Review of Economic Studies (1988) LV, 615-640 0034-6527/88/00380615$02.00

Inference in Nonlinear Econometric Models with Structural Change

DONALD W. K. ANDREWS
and
RAY C. FAIR
Yale University

First version received May 1987; *Final version accepted May* 1988 (*Eds.*)

This paper extends the classical test for structural change in linear regression models (see Chow (1960)) 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.

1. INTRODUCTION

This paper is concerned with testing for structural change in nonlinear models. For the classical linear regression model the F-test discussed by Chow (1960) commonly is used, and for the linear simultaneous equations model the Lo and Newey (1985) or Hodoshima (1986) extensions of this test can be used. Somewhat surprisingly, however, more general cases have received little attention in the literature. An exception is the work of Anderson and Mizon (1983) on the nonlinear simultaneous equations model. In this paper we consider fairly wide classes of models, estimators, and test statistics. 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. Three examples are introduced in Section 2: (1) the single equation nonlinear regression model, (2) the nonlinear simultaneous equations model, and (3) any model estimated by maximum likelihood. General estimation and testing results that cover these examples and others are given in Section 3, with proofs in the Appendix. Section 4 contains a detailed treatment of the application of the general results to the non-linear simultaneous equations example.

The general results of Section 3 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 (1988).[2] In contrast to Gallant (1987, Chapters 3 and 7), least mean distance and method of moment estimators are treated simultaneously. Also in contrast to Gallant and White (1988), a more complete treatment of multi-step procedures is given.[3]

The approach taken in Section 3 is a variant of that of Gallant (1987, Chapter 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. INTRODUCTORY EXAMPLES

This section introduces three examples that are covered by the general results of Section 3. These examples are used in Section 3 to illustrate the way in which the general results can be applied to particular models and estimation procedures. The second example, the nonlinear simultaneous equations model, is considered in more detail in Section 4. See Andrews and Fair (1987) for more discussion of the first and third examples.

First, consider a nonlinear regression model with structural change:

$$\begin{aligned} Y_t &= f_t(X_t, \theta_1, \theta_3) + U_t \quad \text{for } t = -T_1, \ldots, -1, \\ Y_t &= f_t(X_t, \theta_2, \theta_3) + U_t \quad \text{for } t = 1, \ldots, T_2, \end{aligned} \tag{2.1}$$

where Y_t is a scalar dependent variable, X_t is a vector of regressor variables, U_t is a scalar error term, $f_t(\cdot, \cdot, \cdot)$ is a known function, and $\theta = (\theta_1', \theta_2', \theta_3')'$ is an unknown parameter vector. The errors may be heteroskedastic and/or autocorrelated, but must be uncorrelated with the regression function. The regressors X_t may include lagged values of Y_t. The time index is normalized such that structural change occurs at $t = 0$ if such change occurs.

The null hypothesis of no structural change is given by the simple restriction on θ that $\theta_1 = \theta_2$. We are interested in testing this restriction as well as testing joint null hypotheses of no structural change plus additional restrictions $h(\theta) = \mathbf{0}$. In the case of pure structural change, there is no subparameter θ_3 that is constant across periods, and so, θ_3 does not appear in (2.1) or in θ.

Most estimators of θ, such as the least squares (LS) estimator or M-estimators, are extremum estimators. Such estimators are defined as the solution to some minimization problem. The properties of such an estimator (such as consistency and asymptotic normality) can be determined from the properties of the optimand that defines the estimator. Test statistics can be formed using the restricted and unrestricted versions of the estimator and/or the restricted and unrestricted values of the optimand or its derivatives. The properties of the test statistics can also be determined from the properties of the optimand that defines the estimator. In consequence, general results can be obtained for estimation and testing by analyzing general optimization problems without specifying the models from which the optimization problem was obtained. To apply the general results to a particular problem, one links the particular model and estimation procedure with the general results via one's definition of the optimand.

The testing results of Section 3 cover three types of procedures: Wald, LM-like, and LR-like tests. These procedures apply whether or not estimation has been carried out by maximum likelihood. The Wald statistic is defined in the usual way. It is given by a quadratic form based on the difference between the unrestricted estimated value of the restrictions and their value under the null hypothesis. The LM statistic is a quadratic form based on the vector of derivatives with respect to θ of the optimand that defines the estimator, evaluated at the restricted estimator of θ. By suitable choice of a weight matrix for the quadratic form, the Wald and LM statistics have asymptotic chi-square distributions under the null with degrees of freedom given by the number of restrictions. For example, even if the errors are heterogeneous and autocorrelated in (2.1), a weight matrix can be chosen such that the test is valid asymptotically.

The LR-like statistic, on the other hand, only has the desired asymptotic chi-square null distribution under more restrictive conditions. For example, it does in the nonlinear regression case with estimation by LS if the errors are homoskedastic and uncorrelated. Although this condition can be restrictive, it can be circumvented in some cases by transforming a model with correlated errors into one with uncorrelated errors (e.g. see Fair (1970)).

In the case of testing for pure structural change, the LR-like statistic is particularly simple. For example, suppose the nonlinear regression model of (2.1) is estimated by LS. Then, the LR test statistic equals T_1+T_2 times the difference between the sum of squared residuals from the restricted and unrestricted LS regressions divided by the sum of squared residuals from the unrestricted regression. The unrestricted residuals are obtained by doing separate LS regressions on the data with $t<0$ and $t>0$; while the restricted residuals are obtained by doing a single LS regression on the whole data set (with θ_2 set equal to θ_1). The LR test statistic in this case is analogous to the classical F- statistic one obtains in the linear regression model when testing for structural change.

Next, consider a nonlinear dynamic simultaneous equations model with structural change:

$$\begin{aligned} f_{it}(Y_t, X_t, \theta_1, \theta_3) &= U_{it} \quad \text{for } i=1,\ldots,n,\ t=-T_1,\ldots,-1, \\ f_{it}(Y_t, X_t, \theta_2, \theta_3) &= U_{it} \quad \text{for } i=1,\ldots,n,\ t=1,\ldots,T_2, \end{aligned} \tag{2.2}$$

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, $f_{it}(\cdot,\cdot,\cdot,\cdot) \in R^1$ is a known function, $\theta=(\theta_1', \theta_2', \theta_3') \in \Theta \subset R^p$ is an unknown parameter, and n ($\geqq 1$) is the number of equations. The null hypothesis of no structural change is given by $\theta_1=\theta_2$. In the case of pure structural change, no subparameter θ_3 appears in (2.2) or in θ.

In Section 4 a class of nonlinear three stage least squares (3SLS) and two stage least squares (2SLS) estimators introduced by Amemiya (1977) is considered. These estimators are based on instrumental variables (IV). They are examples of extremum estimators. In consequence, their properties and those of the corresponding W, LM-like, and LR-like test statistics can be obtained from the general results of Section 3.

In this example, the conditions needed for the LR statistic to be valid include having each instrument z_t such that $z_t=\mathbf{0}$ for all $t<0$ or $z_t=\mathbf{0}$ for all $t>0$ and having error vectors U_t that are uncorrelated across time, homoskedastic for $t<0$ (i.e. $EU_tU_t'=\Omega_1$ $\forall t<0$), and homoskedastic for $t>0$ (i.e. $EU_tU_t'=\Omega_2\ \forall t>0$). In the case of testing for pure structural change, one simply estimates the restricted value of θ_1 $(=\theta_2)$ using the full data set and one estimates the unrestricted values of θ_1 and θ_2 from the $t<0$ and the $t>0$ data sets respectively. The LR test statistic is $2(T_1+T_2)$ times the value of the optimand based on the whole data set evaluated at the restricted estimator minus the sum of the two values of the optimand for the two sub-samples evaluated at the unrestricted estimators of θ_1 and θ_2 respectively.

As a third example, consider any regular finite dimensional parametric model that is estimated by ML. Such models include a wide variety of dynamic and nondynamic econometric models. (A model is "regular" if its score functions satisfy the conditions of Section 3.) The ML estimator is an extremum estimator whose properties can be determined from the general results of Section 3. The Wald, LM, and LR test statistics are all asymptotically valid in this context.

In the case of testing for pure structural change in an ML situation, the parameter vector θ is of the form $\theta = (\theta_1', \theta_2')'$, where the likelihood function for $t<0$ depends only on θ_1 and the likelihood function for $t>0$ depends only on θ_2. To calculate the LR statistic for testing $\theta_1 = \theta_2$, one needs to compute the restricted estimate of θ_1 $(=\theta_2)$ using the whole data set and then compute the unrestricted estimates of θ_1 and θ_2 using the $t<0$ data and the $t>0$ data respectively. The LR statistic, then, is simply $2(T_1+T_2)$ times the difference between the restricted log-likelihood function and the unrestricted log-likelihood function, where the latter is just the sum of the log-likelihood functions for $t<0$ and $t>0$ evaluated at the unrestricted estimates of θ_1 and θ_2 respectively.

3. GENERAL RESULTS

This section gives general results for estimation and testing in models with structural change. The basic approach 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 (1988). Our approach most closely follows that of BGS (1982) and Gallant (1987). Our notation is chosen to be as compatible as possible with theirs.

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 noncentral chi-square under local alternatives under certain conditions.

3.1. *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 $(\Omega, \mathcal{F}, P)$. 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

$$\pi_{1T} = T_1/T \to \pi_1 \in (0,1) \quad \text{and} \quad \pi_{2T} = T_2/T \to \pi_2 \in (0,1) \quad \text{as } T \to \infty. \tag{3.1}$$

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

$$d(\bar{m}_T(\hat{\theta}, \hat{\tau}), \hat{\tau}) = \inf_{\theta\in\Theta} d(\bar{m}_T(\theta, \hat{\tau}), \hat{\tau}) \tag{3.2}$$

with probability that goes to one as $T\to\infty$, where $\bar{m}_T(\theta, \tau) = 1/T \sum_{t=-T_1}^{T_2} m_t(\theta, \tau)$, $m_t(\theta, \tau) = m_t(W_t, \theta, \tau)$, and $m_t(\cdot, \cdot, \cdot): R^{k_t} \times \Theta \times \mathcal{T}_1 \to R^v$ where $\mathcal{T}_1 \subset R^u$, $m_0(\theta, \tau) \equiv 0$, $\hat{\tau}$ 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).

Note that $\hat{\tau}$ is a preliminary estimator used in the definition of $\hat{\theta}$.

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

In the case of *pure structural change*, 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 *partial structural change*, 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. Consider the nonlinear regression model of (2.1). 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. Correspondingly, for the consistency results for $\hat{\theta}$, we can take either $m_t(\theta, \tau) = (Y_t - f_t(X_t, \theta_j, \theta_3))^2$ and $d(m, \tau) = m$, where $j=1$ for $t<1$ and $j=2$ for $t>1$, or $m_t(\theta, \tau) = (Y_t - f_t(X_t, \theta_j, \theta_3))(\partial/\partial\theta) f_t(X_t, \theta_j, \theta_3)$ and $d(m, \tau) = m'm/2$, whichever is more convenient. For the asymptotic normality and testing results given below, the second definition must be used.

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 $\tilde{\rho}((Y_t - f_t(X_t, \theta_j, \theta_3))/\tau)$ or $\psi((Y_t - f_t(X_t, \theta_j, \theta_3))/\tau)(\partial/\partial\theta) f_t(X_t, \theta_j, \theta_3)$, where the nuisance parameter τ is a scale parameter, $\psi(x) = (d/dx)\tilde{\rho}(x)$, 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, aned 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 $Z_{1t} = \mathbf{0}$ for $t>0$ and $Z_{2t} = \mathbf{0}$ 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

$$D(\tau) = \left(\lim_{T\to\infty} \frac{1}{T} \sum_{-T_1}^{T_2} E Z_t Z_t'\right)^{-1} \quad \text{and} \quad D(\hat{\tau}) = \left(\frac{1}{T} \sum_{-T_1}^{T_2} Z_t Z_t'\right)^{-1}.$$

The definitions of $m_t(\theta, \tau)$ and $d(\cdot, \cdot)$ for the 3SLS estimator are given in Section 4.

For the case of ML estimation of a parametric model, let $f_t(\theta)$ denote the conditional density of the endogenous variables Y_t conditional on the preceding endogenous variables $\{Y_s: s<t\}$ and the exogeneous variables $\{X_s: \forall s\}$. Then, the ML estimator is an extremum estimator with

$$\begin{aligned} &m_t(\theta, \tau) = -\log f_t(\theta) \quad \text{and} \quad d(m, \tau) = m \quad \text{for} \quad m \in R^1 \quad \text{or} \\ &m_t(\theta, \tau) = -\frac{\partial}{\partial\theta} \log f_t(\theta) \quad \text{and} \quad d(m, \tau) = m'm/2 \quad \text{for } m \in R^p. \end{aligned} \tag{3.3}$$

Either set of definitions can be used for establishing consistency of $\hat{\theta}$. The second set must be used for establishing asymptotic normality of $\hat{\theta}$ and obtaining testing results.

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 rv's $\{\hat{\theta}\}$. Such results have content only if such a sequence exists. Clearly, sequences $\{\hat{\theta}\}$ that satisfy (3.2), 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 $\hat{\tau} \to^p \tau_0$ as $T \to \infty$ for some $\tau_0 \in \mathcal{T}_1 \subset R^u$.

(c) There exists a Borel measurable function $m(\cdot, \cdot): \Theta \times \mathcal{T} \to R^v$ such that $\bar{m}_T(\theta, \tau) \to^p m(\theta, \tau)$ uniformly over $(\theta, \tau) \in \Theta \times \mathcal{T}$ as $T \to \infty$, where $\mathcal{T} \subset \mathcal{T}_1$ is some compact neighbourhood of τ_0.

(d) $d(m(\theta, \tau), \tau)$ is continuous in (θ, τ) at all $(\theta, \tau) \in \Theta \times \mathcal{T}$.

(e) $d(m(\theta, \tau_0), \tau_0)$ is uniquely minimized over $\theta \in \Theta$ at θ_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 the application of a weak law of large numbers (WLLN) in some cases (e.g. see Andrews (1988) or McLeish (1975)) 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 $\lim_{T\to\infty} 1/T \sum_{-T_1}^{T_2} Em_t(\theta, \tau)$. Thus, Assumption 1(c) holds if these limits exist and if $\{1/T_1 \sum_{-T_1}^{-1} m_t(\theta, \tau)\}$ and $\{1/T_2 \sum_1^{T_2} m_t(\theta, \tau)\}$ satisfy uniform WLLNs over $\Theta \times \mathcal{T}$. 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, Chapter 7, Theorem 1), Potscher and Prucha (1987), or Bierens (1984, Lemma 2). Assumption 1(d) holds in most applications. 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[4].

Theorem 1. *Under Assumption* 1, *every sequence of extremum estimators* $\{\hat{\theta}\}$ *satisfies* $\hat{\theta} \to^p \theta_0$ *as* $T \to \infty$ *under P.*

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

3.2. *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

$$S = \lim_{T\to\infty} \mathrm{Var}_P(\sqrt{T}\,\bar{m}_T(\theta_0, \tau_0)),$$
$$M = \lim_{T\to\infty} 1/T \textstyle\sum_{-T_1}^{T_2} E(\partial/\partial\theta')m_t(\theta_0, \tau_0),$$
$$D = (\partial^2/\partial m \partial m')d(m(\theta_0, \tau_0), \tau_0),$$
$$\mathcal{J} = M'DM, \; \mathcal{I} = M'DSDM, \text{ and}$$
$$V = \mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1},$$

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 3.1.

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

Assumption 2. (a) $\hat{\theta} \to^p \theta_0 \in R^p$ as $T \to \infty$.

(b) i. $\sqrt{T}(\hat{\tau}-\tau_0) = O_P(1)$ as $T\to\infty$ for some $\tau_0 \in \mathscr{T}_1$,
ii. $(\partial/\partial m)\, d(E\bar{m}_T(\theta_0,\tau_0),\tau_0) = \mathbf{0}\ \forall T$ large, and iii. $(\partial^2/\partial\tau\partial m')\, d(m(\theta_0),\tau_0) = \mathbf{0}$.

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

(d) $\Theta \subset R^p$ and Θ contains a convex neighbourhood Θ_c of θ_0.

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

(f) $m_t(\theta,\tau)$ is once and twice continuously differentiable in τ and θ, respectively, on $\Theta_c\times\mathscr{T}$, $\forall t$, $\forall\omega\in\Omega$. $\{m_t(\theta,\tau)\}$, $(\partial/\partial\theta)m_t(\theta,\tau)\}$, $\{(\partial/\partial\tau)\, m_t(\theta,\tau)\}$, and

$$\left\{\sup_{(\theta^*,\tau^*)\in\Theta\times\mathscr{T}, a=1,\ldots,p}\left\|\frac{\partial^2}{\partial\theta_a\partial\theta'}\, m_t(\theta^*,\tau^*)\right\|\right\}$$

are sequences of $F\backslash$Borel-measurable rv's that satisfy uniform WLLNs over $(\theta,\tau)\in\Theta_c\times\mathscr{T}$. The expectations of the sample averages of the latter sequence are uniformly bounded for $T \geqq 1$.

$$m(\theta,\tau) = \lim_{T\to\infty}\frac{1}{T}\textstyle\sum_{-T_1}^{T_2} Em_t(\theta,\tau),$$

$$M(\theta,\tau) = \lim_{T\to\infty}\frac{1}{T}\textstyle\sum_{-T_1}^{T_2} E\frac{\partial}{\partial\theta'}\, m_t(\theta,\tau),$$

and

$$dm(\theta,\tau) = \lim_{T\to\infty}\frac{1}{T}\textstyle\sum_{-T_1}^{T_2} E\frac{\partial}{\partial\tau}\, m_t(\theta,\tau)$$

exist uniformly for $(\theta,\tau)\in\Theta_c\times\mathscr{T}^5$ and are continuous and $dm(\theta_0,\tau_0)=\mathbf{0}$.

(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 $Em_t(\theta_0,\tau_0)=\mathbf{0}$, $\forall t$, and that $\{m_t(\theta_0,\tau_0)\}$ satisfy standard moment conditions and a condition of asymptotically weak temporal dependence see Gallant (1987, Chapter 7, Theorem 2), McLeish (1977, Theorem 2.4 and Corollary 2.11), Herrndorf (1984, Theorem and Corollaries 1-4), or Withers (1981, Theorems 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 behaviour 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,

$$\sqrt{T}(\hat{\theta}-\theta_0) \rightarrow^d N(\mathbf{0}, V) \quad \text{as } T \rightarrow \infty.$$

Next we consider estimation of the covariance matrix V. Let

$$\hat{M}=\frac{1}{T}\sum_{-T_1}^{T_2}\frac{\partial}{\partial\theta'}m_t(\hat{\theta},\hat{\tau}), \qquad \hat{D}=\frac{\partial^2}{\partial m\partial m'}d(\bar{m}_T(\hat{\theta},\hat{\tau}),\hat{\tau}) \quad \text{and} \quad \hat{\mathscr{J}}=\hat{M}'\hat{D}\hat{M}.$$

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 $\hat{S}=1/T\sum_{-T_1}^{T_2} m_t(\hat{\theta},\hat{\tau})m_t(\hat{\theta},\hat{\tau})'$. 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 Andrews (1987c) and Gallant (1987, pp. 551, 556). Let $\hat{S}=\hat{S}(\hat{\theta})$, where

$$\begin{aligned}
\hat{S}(\theta) &= \pi_{1T}\hat{S}_1(\theta)+\pi_{2T}\hat{S}_2(\theta),\\
\hat{S}_1(\theta) &= \frac{1}{T_1}\sum_{-T_1}^{-1} m_t(\theta)m_t(\theta)'+\sum_{v=1}^{T_1} k\left(\frac{v}{l(T_1)}\right)\frac{1}{T_1}\sum_{-T_1+v}^{-1}[m_t(\theta)m_{t-v}(\theta)'\\
&\quad + m_{t-v}(\theta)m_t(\theta)'],\\
\hat{S}_2(\theta) &= \frac{1}{T_2}\sum_{1}^{T_2} m_t(\theta)m_t(\theta)'+\sum_{v=1}^{T_2} k\left(\frac{v}{l(T_2)}\right)\frac{1}{T_2}\sum_{1+v}^{T_2}[m_t(\theta)m_{t-v}(\theta)'+m_{t-v}(\theta)m_t(\theta)'],
\end{aligned} \tag{3.4}$$

$m_t(\theta)=m_t(\theta,\hat{\tau})$, $l(T_j)$ is a "bandwidth" parameter that satisfies $l(T_j)\rightarrow\infty$ and $l(T_j)=o(T_j)$ as $T_j\rightarrow\infty$ for $j=1,2$, and $k(\cdot)$ is the quadratic spectral (QS) kernel, i.e.

$$k(x)=\frac{25}{12\pi^2x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5}-\cos(6\pi x/5)\right)$$

or $k(\cdot)$ is the Parzen kernel, i.e.

$$k(x)=\begin{cases}1-6x^2+6x^3 & \text{for } 0\leqq x\leqq 1/2\\ 2(1-x)^3 & \text{for } 1/2\leqq x\leqq 1.\\ 0 & \text{for } x\geqq 1\end{cases}$$

See Andrews (1987c) for a detailed analysis of the choice of bandwidth parameter $l(T_j)$ for $j=1, 2$.

Conditions under which this estimator is consistent can be found in the references above or in Newey and West (1987).[6] 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. Given the availability of such conditions, it is straightforward to verify the following assumption.

Assumption 3. $\hat{S}\rightarrow^p S$ as $T\rightarrow\infty$ (where S is as in Assumption 2).

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

Theorem 3. *Under Assumptions* 2 *and* 3, $\hat{M} \to^p M$, $\hat{D} \to^p D$, *and* $\hat{V} \to^p V$ *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.

3.3. *Tests of hypotheses concerning structural change*

We now consider tests of null hypotheses of the form H_0: $h(\theta) = \mathbf{0}$. 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 $h^*(\theta_1) = \mathbf{0}$ when $\theta = (\theta_1', \theta_2')'$ or H_0: $\theta_1 = \theta_2$ and $h^*(\theta_1, \theta_3) = \mathbf{0}$ 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, Chapter 7) and Gallant and White (1988, Chapter 7).

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

Assumption 4. (a) $h(\theta)$ is continuously differentiable in a neighbourhood of θ_0 and $H = (\partial/\partial\theta')h(\theta_0)$ has full rank $r(\leq p)$.

(b) V is nonsingular.

The Wald statistic is defined as

$$\mathrm{W}_T = Th(\hat{\theta})'(\hat{H}\hat{V}\hat{H}')^- h(\hat{\theta}), \tag{3.5}$$

where $\hat{H} = (\partial/\partial\theta')\, h(\hat{\theta})$. Since $\hat{H}\hat{V}\hat{H}' \to^p HVH'$ 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 at $T \to \infty$.[7]

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

$$\mathrm{W}_T = T(\hat{\theta}_1 - \hat{\theta}_2)'(\hat{V}_1/\pi_{1T} + \hat{V}_2/\pi_{2T})^-(\hat{\theta}_1 - \hat{\theta}_2), \tag{3.6}$$

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

$$d(\bar{m}_T(\tilde{\theta}), \hat{\tau}) = \inf\{d(\bar{m}_T(\theta), \hat{\tau}): \theta \in \Theta, h(\theta) = \mathbf{0}\} \tag{3.7}$$

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 $\Theta_0 = \{\theta \in \Theta: h(\theta) = \mathbf{0}\}$, 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 $\tilde{\theta} \to^p \theta_0$ as $T \to \infty$ under the null hypothesis. In consequence, the following assumption is straightforward to verify:

Assumption 5. $\tilde{\theta} \to^p \theta_0$ 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

$$\tilde{M} = 1/T \sum_{-T_1}^{T_2} (\partial/\partial\theta') m_t(\tilde{\theta}, \hat{\tau}),$$

$$\tilde{D} = (\partial^2/\partial m \partial m') d(\bar{m}_T(\tilde{\theta}, \hat{\tau}), \hat{\tau}), \ \tilde{\mathcal{J}} = \tilde{M}'\tilde{D}\tilde{M}, \ \hat{S} = \hat{S}(\tilde{\theta}),$$

and $\tilde{H} = (\partial/\partial\theta')h(\tilde{\theta})$ (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 the notation, we do not need to adjust Assumptions 2(b) or 3 when a restricted estimator of τ_0 is used. Let $\tilde{\mathcal{I}} = \tilde{M}'\tilde{D}\hat{S}\tilde{D}\tilde{M}$ and $\tilde{V} = \tilde{\mathcal{J}}^- \tilde{\mathcal{I}} \tilde{\mathcal{J}}^-$. As above with $\hat{V}$, the estimator $\tilde{V}$ can be simplified when V simplifies, as often occurs in applications of interest.

The LM-like statistic is defined as

$$\mathrm{LM}_T = T \frac{\partial}{\partial\theta'} d(\bar{m}_T(\tilde{\theta}), \hat{\tau}) \tilde{\mathcal{J}}^- \tilde{H}'(\tilde{H}\tilde{V}\tilde{H}')^- \tilde{H} \tilde{\mathcal{J}}^- \frac{\partial}{\partial\theta} d(\bar{m}_T(\tilde{\theta}), \hat{\tau}) \tag{3.8}$$

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

The LR-like 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, $\mathcal{I} = b\mathcal{J}$ for some scalar constant $b \neq 0$ and $\hat{b} \to^p b$ as $T \to \infty$ for some sequence of non-zero rv's $\{\hat{b}\}$ (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

$$m_t(W_t, \theta, \tau) = (\partial/\partial\theta)\rho_t(W_t, \theta, \tau), \forall t.$$

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

$$\bar{\rho}_T(\tilde{\theta}, \hat{\tau}) = \inf\{\bar{\rho}_T(\theta, \hat{\tau}): \theta \in \Theta, h(\theta) = \mathbf{0}\},$$

where

$$\bar{\rho}_T(\theta, \hat{\tau}) = 1/T \sum_{-T_1}^{T_2} \rho_t(W_t, \theta, \hat{\tau}).$$

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

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 when the errors are uncorrelated and homoskedastic.

Note that Assumption 6b(i) is compatible with the definitions of $\hat{\theta}$, because an estimator $\hat{\theta}$ that minimizes $\bar{\rho}_T(\theta, \hat{\tau})$ is in the interior of Θ with probability that goes to one as $T \to \infty$ under Assumption 2, and hence, also minimizes $d(\bar{m}_T(\theta), \hat{\tau})$ with probability that goes to one as $T \to \infty$. Also note that in the definition of LR_T below, when Assumption 6b holds, $\tilde{\theta}$ is as defined in Assumption 6b(i) rather than as in (3.7).

The LR-like statistic is defined as

$$\mathrm{LR}_T = \begin{cases} 2T(d(\bar{m}_T(\tilde{\theta}), \hat{\tau}) - d(\bar{m}_T(\hat{\theta}), \hat{\tau}))/\hat{b} & \text{when 6a holds} \\ 2T(\bar{\rho}_T(\tilde{\theta}, \hat{\tau}) - \bar{\rho}_T(\hat{\theta}, \hat{\tau}))/\hat{c} & \text{when 6b holds.} \end{cases} \tag{3.9}$$

where $m_t(\cdot, \cdot)$ and $d(\cdot, \cdot)$ are as in Assumption 2.[8] 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 $h(\theta) = \mathbf{0}$, respectively.

Theorem 4. *Suppose Assumptions* 2–4 *hold under the null hypothesis,* $h(\theta_0) = \mathbf{0}$, *and the null hypothesis is true. Then the following results hold*:

(a) $\mathrm{W}_T \to^d \chi_r^2$ *as* $T \to \infty$, *where r is the number of restrictions,*
(b) $\mathrm{LM}_T \to^d \chi_r^2$ *as* $T \to \infty$ *provided Assumption* 5 *also holds, and*
(c) $\mathrm{LR}_T \to^d \chi_r^2$ *as* $T \to \infty$ *provided Assumption* 5 *holds and either Assumption* 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 (i.i.d.) errors (see Section 4 below), then we usually have $\hat{\mathcal{I}} = \hat{b}\hat{\mathcal{J}}$ for some scalar rv $\hat{b} \neq 0$. In the latter case, $\hat{V}$ and W_T simplify. We get $\hat{V} = \hat{b}\hat{\mathcal{J}}^-$ and $\mathrm{W}_T = Th(\hat{\theta})'(\hat{H}\hat{\mathcal{J}}^-\hat{H}')^- h(\hat{\theta})/\hat{b}$.

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

2. When Assumption 6b(i) holds, both W_T and LM_T simplify. In this case,

$$D = I_p, \qquad M = \lim_{T\to\infty} \frac{1}{T} \sum_{-T_1}^{T_2} E \frac{\partial^2}{\partial\theta\partial\theta'} \rho_t(W_t, \theta, \tau), \qquad \mathcal{J} = M^2, \qquad \mathcal{I} = MSM,$$

$(\partial/\partial\theta)d(\bar{m}_T(\tilde{\theta}), \hat{\tau}) = \tilde{M}\bar{m}_T(\tilde{\theta})$, and by 2(g), M is nonsingular. We get $\mathrm{W}_T \doteq Th(\hat{\theta})'(\hat{H}\hat{M}^-\hat{S}\hat{M}^-\hat{H}')^- h(\hat{\theta})$ and $\mathrm{LM}_T \doteq T\bar{m}_T(\tilde{\theta})'\tilde{M}^-\tilde{H}'(\tilde{H}\tilde{M}^-\tilde{S}\tilde{M}^-\tilde{H}')^-\tilde{H}\tilde{M}^-\bar{m}_T(\tilde{\theta})$.

If, in addition, $\hat{S} = \hat{c}\hat{M}$ or $\hat{S} = \hat{c}\tilde{M}$ for some scalar rv's $\hat{c} \neq 0$ (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 $LM_T \doteq T\bar{m}_T(\tilde{\theta})'\tilde{M}^{-}\bar{m}_T(\tilde{\theta})/\hat{c}$, 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 divisors 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 $(\Omega, \mathscr{F})$ 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 of $\{(Y_{Tt}, X_t, U_t)$: $t = \ldots, -1, 1, \ldots\}$ for $T = 1, 2, \ldots$.

Verificiation 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. $\tilde{\theta} \to^p \theta_0$ under $\{P_T\}$ as $T \to \infty$.

Assumption 10a. Assumption 6a holds and $\hat{b} \to^p b$ under $\{P_T\}$ as $T \to \infty$.

Assumption 10b. Assumption 6b holds and $\hat{c} \to^p c$ 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) $W_T \to^d \chi_r^2(\delta^2)$, *where* $\delta^2 = \eta' H'(HVH')^{-1}H\eta$,

(b) $LM_T \to^d \chi_r^2(\delta^2)$ *provided Assumption* 9 *also holds, and*

(c) $LR_T \to^d \chi_r^2(\delta^2)$ *provided Assumption* 9 *holds and either Assumption* 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 alteratives in Theorem 5, the approximations described in Comment 1 usually are more accurate for close alternatives to the null hypothesis than for distant alternatives.

4. NONLINEAR SIMULTANEOUS EQUATIONS

In this section we consider structural change in the nonlinear simultaneous equations model of (2.2). Let $\theta_0 \in \Theta$ denote the true parameter vector in this model.

4.1. *Three stage least squares estimation*

We consider Amemiya's (1977) class of 3SLS estimators generalized to the structural change problem considered here. A special case of the 3SLS estimator is the 2SLS estimator.

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

$$f_1(\theta) = (f_{1,-T_1}(\theta), \ldots, f_{1,-1}(\theta), f_{2,-T_1}(\theta), \ldots, f_{n,-1}(\theta))'_{nT_1\times 1}. \tag{4.1}$$

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 $i = 1, \ldots, n$, 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

$$Z_1 = \operatorname{diag}\{Z_1^1, \ldots, Z_1^n\}_{nT_1\times v}, \quad \text{where } v = \textstyle\sum_{i=1}^n v_i. \tag{4.2}$$

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 $n \times n$ nuisance parameter estimators. Either $\hat{\Omega}_1$ and $\hat{\Omega}_2$ are estimators of $\Omega_1 = \lim_{T_1\to\infty} 1/T_1 \sum_{-T_1}^{-1} EU_tU_t'$ and $\Omega_2 = \lim_{T_2\to\infty} 1/T_2 \sum_1^{T_2} EU_tU_t'$, 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 $\Omega_1 = \Omega_2 = \lim_{T\to\infty} 1/T \sum_{-T_1}^{T_2} EU_tU_t'$. 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.

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 *3SLS estimators* of θ_0 for $T = 1, 2, \ldots$ is defined to be any sequence of rv's $\{\hat{\theta}\}$ such that $\hat{\theta}$ minimizes

$$(f_1(\theta)'\hat{\Lambda}_1^- Z_1 + f_2(\theta)'\hat{\Lambda}_2^- Z_2)(Z_1'\hat{\Lambda}_1^- Z_1 + Z_2'\hat{\Lambda}_2^- Z_2)^-(Z_1'\hat{\Lambda}_1^- f_1(\theta) + Z_2'\hat{\Lambda}_2^- f_2(\theta)) \tag{4.3}$$

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 (4.3) 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 (4.3) 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 3 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

$$\hat{D} = T(Z_1'\hat{\Lambda}_1^- Z_1 + Z_2'\hat{\Lambda}_2^- Z_2)^- = \left(\frac{1}{T}\sum_{-T_1}^{T_2} Z_t'\hat{\Omega}_j^- Z_t\right)^-_{v\times v} \tag{4.4}$$

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 $\mathscr{F}\backslash$Borel-measurable.

Assumption S1. (a) Θ is a compact subset of R^p.

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

(c) $\pi_1 = \lim_{T\to\infty}\pi_{1T}$, $\lim_{T_1\to\infty} 1/T_1 \sum_{-T_1}^{-1} Ef_{it}(\theta)Z_{rt}$, and $\lim_{T_2\to\infty} 1/T_2\sum_1^{T_2} Ef_{it}(\theta)Z_{rt}$ exist uniformly for $\theta\in\Theta$ and are continuous in θ for all $\theta\in\Theta$ for $i, r = 1, \ldots, n$. $\lim_{T\to\infty} 1/T\sum_{-T_1}^{T_2} EZ_t'\Omega_j^{-1}f_t(\theta) = \mathbf{0}$ if and only if $\theta=\theta_0$. $D = \lim_{T\to\infty}(1/T\sum_{-T_1}^{T_2} EZ_t'\Omega_j^{-1}Z_t)^{-1}_{v\times v}$ exists and is positive definite.

(d) $\{(Y_t, X_t, Z_t)\}$ is strong mixing.[9]

(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 $\xi > 1$.

(f) $f_{it}(\theta)$ is 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 $\overline{\lim}_{T\to\infty} 1/T\sum_{-T_1}^{T_2} E\sup_{\theta\in\Theta^*}\|(\partial/\partial\theta)f_{it}(\theta)Z_{rt}'\| < \infty$, $\forall i$, $r = 1, \ldots, n$.

The strong mixing Assumption S1(d) is used to ensure that an 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 3) and use the results of Section 3 to establish consistency and asymptotic normality of $\hat{\theta}$.

Nuisance parameter estimators $\hat{\Omega}_1$ and $\hat{\Omega}_2$ that satisfy Assumption S1(b) can be obtained as follows. Let $\bar{\theta}$ 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

$$\hat{\Omega}_1 = \frac{1}{T_1}\sum_{-T_1}^{-1} f_t(\bar{\theta})f_t(\bar{\theta})' \quad\text{and}\quad \hat{\Omega}_2 = \frac{1}{T_2}\sum_1^{T_2} f_t(\bar{\theta})f_t(\bar{\theta})'. \tag{4.5}$$

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

$$\hat{\Omega}_1 = \hat{\Omega}_2 = \frac{1}{T}\sum_{-T_1}^{T_2} f_t(\bar{\theta})f_t(\bar{\theta})'. \tag{4.6}$$

Next, we introduce an Assumption S2 such that Assumptions S1 and S2 imply Assumption 2 of Section 3 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 $N(\mathbf{0}, V)$ distribution as $T\to\infty$, where $V = (M'DM)^{-1}M'DSDM(M'DM)^{-1}$,

$$M = \lim_{T\to\infty}\frac{1}{T}\sum_{-T_1}^{T_2} EZ_t'\Omega_j^{-1}\frac{\partial}{\partial\theta'}f_t(\theta_0)_{v\times p}, \tag{4.7}$$

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

Assumption S2. (a) Θ contains a convex compact neighbourhood Θ_c of θ_0.
(b) $EU_{it}Z_{rt}=0$, $\forall i$, $r=1,\ldots,n$.
(c) $S=\lim_{T\to\infty}\operatorname{Var}(1/\sqrt{T}\sum_{-T_1}^{T_2}Z_t'\Omega_j^{-1}U_t)$ 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 $\lim_{T\to\infty}\operatorname{Var}(1/\sqrt{T}\sum_{-T_1}^{T_2}Z_{it}'Z_{it})$ exists for all $i=1,\ldots,n$.
(e) $\lim_{T_1\to\infty}1/T_1\sum_{-T_1}^{-1}E(\partial/\partial\theta)f_{it}(\theta)Z_{rt}'$ and $\lim_{T_2\to\infty}1/T_2\sum_1^{T_2}E(\partial/\partial\theta)f_{it}(\theta)Z_{rt}'$ exist uniformly for $\theta\in\Theta$ and are continuous for $\theta\in\Theta$, $\forall i$, $r=1,\ldots,n$, and M is full column rank.
(f) The strong mixing numbers $\{\alpha(s)\}$ of $\{(Y_t,X_t,Z_t)\}$ satisfy $\alpha(s)=o(s^{-\alpha/(\alpha-1)})$ as $s\to\infty$ for some $\alpha>1$.
(g) $f_{it}(\theta)$ is twice differentiable in θ, $\forall\theta\in\Theta_c$, $\forall i=1,\ldots,n$, $\forall t$, for all realizations of $\{Y_t,X_t\}$, S1(e) holds for some $\xi>\alpha$, and

$$\sup_t E\sup_{\theta\in\Theta_c}\left(\left\|\frac{\partial}{\partial\theta}f_{it}(\theta)Z_{rt}\right\|^{\zeta}+\left\|\frac{\partial^2}{\partial\theta_a\partial\theta}f_{it}(\theta)Z_{rt}'\right\|^{\zeta}+\|U_{it}Z_{rt}\|^{2\xi}+(Z_{rt}'Z_{rt})^{2\xi}\right)$$
$<\infty$
$\forall i$, $r=1,\ldots,n$, $\forall a=1,\ldots,p$, for some $\xi>\alpha$ and some $\zeta>1$.

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

$$E(U_tU_t'|Z_t)=\Omega_j \quad \text{a.s.,} \quad \forall t,$$

and (4.8)

$$EZ_t'U_tU_{t-k}'Z_{t-k}=\mathbf{0}, \quad \forall t, \quad \forall k=1,2,\ldots.$$

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 $\hat{S}(\theta)$ defined in equation (3.4) with $m_t(\cdot,\cdot)$ defined just above equation (4.4), with $l(T_j)$ such that $l(T_j)\to\infty$ and $l(T_j)=o(T^{1/4})$ as $T_j\to\infty$, and with $k(\cdot)$ corresponding to the QS or Parzen kernel.[10] (See Andrews (1987c) for results regarding the optimal choice of the bandwidth parameters $l(T_j)$ and the kernel $k(\cdot)$.) If the second condition of equation (4.8) holds, then $\hat{S}$ can be simplified by taking $l(T_1)=l(T_2)=0$ in its definition. This yields

$$\hat{S}=\frac{1}{T}\sum_{-T_1}^{T_2}Z_t'\hat{\Omega}_j^-f_t(\hat{\theta})f_t(\hat{\theta})'\hat{\Omega}_j^-Z_t. \tag{4.9}$$

If both of the conditions of (4.8) hold, then take

$$\hat{S}=\hat{D}^- \quad\text{and}\quad \hat{V}=(\hat{M}'\hat{D}\hat{M})^-. \tag{4.10}$$

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

Assumption S3. $\sup_t E\|Z_t'U_t\|^{4\xi}<\infty$ for some $\xi>\alpha$. $\zeta\geqq 2$ in S2(g).

Theorem 6. (a) *Under Assumptions S1–S3, $\hat{S}\to^p S$, $\hat{M}\to^p M$, and $\hat{V}\to^p V$ as $T\to\infty$ for $\hat{S}$ as defined just below equation* (4.8).

(b) *Under Assumptions S1 and S2, $\hat{S}\to^p S$, $\hat{M}\to^p M$, and $\hat{V}\to^p V$ as $T\to\infty$ for $\hat{S}$ as defined in* (4.9) *or* (4.10), *provided the additional conditions outlined above* (4.9) *or* (4.10) *are satisfied, respectively.*

4.2. *Tests of structural change*

We now consider tests of nonlinear restrictions H_0: $h(\theta)=\mathbf{0}$.

A sequence of *restricted 3SLS* estimators of θ_0 is any sequence of rv's $\{\tilde{\theta}\}$ such that $\tilde{\theta}$ minimizes equation (4.3) over $\theta\in\Theta_0=\{\theta\in\Theta: h(\theta)=\mathbf{0}\}$. Assumptions S1 and S5 (below) guarantee the existence and consistency of sequences of restricted 3SLS estimators, since they imply that Assumption 1 of Section 3 holds with parameter space Θ_0.

Assumption S5. Θ_0 is compact.

The LM test statistic of equation (3.8) 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 (4.8). 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 (4.9) and (4.10), $\tilde{S}$ can be replaced by the simpler estimator

$$\tilde{S}=\frac{1}{T}\sum_{-T_1}^{T_2} Z_t'\hat{\Omega}_j^-f_t(\tilde{\theta})f_t(\tilde{\theta})'\hat{\Omega}_j^-Z_t \quad \text{or} \quad \tilde{S}=\hat{D}^- \tag{4.11}$$

when the conditions outlined above (4.9) or (4.10), 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 6 and Assumption S5.

The following Assumption S6a implies Assumption 6a of Section 3. It is used to obtain the asymptotic null distribution of the LR statistic.

Assumption S6a. Under the null hypothesis,

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

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

Assumption S6a implies that $S=D^{-1}$ and $\mathscr{I}=\mathscr{J}$. S6a holds under (4.8).

Assumptions S1–S3, 4, S5, and S6a for the 3SLS estimator imply Assumptions 1–5 and 6a of Section 3. Thus, Theorem 4 holds and the W, LM, and LR statistics of equations (3.5), (3.8), and (3.9) 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 $\{(Y_{Tt},X_t,Z_t): -T_1\leqq t\leqq T_2, T=1,2,\ldots\}$ in S1(d) and S2(f), and with $\sup_t$ replaced by $\sup_{t\leqq T, T=1,2,\ldots}$ in S1(e) and S2(g).

Assumptions 7, 8, 9, and 10a (with $\hat{b}=1$) of Section 3 are implied by Assumptions S7, S3 and S7, S5 and S7, and S6a, respectively. Thus, Theorem 5 of Section 3 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 (3.5). If Assumption S6a holds, then $\hat{S}$ can be taken as in equation (4.10), $\hat{S}=\hat{D}^{-}$, $\hat{\mathcal{I}}=\hat{\mathcal{J}}$, and W_T is given by the simplified formulae of Comment 1 to Theorem 4 with $\hat{\mathcal{J}}=\hat{M}'\hat{D}\hat{M}$ and $\hat{b}=1$:

$$\mathrm{W}_T = Th(\hat{\theta})'[\hat{H}(\hat{M}'\hat{D}\hat{M})^{-}\hat{H}']^{-}h(\hat{\theta}). \tag{4.12}$$

When testing for pure structural change, we assume that the IV's 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) with blocks $\hat{D}_1$ and $\hat{D}_2$, say. It also ensures that $m_t(\hat{\theta},\hat{\tau})$ has elements corresponding to $\hat{D}_1$ that are non-zero only if $t<0$ and elements corresponding to $\hat{D}_2$ that are non-zero only if $t>0$. Hence, the Wald statistic for testing pure structural change is given by (3.6):

$$\mathrm{W}_T = T(\hat{\theta}_1-\hat{\theta}_2)'(\hat{V}_1/\pi_{1T}+\hat{V}_2/\pi_{2T})^{-}(\hat{\theta}_1-\hat{\theta}_2), \tag{4.13}$$

where $\hat{V}_j$ is analogous to $\hat{V}$ but is based on the j-th sub-sample of the data for $j=1,2$. When Assumption S6a holds, $\hat{V}_1$ and $\hat{V}_2$ of (4.13) can be simplified as in (4.9) or (4.10).

The LM statistic corresponding to 3SLS estimation is given by

$$\mathrm{LM}_T = T\bar{m}_T(\tilde{\theta})'\hat{D}\tilde{M}\tilde{\mathcal{J}}^{-}\tilde{H}'(\tilde{H}\tilde{V}\tilde{H}')^{-}\tilde{H}\tilde{\mathcal{J}}^{-}\tilde{M}'\hat{D}\bar{m}_T(\tilde{\theta}), \tag{4.14}$$

where $\tilde{\mathcal{J}}=\tilde{M}'\hat{D}\tilde{M}$. 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

$$\begin{aligned}\mathrm{LM}_T = T(&\bar{m}_{1T}(\tilde{\theta})'\hat{D}_1\tilde{M}_1\tilde{\mathcal{J}}_1^{-}-\bar{m}_{2T}(\tilde{\theta})'\hat{D}_2\tilde{M}_2\tilde{\mathcal{J}}_2^{-})[\tilde{V}_1/\pi_{1T}+\tilde{V}_2/\pi_{2T}]^{-}\\ &\cdot(\tilde{\mathcal{J}}_1^{-}\tilde{M}_1'\hat{D}_1\bar{m}_{1T}(\tilde{\theta})-\tilde{\mathcal{J}}_2\tilde{M}_2'\hat{D}_2\bar{m}_{2T}(\tilde{\theta})),\end{aligned} \tag{4.15}$$

where

$$\bar{m}_{jT}(\tilde{\theta})=\frac{1}{T_j}Z_j'\hat{\Lambda}_j^{-}f_j(\tilde{\theta}),\qquad \tilde{M}_j=\frac{1}{T_j}Z_j'\hat{\Lambda}_j^{-}\frac{\partial}{\partial\theta'}f_j(\tilde{\theta}),$$

$$\tilde{\mathcal{J}}_j=\tilde{M}_j'\hat{D}_j\tilde{M}_j,\ \tilde{V}_j=\tilde{\mathcal{J}}_j^{-}\tilde{\mathcal{I}}_j\tilde{\mathcal{J}}_j^{-},\qquad \tilde{\mathcal{I}}_j=\tilde{M}_j'\hat{D}_j\tilde{S}_j\hat{D}_j\tilde{M}_j,$$

and $\tilde{S}_j=\hat{S}_j(\tilde{\theta})$ for $\hat{S}_j(\theta)$ defined in equation (3.4) for $j=1,2$.

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

$$\mathrm{LM}_T \doteq T\bar{m}_T(\tilde{\theta})'\hat{D}\tilde{M}\tilde{\mathcal{J}}^{-}\tilde{M}'\hat{D}\bar{m}_T(\tilde{\theta}). \tag{4.16}$$

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

$$\mathrm{LM}_T \doteq T_1\bar{m}_{1T}(\tilde{\theta})'\hat{D}_1\tilde{M}_1\tilde{\mathcal{J}}_1^{-}\tilde{M}_1'\hat{D}_1\bar{m}_{1T}(\tilde{\theta})+T_2\bar{m}_{2T}(\tilde{\theta})'\hat{D}_2\tilde{M}_2\tilde{\mathcal{J}}_2^{-}\tilde{M}_2'\hat{D}_2\bar{m}_{2T}(\tilde{\theta}). \tag{4.17}$$

The LR statistic in the 3SLS case is given by

$$\mathrm{LR}_T = 2T(d(\bar{m}_T(\tilde{\theta}),\tilde{\tau})-d(\bar{m}_T(\hat{\theta}),\hat{\tau})), \tag{4.18}$$

where $d(\bar{m}_T(\theta), \hat{\tau})$ is the expression given in (4.3), 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:

$$d(\bar{m}_T(\theta), \hat{\tau}) = d_1(\bar{m}_{1T}(\theta), \hat{\tau}) + d_2(\bar{m}_{2T}(\theta), \hat{\tau}), \tag{4.19}$$

where $d_j(\bar{m}_{jT}(\theta), \hat{\tau}) = f_j(\theta)'\hat{\Lambda}_j^- Z_j(Z_j'\hat{\Lambda}_j^- Z_j)^- Z_j'\hat{\Lambda}_j^- f_j(\theta)$ for $j = 1, 2$.

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 *and* 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 (4.5) and (4.6). With these definitions, Assumption S6a only requires the errors to be homoskedastic and uncorrelated conditional on the IVs. 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 (4.5) and (4.6) (i.e. with $\bar{\theta} = \hat{\theta}$) for the purposes of generating the W, LM, and LR test statistics.

APPENDIX

Proof of Theorem 1. Let $d(\theta, \tau)$ and $d_T(\theta, \tau)$ abbreviate $d(m(\theta, \tau), \tau)$ and $d(\bar{m}_T(\theta, \tau), \tau)$ respectively. Let Θ_0 be any open neighbourhood of θ_0. Then, $\Theta_1 = \Theta - \Theta_0$ is compact, using Assumption 1(a). We show below that there exists a constant $\delta > 0$ and a compact neighbourhood $\mathcal{T}_\delta$ of τ_0 such that

$$\min_{\theta \in \Theta_1, \tau \in \mathcal{T}_\delta} d(\theta, \tau) - d(\theta_0, \tau_0) \geqq \delta > 0. \tag{A.1}$$

We also show that

$$d(\hat{\theta}, \hat{\tau}) \to^p d(\theta_0, \tau_0) \quad \text{as } T \to \infty. \tag{A.2}$$

Combining (A.1), (A.2), and Assumption 1(b) gives

$$P(\hat{\theta} \in \Theta_0) \geqq P(d(\hat{\theta}, \hat{\tau}) - d(\theta_0, \tau_0) < \delta, \hat{\tau} \in \mathcal{T}_\delta) \to 1 \tag{A.3}$$

as $T \to \infty$, which is the desired result.

First, we establish (A.1). By the compactness of Θ_1 and Assumption 1(e),

$$\delta \equiv [\min_{\theta \in \Theta_1} d(\theta, \tau_0) - d(\theta_0, \tau_0)]/2 \tag{A.4}$$

exists and is positive. By the uniform continuity of $d(\theta, \tau)$ on $\Theta \times \mathcal{T}$ (Assumption 1(a) and 1(d)), given $\delta > 0$ these exists a compact neighbourhood $\mathcal{T}_\delta (\subset \mathcal{T})$ of τ_0 such that for all $\tau \in \mathcal{T}_\delta$

$$|d(\theta, \tau) - d(\theta, \tau_0)| < \delta \quad \forall \theta \in \Theta. \tag{A.5}$$

Let (θ^*, τ^*) be some element of $\Theta_1 \times \mathcal{T}_\delta$ such that $d(\theta^*, \tau^*) = \min_{\theta \in \Theta_1, \tau \in \mathcal{T}_\delta} d(\theta, \tau)$. Using (A.4), (A.5), and Assumption 1(e), we now have

$$\begin{aligned} \min_{\theta \in \Theta_1, \tau \in \mathcal{T}_\delta} d(\theta, \tau) - d(\theta_0, \tau_0) &= [d(\theta^*, \tau^*) - d(\theta^*, \tau_0)] + [d(\theta^*, \tau_0) - d(\theta_0, \tau_0)] \\ &\geqq -\delta + \min_{\theta \in \Theta_1} d(\theta, \tau_0) - d(\theta_0, \tau_0) = \delta \end{aligned} \tag{A.6}$$

and (A.1) is established.

To show (A.2), we write

$$\begin{aligned}
&d(\hat{\theta}, \hat{\tau}) - d(\theta_0, \tau_0) \\
&\quad = [d(\hat{\theta}, \hat{\tau}) - d_T(\hat{\theta}, \hat{\tau})] + [d_T(\hat{\theta}, \hat{\tau}) - d(\theta_0, \hat{\tau})] + [d(\theta_0, \hat{\tau}) - d(\theta_0, \tau_0)] \\
&\quad \leqq [d(\hat{\theta}, \hat{\tau}) - d_T(\hat{\theta}, \hat{\tau})] + [d_T(\theta_0, \hat{\tau}) - d(\theta_0, \hat{\tau})] + [d(\theta_0, \hat{\tau}) - d(\theta_0, \tau_0)] \\
&\quad \leqq 2 \sup_{\theta \in \Theta, \tau \in \mathcal{T}} |d_T(\theta, \tau) - d(\theta, \tau)| + [d(\theta_0, \hat{\tau}) - d(\theta_0, \tau_0)] \\
&\quad \to^p 0 \quad \text{as } T \to \infty,
\end{aligned} \tag{A.7}$$

where the second inequality holds with probability that goes to one as $T \to \infty$ since $\hat{\tau} \to^p \tau_0 \in \mathcal{T}$ and the convergence to zero uses Assumptions 1(b), (c), and (d).

In addition, we have

$$\begin{aligned}
d(\hat{\theta}, \hat{\tau}) - d(\theta_0, \tau_0) &= [d(\hat{\theta}, \hat{\tau}) - d(\hat{\theta}, \tau_0)] + [d(\hat{\theta}, \tau_0) - d(\theta_0, \tau_0)] \\
&\geqq d(\hat{\theta}, \hat{\tau}) - d(\hat{\theta}, \tau_0) \\
&\to^p 0 \quad \text{as } T \to \infty,
\end{aligned} \tag{A.8}$$

where the inequality uses Assumption 1(e), and the convergence to zero uses the fact that Assumptions 1(a), (b), and (d) imply that $\sup_{\theta \in \Theta} |d(\theta, \hat{\tau}) - d(\theta, \tau_0)| \to^p 0$ as $T \to \infty$. Equations (A.7) and (A.8) give (A.2) and the proof is complete. ‖

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 $\sqrt{T}(\partial/\partial\theta) d(\bar{m}_T(\hat{\theta}), \hat{\tau})$ about θ_0 give: $\forall a = 1, \ldots, p$,

$$o_p(1) = \sqrt{T} \frac{\partial}{\partial \theta_a} d(\bar{m}_T(\hat{\theta}), \hat{\tau}) = \sqrt{T} \frac{\partial}{\partial \theta_a} d(\bar{m}_T(\theta_0), \hat{\tau}) + \frac{\partial^2}{\partial \theta_a \partial \theta'} d(\bar{m}_T(\theta^*), \hat{\tau}) \sqrt{T}(\hat{\theta} - \theta_0), \tag{A.9}$$

where θ^* is a rv on the line segment joining $\hat{\theta}$ and θ_0, and hence, $\theta^* \to^p \theta_0$. (See Jennrich (1969) Lemma 3 for the mean value theorem for random functions.) The first equality holds because $\hat{\theta}$ minimizes $d(\bar{m}_T(\theta), \hat{\tau})$ 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

$$\frac{\partial^2}{\partial \theta_a \, \partial \theta'} d(\bar{m}_T(\theta^*), \hat{\tau}) = \frac{\partial^2}{\partial \theta_a \partial \theta'} d(m(\theta_0), \tau_0) + o_p(1), \tag{A.10}$$

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

$$\sqrt{T} \frac{\partial}{\partial \theta} d(\bar{m}_T(\theta_0), \hat{\tau}) \to^d N(\mathbf{0}, M'DSDM) \quad \text{as } T \to \infty. \tag{A.11}$$

These results, equation (A.9), and the nonsingularity of $M'DM$ give

$$\sqrt{T}(\hat{\theta} - \theta_0) = -(M'DM)^{-1} \frac{\partial}{\partial \theta} d(\bar{m}_T(\theta_0), \hat{\tau}) + o_p(1) \to^d N(\mathbf{0}, V) \quad \text{as } T \to \infty. \tag{A.12}$$

To show (A.10), we proceed as follows:

$$\begin{aligned}
\frac{\partial^2}{\partial \theta_a \partial \theta_l} d(\bar{m}_T(\theta^*), \hat{\tau}) &= \frac{\partial^2}{\partial \theta_a \partial \theta_l} \bar{m}_T(\theta^*)' \frac{\partial}{\partial m} d(\bar{m}_T(\theta^*), \hat{\tau}) \\
&\quad + \frac{\partial}{\partial \theta_a} \bar{m}_T(\theta^*)' \frac{\partial^2}{\partial m \partial m'} d(\bar{m}_T(\theta^*), \hat{\tau}) \frac{\partial}{\partial \theta_l} \bar{m}_T(\theta^*).
\end{aligned} \tag{A.13}$$

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

$$\begin{aligned}
\|\bar{m}_T(\theta^*) - m(\theta_0)\| &\leq \|\bar{m}_T(\theta^*, \hat{\tau}) - E\bar{m}_T(\theta, \tau)|_{\theta = \theta^*, \tau = \hat{\tau}}\| \\
&\quad + \|E\bar{m}_T(\theta, \tau)|_{\theta = \theta^*, \tau = \hat{\tau}} - m(\theta^*, \hat{\tau})\| + \|m(\theta^*, \hat{\tau}) - m(\theta_0, \tau_0)\| \to^p 0
\end{aligned} \tag{A.14}$$

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

$$\frac{\partial}{\partial m} d(\bar{m}_T(\theta^*), \hat{\tau}) \to^p \frac{\partial}{\partial m} d(m(\theta_0), \tau_0) = \mathbf{0} \quad \text{as } T \to \infty, \tag{A.15}$$

where the equality holds by 2(b), (e), and (f). Using Assumption 2(f), it is straightforward to show that $(\partial^2/\partial\theta_a\partial\theta_l)\bar{m}_T(\theta^*) = O_p(1)$ as $T \to \infty$. This result and (A.15) imply that the first term of (A.13) is $o_p(1)$ as $T \to \infty$.

Similarly, the continuity of $(\partial^2/\partial m\partial m')\, d(m, \tau)$ over $\mathcal{M}\times\mathcal{T}$ (Assumption 2(e)), equation (A.14), $\hat{\tau}\to^p \tau_0$, and the continuous mapping theorem gives

$$\frac{\partial^2}{\partial m\partial m'} d(\bar{m}_T(\theta^*), \hat{\tau}) \to^p \frac{\partial^2}{\partial m\partial m'} d(m(\theta_0), \tau_0) \equiv D \quad \text{as } T\to\infty. \tag{A.16}$$

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

$$\left\|\frac{\partial}{\partial\theta'}\bar{m}_T(\theta^*) - M\right\| \leq \left\|\frac{\partial}{\partial\theta'}\bar{m}_T(\theta^*) - M(\theta^*, \hat{\tau})\right\| + \|M(\theta^*, \hat{\tau}) - M(\theta_0, \tau_0)\| \to^p 0 \tag{A.17}$$

as $T\to\infty$. Equations (A.16) and (A.17) imply that the second term of (A.13) equals $[M'DM]_{al} + o_p(1)$, and hence, (A.10) is established.

To establish equation (A.11), we write

$$\sqrt{T}\frac{\partial}{\partial\theta} d(\bar{m}_T(\theta_0), \hat{\tau}) = \sqrt{T}\frac{\partial}{\partial\theta'}\bar{m}_T(\theta_0)\frac{\partial}{\partial m} d(\bar{m}_T(\theta_0), \hat{\tau}) = M'\sqrt{T}\frac{\partial}{\partial m} d(\bar{m}_T(\theta_0), \hat{\tau}) + o_p(1) \tag{A.18}$$

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

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

$$\sqrt{T}\frac{\partial}{\partial m_a} d(\bar{m}_T(\theta_0, \hat{\tau}), \hat{\tau}) = \sqrt{T}\frac{\partial}{\partial m_a} d(E\bar{m}_T(\theta_0, \tau_0), \tau_0)$$
$$+\frac{\partial^2}{\partial m'\partial m_a} d(m^*, \tau^*)\sqrt{T}(\bar{m}_T(\theta_0, \hat{\tau}) - E\bar{m}_T(\theta_0, \tau_0)) + \frac{\partial^2}{\partial\tau'\partial m_a} d(m^*, \tau^*)\sqrt{T}(\hat{\tau}-\tau_0) \tag{A.19}$$

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, $m^*\to^p m(\theta_0)$ and $\tau^*\to^p\tau_0$ as $T\to\infty$. (More precisely, (A.19) holds with probability that goes to one as $T\to\infty$.)

The first term of the right-hand side of (A.19) is zero for T large by Assumption 2(b). Also, since $\sqrt{T}(\hat{\tau}-\tau_0) = O_p(1)$ (Assumption 2(b)) and $(\partial^2/\partial\tau'\partial m_a)d(m, \tau)$ is continuous over $\mathcal{M}\times\mathcal{T}$ (Assumption 2(e)), we have:

$$\frac{\partial^2}{\partial\tau'\partial m_a} d(m^*, \tau^*)\sqrt{T}(\hat{\tau}-\tau_0) = \frac{\partial^2}{\partial\tau'\partial m_a} d(m(\theta_0), \tau_0)\sqrt{T}(\hat{\tau}-\tau_0) + o_p(1) = o_p(1), \tag{A.20}$$

where the second equality follows from 2(b). Similarly, using Assumption 2(e), $(\partial^2/\partial m_a\partial m')d(m^*, \tau^*) = [D]'_a + o_p(1)$ 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.19) yield

$$\sqrt{T}\frac{\partial}{\partial m} d(\bar{m}_T(\theta_0, \hat{\tau}), \hat{\tau}) = D\sqrt{T}(\bar{m}_T(\theta_0, \hat{\tau}) - E\bar{m}_T(\theta_0, \tau_0)) + o_p(1). \tag{A.21}$$

The proof is complete once we show that

$$\sqrt{T}(\bar{m}_T(\theta_0, \hat{\tau}) - E\bar{m}_T(\theta_0, \tau_0)) \to^d N(\mathbf{0}, S) \quad \text{as } T\to\infty, \tag{A.22}$$

since this implies that (A.21) and (A.18) hold, which establishes (A.11).

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

$$\sqrt{T}(\bar{m}_{Ta}(\theta_0, \hat{\tau}) - E\bar{m}_{Ta}(\theta_0, \tau_0))$$
$$= \sqrt{T}(\bar{m}_{Ta}(\theta_0, \tau_0) - E\bar{m}_{Ta}(\theta_0, \tau_0)) + \frac{\partial}{\partial\tau'}\bar{m}_{Ta}(\theta_0, \tau^*)\sqrt{T}(\hat{\tau}-\tau_0)$$
$$= \sqrt{T}(\bar{m}_{Ta}(\theta_0, \tau_0) - E\bar{m}_{Ta}(\theta_0, \tau_0)) + o_p(1), \tag{A.23}$$

where τ^* lies on the line segment joining $\hat{\tau}$ and τ_0, using Assumption 2(b), since $(\partial/\partial\tau)\bar{m}_T(\theta_0, \tau^*) \to^p dm(\theta_0, \tau_0) = \mathbf{0}$ by Assumption 2(f). Stacking equation (A.23) for $a = 1, \ldots, p$ and using Assumption 2(c) gives (A.22). ‖

Proof of Theorem 3. $\hat{M}\to^p M$ and $\hat{D}\to^p D$ as $T\to\infty$ by the arguments used in equations (A.16) and (A.17), respectively. Thus, $\hat{\mathcal{J}}\to^p\mathcal{J}$ and $\hat{\mathcal{J}}^-\to^p\mathcal{J}^{-1}$, since $\mathcal{J}$ is nonsingular (Assumption 2(g)). ‖

Proof of Theorem 4. To prove part (a), the delta method gives

$$\sqrt{T}(h(\hat{\theta}) - h(\theta_0)) \to^d N(\mathbf{0}, HVH') \quad \text{as } T\to\infty \tag{A.24}$$

using Assumption 4. By Theorem 3, $\hat{V}\to^p V$ and by the continuous mapping theorem and Assumption 2(a), $\hat{H}\to^p H$ as $T\to\infty$. Since HVH' is nonsingular, this implies that $(\hat{H}\hat{V}\hat{H}')^-\to^p (HVH')^{-1}$ as $T\to\infty$. This result, (A.24), and the continuous mapping theorem give the desired result.

Next we establish part (b). Standard arguments gives

$$\tilde{\mathscr{J}} \to^p \mathscr{J}, \quad \tilde{H} \to^p H, \quad \text{and} \quad \tilde{V} \to^p V \quad \text{as} \quad T \to \infty. \tag{A.25}$$

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

$$\sqrt{T}\frac{\partial}{\partial\theta_a} d(\bar{m}_T(\tilde{\theta}), \hat{\tau}) = \sqrt{T}\frac{\partial}{\partial\theta_a} d(\bar{m}_T(\theta_0), \hat{\tau}) + \frac{\partial^2}{\partial\theta_a\partial\theta'} d(\bar{m}_T(\dot{\theta}), \hat{\tau})\sqrt{T}(\tilde{\theta} - \theta_0), \tag{A.26}$$

$$\sqrt{T}h_a(\tilde{\theta}) = \sqrt{T}h_a(\theta_0) + \frac{\partial}{\partial\theta'} h_a(\theta^*)\sqrt{T}(\tilde{\theta} - \theta_0), \tag{A.27}$$

where $\dot{\theta}$ and θ^* lie on the line segment joining $\tilde{\theta}$ and θ_0, and hence, satisfy $\dot{\theta} \to^p \theta_0$ and $\theta^* \to^p \theta_0$ as $T \to \infty$. We stack equations (A.26) and (A.27) for $a = 1, \ldots, p$ and write them as

$$\sqrt{T}\frac{\partial}{\partial\theta} d(\bar{m}_T(\tilde{\theta}), \hat{\tau}) = \sqrt{T}\frac{\partial}{\partial\theta} d(\bar{m}_T(\theta_0), \hat{\tau}) + \dot{\mathscr{J}}\sqrt{T}(\tilde{\theta} - \theta_0) \tag{A.28}$$

and

$$\mathbf{0} = H^*\sqrt{T}(\tilde{\theta} - \theta_0) \tag{A.29}$$

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

By equation (A.11), $\sqrt{T}(\partial/\partial\theta)d(\bar{m}_T(\theta_0), \hat{\tau}) \to^d N(\mathbf{0}, \mathscr{I})$ as $T \to \infty$. By equation (A.10), $\dot{\mathscr{J}} \to^p \mathscr{J}$ as $T \to \infty$. Hence, using the nonsingularity of $\mathscr{J}$, we get $\dot{\mathscr{J}}^-\dot{\mathscr{J}} \doteq \mathscr{I}_p$, where $\doteq$ is defined in Comment 1 of Theorem 4. By Assumptions 4 and 5, $H^* \to^p H$ as $T \to \infty$. Pre-multiplication of (A.28) by $H^*\dot{\mathscr{J}}^-$ now gives

$$\begin{aligned} H^*\dot{\mathscr{J}}^-\sqrt{T}(\partial/\partial\theta)d(\bar{m}_T(\tilde{\theta}), \hat{\tau}) &= H^*\dot{\mathscr{J}}^-\sqrt{T}\frac{\partial}{\partial\theta} d(\bar{m}_T(\theta_0), \hat{\tau}) + H^*\dot{\mathscr{J}}^-\dot{\mathscr{J}}\sqrt{T}(\tilde{\theta} - \theta_0) \\ &\doteq H^*\dot{\mathscr{J}}^-\sqrt{T}\frac{\partial}{\partial\theta} d(\bar{m}_T(\theta_0), \hat{\tau}) \to^d N(\mathbf{0}, H\mathscr{J}^{-1}\mathscr{I}\mathscr{J}^{-1}H') \quad \text{as } T \to \infty. \end{aligned} \tag{A.30}$$

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

$$\frac{\partial}{\partial\theta} d(\bar{m}_T(\tilde{\theta}), \hat{\tau}) + \tilde{H}'\tilde{\lambda} = \mathbf{0}, \tag{A.31}$$

where $\tilde{H} = (\partial/\partial\theta')h(\tilde{\theta})$. Equations (A.30) and (A.31) combine to give

$$-H^*\dot{\mathscr{J}}^-\tilde{H}'\sqrt{T}\tilde{\lambda} \doteq H^*\dot{\mathscr{J}}^-\sqrt{T}\frac{\partial}{\partial\theta} d(\bar{m}_T(\tilde{\theta}), \hat{\tau}) = O_p(1). \tag{A.32}$$

Since $H^*\dot{\mathscr{J}}^-\tilde{H}' \to^p H\mathscr{J}^{-1}H'$ and $H\mathscr{J}^{-1}H'$ is nonsingular, equations (A.32) and (A.31) imply that $\sqrt{T}\tilde{\lambda} = O_p(1)$ and

$$\sqrt{T}\frac{\partial}{\partial\theta} d(\bar{m}_T(\tilde{\theta}), \hat{\tau}) = O_p(1). \tag{A.33}$$

Equations (A.25), (A.30), and (A.33) yield

$$\tilde{H}\tilde{\mathscr{J}}\sqrt{T}\frac{\partial}{\partial\theta} d(\bar{m}_T(\tilde{\theta}), \hat{\tau}) \to^d N(\mathbf{0}, HVH') \quad \text{as } T \to \infty. \tag{A.34}$$

The desired result now follows from equations (A.25) and (A.34) 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

$$\begin{aligned} \mathrm{LR}_T &= 2T(d(\bar{m}_T(\tilde{\theta}), \hat{\tau}) - d(\bar{m}_T(\hat{\theta}), \hat{\tau}))/\hat{b} = 2T\frac{\partial}{\partial\theta'} d(\bar{m}_T(\hat{\theta}), \hat{\tau})(\tilde{\theta} - \hat{\theta})/\hat{b} \\ &\quad + T(\tilde{\theta} - \hat{\theta})'\frac{\partial^2}{\partial\theta\partial\theta'} d(\bar{m}_T(\theta^*), \hat{\tau})(\tilde{\theta} - \hat{\theta})/\hat{b} \\ &\doteq T(\tilde{\theta} - \hat{\theta})'\mathscr{J}^*(\tilde{\theta} - \hat{\theta})/\hat{b}, \end{aligned} \tag{A.35}$$

where θ^* lies on the line segment joining $\tilde{\theta}$ and $\hat{\theta}$, and hence, $\theta^* \to^p \theta_0$ as $T \to \infty$, $\mathscr{J}^*$ 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

$$\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\hat{\tau})=\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\hat{\theta}),\hat{\tau})+\mathring{\mathcal{J}}\sqrt{T}(\tilde{\theta}-\hat{\theta})$$
$$\doteq\mathring{\mathcal{J}}\sqrt{T}(\tilde{\theta}-\hat{\theta}) \tag{A.36}$$

for a matrix $\mathring{\mathcal{J}}$ that satisfies $\mathring{\mathcal{J}}\to^p\mathcal{J}$ as $T\to\infty$. Pre-multiplying (A.36) by $\mathcal{J}^*\mathring{\mathcal{J}}^-$ and substituting the result in (A.35) gives

$$\begin{aligned}\mathrm{LR}_T&\doteq T\frac{\partial}{\partial\theta'}d(\bar{m}_T(\tilde{\theta}),\hat{\tau})(\mathcal{J}^*\mathring{\mathcal{J}}^-)'(\mathcal{J}^*)^-\mathcal{J}^*\mathring{\mathcal{J}}^-\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\hat{\tau})/\hat{b}\\&=T\frac{\partial}{\partial\theta'}d(\bar{m}_T(\tilde{\theta}),\hat{\tau})\tilde{\mathcal{J}}^-\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\hat{\tau})/\hat{b}+o_p(1),\end{aligned} \tag{A.37}$$

because $\mathring{\mathcal{J}}^-\mathring{\mathcal{J}}\doteq I_p$, $\sqrt{T}(\partial/\partial\theta)d(\bar{m}_T(\tilde{\theta}),\hat{\tau})=O_p(1)$, $\mathcal{J}^*\mathring{\mathcal{J}}^-\to^p I_p$, and $\tilde{\mathcal{J}}^-\mathcal{J}^*\to^p\mathbf{0}$ as $T\to\infty$, by (A.10), (A.25), and (A.33).

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

$$\mathrm{LM}_T=T\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\hat{\tau})'\tilde{\mathcal{J}}^-\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\hat{\tau})/\hat{b}+o_p(1)=\mathrm{LR}_T+o_p(1) \tag{A.38}$$

using $(\partial/\partial\theta)d(\bar{m}_T(\tilde{\theta}),\hat{\tau})\doteq-\tilde{H}'\tilde{\lambda}$, 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 Assumption 6a. ||

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

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

$$\sqrt{T}h(\hat{\theta})=\sqrt{T}h(\theta_T)+H^*\sqrt{T}(\hat{\theta}-\theta_T) \tag{A.39}$$

for an $r\times p$ matrix H^* that satisfies $H^*\to^p H$ 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

$$\sqrt{T}(\hat{\theta}-\theta_T)\to^d N(\mathbf{0},V)\quad\text{under }\{P_T\}\text{ as }T\to\infty. \tag{A.40}$$

This follows using Assumption 7 by the proof of Theorem 2 with θ_0 replaced by θ_T in all equations but (A.10), (A.13)-(A.17), and (A.20).

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.26)-(A.28) and equations (A.29), (A.30), and (A.34) are replaced by

$$\mathbf{0}=\sqrt{T}h(\theta_T+H^*\sqrt{T}(\tilde{\theta}-\theta_T), \tag{A.41}$$

$$\begin{aligned}H^*\dot{\mathcal{J}}^-\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\hat{\tau})&\doteq H^*\dot{\mathcal{J}}^-\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\theta_T),\hat{\tau})+\sqrt{T}h(\theta_T)\\&\to^d N(H\eta,HVH')\quad\text{as }T\to\infty,\end{aligned} \tag{A.42}$$

and

$$\tilde{H}\tilde{\mathcal{J}}^-\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}),\hat{\tau})\to^d N(H\eta),HVH')\quad\text{under }\{P_T\}\text{ as }T\to\infty, \tag{A.43}$$

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 $S1\Rightarrow 1$ *and* $S1$ *plus* $S2\Rightarrow 2$. 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 (4.4).

Assumption 1(a) is implied by S1(a). Assumption 1(b) follows from S1(b), the fact that $\{Z_t Z_t': t = -1, \ldots, -T_1\}$ and $\{Z_t Z_t': t = 1, \ldots, T_2\}$ satisfy weak LLNs as $T_1 \to \infty$ and $T_2 \to \infty$, respectively (which follows from Andrews (1988) Theorem 1 and Example 4 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 $(\theta, \tau) \in \Theta \times \mathscr{T}$. 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, $r = 1, \ldots, n$. The latter follow using the Theorem and Corollary 2 of Andrews (1987b), since Assumptions S1(a), S1(d) and (e), and S1(a) and (f) imply Assumptions A1, A2, and A5 of Andrews (1987b), respectively, where A2 is verified using Theorem 1 and Example 4 of Andrews (1988). Assumption S1(c) guarantees that the function $m(\theta, \tau) = \lim_{T\to\infty} 1/T \sum_{-T_1}^{T_2} Em_t(\theta, \tau)$ exists uniformly for $(\theta, \tau) \in \Theta \times \mathscr{T}$.

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

Assumption 1(e) holds 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 $1/\sqrt{T} \sum_{-T_1}^{T_2} (Z_{it}' Z_{it} - EZ_{it}' Z_{it})$ satisfies a CLT for all $i = 1, \ldots, n$. The latter holds using Assumptions S1(d), S2(d), S2(f), and S2(g) and Herrndorf's (1984) Corollary 1 or Withers' (1981) Theorem 2.1A and equations (6.1)–(6.3). The second and third parts of Assumptions 2(b) hold by Assumptions S2(b) and S1(c), respectively.

Assumption 2(c) follows from Herrndorf's (1984, Corollary 1) or Withers' (1981, Theorem 2.1A) CLT using S1(d), S2(c), S2(f), and S2(g). 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(g). $\{m_t(\theta, \tau)\}$ satisfies a uniform LLN using Assumption S1 by the above proof that $S1 \Rightarrow 1$. $\{(\partial/\partial\theta) m_t(\theta, \tau)\}$ and $\{(\partial/\partial\tau) m_t(\theta, \tau)\}$ satisfy uniform LLNs by the Theorem and Corollary 2 of Andrews (1987b) since Assumptions S1(a), S1(d) and S2(g), and S2(g) imply Assumptions A1, A2, and A5 of Andrews (1987b), respectively, where A2 is verified using Theorem 1 and Example 4 of Andrews (1988). $m(\theta, \tau)$ 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 $E(\partial/\partial\theta) m_t(\theta_0, \tau) = 0$, $\forall t, \forall \tau$, by S2(b). $\{\sup_{(\theta,\tau)\in\Theta_c \times \mathscr{T}} (\partial^2/\partial\theta_a \partial\theta) m_t(\theta, \tau)\}$ satisfies a weak LLN for all $a = 1, \ldots, p$ by assumptions S1(d) and S2(g).

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

Proof of Theorem 6. If $\hat{S} \to^p S$ as $T \to \infty$, then $\hat{M} \to^p M$ and $\hat{V} \to^p V$ as $T \to \infty$ in parts (a) and (b) of Theorem 6 by Theorem 3, since Assumptions S1 and S2 imply Assumption 2 (as shown immediately above). In part (b), the proof of $\hat{S} \to^p S$ is analogous to that of $\hat{M} \to^p M$.

It remains to show $\hat{S} \to^p S$ 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 S2(g), S2(g) and S3, and S2(b) and the asymptotic normality of $\sqrt{T}(\hat{\theta} - \theta_0)$, respectively. Their assumption (iii) is stronger than our assumption S2(f). Their proof still works with the weaker assumption S2(f), 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. The fact that our observations are indexed by a doubly infinite sequence only requires a slight alteration of their proof. ‖

NOTES

1. The authors thank Charles Coleman, Douglas Rivers, Quang Vuong, Guofu Tan, and two referees 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. These books do, however, have much more detail than the present paper.

3. Gallant and White (1988, Chapter 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 (Chapter 5, p. 81), 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, theiry 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, various M-estimators, and ML estimators can be defined in two ways. 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, M-, and ML 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 (3.2) 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 $(\theta, \tau) \in \Theta_c \times \mathscr{T}$ means that

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

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

6. In addition to the conditions given in these references, one needs the limiting covariance of $1/\sqrt{T} \sum_t m_t(\theta_0, \tau_0)$ between the two samples to be zero, i.e.

$$\lim_{T \to \infty} E\left(\frac{1}{\sqrt{T}} \sum_{-T_1}^{1} m_t(\theta_0, \tau_0)\right)\left(\frac{1}{\sqrt{T}} \sum_{u=1}^{T_2} m_u(\theta_0, \tau_0)'\right) = 0.$$

This follows under standard conditions of asymptotic weak dependence. For example, if $\{m_t(\theta_0, \tau_0)\}$ is strong mixing with mixing numbers $\{\alpha(s)\}$ that satisfy $\alpha(s) = O(s^{-q})$ as $s \to \infty$ for some $q > 1$, then this condition holds.

7. 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 matrix—see Andrews (1987*a*).

8. 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.

9. Strong mixing is a condition of asymptotic weak dependence. A sequence of rv's $\{W_t\}$ is *strong mixing* if

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

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

10. Strictly speaking, the consistency result for $\hat{S}$ given by Theorem 6(a) below only applies to $\hat{S}$ when $\hat{S}$ is defined using the Parzen kernel. When defined using the QS kernel, the consistency of $\hat{S}$ can be established under somewhat different assumptions regarding the asymptotic weak dependence of $\{m_t(\theta_0, \tau_0)\}$ than the strong mixing assumptions used here, see Andrews (1987*c*).

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.

ANDERSON, G. J. and MIZON, G. E. (1983), "Parameter Constancy Tests: Old and New" (Discussion Paper in Econometrics No. 8325, University of Southampton).

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**, 348-358.

ANDREWS, D. W. K. (1987b), "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers", *Econometrica*, **55**, 1465-1472.

ANDREWS, D. W. K. (1987c), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation" (Cowles Foundation Discussion Paper No. 877, Yale University)

ANDREWS, D. W. K. (1988), "Laws of Large Numbers for Dependent Non-identically Distributed Random Variables", *Econometric Theory*, (forthcoming).

ANDREWS, D. W. K. and FAIR, R. C. (1987), "Inference in Econometric Models with Structural Change" (Cowles Foundation Discussion Paper No. 832, Yale University).

BARTEN, A. P. and SALVAS BRONSARD, L. (1970), "Two-Stage Least Squares Estimaton 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 WHITE, H. (1988), *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models* (New York: Basil Blackwell).

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

HODOSHIMA, J. (1986), "The LM and LR Tests for Structural Change of a Single Structural Equation" (unpublished manuscript, Faculty of Economics, Nanzan University, Nagoya, Japan).

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

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. (1975), "A Maximal Inequality and Dependent Strong Laws", *Annals of Probability*, **3**, 826–836.

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**, 703–708.

POLLARD, D. (1985), "New Ways to Prove Central Limit Theorems", *Econometric Theory*, **1**, 295–313.

POTSCHER, B. M. and PRUCHA, I. R. (1987), "A Uniform Law of Large Numbers for Dependent and Heterogeneous Data Processes" (manuscript, Department of Economics, 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).

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