2,\ldots}$ *in* R2(d), *and with* $m_t(\theta)$ *defined by* $(Y_{Tt} - f_t(\theta))\frac{\partial}{\partial\theta}f_t(\theta)$ *in* R2(d).

ASSUMPTION R8: $\sup_{t\le T, T=1,2,\ldots} E\left\|U_t \frac{\partial}{\partial\theta}f_t(\theta_T)\right\|^{4\xi} < \infty$ *for some* $\xi > \alpha$.

ASSUMPTION 10b: *Assumption* R6b *holds with* θ_0 *replaced by* θ_T.

It is straightforward to show that assumption R7 implies assumption 7 of Section 2. Assumption R7 is not much stronger than R2 because $f_t(\theta_T) = f_t(\theta_0) + \frac{\partial}{\partial\theta'}f_t(\theta^*)\eta/\sqrt{T}$ and so the replacement of $f_t(\theta_0)$ by $f_t(\theta_T)$ only causes a change of order $O_p(1/\sqrt{T})$. If desired, assumption R2 plus a

moment assumption on $\frac{\partial}{\partial\theta}f_t(\theta)$ and $\frac{\partial^2}{\partial\theta\partial\theta'}f_t(\theta)$ can be used in place of assumption R7.

Assumptions R7 and R8 imply assumption 8 of Section 2. This follows by the proof of Newey and West (1987) applied to a triangular array of strong mixing rv's rather than to a sequence of strong mixing rv's. Also, assumptions R5 and R7 imply assumption 9 of Section 2 by Theorem 1 with the parameter Θ replaced by Θ_0. Assumptions R7 and R10b imply assumption 10b of Section 2 with $\hat{c} = \frac{1}{T}\sum_1^T (Y_{Tt} - f_t(\hat{\theta}))^2$ or $\tilde{c} = \frac{1}{T}\sum_1^T (Y_{Tt} - f_t(\tilde{\theta}))^2$.

In sum, assumptions 4, 7, 8, 9, and 10b of Section 2 are implied by assumptions 4, R7, R7 and R8, R5 and R7, and R7 and R10b, respectively. In consequence, Theorem 5 holds and the W, LM, and LR statistics have non-central chi-square distributions under sequences of local alternatives.

We now provide some simplified formulae for the W, LM, and LR statistics in the nonlinear regression context and in particular for the special case of testing for pure structural change. The general formula for the W statistic is given in (2.4) and its covariance matrix estimator $\hat{V}$ is given for the nonlinear regression context by equations (3.3)-(3.5). For the case of pure structural change, the W statistic is given in (2.5).

When assumption R6b holds, the W statistic simplifies considerably by taking $\hat{S}$ as in (3.5). In this case, W_T is given by the formula in Comment 2 to Theorem 4 with $\hat{c} = \hat{\sigma}^2$. For example, when testing for pure structural change under assumption R6b, W_T becomes

$$W_T \doteq T(\hat{\theta}_1 - \hat{\theta}_2)'\left[\hat{M}_1^-/\pi_{1T} + \hat{M}_2^-/\pi_{2T}\right]^-(\hat{\theta}_1 - \hat{\theta}_2) \ . \tag{3.8}$$

For general null hypotheses, the LM statistic is given by

$$LM_T = T\bar{m}_T(\tilde{\theta})'\tilde{M}^-\tilde{H}'(\tilde{H}\tilde{V}\tilde{H}')^-\tilde{H}\tilde{M}^-\bar{m}_T(\tilde{\theta}) \ , \tag{3.9}$$

where $\bar{m}_T(\tilde{\theta}) = \frac{1}{T}\Sigma_{-T_1}^{T_2}(Y_t - f_t(\tilde{\theta}))\frac{\partial}{\partial\theta}f_t(\tilde{\theta})$. Note that LM_T is a quadratic form in the first order conditions for the unrestricted problem of minimizing the sum of squared residuals, evaluated at the restricted estimator $\tilde{\theta}$.

When testing for pure structural change, LM_T becomes

$$LM_T = T\left[\tilde{M}_1^-\bar{m}_{1T}(\tilde{\theta}) - \tilde{M}_2^-\bar{m}_{2T}(\tilde{\theta})\right]'\left[\tilde{V}_1/\pi_{1T} + \tilde{V}_2/\pi_{2T}\right]^-\left[\tilde{M}_1^-\bar{m}_{1T}(\tilde{\theta}) - \tilde{M}_2^-\bar{m}_{2T}(\tilde{\theta})\right] \ , \tag{3.10}$$

where $\bar{m}_{1T}(\tilde{\theta}) = \frac{1}{T_1}\Sigma_{-T_1}^{-1}(Y_t - f_t(\tilde{\theta}))\frac{\partial}{\partial\theta_1}f_t(\tilde{\theta})$, $\tilde{V}_1 = \tilde{M}_1^-\tilde{S}_1\tilde{M}_1^-$, $\tilde{M}_1$ and $\tilde{S}_1$ are defined analogously to $\tilde{M}$ and $\tilde{S}$ using only the observations indexed by $t = -T_1, \ldots, -1$, and $\bar{m}_{2T}(\tilde{\theta})$, $\tilde{V}_2$, $\tilde{M}_2$, and $\tilde{S}_2$ are defined analogously.

When assumption R6b holds, LM_T simplifies by taking $\tilde{S} = \tilde{\sigma}^2\tilde{M}$:

$$LM_T \doteq T\bar{m}_T(\tilde{\theta})'\tilde{M}^-\bar{m}_T(\tilde{\theta})/\tilde{\sigma}^2 \ . \tag{3.11}$$

For example, when testing for pure structural change under assumption R6b,

$$LM_T \doteq T_1\bar{m}_{1T}(\tilde{\theta})'\tilde{M}_1^-\bar{m}_{1T}(\tilde{\theta}) + T_2\bar{m}_{2T}(\tilde{\theta})'\tilde{M}_2^-\bar{m}_{2T}(\tilde{\theta}) \ . \tag{3.12}$$

In the nonlinear regression context, the LR statistic is defined as

$$LR_T = \left[\Sigma_{-T_1}^{T_2}(Y_t - f_t(\tilde{\theta}))^2 - \Sigma_{-T_1}^{T_2}(Y_t - f_t(\hat{\theta}))^2\right]/\hat{\sigma}^2 \ . \tag{3.13}$$

Recall from Section 2 that LR_T has the desired asymptotic null distribution only if assumption R6b holds. In the case of testing for pure structural change, the first term above equals the sum of squared residuals (SSR) from the regression of Y_t on $f_t(\cdot)$ with $t = -T_1, \ldots, T_2$ (and no structural change), while the second term equals the SSR from the regression of Y_t on

$f_t(\cdot)$ with $t = -T_1, \ldots, -1$ plus the SSR from the same regression with $t = 1, \ldots, T_2$.

Computationally, the relative attributes of the W, LM, and LR statistics can be summarized as follows: The Wald statistic W only requires calculation of the unrestricted estimator $\hat{\theta}$ and not the restricted estimator $\tilde{\theta}$. Once one has calculated $\hat{\theta}$ and a consistent estimator of its covariance matrix, the Wald statistic can be computed by simple matrix manipulations.

The LM statistic only requires calculation of the restricted estimator $\tilde{\theta}$ and not $\hat{\theta}$. Thus, if the latter is difficult to compute, which may occur in some models of partial structural change, the LM statistic is the easiest of the three test statistics to compute.

The LR statistic requires computation of both $\hat{\theta}$ and $\tilde{\theta}$. Once these estimators have been computed, however, the LR statistic can be calculated directly from information provided by standard software packages.

The small sample properties of W_T, LM_T, and LR_T may be improved if the divisors T, T_1, and T_2 of the various sample averages that arise in the statistics' definitions are replaced by their counterparts with the estimated number of parameters subtracted off.

4. NONLINEAR SIMULTANEOUS EQUATIONS

In this section we consider structural change in the nonlinear simultaneous equations model

$$f_{it}(Y_t, X_t, \theta_0) = U_{it} \ , \quad i = 1, \ldots, n \ , \quad t = -T_1, \ldots, T_2 \ , \tag{4.1}$$

where $Y_t \in R^G$ and $X_t \in R^K$ are observed endogenous and predetermined variables, respectively, $U_{it} \in R^1$ is an unobserved error, $f_{it}(\cdot,\cdot,\cdot) \in R^1$ is a known function, $\theta_0 \in \Theta \subset R^p$ is an unknown parameter, and n (≥ 1) is the number of equations. As above, in the cases of pure and partial structural change the parameter vector θ_0 can be partitioned as $(\theta_1', \theta_2')'$ and $(\theta_1', \theta_2', \theta_3')'$, respectively.

4.1. *Three Stage Least Squares Estimation*

We consider a class of nonlinear three stage least squares (3SLS) estimators introduced by Amemiya (1977) and generalized to the structural change problem considered here. A special case of the 3SLS estimator is the two stage least squares (2SLS) estimator.

Let $f_{it}(\theta)$ abbreviate $f_{it}(Y_t, X_t, \theta)$ and take

$$f_1(\theta) = \left(f_{1,-T_1}(\theta), \ \ldots, \ f_{1,-1}(\theta), \ f_{2,-T_1}(\theta), \ \ldots, \ f_{n,-1}(\theta)\right)'_{nT_1\times 1} . \quad (4.2)$$

Let Z_{it} be a column v_i-vector of instrumental variables (IVs) for the i^{th} equation and the t^{th} time period. For i = 1, ..., n, let Z_1^i be a $T_1 \times v_i$ matrix whose rows are given by Z_{it}' for $t = -T_1, \ldots, -1$. Define

$$Z_1 = \operatorname{diag}\left\{Z_1^1, \ \ldots, \ Z_1^n\right\}_{nT_1\times v} , \quad \text{where} \quad v = \sum_{i=1}^{n} v_i . \quad (4.3)$$

Define $f_2(\theta)$ and Z_2 analogously with the time periods $t = -T_1, \ldots, -1$ replaced by $t = 1, \ldots, T_2$.

Let $\hat{\Omega}_1$ and $\hat{\Omega}_2$ denote n × n nuisance parameter estimators. Either $\hat{\Omega}_1$ and $\hat{\Omega}_2$ are estimators of $\Omega_1 = \lim_{T_1\to\infty} \frac{1}{T_1} \Sigma_{-T_1}^{-1} EU_tU_t'$ and $\hat{\Omega}_2 = \lim_{T_2\to\infty} \frac{1}{T_2} \Sigma_1^{T_2} EU_tU_t'$, respectively, where $U_t = (U_{1t}, \ldots, U_{nt})'$ or $\hat{\Omega}_1 = \hat{\Omega}_2$ and $\hat{\Omega}_1$ and $\hat{\Omega}_2$ are estimators of $\Omega_1 = \Omega_2 = \lim_{T\to\infty} \frac{1}{T} \Sigma_{-T_1}^{T_2} EU_tU_t'$. The former case corresponds to the common situation where one believes that structural change may affect both θ_0 and the distribution of U_t. The latter case corresponds to the less likely situation where one believes that structural change may affect θ_0 but not the distribution of U_t.[11]

Let $\hat{\Lambda}_j = \hat{\Omega}_j \otimes I_{T_j}$ and $\Lambda_j = \Omega_j \otimes I_{T_j}$ for j = 1, 2.

A sequence of <u>3SLS estimators</u> of θ_0 for T = 1, 2, ... is defined to be any sequence of rv's $\{\hat{\theta}\}$ such that $\hat{\theta}$ minimizes

$$\left(f_1(\theta)'\hat{\Lambda}_1^- Z_1 + f_2(\theta)'\hat{\Lambda}_2^- Z_2\right)\left(Z_1'\hat{\Lambda}_1^- Z_1 + Z_2'\hat{\Lambda}_2^- Z_2\right)^-\left(Z_1'\hat{\Lambda}_1^- f_1(\theta) + Z_2'\hat{\Lambda}_2^- f_2(\theta)\right) \quad (4.4)$$

over $\theta \in \Theta$ with probability that goes to one as $T \to \infty$.

In the special case where one takes $\hat{\Omega}_1 = \hat{\Omega}_2 = I_n$, the estimator $\hat{\theta}$ defined by equation (4.4) is the 2SLS estimator of θ_0. In this case, the objective function can be written as the sum of n terms, each involving a separate equation. If the parameter space Θ does not impose any cross equation restrictions, then the 2SLS estimators of the n sub-vectors of θ_0 can be estimated one at a time.

When only one equation is estimated (n = 1), equation (4.4) simplifies. In particular, in the case of pure structural change, it can be written as the sum of two terms, the first of which corresponds to the ordinary 2SLS estimator using the $t < 0$ data and the second to the 2SLS estimator using the $t > 0$ data. The scalars $\hat{\Omega}_1$ and $\hat{\Omega}_2$ become redundant in this case and need not be calculated.

The following assumption S1 guarantees the existence of a sequence of 3SLS estimators $\{\hat{\theta}\}$. Also, it implies assumption 1 of Section 2 with $W_t = (Y_t, X_t, Z_t)$, $m_t(\theta,\hat{\tau}) = Z_t'\hat{\Omega}_j^- f_t(\theta)$, where $Z_t = \operatorname{diag}(Z_{1t}', \ldots, Z_{nt}')_{n\times v}$, $f_t(\theta) = (f_{1t}(\theta), \ldots, f_{nt}(\theta))'_{n\times 1}$, j = 1 for $t < 0$, and j = 2 for $t > 0$, and $d(m,\hat{\tau}) = m'\hat{D}m/2$, where

$$\hat{D} = T\left(Z_1'\hat{\Lambda}_1^- Z_1 + Z_2'\hat{\Lambda}_2^- Z_2\right)^- = \left(\frac{1}{T}\ \Sigma_{T_1}^{T_2} Z_t'\hat{\Omega}_j^- Z_t\right)^-_{v\times v} \tag{4.5}$$

and $\hat{\tau}$ is a u-vector comprised of the non-redundant elements of $\hat{\Omega}_1$, $\hat{\Omega}_2$, and $\hat{D}$. Using Theorem 1, assumption S1 guarantees the consistency of every sequence of 3SLS estimators. We note that each variable and vector that appears in this assumption and the others below is assumed implicitly to be $\mathcal{F}\backslash$Borel-measurable.

ASSUMPTION S1: (a) *Θ is a compact subset of* R^p.

(b) $\hat{\Omega}_1 \xrightarrow{p} \Omega_1$ *and* $\hat{\Omega}_2 \xrightarrow{p} \Omega_2$ *as* $T \to \infty$ *for some* $n \times n$ *nonsingular matrices* Ω_1 *and* Ω_2.

(c) $\pi_1 = \lim_{T\to\infty} \pi_{1T}$, $\lim_{T_1\to\infty} \frac{1}{T_1}\ \Sigma_{T_1}^{-1} Ef_{it}(\theta)Z_{rt}$, *and* $\lim_{T_2\to\infty} \frac{1}{T_2}\ \Sigma_1^{T_2} Ef_{it}(\theta)Z_{rt}$ *exist uniformly for $\theta \in \Theta$ and are continuous in θ for all $\theta \in \Theta$ for* $i, r = 1, \ldots, n$. $\lim_{T\to\infty} \frac{1}{T}\ \Sigma_{T_1}^{T_2} EZ_t'\Omega_j^{-1} f_t(\theta) = \underset{\sim}{0}$ *if and only if* $\theta = \theta_0$.

$D = \lim_{T\to\infty}\left(\frac{1}{T}\ \Sigma_{T_1}^{T_2} EZ_t'\Omega_j^{-1} Z_t\right)^{-1}_{v\times v}$ *exists and is positive definite.*

(d) $\{(Y_t, X_t, Z_t)\}$ *is strong mixing with strong mixing numbers* $\{\alpha(s)\}$ *that satisfy* $\alpha(s) = o(s^{-\alpha/(\alpha-1)})$ *for some* $\alpha > 1$.

(e) $\sup_t E[\sup_{\theta\in\Theta} \|f_{it}(\theta)Z_{rt}\|^{\xi} + |Z_{rt}'Z_{rt}|^{\xi}] < \infty$, $\forall i, r = 1, \ldots, n$, *for some* $\xi > \alpha$.

(f) $f_{it}(\theta)$ *is defined and differentiable in* θ, $\forall i = 1, \ldots, n$, $\forall t$, *for all realizations of* $\{(Y_t, X_t)\}$, $\forall \theta \in \Theta^*$, *where* Θ^* *is some convex or open set that contains* Θ, *and* $\overline{\lim}_{T\to\infty} \frac{1}{T}\ \Sigma_{T_1}^{T_2} E \sup_{\theta\in\Theta^*} \left\|\frac{\partial}{\partial\theta} f_{it}(\theta)Z_{rt}'\right\| < \infty$, $\forall i, r = 1, \ldots, n$.

Nuisance parameter estimators $\hat{\Omega}_1$ and $\hat{\Omega}_2$ that satisfy assumption S1(b) can be obtained as follows: Let $\tilde{\theta}$ be some consistent preliminary estimator of θ_0, such as the 2SLS estimator. Then, for the case where $\hat{\Omega}_1$ and $\hat{\Omega}_2$ are allowed to differ, take

$$\hat{\Omega}_1 = \frac{1}{T_1}\ \Sigma_{T_1}^{-1} f_t(\tilde{\theta})f_t(\tilde{\theta})' \quad \text{and} \quad \hat{\Omega}_2 = \frac{1}{T_2}\ \Sigma_1^{T_2} f_t(\tilde{\theta})f_t(\tilde{\theta})' \ . \tag{4.6}$$

For the case where $\hat{\Omega}_1$ and $\hat{\Omega}_2$ are constrained to be equal, take

$$\hat{\Omega}_1 = \hat{\Omega}_2 = \frac{1}{T}\ \Sigma_{T_1}^{T_2} f_t(\tilde{\theta})f_t(\tilde{\theta})' \ . \tag{4.7}$$

Next, we introduce an assumption S2 such that assumptions S1 and S2 imply assumption 2 of Section 2 with $m_t(\theta,\hat{\tau})$ and $d(m,\hat{\tau})$ as above. Hence, by Theorem 2, under assumptions S1 and S2, $\sqrt{T}(\hat{\theta} - \theta_0)$ has an asymptotic $N(\underset{\sim}{0},V)$ distribution as $T \to \infty$, where $V = (M'DM)^{-1}M'DSDM(M'DM)^{-1}$,

$$M = \lim_{T\to\infty} \frac{1}{T}\ \Sigma_{T_1}^{T_2} EZ_t'\Omega_j^{-1}\ \frac{\partial}{\partial\theta'} f_t(\theta_0)_{v\times p} \ , \tag{4.8}$$

D is as in S1(c), and S is as in S2(c) below.

ASSUMPTION S2: (a) *Θ contains a convex compact neighborhood* Θ_c *of* θ_0.

(b) $EU_{it}Z_{rt} = 0$, $\forall i, r = 1, \ldots, n$.

(c) $S = \lim_{T\to\infty} \mathrm{Var}\left(\frac{1}{\sqrt{T}}\ \Sigma_{T_1}^{T_2} Z_t'\Omega_j^{-1} U_t\right)$ *exists where* $U_t = (U_{1t}, \ldots, U_{nt})'$.

(d) $\sqrt{T_1}(\hat{\Omega}_1 - \Omega_1) = O_p(1)$ *as* $T_1 \to \infty$, $\sqrt{T_2}(\hat{\Omega}_2 - \Omega_2) = O_p(1)$ *as* $T_2 \to \infty$, *and* $\lim_{T\to\infty} \mathrm{Var}\left(\frac{1}{\sqrt{T}}\ \Sigma_{T_1}^{T_2} Z_{it}'Z_{it}\right)$ *exists for all* $i = 1, \ldots, n$.

(e) $\lim_{T_1\to\infty} \frac{1}{T_1}\ \Sigma_{T_1}^{-1} E\frac{\partial}{\partial\theta} f_{it}(\theta)Z_{rt}'$ *and* $\lim_{T_2\to\infty} \frac{1}{T_2}\ \Sigma_1^{T_2} E\frac{\partial}{\partial\theta} f_{it}(\theta)Z_{rt}'$ *exist uniformly for $\theta \in \Theta$ and are continuous for $\theta \in \Theta$*, $\forall i, r = 1, \ldots, n$, *and M is nonsingular.*

(f) $f_{it}(\theta)$ *is twice differentiable in* θ, $\forall \theta \in \Theta_c$, $\forall i = 1, \ldots, n$, $\forall t$, *for all realizations of* $\{Y_t, X_t\}$ *and*

$$\sup_t E \sup_{\theta\in\Theta_c}\left(\left\|\frac{\partial}{\partial\theta}f_{it}(\theta)Z_{rt}\right\|^{\xi} + \left\|\frac{\partial^2}{\partial\theta_a\partial\theta}f_{it}(\theta)Z'_{rt}\right\| + \|U_{it}Z_{rt}\|^{2\xi} + (Z'_{rt}Z_{rt})^{2\xi}\right) < \infty$$

$\forall i, r = 1, \ldots, n$, $\forall a = 1, \ldots, p$, *for some* $\xi > \alpha$.

In cases where $S = D^{-1}$, the covariance matrix V simplifies to $V = (M'DM)^{-1}$. This occurs when

$$E(U_tU_t'|Z_t) = \Omega_j \quad \text{a.s.}, \ \forall t, \quad \text{and}$$
$$EZ_t'U_tU_{t-k}'Z_{t-k} = \underset{\sim}{0}, \quad \forall t, \quad \forall k = 1, 2, \ldots . \tag{4.9}$$

A consistent estimator of the covariance matrix V is given by $\hat{V} = (\hat{M}'\hat{D}\hat{M})^{-}\hat{M}'\hat{D}\hat{S}\hat{D}\hat{M}(\hat{M}'\hat{D}\hat{M})^{-}$, where $\hat{M} = \hat{M}(\hat{\theta})$, $\hat{M}(\theta) = \frac{1}{T}\left[Z_1'\hat{\Lambda}_1^{-}\frac{\partial}{\partial\theta'}f_1(\theta) + Z_2'\hat{\Lambda}_2^{-}\frac{\partial}{\partial\theta'}f_2(\theta)\right]$, and $\hat{S} = \hat{S}(\hat{\theta})$ for $\hat{S}(\theta)$ as defined in equation (2.3) with $m_t(\cdot,\cdot)$ defined just above equation (4.5) and $w(\cdot)$ corresponding to the Parzen or Bartlett weights. If the second condition of equation (4.9) holds, then $\hat{S}$ can be simplified by taking $\ell(T_1) = \ell(T_2) = 0$ in its definition. This yields

$$\hat{S} = \frac{1}{T}\sum_{T_1}^{T_2} Z_t'\hat{\Omega}_j^{-}f_t(\hat{\theta})f_t(\hat{\theta})'\hat{\Omega}_j^{-}Z_t . \tag{4.10}$$

If both of the conditions of (4.9) hold, then take

$$\hat{S} = \hat{D}^{-} \quad \text{and} \quad \hat{V} = (\hat{M}'\hat{D}\hat{M})^{-} . \tag{4.11}$$

To establish consistency of $\hat{V}$ we assume:

ASSUMPTION S3: $\sup_t E\|Z_t'U_t\|^{4\xi} < \infty$ *for some* $\xi > \alpha$.

THEOREM 7: (a) *Under assumptions* S1-S3, $\hat{S} \xrightarrow{p} S$, $\hat{M} \xrightarrow{p} M$, *and* $\hat{V} \xrightarrow{p} V$ *as* $T \to \infty$ *for* $\hat{S}$ *as defined just below equation* (4.10).

(b) *Under assumptions* S1 *and* S2, $\hat{S} \xrightarrow{p} S$, $\hat{M} \xrightarrow{p} M$, *and* $\hat{V} \xrightarrow{p} V$ *as* $T \to \infty$ *for* $\hat{S}$ *as defined in* (4.10) *or* (4.11), *provided the additional conditions outlined above* (4.10) *or* (4.11) *are satisfied, respectively.*

4.2. *Tests of Structural Change*

We now consider tests of nonlinear restrictions $H_0 : h(\theta) = \underset{\sim}{0}$. For brevity we omit many comments of Section 3.2 that apply here as well.

A sequence of <u>restricted 3SLS</u> estimators of θ_0 is any sequence of rv's $\{\tilde{\theta}\}$ such that $\tilde{\theta}$ minimizes equation (4.4) over $\theta \in \Theta_0 = \{\theta \in \Theta : h(\theta) = \underset{\sim}{0}\}$. Assumptions S1 and S5 (below) guarantee the existence and consistency of sequences of restricted 3SLS estimators, since they imply that assumption 1 of Section 2 holds with parameter space Θ_0.

ASSUMPTION S5: Θ_0 *is compact.*

The LM test statistic of equation (2.7) uses a restricted covariance matrix estimator given by $\tilde{V} = (\tilde{M}'\hat{D}\tilde{M})^{-}\tilde{M}'\hat{D}\tilde{S}\hat{D}\tilde{M}(\tilde{M}'\hat{D}\tilde{M})^{-}$, where $\tilde{M} = \hat{M}(\tilde{\theta})$, $\tilde{S} = \hat{S}(\tilde{\theta})$, and $\hat{M}(\theta)$ and $\hat{S}(\theta)$ are as defined just below equation (4.9). The estimator $\hat{D}$ is a preliminary estimator that does not depend on $\hat{\theta}$ or $\tilde{\theta}$. If desired, the preliminary estimator of θ_0 that is used in forming $\hat{D}$ can be chosen to be a restricted estimator of θ_0. As in equations (4.10) and (4.11), $\hat{S}$ can be replaced by the simpler estimator

$$\tilde{S} = \frac{1}{T}\sum_{T_1}^{T_2} Z_t'\hat{\Omega}_j^{-}f_t(\tilde{\theta})f_t(\tilde{\theta})'\hat{\Omega}_j^{-}Z_t \quad \text{or} \quad \tilde{S} = \hat{D}^{-} \tag{4.12}$$

when the conditions outlined above (4.10) or (4.11), respectively, hold under the null hypothesis. By the same argument as in the proof of Theorem 6, $\tilde{S}$,

$\tilde{M}$, and $\tilde{V}$ are consistent for S, M, and V, respectively, under the null hypothesis under the conditions of Theorem 7 and assumption S5.

The following assumption S6a implies assumption 6a of Section 2. It is used to obtain the asymptotic null distribution of the LR statistic.

ASSUMPTION S6a: *Under the null hypothesis,*

$$EZ_t'\Omega_j^{-1}U_tU_s'\Omega_j^{-1}Z_s = \begin{cases} EZ_t'\Omega_j^{-1}Z_t & \text{if } t = s \\ \underset{\sim}{0} & \text{if } t \neq s \end{cases} \quad \textit{for all } t,\ s = \ldots -1,\ 1,\ 2,\ \ldots,$$

where $j = 1$ *for* $t < 0$ *and* $j = 2$ *for* $t > 0$.

Assumption S6a implies that $S = D^{-1}$ and $I = J$. S6a holds under (4.9).

Assumptions S1-S3, 4, S5, and S6a for the 3SLS estimator imply assumptions 1-5 and 6a of Section 2. Thus, Theorem 4 holds and the W, LM, and LR statistics of equations (2.4), (2.7), and (2.8) are asymptotically chi-square with r degrees of freedom under the null hypothesis (where assumption S6a is needed only for the LR statistic).

The next assumption is used to obtain local power results:

ASSUMPTION S7: *Given* $\eta \in R^p$, *let* $\theta_T = \theta_0 + \eta/\sqrt{T}$ *and* $f_{it}(Y_{Tt}, X_t, \theta_T) = U_{it}$. *Let* P_T *denote the distribution of* $\{(Y_{Tt}, X_t, U_t, Z_t)\}$ *for* $T = 1, 2, \ldots$. *Suppose assumptions* S1 *and* S2 *hold with* Y_t *and* $f_{it}(\theta)$ *replaced by* Y_{Tt} *and* $f_{it}(Y_{Tt}, X_t, \theta)$ *throughout, with* S1(b) *and* S1(d) *holding under* $\{P_T\}$, *with the sequence* $\{(Y_t, X_t, Z_t)\}$ *replaced by the triangular array* $\{(Y_{Tt}, X_t, Z_t) : -T_1 \leq t \leq T_2,\ T = 1, 2, \ldots\}$ *in* S1(d), *and with* $\sup_t$ *replaced by* $\sup_{t \leq T, T=1,2,\ldots}$ *in* S1(e) *and* S2(f).

Assumptions 7, 8, 9, and 10a (with $\hat{b} = \bar{b} = 1$) of Section 2 are implied by assumptions S7, S3 and S7, S5 and S7, and S6a, respectively. Thus, Theorem 5 of Section 2 applies and the W, LM, and LR statistics have noncentral chi-square distributions under local alternatives. Their large sample power functions can be approximated accordingly.

We now provide some simplified formulae for the W, LM, and LR test statistics in the nonlinear simultaneous equations context. The general form for the Wald statistic is given in equation (2.4). If assumption S6a holds, then $\hat{S}$ can be taken as in equation (4.11), $\hat{S} = \hat{D}^-$, $\hat{I} = \hat{J}$, and W_T is given by the simplified formulae of Comment 1 to Theorem 4 with $\hat{J} = \hat{M}'\hat{D}\hat{M}$.

When testing for pure structural change, we assume that the IVs are taken such that each IV is non-zero only for observations with $t < 0$ or only for observations with $t > 0$. This condition ensures that the matrix $\hat{D}$ is block diagonal (after appropriate permutation of its rows and columns) and that $m_t(\hat{\theta}, \hat{\tau})$ satisfies the condition following equation (2.5). Hence, the Wald statistic for testing pure structural change is given by (2.5). When assumption S6a holds, $\hat{V}_1$ and $\hat{V}_2$ of (2.5) can be simplified as in (4.10) or (4.11).

The LM statistic corresponding to 3SLS estimation is given by

$$LM_T = T\bar{m}_T(\tilde{\theta})'\hat{D}\tilde{M}\tilde{J}^-\tilde{H}'(\tilde{H}\tilde{V}\tilde{H}')^-\tilde{H}\tilde{J}^-\tilde{M}'\hat{D}\bar{m}_T(\tilde{\theta}) \ , \tag{4.13}$$

where $\tilde{J} = \tilde{M}'\hat{D}\tilde{M}$. Note that the LM statistic is a quadratic form in the vector of orthogonality conditions between the IVs and the model evaluated at the restricted estimator $\tilde{\theta}$.

When testing for pure structural change (with IVs as in the second paragraph above), the LM statistic becomes

$$\begin{aligned} LM_T = T\left(\bar{m}_{1T}(\tilde{\theta})'\hat{D}_1\tilde{M}_1\tilde{J}_1^- - \bar{m}_{2T}(\tilde{\theta})'\hat{D}_2\tilde{M}_2\tilde{J}_2^-\right)\left(\tilde{V}_1/\pi_{1T} + \tilde{V}_2/\pi_{2T}\right)^- \\ \cdot\left(\tilde{J}_1^-\tilde{M}_1'\hat{D}_1\bar{m}_{1T}(\tilde{\theta}) - \tilde{J}_2^-\tilde{M}_2'\hat{D}_2\bar{m}_{2T}(\tilde{\theta})\right) \ , \end{aligned} \tag{4.14}$$

where $\bar{m}_{jT}(\tilde{\theta}) = \frac{1}{T_j} Z_j'\hat{\Lambda}_j^- f_j(\tilde{\theta})$, $\tilde{M}_j = \frac{1}{T_j} Z_j'\hat{\Lambda}_j^- \frac{\partial}{\partial\theta'} f_j(\tilde{\theta})$, $\tilde{\mathcal{J}}_j = \tilde{M}_j'\hat{D}_j\tilde{M}_j$, $\tilde{V}_j = \tilde{\mathcal{J}}_j^-\tilde{\mathcal{I}}_j\tilde{\mathcal{J}}_j^-$, $\tilde{\mathcal{I}}_j = \tilde{M}_j'\hat{D}_j\tilde{S}_j\hat{D}_j\tilde{M}_j$, and $\tilde{S}_j = \tilde{S}_j(\tilde{\theta})$ for $\hat{S}_j(\theta)$ defined in equation (2.3) for $j = 1, 2$.

When assumption S6a holds, LM_T simplifies by taking $\tilde{S} = \hat{D}^-$:

$$LM_T \doteq \bar{m}_T(\tilde{\theta})'\hat{D}\tilde{M}\tilde{\mathcal{J}}^-\tilde{M}'\hat{D}\bar{m}_T(\tilde{\theta}) \quad . \tag{4.15}$$

In particular, when testing for pure structural change under assumption S6a,

$$LM_T \doteq T_1\bar{m}_{1T}(\tilde{\theta})'\hat{D}_1\tilde{M}_1\tilde{\mathcal{J}}_1^-\tilde{M}_1'\hat{D}_1\bar{m}_{1T}(\tilde{\theta}) + T_2\bar{m}_{2T}(\tilde{\theta})'\hat{D}_2\tilde{M}_2\tilde{\mathcal{J}}_2^-\tilde{M}_2'\hat{D}_2\bar{m}_{2T}(\tilde{\theta}) \quad . \tag{4.16}$$

The LR statistic in the 3SLS case is given by

$$LR_T = 2T\left[d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau}) - d(\bar{m}_T(\hat{\theta}),\ \hat{\tau})\right] \quad , \tag{4.17}$$

where $d(\bar{m}_T(\theta),\ \hat{\tau})$ is the expression given in (4.4), i.e., the objective function for the 3SLS estimator. When testing for pure structural change (with IVs as above), the objective function factors as follows:

$$\begin{aligned} d(\bar{m}_T(\theta),\ \hat{\tau}) &= d_1(\bar{m}_{1T}(\theta),\ \hat{\tau}) + d_2(\bar{m}_{2T}(\theta),\ \hat{\tau}), \text{ where} \\ d_j(\bar{m}_{jT}(\theta),\ \hat{\tau}) &= f_j(\theta)'\hat{\Lambda}_j^-Z_j\left(Z_j'\hat{\Lambda}_j^-Z_j\right)^-Z_j'\hat{\Lambda}_j^-f_j(\theta) \quad \text{for} \quad j = 1, 2 \ . \end{aligned} \tag{4.18}$$

Thus, LR_T is obtained quite simply by performing 3SLS estimation on the observations indexed by $\{-T_1, \ldots, -1\}$, $\{1, \ldots, T_2\}$, and $\{-T_1, \ldots, T_2\}$.

When carrying out 2SLS estimation by setting $\hat{\Omega}_1 = \hat{\Omega}_2 = \Omega_1 = \Omega_2 = I_n$, the simplifying assumption S6a generally will not hold because it requires $EZ_t'U_tU_sZ_s = \begin{cases} EZ_t'Z_t & \forall t = s \\ 0 & \forall t \neq s \end{cases}$. The latter holds if the errors have variance one and are uncorrelated across time periods <u>and</u> equations conditional on the IVs--unrealistic assumptions in most applications. This problem can be

avoided by calculating the 2SLS estimator one equation at a time and by defining the scalars $\hat{\Omega}_1$ and $\hat{\Omega}_2$ as in (4.6) and (4.7). With these definitions, assumption S6a only requires the errors to be homoskedastic and uncorrelated conditional on the IV's. In the case of testing for pure structural change, the 2SLS estimator is the same regardless of the values of the scalars $\hat{\Omega}_1$ and $\hat{\Omega}_2$. Thus, the latter can be defined using the 2SLS estimator itself in (4.6) and (4.7) (i.e., with $\bar{\theta} = \hat{\theta}$) for the purposes of generating the W, LM, and LR test statistics.

5. MAXIMUM LIKELIHOOD ESTIMATION

This section considers ML estimators and corresponding tests for dynamic heterogeneous models that may exhibit structural change. For brevity, we do not give formal assumptions that imply assumptions 1-5 and 6b of Section 2. Such assumptions can be obtained in the same manner as is done in Sections 3 and 4.

Let $Y_t \in R^G$ and $X_t \in R^K$ denote endogenous and exogenous variables, respectively. Let

$$\{f_t(\theta) : \theta \in \Theta\} = \{f_t(Y_t|Y_{-T_1}, \ldots, Y_{t-1};\ X_{-T_1}, \ldots, X_{T_2};\ \theta) : \theta \in \Theta\} \tag{5.1}$$

denote a parametric family of conditional densities (with respect to some measure μ) of Y_t given $Y_{-T_1}, \ldots, Y_{t-1}$ and $X_{-T_1}, \ldots, X_{T_2}$, evaluated at the rv's $Y_{-T_1}, \ldots, Y_t$ and $X_{-T_1}, \ldots, X_{T_2}$. The conditional log-likelihood function of $\{Y_t\}$ given $\{X_t\}$ is $\Sigma_{-T_1}^{T_2} \log f_t(\theta)$. The analysis can be carried out conditionally on $\{X_t\}$ or unconditionally. In the latter case, the marginal distribution of $\{X_t\}$ is assumed not to depend on θ. The R^p-valued parameter θ may reflect pure or partial structural change, as described in

Section 2.

To apply the consistency result of Theorem 1 Section 2, let

$$m_t(\theta,\tau) = -\log f_t(\theta) \quad \text{and} \quad d(m,\tau) = m \quad \text{for} \quad m \in R^1 \quad \text{or} \tag{5.2}$$

$$m_t(\theta,\tau) = -\frac{\partial}{\partial\theta}\log f_t(\theta) \quad \text{and} \quad d(m,\tau) = m'm/2 \quad \text{for} \quad m \in R^p\,.[12] \tag{5.3}$$

Define a sequence of <u>ML estimators</u> $\{\hat\theta\}$ to be any sequence of extremum estimators that satisfies the definition of Section 2.1 with $m_t(\cdot,\cdot)$ and $d(\cdot,\cdot)$ as above. The "first-order conditions" definition of (5.3) is preferred to that of (5.2), since it must be used in the application of Theorems 2-5 anyway. In some contexts, however, the limit of the expectation of the normalized likelihood equations is not solved uniquely by $\theta = \theta_0$, whereas the limit of the expectation of the normalized log-likelihood function is maximized uniquely at $\theta = \theta_0$. In such cases, the definition of (5.2) needs to be used to establish consistency of the ML estimator $\hat\theta$ (via Theorem 1) and the definition of (5.3) needs to be used to establish various asymptotic distributional results (via Theorems 2-5).

The asymptotic covariance matrix of $\{\hat\theta\}$ simplifies as follows:
$V = (M'DM)^{-1}M'DSDM(M'DM)^{-1} = M^{-1}$, because $D = I_p$,

$$M = \lim_{T\to\infty} -\frac{1}{T}\Sigma_{-T_1}^{T_2} E\frac{\partial^2}{\partial\theta\partial\theta'}\log f_t(\theta_0) = \lim_{T\to\infty}\frac{1}{T}\Sigma_{-T_1}^{T_2} E\frac{\partial}{\partial\theta}\log f_t(\theta_0)\frac{\partial}{\partial\theta'}\log f_t(\theta_0)\ , \text{ and}$$

$$S = \lim_{T\to\infty} \mathrm{Var}\left(\frac{1}{\sqrt{T}}\Sigma_{-T_1}^{T_2}\frac{\partial}{\partial\theta}\log f_t(\theta)\right) = \lim_{T\to\infty}\frac{1}{T}\Sigma_{-T_1}^{T_2} E\frac{\partial}{\partial\theta}\log f_t(\theta_0)\frac{\partial}{\partial\theta'}\log f_t(\theta_0) = M\ , \tag{5.4}$$

provided the conditional information matrix inequality holds, using the fact that $\left\{\frac{\partial}{\partial\theta}\log f_t(\theta)\right\}$ is a martingale difference triangular array with respect to the triangular array of σ-fields generated by the conditioning variables in $f_t(\theta)$. Note that M is the limiting average of the conditional information matrices for each observation evaluated at the true parameter θ_0.

Assumption 6b of Section 2 holds with $\rho(W_t, \theta, \tau) = -\log f_t(\theta)$ and $c = 1$. In consequence, the LR statistic for testing $H_0 : h(\theta) = \underset{\sim}{0}$, viz.,

$$LR_T = -2\left(\Sigma_{-T_1}^{T_2}\log f_t(\tilde\theta) - \Sigma_{-T_1}^{T_2}\log f_t(\hat\theta)\right)\ , \tag{5.5}$$

has the desired χ_r^2 asymptotic distribution (under assumptions 1, 2, 4, 5, and 6b), where r is the number of restrictions and $\tilde\theta$ is the restricted ML estimator of θ_0. Furthermore, the Wald and LM statistics are given by the simplified formulae of Comment 2 to Theorem 4:

$$\begin{aligned} W_T &= Th(\hat\theta)'(\hat H\hat M^{-}\hat H')^{-}h(\hat\theta) \quad \text{and} \\ LM_T &= T\left(\frac{1}{T}\Sigma_{-T_1}^{T_2}\frac{\partial}{\partial\theta}\log f_t(\tilde\theta)\right)'\tilde M^{-}\left(\frac{1}{T}\Sigma_{-T_1}^{T_2}\frac{\partial}{\partial\theta}\log f_t(\tilde\theta)\right)\ , \end{aligned} \tag{5.6}$$

where $\hat M$ is defined to be $-\frac{1}{T}\Sigma_{-T_1}^{T_2}\frac{\partial^2}{\partial\theta\partial\theta'}\log f_t(\hat\theta)$, $\frac{1}{T}\Sigma_{-T_1}^{T_2}\frac{\partial}{\partial\theta}\log f_t(\hat\theta)\frac{\partial}{\partial\theta'}\log f_t(\hat\theta)$, or $-\frac{1}{T}\Sigma_{-T_1}^{T_2}E\frac{\partial^2}{\partial\theta\partial\theta'}\log f_t(\hat\theta)$ and $\tilde M$ is defined analogously with $\hat\theta$ replaced by $\tilde\theta$.

In the case of testing for pure structural change, the LR statistic is obtained quite simply by calculating the ML estimators for the data indexed by $\{-T_1, \ldots, -1\}$, $\{1, \ldots, T_2\}$ and $\{-T_1, \ldots, T_2\}$. The W and LM statistics are given in this case by

$$\begin{aligned} W_T &= T(\hat\theta_1 - \hat\theta_2)'(\hat M_1^{-}/\pi_{1T} + \hat M_2^{-}/\pi_{2T})^{-}(\hat\theta_1 - \hat\theta_2) \quad \text{and} \\ LM_T &= T_1\bar m_{1T}(\tilde\theta)'\tilde M_1^{-}\,\bar m_{1T}(\tilde\theta) + T_2\bar m_{2t}(\tilde\theta)'\tilde M_2^{-}\,\bar m_{2T}(\tilde\theta)\ , \end{aligned} \tag{5.7}$$

where $\theta = (\theta_1', \theta_2')'$, $\hat\theta = (\hat\theta_1', \hat\theta_2')'$, $\hat M_1$ equals $-\frac{1}{T_1}\Sigma_{-T_1}^{-1}\frac{\partial^2}{\partial\theta_1\partial\theta_1'}\log f_t(\hat\theta)$,

$\frac{1}{T_1}\Sigma_{T_1}^{-1}\frac{\partial}{\partial\theta_1}\log f_t(\hat{\theta})\frac{\partial}{\partial\theta_1'}\log f_t(\hat{\theta})$, or $-\frac{1}{T_1}\Sigma_{T_1}^{-1}E\frac{\partial^2}{\partial\theta_1\partial\theta_1'}\log f_t(\hat{\theta})$, $\hat{M}_2$ is defined analogously, $\bar{m}_{1T}(\tilde{\theta}) = \frac{1}{T_1}\Sigma_{T_1}^{-1}\frac{\partial}{\partial\theta_1}\log f_t(\tilde{\theta})$, and $\tilde{M}_1$, $\tilde{M}_2$, and $\bar{m}_{2T}(\tilde{\theta})$ are defined analogously with $\hat{\theta}$ replaced by $\tilde{\theta}$.

APPENDIX

PROOF OF THEOREM 1: We show that $d(\bar{m}_T(\hat{\theta},\hat{\tau}), \hat{\tau}) \xrightarrow{p} d(m(\theta_0, \tau_0), \tau_0)$ as $T \to \infty$. In view of assumption 1(e), a standard argument (using a Skorokhod representation, e.g., see Serfling (1980, Sec. 1.6.3), and a subsequence argument) then gives the desired result.

Since $d(\bar{m}_T(\hat{\theta},\hat{\tau}), \hat{\tau}) \le d(\bar{m}_T(\theta_0, \hat{\tau}), \hat{\tau})$, we get

$$\begin{aligned} &d(m(\hat{\theta},\hat{\tau}), \hat{\tau}) - d(m(\theta_0, \tau_0), \tau_0) \le d(m(\hat{\theta},\hat{\tau}), \hat{\tau}) - d(\bar{m}_T(\hat{\theta},\hat{\tau}), \hat{\tau}) \\ &+ d(\bar{m}_T(\theta_0, \hat{\tau}), \hat{\tau}) - d(m(\theta_0, \hat{\tau}), \hat{\tau}) + d(m(\theta_0, \hat{\tau}), \hat{\tau}) - d(m(\theta_0, \tau_0), \tau_0) \quad (A.1) \\ &\le 2 \sup_{(\theta,\tau)\in\Theta\times T} |d(\bar{m}_T(\theta,\tau),\tau) - d(m(\theta,\tau),\tau)| + d(m(\theta_0, \hat{\tau}), \hat{\tau}) - d(m(\theta_0, \tau_0), \tau_0) \\ &\xrightarrow{p} 0 \quad \text{as} \quad T \to \infty \end{aligned}$$

using assumptions 1(b)-1(e) (where the inequalities hold with probability that tends to one as $T \to \infty$).

Next, let $\{\tau_i\}$ be any non-random sequence such that $\tau_i \to \tau_0$ as $i \to \infty$. Suppose there exists a non-random sequence $\{\theta_i\}$ such that $\theta_i \in \Theta$, $\forall i$, and for some $\varepsilon > 0$,

$$d(m(\theta_i,\tau_i), \tau_i) - d(m(\theta_0,\tau_0), \tau_0) \le -\varepsilon \quad (A.2)$$

for infinitely many i. This is impossible, because there exists a subsequence $\{\theta_{i_\ell}\}$ of $\{\theta_i\}$ such that (A.2) holds for all $i = i_\ell$ and $\theta_{i_\ell} \to \theta_+$ as $\ell \to \infty$ for some $\theta_+ \in \Theta$ by compactness of Θ. By assumption 1(e), $\underline{\lim}_{\ell\to\infty} d(m(\theta_{i_\ell}, \tau_{i_\ell}), \tau_{i_\ell}) \ge d(m(\theta_0, \tau_0), \tau_0)$, which yields a contradiction. Thus, for any fixed sequences $\{\tau_i\}$ and $\{\theta_i\}$ as above, $d(m(\theta_i, \tau_i), \tau_i) - d(m(\theta_0, \tau_0), \tau_0) \ge -\varepsilon$ for all i large, for any $\varepsilon > 0$. By standard arguments (using a Skorokhod representation), this implies

$$d(m(\hat{\theta},\hat{\tau}),\ \hat{\tau}) - d(m(\theta_0,\ \tau_0),\ \tau_0) \geq v_T \qquad (A.3)$$

for some sequence of rv's $\{v_T\}$ such that $v_T = o_p(1)$ as $T \to \infty$. □

The proofs of Theorems 2 and 4 are similar to proofs in Gallant (1987).

PROOF OF THEOREM 2: Element by element mean value expansions of $\sqrt{T}\,\frac{\partial}{\partial\theta}d(\bar{m}_T(\hat{\theta}),\ \hat{\tau})$ about θ_0 give: $\forall a = 1, \ldots, p$,

$$o_p(1) = \sqrt{T}\frac{\partial}{\partial\theta_a}d(\bar{m}_T(\hat{\theta}),\hat{\tau}) = \sqrt{T}\frac{\partial}{\partial\theta_a}d(\bar{m}_T(\theta_0),\hat{\tau}) + \frac{\partial^2}{\partial\theta_a\partial\theta'}d(\bar{m}_T(\theta^*),\hat{\tau})\sqrt{T}(\hat{\theta}-\theta_0), \quad (A.4)$$

where θ^* is a rv on the line segment joining $\hat{\theta}$ and θ_0, and hence, $\theta^* \xrightarrow{p} \theta_0$. (See Jennrich (1969) Lemma 3 for the mean value theorem for random functions.) The first equality holds because $\hat{\theta}$ minimizes $d(\bar{m}_T(\hat{\theta}),\ \hat{\tau})$ and $\hat{\theta}$ is in the interior of Θ with probability that goes to one as $T \to \infty$ by assumptions 2(a) and (d).

Below we show that

$$\frac{\partial^2}{\partial\theta_a\partial\theta'}d(\bar{m}_T(\theta^*),\ \hat{\tau}) = \frac{\partial^2}{\partial\theta_a\partial\theta'}d(m(\theta_0),\ \tau_0) + o_p(1), \qquad (A.5)$$

where $\frac{\partial^2}{\partial\theta\partial\theta'}d(m(\theta_0),\ \tau_0) = M'DM$ and

$$\sqrt{T}\ \frac{\partial}{\partial\theta}d(\bar{m}_T(\theta_0),\ \hat{\tau}) \xrightarrow{d} N(\underset{\sim}{0},\ M'DSDM) \quad \text{as} \quad T \to \infty\ . \qquad (A.6)$$

These results, equation (A.4), and the nonsingularity of M'DM give

$$\sqrt{T}(\hat{\theta} - \theta_0) = -(M'DM)^{-1}\ \frac{\partial}{\partial\theta}d(\bar{m}_T(\theta_0),\ \hat{\tau}) + o_p(1) \xrightarrow{d} N(\underset{\sim}{0},\ V) \quad \text{as} \quad T \to \infty\ . \qquad (A.7)$$

To show (A.5), we proceed as follows:

$$\frac{\partial^2}{\partial\theta_a\partial\theta_\ell}d(\bar{m}_T(\theta^*),\ \hat{\tau}) = \frac{\partial^2}{\partial\theta_a\partial\theta_\ell}\bar{m}_T(\theta^*)'\frac{\partial}{\partial m}d(\bar{m}_T(\theta^*),\ \hat{\tau})$$
$$+ \frac{\partial}{\partial\theta_a}\bar{m}_T(\theta^*)'\frac{\partial^2}{\partial m\partial m'}d(\bar{m}_T(\theta^*),\ \hat{\tau})\frac{\partial}{\partial\theta_\ell}\bar{m}_T(\theta^*)\ . \qquad (A.8)$$

By assumptions 2(a), (b), and (f),

$$\|\bar{m}_T(\theta^*) - m(\theta_0)\| \leq \left\|\bar{m}_T(\theta^*,\hat{\tau}) - E\bar{m}_T(\theta,\tau)\Big|_{\theta=\theta^*,\tau=\hat{\tau}}\right\| \qquad (A.9)$$
$$+ \left\|E\bar{m}_T(\theta,\tau)\Big|_{\theta=\theta^*,\tau=\hat{\tau}} - m(\theta^*,\hat{\tau})\right\| + \|m(\theta^*,\hat{\tau}) - m(\theta_0,\ \tau_0)\| \xrightarrow{p} 0$$

as $T \to \infty$, where $\|\cdot\|$ denotes the Euclidean norm. Using this result, the continuity of $\frac{\partial}{\partial m}d(m,\tau)$ over $\mathcal{M} \times \mathrm{T}$ (assumption 2(e)), the assumption 2(b) that $\hat{\tau} \xrightarrow{p} \tau_0$, and the continuous mapping theorem, we get

$$\frac{\partial}{\partial m}d(\bar{m}_T(\theta^*),\ \hat{\tau}) \xrightarrow{p} \frac{\partial}{\partial m}d(m(\theta_0),\ \tau_0) = \underset{\sim}{0} \quad \text{as} \quad T \to \infty\ , \qquad (A.10)$$

where the equality holds by 2(b), (e), and (f). Using assumption 2(f), it is straightforward to show that $\frac{\partial^2}{\partial\theta_a\partial\theta_\ell}\bar{m}_T(\theta^*) = O_p(1)$ as $T \to \infty$. This result and (A.10) imply that the first term of (A.8) is $o_p(1)$ as $T \to \infty$.

Similarly, the continuity of $\frac{\partial^2}{\partial m\partial m'}d(m,\tau)$ over $\mathcal{M} \times \mathrm{T}$ (assumption 2(e)), equation (A.9), $\hat{\tau} \xrightarrow{p} \tau_0$, and the continuous mapping theorem give

$$\frac{\partial^2}{\partial m\partial m'}d(\bar{m}_T(\theta^*),\ \hat{\tau}) \xrightarrow{p} \frac{\partial^2}{\partial m\partial m'}d(m(\theta_0),\ \tau_0) = D \quad \text{as} \quad T \to \infty\ . \qquad (A.11)$$

It follows from assumptions 2(a), (b), and (f) that

$$\left\|\frac{\partial}{\partial\theta'}\bar{m}_T(\theta^*) - M\right\| \leq \left\|\frac{\partial}{\partial\theta'}\bar{m}_T(\theta^*) - M(\theta^*,\hat{\tau})\right\| + \|M(\theta^*,\hat{\tau}) - M(\theta_0,\tau_0)\| \xrightarrow{p} 0 \quad (A.12)$$

as $T \to \infty$. Equations (A.11) and (A.12) imply that the second term of (A.8) equals $[M'DM]_{a\ell} + o_p(1)$, and hence, (A.5) is established.

To establish equation (A.6), we write

$$\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\theta_0),\hat{\tau}) = \sqrt{T}\frac{\partial}{\partial\theta}\bar{m}_T'(\theta_0)\frac{\partial}{\partial m}d(\bar{m}_T(\theta_0),\hat{\tau}) = M'\sqrt{T}\frac{\partial}{\partial m}d(\bar{m}_T(\theta_0),\hat{\tau}) + o_p(1) \quad (A.13)$$

using 2(f) provided $\sqrt{T}\frac{\partial}{\partial m}d(\bar{m}_T(\theta_0),\hat{\tau}) = O_p(1)$, as we now demonstrate.

By the mean value theorem, the a^th element of $\sqrt{T}\frac{\partial}{\partial m}d(\bar{m}_T(\theta_0,\hat{\tau}),\hat{\tau})$ can be expanded about $(E\bar{m}_T(\theta_0, \tau_0), \tau_0)$ to get:

$$\sqrt{T}\frac{\partial}{\partial m_a}d(\bar{m}_T(\theta_0,\hat{\tau}),\hat{\tau}) = \sqrt{T}\frac{\partial}{\partial m_a}d(E\bar{m}_T(\theta_0,\tau_0), \tau_0)$$
$$+ \frac{\partial^2}{\partial m'\partial m_a}d(m^*,\tau^*)\sqrt{T}(\bar{m}_T(\theta_0,\hat{\tau}) - E\bar{m}_T(\theta_0,\tau_0)) + \frac{\partial^2}{\partial\tau'\partial m_a}d(m^*,\tau^*)\sqrt{T}(\hat{\tau} - \tau_0), \quad (A.14)$$

where (m^*,τ^*) is on the line segment joining $(\bar{m}_T(\theta_0,\hat{\tau}), \hat{\tau})$ and $(E\bar{m}_T(\theta_0,\tau_0), \tau_0)$, and hence, $m^* \xrightarrow{p} m(\theta_0)$ and $\tau^* \xrightarrow{p} \tau_0$ as $T \to \infty$. (More precisely, (A.14) holds with probability that goes to one as $T \to \infty$.)

The first term of the right-hand-side of (A.14) is zero for T large by assumption 2(b). Also, since $\sqrt{T}(\hat{\tau}-\tau_0) = O_p(1)$ (assumption 2(b)) and $\frac{\partial^2}{\partial\tau'\partial m_a}d(m,\tau)$ is continuous over $M \times T$ (assumption 2(e)), we have:

$$\frac{\partial^2}{\partial\tau'\partial m_a}d(m^*,\tau^*)\sqrt{T}(\hat{\tau}-\tau_0) = \frac{\partial^2}{\partial\tau'\partial m_a}d(m(\theta_0),\tau_0)\sqrt{T}(\hat{\tau}-\tau_0) + o_p(1) = o_p(1), \quad (A.15)$$

where the second equality follows from 2(b). Similarly, using assumption 2(e), $\frac{\partial^2}{\partial m_a\partial m'}d(m^*,\tau^*) = [D]_a' + o_p(1)$ where $[D]_a$ denotes the a^th column of D. Hence, if $\sqrt{T}(\bar{m}_T(\theta_0,\hat{\tau}) - E\bar{m}_T(\theta_0,\tau_0)) = O_p(1)$, the above results and (A.14) yield

$$\sqrt{T}\frac{\partial}{\partial m}d(\bar{m}_T(\theta_0, \hat{\tau}), \hat{\tau}) = D\sqrt{T}(\bar{m}_T(\theta_0, \hat{\tau}) - E\bar{m}_T(\theta_0, \tau_0)) + o_p(1) . \quad (A.16)$$

The proof is complete once we show that

$$\sqrt{T}(\bar{m}_T(\theta_0, \hat{\tau}) - E\bar{m}_T(\theta_0, \tau_0)) \xrightarrow{d} N(\underset{\sim}{0},S) \quad \text{as} \quad T \to \infty , \quad (A.17)$$

since this implies that (A.16) and (A.13) hold, which establishes (A.6).

A mean value expansion of the a^th element of $\bar{m}_T(\theta_0, \hat{\tau})$ yields

$$\sqrt{T}(\bar{m}_{Ta}(\theta_0,\hat{\tau})-E\bar{m}_{Ta}(\theta_0,\tau_0)) = \sqrt{T}(\bar{m}_{Ta}(\theta_0,\tau_0)-E\bar{m}_{Ta}(\theta_0,\tau_0))$$
$$+ \frac{\partial}{\partial\tau'}\bar{m}_{Ta}(\theta_0,\tau^*)\sqrt{T}(\hat{\tau}-\tau_0) = \sqrt{T}(\bar{m}_{Ta}(\theta_0,\tau_0) - E\bar{m}_{Ta}(\theta_0,\tau_0)) + o_p(1), \quad (A.18)$$

where τ^* lies on the line segment joining $\hat{\tau}$ and τ_0, using assumption 2(b), since $\frac{\partial}{\partial\tau}\bar{m}_T(\theta_0,\tau^*) \xrightarrow{p} dm(\theta_0,\tau_0) = \underset{\sim}{0}$ by assumption 2(f). Stacking equation (A.18) for a = 1, ..., p and using assumption 2(c) gives (A.17). □

PROOF OF THEOREM 3: $\hat{M} \xrightarrow{p} M$ and $\hat{D} \xrightarrow{p} D$ as $T \to \infty$ by the arguments used in equations (A.11) and (A.12), respectively. Thus, $\hat{I} \xrightarrow{p} I$ and $\hat{J}^- \xrightarrow{p} J^{-1}$, since J is nonsingular (assumption 2(g)). □

PROOF OF THEOREM 4: To prove part (a), the delta method gives

$$\sqrt{T}(h(\hat{\theta}) - h(\theta_0)) \xrightarrow{d} N(\underset{\sim}{0}, HVH') \quad \text{as} \quad T \to \infty \quad (A.19)$$

using assumption 4. By Theorem 3, $\hat{V} \xrightarrow{p} V$ and by the continuous mapping theorem and assumption 2(a), $\hat{H} \xrightarrow{p} H$ as $T \to \infty$. Since HVH' is nonsingular, this implies that $(\hat{H}\hat{V}\hat{H}')^- \xrightarrow{p} (HVH')^{-1}$ as $T \to \infty$. This result, (A.19), and the continuous mapping theorem give the desired result.

Next we establish part (b). Standard arguments give

$$\tilde{J} \xrightarrow{p} J , \quad \tilde{H} \xrightarrow{p} H , \quad \text{and} \quad \tilde{V} \xrightarrow{p} V \quad \text{as} \quad T \to \infty . \quad (A.20)$$

Mean value expansions about θ_0 yield: $\forall a = 1, \ldots, p$,

$$\sqrt{T}\frac{\partial}{\partial\theta_a}d(\bar{m}_T(\tilde{\theta}), \hat{\tau}) = \sqrt{T}\frac{\partial}{\partial\theta_a}d(\bar{m}_T(\theta_0), \hat{\tau}) + \frac{\partial^2}{\partial\theta_a\partial\theta'}d(\bar{m}_T(\dot{\theta}), \hat{\tau})\sqrt{T}(\tilde{\theta} - \theta_0), \quad (A.21)$$

$$\sqrt{T}\ h_a(\tilde{\theta}) = \sqrt{T}\ h_a(\theta_0) + \frac{\partial}{\partial\theta'}h_a(\theta^*)\sqrt{T}(\tilde{\theta} - \theta_0)\ , \tag{A.22}$$

where $\dot{\theta}$ and θ^* lie on the line segment joining $\tilde{\theta}$ and θ_0, and hence, satisfy $\dot{\theta} \xrightarrow{p} \theta_0$ and $\theta^* \xrightarrow{p} \theta_0$ as $T \to \infty$. We stack equations (A.21) and (A.22) for $a = 1, \ldots, p$ and write them as

$$\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau}) = \sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\theta_0),\ \hat{\tau}) + \dot{J}\sqrt{T}(\tilde{\theta} - \theta_0) \quad \text{and} \tag{A.23}$$

$$\underset{\sim}{0} = H^*\sqrt{T}(\tilde{\theta} - \theta_0) \tag{A.24}$$

using the fact that $h(\tilde{\theta}) = h(\theta_0) = \underset{\sim}{0}$.

By equation (A.6), $\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\theta_0),\ \hat{\tau}) \xrightarrow{d} N(\underset{\sim}{0}, I)$ as $T \to \infty$. By equation (A.5), $\dot{J} \xrightarrow{p} J$ as $T \to \infty$. Hence, using the nonsingularity of J, we get $\dot{J}^-\dot{J} \doteq I_p$, where $\doteq$ is defined in Comment 1 of Theorem 4. By assumptions 4 and 5, $H^* \xrightarrow{p} H$ as $T \to \infty$. Pre-multiplication of (A.23) by $H^*\dot{J}^-$ now gives

$$\begin{aligned} H^*\dot{J}^-\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau}) &= H^*\dot{J}^-\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\theta_0),\ \hat{\tau}) + H^*\dot{J}^-\dot{J}\sqrt{T}(\tilde{\theta} - \theta_0) \\ &\doteq H^*\dot{J}^-\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\theta_0),\ \hat{\tau}) \xrightarrow{d} N(\underset{\sim}{0},\ HJ^{-1}IJ^{-1}H') \quad \text{as} \quad T \to \infty\ . \end{aligned} \tag{A.25}$$

With probability that tends to one as $T \to \infty$, $\tilde{\theta}$ is in the interior of Θ and there exists a rv $\tilde{\lambda}$ of Lagrange multipliers such that

$$\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau}) + \tilde{H}'\tilde{\lambda} = \underset{\sim}{0}\ , \tag{A.26}$$

where $\tilde{H} = \frac{\partial}{\partial\theta'}h(\tilde{\theta})$. Equations (A.25) and (A.26) combine to give

$$-H^*\dot{J}^-\tilde{H}'\sqrt{T}\tilde{\lambda} \doteq H^*\dot{J}^-\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau}) = O_p(1)\ . \tag{A.27}$$

Since $H^*\dot{J}^-\tilde{H}' \xrightarrow{p} HJ^{-1}H'$ and $HJ^{-1}H'$ is nonsingular, equations (A.27) and (A.26) imply that $\sqrt{T}\ \tilde{\lambda} = O_p(1)$ and

$$\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau}) = O_p(1)\ . \tag{A.28}$$

Equations (A.20), (A.25), and (A.28) yield

$$\tilde{H}J^-\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau}) \xrightarrow{d} N(\underset{\sim}{0},\ HVH') \quad \text{as} \quad T \to \infty\ . \tag{A.29}$$

The desired result now follows from equations (A.20) and (A.29) and the continuous mapping theorem.

We now prove part (c). Suppose that assumption 6a holds. A two-term Taylor expansion of $d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau})$ about $\hat{\theta}$ gives

$$\begin{aligned} LR_T &= 2T\left[d(\bar{m}_T(\tilde{\theta}),\hat{\tau}) - d(\bar{m}_T(\hat{\theta}),\hat{\tau})\right]/\hat{b} = 2T\frac{\partial}{\partial\theta'}d(\bar{m}_T(\hat{\theta}),\ \hat{\tau})(\tilde{\theta}-\hat{\theta})/\hat{b} \\ &\quad + T(\tilde{\theta}-\hat{\theta})'\frac{\partial^2}{\partial\theta\partial\theta'}d(\bar{m}_T(\theta^*),\ \hat{\tau})(\tilde{\theta}-\hat{\theta})/\hat{b} \\ &\doteq T(\tilde{\theta}-\hat{\theta})'J^*(\tilde{\theta}-\hat{\theta})/\hat{b}\ , \end{aligned} \tag{A.30}$$

where θ^* lies on the line segment joining $\tilde{\theta}$ and $\hat{\theta}$, and hence, $\theta^* \xrightarrow{p} \theta_0$ as $T \to \infty$, J^* is defined implicitly, and "$\doteq$" holds by the first order conditions for the estimator $\hat{\theta}$.

Applying the mean value theorem element by element and stacking the equations yields

$$\begin{aligned} \sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau}) &= \sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\hat{\theta}),\ \hat{\tau}) + \mathring{J}\sqrt{T}(\tilde{\theta}-\hat{\theta}) \\ &\doteq \mathring{J}\sqrt{T}(\tilde{\theta}-\hat{\theta}) \end{aligned} \tag{A.31}$$

for a matrix $\mathring{J}$ that satisfies $\mathring{J} \xrightarrow{p} J$ as $T \to \infty$. Pre-multiplying (A.31) by $J^*\mathring{J}^-$ and substituting the result in (A.30) gives

$$LR_T \doteq T\frac{\partial}{\partial\theta'}d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau})(J^*\mathring{J}^-)'(J^*)^-J^*\mathring{J}^-\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau})/\hat{b}$$
$$= T\frac{\partial}{\partial\theta'}d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau})\tilde{J}^-\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau})/\hat{b} + o_p(1)\ , \qquad (A.32)$$

because $\mathring{J}^-\mathring{J} \doteq I_p$, $\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau}) = O_p(1)$, $J^*\mathring{J}^- \xrightarrow{p} I_p$, *and* $\tilde{J} - J^* \xrightarrow{p} \underset{\sim}{0}$ as $T \to \infty$, by (A.5), (A.20), and (A.28).

Since $I = bJ$ and $\hat{b} \xrightarrow{p} b$ by assumption 6a, $\tilde{V} = \tilde{J}^- + o_p(1)$. In this case, LM_T simplifies to

$$LM_T = T\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau})'\tilde{J}^-\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau})/\tilde{b} + o_p(1) = LR_T + o_p(1) \qquad (A.33)$$

using $\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau}) \doteq \tilde{H}'\tilde{\lambda}$, as above. The desired result now follows from part (b) of the Theorem. The proof of part (c) when assumption 6b holds is analogous to the above proof under 6a. □

PROOF OF THEOREM 5: First we prove part (a). The proof of Theorem 3 shows that $\hat{M} \xrightarrow{p} M$ and $\hat{D} \xrightarrow{p} D$ under $\{P_T\}$, since $\hat{\theta} \xrightarrow{p} \theta_0$, $\hat{\tau} \xrightarrow{p} \tau_0$, and assumption 2(f) holds under $\{P_T\}$. We have HVH′ is nonsingular, $\hat{S} \xrightarrow{p} S$, and $\hat{H} \xrightarrow{p} H$ under $\{P_T\}$, by assumptions 4, 8, and 4 and 7, respectively. Thus, $(\hat{H}\hat{V}\hat{H}')^- \xrightarrow{p} (HVH')^{-1}$ under $\{P_T\}$.

Mean value expansions of $h_a(\hat{\theta})$ about $h_a(\theta_T)$, stacked for a = 1, ..., p, yield

$$\sqrt{T}h(\hat{\theta}) = \sqrt{T}h(\theta_T) + H^*\sqrt{T}(\hat{\theta} - \theta_T) \qquad (A.34)$$

for an r × p matrix H* that satisfies $H^* \xrightarrow{p} H$ under $\{P_T\}$. Assumption 4 and element by element mean value expansions give $\sqrt{T}h(\theta_T) \to H\eta$ as $T \to \infty$. Part (a) now follows by the continuous mapping theorem once we show that

$$\sqrt{T}(\hat{\theta} - \theta_T) \xrightarrow{d} N(\underset{\sim}{0},V) \quad \text{under} \quad \{P_T\} \quad \text{as} \quad T \to \infty\ . \qquad (A.35)$$

This follows using assumption 7 by the proof of Theorem 2 with θ_0 replaced by θ_T in all equations but (A.5), (A.8)-(A.12), and (A.15).

To prove part (b), note that under assumptions 4 and 7-9 the proof of Theorem 4(b) goes through with the following changes. The parameter θ_0 is replaced by θ_T in equations (A.21)-(A.23) and equations (A.24), (A.25), and (A.29) are replaced by

$$\underset{\sim}{0} = \sqrt{T}h(\theta_T) + H^*\sqrt{T}(\tilde{\theta} - \theta_T)\ , \qquad (A.36)$$

$$H^*\dot{J}^-\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau}) \doteq H^*\dot{J}^-\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\theta_T),\ \hat{\tau}) + \sqrt{T}h(\theta_T)$$
$$\xrightarrow{d} N(H\eta,\ HVH') \quad \text{as} \quad T \to \infty\ , \text{ and} \qquad (A.37)$$

$$\tilde{H}\tilde{J}^-\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\ \hat{\tau}) \xrightarrow{d} N(H\eta,\ HVH') \quad \text{under} \quad \{P_T\} \quad \text{as} \quad T \to \infty\ , \qquad (A.38)$$

respectively.

Part (c) is proved by the proof of Theorem 4(c). The latter goes through under assumptions 4, 6a or 6b, 7, 9, and 10 with the only change being an appeal to Theorem 5(b) rather than Theorem 4(b). □

PROOF THAT ASSUMPTION R1 ⇒ 1 AND R1 PLUS R2 ⇒ 2: First, we note that assumption R1(f) and Lemma 2 of Jennrich (1969) guarantee the existence of a sequence of LS estimators $\{\hat{\theta}\}$. Next, the notation of assumptions 1 and R1 are linked via the definitions given just below equation (3.2).

Assumptions 1(a), (b), and (d) follow immediately from R1. Assumption 1(c) follows from R1 using a uniform LLN of Andrews (1987b). In particular, assumptions R1(a), R1(d), R1(e), and R1(a) and (f) imply assumptions A1, B1, B2, and A5, respectively, of Andrews (1987b). Corollaries 1 and 2 of Andrews (1987b) (the former of which follows from Theorem 2.10 of McLeish (1975a)) then imply that $\{m_t(\theta)\}$ satisfies a uniform LLN over Θ. The existence of the limit function m(·) of assumption 1(c) follows from R1(b) and (c). The latter two conditions also imply 1(e).

The notation of assumptions 2 and R2 are linked via $W_t = (Y_t, X_t')'$, $m_t(\theta,\tau) = (Y_t - f_t(\theta))\frac{\partial}{\partial\theta}f_t(\theta)$, and $d(m,\tau) = m'm/2$, for $m \in R^p$. Assumption 2(a) follows from Theorem 1 under assumption R1. Assumption 2(b) holds, because τ does not arise in this case and $E\bar{m}_T(\theta_0) = \frac{1}{T}\sum_{-T_1}^{T_2} EU_t \frac{\partial}{\partial\theta}f_t(\theta_0) = \underset{\sim}{0}$, $\forall T$, by R1(b) and (f) (since R1(f) allows an interchange of the integral and derivative operations).

Assumption 2(c) is verified using Theorem 2.1(A) and equations (6.1)-(6.3) of Withers (1981) or Corollary 1 of Herrndorf (1984). We refer to Herrndorf's CLT, since his conditions are simpler, and hence, easier to verify in this case. The summands for the CLT are $\left\{U_t \frac{\partial}{\partial\theta}f_t(\theta_0)\right\}$. Using the Cramer-Wold device, it suffices to establish CLTs for arbitrary linear combinations of these vector-valued summands. For any such linear combination, equations (1.1), (1.2), and (1.6) of Herrndorf hold by R1(b) and (f), R2(c), and R1(d) and R2(d), respectively. Thus, assumption 2(c) holds.

Assumption 2(d) follows directly from R1(a) and R2(a). Assumption 2(e) holds because m'm/2 is twice differentiable.

To establish 2(f), note that the differentiability of $m_t(\theta)$ follows by R2(d), and $\{m_t(\theta)\}$ and $\left\{\frac{\partial}{\partial\theta}m_t(\theta)\right\}$ satisfy uniform LLNs over Θ_c because R1(d), R2(d), and R2(d) imply assumptions B1, B2, and A5 of Andrews (1987b), respectively. The limits $m(\theta)$ and $M(\theta)$ exist uniformly for $\theta \in \Theta_c$ and are continuous on Θ_c by R2(b). $\left\{\frac{\partial^2}{\partial\theta_a\partial\theta}m_t(\theta)\right\}$ satisfies the conditions of 2(f) by R1(d) and R2(d). Hence, assumption 2(f) holds. Assumption 2(g) follows from R2(b). □

PROOF OF THEOREM 6: If $\hat{S} \xrightarrow{p} S$ as $T \to \infty$, then $\hat{M} \xrightarrow{p} M$ and $\hat{V} \xrightarrow{p} V$ as $T \to \infty$ in parts (a) and (b) of Theorem 6 by Theorem 3, since assumptions R1 and R2 imply assumption 2 (as shown immediately above). In part (b), the proof of $\hat{S} \xrightarrow{p} S$ is analogous to that of $\hat{M} \xrightarrow{p} M$.

It remains to show $\hat{S} \xrightarrow{p} S$ in part (a). This follows by the method of proof of Theorem 2 of Newey and West (1987), noting that their assumptions (i), (ii), and (iv) are implied by R2(d), R2(d) and R3, and R1(a) and the consistency of $\hat{\theta}$, respectively. Their assumption (iii) is stronger than our assumption R1(d). Their proof still works with the weaker assumption R1(d), however, by using the mixing inequality of Lemma 2.1 of Herrndorf (1984) in place of that of White's (1984) Corollary 6.16 in the proof of White's (1984) Lemmas 6.17 and 6.19, which are used in Newey and West's (1987) proof. (Note that the use of Lemma A1 very conveniently allows the same mixing condition to be used to obtain consistency of the covariance matrix estimator as is used for consistency and asymptotic normality of $\hat{\theta}$.) The fact that our observations are indexed by a doubly infinite sequence only requires a slight alteration of their proof. □

PROOF THAT ASSUMPTIONS R1 PLUS R5 ⇒ 5 AND R6b ⇒ 6b: Assumption 5 holds by applying Theorem 1 with the parameter space Θ_0 instead of Θ. Since Θ_0 is compact, the proof that R1 ⇒ 1 goes through without change.

Assumption R6b implies 6b with $\rho_t(W_t, \theta, \tau) = (Y_t - f_t(\theta))^2/2$ and $S = \sigma^2 \cdot \lim_{T\to\infty} \frac{1}{T}\sum_{-T_1}^{T_2} E\,\frac{\partial}{\partial\theta}f_t(\theta_0)\frac{\partial}{\partial\theta'}f_t(\theta_0) = cM$, where $c = -\sigma^2$. □

PROOF THAT S1 ⇒ 1 AND S1 PLUS S2 ⇒ 2: Assumption S1(f) and Lemma 2 of Jennrich (1969) guarantee the existence of a sequence of 3SLS estimators $\{\hat{\theta}\}$. Next, the notation of assumptions 1 and 2 and S1 and S2 are linked via the definitions of $m_t(\cdot,\cdot)$ and $d(\cdot,\cdot)$ given just above equation (4.5).

Assumption 1(a) is implied by S1(a). Assumption 1(b) follows from S1(b), the fact that $\{Z_tZ_t' : t = -1, \ldots, -T_1\}$ and $\{Z_tZ_t' : t = 1, \ldots, T_2\}$ satisfy weak LLNs as $T_1 \to \infty$ and $T_2 \to \infty$, respectively (which follows from McLeish's (1975a) Theorem 2.10 using assumptions S1(d) and (e)) and the assumption S1(c) that the appropriate limits exist.

To establish assumption 1(c), we need $\{m_t(\theta,\tau) : t = -T_1, \ldots, T_2\}$ to satisfy a uniform LLN over $(\theta,\tau) \in \Theta \times T$. Due to the multiplicative way in which τ (i.e., Ω_j) enters $m_t(\theta,\tau)$ and the assumption that $\lim_{T\to\infty} \pi_{1T} = \pi_1$ exists, this reduces to obtaining uniform LLNs for $\{f_{it}(\theta)Z_{rt} : t = -1, \ldots, -T_1\}$ and $\{f_{it}(\theta)Z_{rt} : t = 1, \ldots, T_2\}$ over $\theta \in \Theta$ as $T_1 \to \infty$ and $T_2 \to \infty$, respectively, for each i, r = 1, ..., n. The latter follows using Corollaries 1 and 2 of Andrews (1987b), since assumptions S1(a), S1(d), S1(e), and S1(a) and (f) imply assumptions A1, B1, B2, and A5 of Andrews (1987b), respectively. Assumption S1(c) guarantees that the function $m(\theta,\tau) = \lim_{T\to\infty} \frac{1}{T} \Sigma_{-T_1}^{T_2} Em_t(\theta,\tau)$ exists uniformly for $(\theta,\tau) \in \Theta \times T$.

Assumption 1(d) holds because (1) $d(\cdot,\cdot)$ is a quadratic form and (2) $m(\theta,\tau)$ is continuous on the compact set $\Theta \times T$ by a subsidiary result of the uniform LLN used above (which utilizes assumption S1(f)) and by the fact that τ enters multiplicatively.

Since $d(\cdot,\cdot)$ and $m(\cdot,\cdot)$ are continuous, assumption 1(e) reduces to: $d(m(\theta, \tau_0), \tau_0)$ is minimized uniquely at $\theta = \theta_0$. This follows because D is nonsingular and $m(\theta, \tau_0)$ has a unique zero at $\theta = \theta_0$ by S1(c).

Assumption S1 and Theorem 1 imply that assumption 2(a) holds. The first part of assumption 2(b) holds by assumption S2(d) and the fact that

$\frac{1}{\sqrt{T}} \Sigma_{-T_1}^{T_2}(Z_{it}'Z_{it} - EZ_{it}'Z_{it})$ satisfies a CLT for all i = 1, ..., n. The latter holds using assumptions S1(d), S2(d), and S2(f) and Herrndorf's (1984) Corollary 1. The second and third parts of assumption 2(b) hold by assumptions S2(b) and S1(c), respectively.

Assumption 2(c) follows from Herrndorf's (1984, Corollary 1) CLT using S1(d), S2(c), and S2(f). Assumption 2(d) follows directly from S1(a) and S2(a). Assumption 2(e) holds because $d(\cdot,\cdot)$ is a quadratic form.

Assumption 2(f) is established as follows: The differentiability of $m_t(\theta,\tau)$ holds by S2(f). $\{m_t(\theta,\tau)\}$ satisfies a uniform LLN using assumption S1 by the above proof that S1 $\Rightarrow$ 1. $\left\{\frac{\partial}{\partial\theta}m_t(\theta,\tau)\right\}$ and $\left\{\frac{\partial}{\partial\tau}m_t(\theta,\tau)\right\}$ satisfy uniform LLNs by Corollaries 1 and 2 of Andrews (1987b) since assumptions S1(a), S1(d), S2(f), and S2(f) imply assumptions A1, B1, B2, and A5 of Andrews (1987b), respectively. $m(\theta,\tau)$ and $M(\theta,\tau)$ exist by assumptions S1(c) and S2(e), respectively. $dm(\theta,\tau)$ exists and $dm(\theta_0, \tau_0) = 0$ because $E\frac{\partial}{\partial\theta}m_t(\theta_0, \tau) = 0$, $\forall t$, $\forall\tau$, by S2(b). $\left\{\sup_{(\theta,\tau)\in\Theta_c\times T} \frac{\partial^2}{\partial\theta_a\partial\theta}m_t(\theta,\tau)\right\}$ satisfy a weak LLN for all a = 1, ..., p by assumptions S1(d) and S2(f).

Assumption 2(g) follows immediately from S1(c) and S2(e). □

PROOF OF THEOREM 7: The proof is analogous to that of Theorem 6 using S1-S3 in place of R1-R3. □

FOOTNOTES

1. The authors thank Charles Coleman, Douglas Rivers, Quang Vuong, and Guofu Tan for helpful comments. The first author thanks the California Institute of Technology for its hospitality while part of this research was carried out and the Alfred P. Sloan Foundation and the National Science Foundation for research support provided through a Research Fellowship and grant No. SES-8419789 respectively.

2. The comparative advantage of these two books is their depth and detail, in which they dominate the present paper.

3. Gallant and White (1987, Ch. 2, pp. 11-12) accommodate multi-stage estimation procedures by elongating the parameter vector θ to include preliminary estimators. If both a preliminary estimator and the final estimator are asymptotically efficient, however, then their assumption PD (Ch. 5, p. 82), which requires the two estimators to have nonsingular asymptotic joint covariance matrix, is not satisfied. For example, this occurs with the 2SLS and 3SLS estimators in a simultaneous equations model when the errors are uncorrelated across equations. In consequence, their asymptotic distributional results for multi-stage estimators and test statistics do not apply in certain important contexts.

   In addition, when misspecification occurs, the estimator obtained by elongating the parameter vector does not necessarily equal the multi-stage estimator of interest.

4. As mentioned above, the nonlinear LS estimator and various M-estimators can be defined in two ways. The same is true of the ML estimator (see Section 5.) The choice between the two definitions depends on assumption 1(e). If the limit function $d(m(\theta, \tau_0), \tau_0)$ is minimized uniquely at $\theta = \theta_0$ when $m_t(\cdot,\cdot)$ and $d(\cdot,\cdot)$ are defined in terms of the first order conditions (i.e., the second definition given above for the LS and M-estimators), then this is the most convenient definition. The reason is that this definition must be used in any event to establish asymptotic normality by Theorem 2 below.

   On the other hand, the limiting first order conditions may have multiple solutions, even though the function $d(m(\theta, \tau_0), \tau_0)$ that corresponds to the underlying minimization problem (i.e., the function that corresponds to the first definition of $m_t(\cdot,\cdot)$ and $d(\cdot,\cdot)$ for the LS example) has a unique minimum at θ_0. In this case, we need to use the first definition of $m_t(\cdot,\cdot)$ and $d(\cdot,\cdot)$ to establish consistency of $\{\hat{\theta}\}$. Then, given consistency, we use the second definition to establish asymptotic normality. Since θ_0 is assumed to lie in the interior of Θ for the proof of asymptotic normality, a sequence of estimators defined using the first definition also solves equation (2.2) for the second definition with probability that goes to one as $T \to \infty$.

   The advantage of proceeding as above is that one need not treat the classes of least mean distance and method of moments estimators separately (as is done by BGS (1982) and Gallant (1987)). This results in considerable economy of presentation without sacrificing the generality of the consistency results.

5. The existence of the limits uniformly for $(\theta,\tau) \in \Theta_c \times T$ means that

$$\sup_{(\theta,\tau)\in\Theta_c\times T}\left|\frac{1}{T}\sum_{T_1}^{T_2} Em_t(\theta,\tau) - m(\theta,\tau)\right| \to 0 \quad \text{as} \quad T \to \infty$$

   and likewise for $M(\theta,\tau)$ and $dm(\theta,\tau)$.

6. If necessary, the nonsingularity of HVH' can be avoided by using asymptotic distributional results for quadratic forms with g-inverted weighting matrices and singular limiting weight matrix--see Andrews (1987a).

7. As defined, LR_T is unique except in the very rare case that M is proportional to the identity matrix. In this case, LR_T can be taken as either of the two expressions above.

8. Strong mixing is a condition of asymptotic weak dependence. A sequence of rv's $\{W_t\}$ is <u>strong mixing</u> if

$$\alpha(s) = \sup_t \inf_{A\in F_{-\infty}^t,\, B\in F_{t+s}^{\infty}} |P(A\cap B) - P(A)P(B)| \to 0 \quad \text{as} \quad s \to \infty ,$$

   where $F_{-\infty}^t$ denotes the smallest σ-field in F that is generated by the rv's $\{\ldots, W_{t-1}, W_t\}$ and likewise for F_{t+s}^{∞}.

9. Assumption R1 is weaker than the consistency assumptions of White and Domowitz (1984) in terms of the moment assumption placed on the errors. Also, it replaces the deceptively restrictive assumption of continuity of $f_t(X_t, \theta)$ in θ uniformly in t almost surely (a.s.) (see the discussion in Andrews (1987b)) by the smoothness condition R1(f). On the other hand, R1(b) assumes the existence of a certain limit, which is avoided in White and Domowitz (1984).

10. Assumption R2 is similar to assumptions in the literature. It requires $f_t(\theta)$ to be three times differentiable, however, rather than just twice differentiable, as often is assumed. This added smoothness constitutes the price one pays for treating LS estimators of nonlinear regression models, 2SLS and 3SLS estimators of nonlinear simultaneous equations models, ML estimators, and various other procedures as particular examples of a single general method of estimation, as is done here. Clearly, if one treats each estimation problem separately, weaker

conditions can be obtained. The conditions given here, however, are sufficiently weak to cover most nonlinear regression models encountered in practice.

11. For the linear simultaneous equations model, Hodoshima (1985) explores the consequences for estimation of these differing scenarios.

12. With these definitions, the function $m_t(\theta,\tau)$ actually depends on both t and T. This double subscripting does not affect the results of Theorems 1-5. All that is required is that the appropriate sequences of rv's are replaced by triangular arrays of rv's.

REFERENCES

Amemiya, T. (1977), "The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model," Econometrica, 45, 955-968.

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

Andrews, D. W. K. (1985), "A Nearly Independent, But Non-strong Mixing, Triangular Array," Journal of Applied Probability, 22, 729-731.

Andrews, D. W. K. (1987a), "Asymptotic Results for Generalized Wald Tests," Econometric Theory, 3, forthcoming.

Andrews, D. W. K. (1987b), "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," Econometrica, 55, forthcoming.

Barten, A. P. and Salvas Bronsard, L. (1970), "Two-Stage Least Squares Estimation with Shifts in the Structural Form," Econometrica, 38, 938-941.

Bates, C. and White, H. (1985), "A Unified Theory of Consistent Estimation for Parametric Models," Econometric Theory, 1, 151-178.

Bierens, H. J. (1984), "Model Specification Testing of Time Series Regressions," Journal of Econometrics, 26, 323-353.

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

Chow, G. C. (1960), "Tests of Equality between Sets of Coefficients in Two Linear Regressions," Econometrica, 28, 591-605.

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

Fair, R. C. (1970), "The Estimation of Simultaneous Equations Models with Lagged Endogenous Variables and First Order Serially Correlated Errors," Econometrica, 38, 507-516.

Gallant, A. R. (1987), Nonlinear Statistical Models (New York: Wiley).

Gallant, A. R. and Goebel, J. J. (1976), "Nonlinear Regression with Autocorrelated Errors," Journal of the American Statistical Association, 71, 961-967.

Gallant, A. R. and White, H. (1987), A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models, book manuscript, to be published.

Herrndorf, N. (1984), "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables," Annals of Probability, 12, 141-153.

Hodoshima, J. (1985), "Identification and Estimation in Linear Simultaneous Equations Models with Structural Change under Limited Information: Gains by Homoskedasticity," CORE Discussion Paper No. 8531.

Huber, P. J. (1967), "The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions," Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, I, 221-233 (Berkeley, CA: University of California Publications).

Huber, P. J. (1981), Robust Statistics (New York: Wiley).

Jennrich, R. I. (1969), "Asymptotic Properties of Non-Linear Least Squares Estimation," Annals of Mathematical Statistics, 40, 633-643.

Lo, A. W. and Newey, W. K. (1985), "A Large-Sample Chow Test for the Linear Simultaneous Equation," Economics Letters, 18, 351-353.

McLeish, D. L. (1975a), "A Maximal Inequality and Dependent Strong Laws," Annals of Probability, 3, 826-836.

McLeish, D. L. (1975b), "Invariance Principles for Dependent Variables," Z. Wahrscheinlichkeitstheorie und. Verw. Gebiete, 32, 165-178.

McLeish, D. L. (1977), "On the Invariance Principle for Nonstationary Mixingales," Annals of Probability, 5, 616-621.

Newey, W. K. and West, K. D. (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," Econometrica, 55, forthcoming.

Pollard, D. (1985), "New Ways to Prove Central Limit Theorems," Econometric Theory, 1, 295-313.

Potscher, B. M. and Prucha, I. R. (1986), "Consistency in Nonlinear Econometrics: A Generic Uniform Law of Large Numbers and Some Comments on Recent Results," Department of Economics Working Paper No. 86-9, University of Maryland.

Richard, J. F. (1980), "Models with Several Regimes and Changes in Exogeneity," Review of Economic Studies, 47, 1-20.

Serfling, R. J. (1980), Approximation Theorems of Mathematical Statistics (New York: Wiley).

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

White, H. (1984), Asymptotic Theory for Econometricians (New York: Academic Press).

White, H. and Domowitz, I. (1984), "Nonlinear Regression with Dependent Observations," Econometrica, 52, 143-161.

Withers, C. S. (1981), "Central Limit Theorems for Dependent Variables. I," Z. Wahrscheinlichkeitstheorie verw. Gebiete, 57, 509-534.