Econometrica, Vol. 61, No. 4 (July, 1993), 821–856

TESTS FOR PARAMETER INSTABILITY AND STRUCTURAL CHANGE WITH UNKNOWN CHANGE POINT

BY DONALD W. K. ANDREWS[1]

This paper considers tests for parameter instability and structural change with unknown change point. The results apply to a wide class of parametric models that are suitable for estimation by generalized method of moments procedures. The paper considers Wald, Lagrange multiplier, and likelihood ratio-like tests. Each test implicitly uses an estimate of a change point. The change point may be completely unknown or it may be known to lie in a restricted interval. Tests of both "pure" and "partial" structural change are discussed.

The asymptotic distributions of the test statistics considered here are nonstandard because the change point parameter only appears under the alternative hypothesis and not under the null. The asymptotic null distributions are found to be given by the supremum of the square of a standardized tied-down Bessel process of order $p \geqslant 1$, as in D. L. Hawkins (1987). Tables of critical values are provided based on this asymptotic null distribution.

As tests of parameter instability, the tests considered here are shown to have nontrivial asymptotic local power against all alternatives for which the parameters are nonconstant. As tests of one-time structural change, the tests are shown to have some weak asymptotic local power optimality properties for large sample size and small significance level. The tests are found to perform quite well in a Monte Carlo experiment reported elsewhere.

KEYWORDS: Asymptotic distribution, change point, Bessel process, Brownian bridge, Brownian motion, generalized method of moments estimator, Lagrange multiplier test, likelihood ratio test, parameter instability, structural change, Wald test, weak convergence.

1. INTRODUCTION

THIS PAPER CONSIDERS TESTS for parameter instability and structural change with unknown change point in nonlinear parametric models. The proposed tests are designed for a one-time change in the value of a parameter vector, but are shown to have power against more general forms of parameter instability. Tests are considered both for the case where the change point can be specified to lie in a particular interval and for the case where the change point is completely unknown. Tests are considered for the case of "pure" structural change, in which the entire parameter vector is subject to change under the alternative, and for the case of "partial" structural change, in which only a component of the parameter vector is subject to change under the alternative.

The results given here cover Wald, Lagrange multiplier (LM), and likelihood-ratio (LR)-like tests based on generalized method of moments (GMM) estimators. Included in this class are tests based on various least squares, nonlinear instrumental variables, maximum likelihood (ML), and pseudo-ML

[1] I thank Inpyo Li for computing the critical values reported in Section 5. I also thank two referees, a co-editor, Jean-Marie Dufour, Bruce Hansen, Werner Ploberger, and the participants of the Princeton econometrics workshop for helpful comments. I gratefully acknowledge research support from the National Science Foundation through Grant Numbers SES-8821021 and SES-9121914. The first version of this paper appeared as the discussion paper Andrews (1989c).

estimators among others. See L. P. Hansen (1982) for further discussion of GMM estimators. The data may be stationary or nonstationary under the null hypothesis of parameter stability, provided they do not exhibit deterministic or stochastic time trends. For results based on a more general class of extremum estimators, see Andrews (1989c).

The statistical literature on change point problems is extensive. (See the review papers by Zacks (1983) and Krishnaiah and Miao (1988).) The econometric literature, on the other hand, is relatively small but growing rapidly. Most of the results in the statistical literature concern models that are too simple for economic applications. Most, but not all, cover scalar parameter models and/or models with independent observations. For example, the recent papers by James, James, and Siegmund (1987), D. L. Hawkins (1987), and Kim and Siegmund (1989) fall into this category. Few econometric models are covered by such results. In addition, results in the econometric literature focus entirely on linear regression models, e.g., see Chu (1989), Banerjee, Lumsdaine, and Stock (1992), Zivot and Andrews (1992), and B. E. Hansen (1992).[2]

The contribution of this paper is to provide results for a wide variety of nonlinear models that arise in econometric applications and to provide tests that can accommodate different restrictions on the change point. The results allow for multiple parameters, temporally dependent data, and nonlinear models estimated by a variety of different methods.

The closest results in the literature to those given here are those of D. L. Hawkins (1987). Hawkins considers Wald tests of pure structural change based on ML estimators for independent identically distributed (iid) data when no information is available regarding the change point. When specialized to this case, the Wald test statistic considered here is identical to Hawkins' statistic. The method used here to obtain the asymptotic distributional results is essentially the same as that used by Hawkins (1987). (The proofs are different, however, because the present paper applies in a more general context.)

The remainder of this paper is organized as follows: Section 2 describes the null hypothesis and various alternative hypotheses that are of interest. Section 3 introduces a class of partial-sample GMM (PS-GMM) estimators and establishes their asymptotic distributions. Section 4 defines the Wald, LM, and LR-like test statistics which are based on the PS-GMM estimators. Section 5 determines the asymptotic null distributions of the Wald, LM, and LR-like test statistics and provides tables of critical values for them. Section 5 also establishes the asymptotic distributions of these test statistics under local alternatives and obtains two local power optimality results. Section 6 contains some concluding comments. An Appendix provides proofs of the results given in the paper.

Lastly, we mention notational conventions that are used throughout the paper: Unless specified otherwise, all limits are taken as $T \to \infty$, where T is the sample size. The symbol $\Rightarrow$ denotes weak convergence as defined by Pollard

[2] One paper in the econometrics literature that does consider nonlinear models is B. E. Hansen (1990). This paper was written subsequent to the first version of the present paper.

(1984, pp. 64–66) for sequences of (measurable) random elements of a space of bounded Euclidean-valued cadlag functions on $[0,1]$ or on $\Pi \subset (0,1)$ equipped with the uniform metric and the σ-field generated by the closed balls under this metric, $\rightarrow_d$ denotes convergence in distribution, $\rightarrow_p$ denotes convergence in probability, Σ_a^b abbreviates $\Sigma_{t=a}^b$, $\|\cdot\|$ denotes the Euclidean norm of a vector or matrix, $\|\cdot\|_2$ denotes the L^2 norm of a random vector (i.e., $\|X\|_2 = (E\|X\|^2)^{1/2}$), and for simplicity $T\pi$ denotes $[T\pi]$, where $[\cdot]$ is the integer part operator. Π denotes a set whose closure lies in $(0,1)$. Throughout, it is implicitly assumed that any sequence of random variables or vectors that converges in probability or almost surely to zero is Borel measurable.

2. HYPOTHESES OF INTEREST

In this section we discuss the null and alternative hypotheses of interest and provide a general discussion of the choice and use of the test statistics that are considered in the paper.

We consider a parametric model indexed by parameters (β_t, δ_0) for $t = 1, 2, \ldots$. The null hypothesis of interest here is one of parameter stability:

$$(2.1) \qquad H_0\colon \beta_t = \beta_0 \quad \text{for all } t \geqslant 1 \text{ for some } \beta_0 \in B \subset R^p.$$

In the case of tests of pure structural change, no parameter δ_0 appears and the whole parameter vector is subject to change under the alternative hypothesis. In the case of tests of partial structural change, the parameter δ_0 appears and is taken to be constant under the null hypothesis and the alternative.

The alternative hypothesis of interest may be of several forms. First, consider a one-time structural change alternative with change point $\pi \in (0,1)$. Here, T is the sample size, $T\pi$ is the time of change, and for simplicity π, rather than $T\pi$, is referred to as the change point or point of structural change. The one-time change alternative with change point π is given by

$$(2.2) \qquad H_{1T}(\pi)\colon \beta_t = \begin{cases} \beta_1(\pi) & \text{for } t = 1, \ldots, T\pi \\ \beta_2(\pi) & \text{for } t = T\pi + 1, \ldots \end{cases}$$

$$\text{for some constants } \beta_1(\pi),\ \beta_2(\pi) \in B \subset R^p.$$

For the case where π is known, one can form a Wald, LM, or LR-like test for testing H_0 versus $H_{1T}(\pi)$ (e.g., see Andrews and Fair (1988) for such tests in nonlinear models). For specificity, let $W_T(\pi)$, $LM_T(\pi)$, and $LR_T(\pi)$, respectively, denote the test statistics that correspond to these tests. For a normal linear regression model (with β_t equal to the regression parameter), these tests are equivalent F tests and are often referred to in the literature as Chow tests.

In the present paper, we are interested in cases where the change point π is unknown. In such cases, one has to construct test statistics that do not take π as given. Doing so is complicated by the fact that the problem of testing for structural change with an unknown change point does not fit into the standard "regular" testing framework; see Davies (1977, 1987). The reason is that the parameter π only appears under the alternative hypothesis and not under the null. In consequence, Wald, LM, and LR-like tests constructed with π treated as a parameter do not possess their standard large sample asymptotic distributions.

Here we adopt a common method used in this scenario and consider test statistics of the form

$$\sup_{\pi \in \Pi} W_T(\pi), \quad \sup_{\pi \in \Pi} LM_T(\pi), \quad \text{and} \quad \sup_{\pi \in \Pi} LR_T(\pi), \tag{2.3}$$

where Π is some pre-specified subset of $[0,1]$ whose closure lies in $(0,1)$. (The specification of Π is discussed below.) Other papers that consider tests of this form include Davies (1977, 1987) and D. L. Hawkins (1987) among many others in the statistical literature. Tests of this form can be motivated or justified on several grounds. First, $\sup_{\pi \in \Pi} LR_T(\pi)$ is the LR (or LR-like) test statistic for the case of an unspecified parameter π with parameter space Π. In addition, the test statistics $\sup_{\pi \in \Pi} W_T(\pi)$ and $\sup_{\pi \in \Pi} LM_T(\pi)$ are asymptotically equivalent to $\sup_{\pi \in \Pi} LR_T(\pi)$ under the null and local alternatives under suitable assumptions. Second, the test statistics $\sup_{\pi \in \Pi} W_T(\pi), \ldots, \sup_{\pi \in \Pi} LR_T(\pi)$ correspond to the tests derived from Roy's type I (or union-intersection) principle; see Roy (1953) and Roy, Gnanadesikan, and Srivastava (1971, pp. 36–46). Third, the above test statistics can be shown to possess certain (weak) asymptotic optimality properties against local alternatives for large sample size and small significance level. These results are due to Davies (1977, Thm. 4.2) for scalar parameter one-sided tests and are extended below to multi-parameter two-sided tests.

We note that although the paper concentrates on statistics of the form (2.3), the results of the paper apply more generally to statistics of the form $g(\{W_T(\pi): \pi \in \Pi\})$ for arbitrary continuous function g (and likewise for $LM_T(\cdot)$ and $LR_T(\cdot)$). Depending upon the alternatives of interest, one may want to use a function g that differs from the "sup" function. For example, in the maximum likelihood case analyzed recently by Andrews and Ploberger (1992), test statistics of the form $\int_\Pi h(W_T(\pi), \pi)\, d\lambda(\pi)$ are considered. Test statistics of this form are found to have some advantages in terms of weighted average power, for a certain weight function, over test statistics of the "sup" form.

We now return to the discussion of the alternative hypotheses of interest. Two distinct cases arise. The first is the case where interest centers on change points in a known restricted interval, say $\Pi \subset (0,1)$. The second is the case where no information is available regarding the time of possible structural change and hence all change points in $(0,1)$ are of some interest.

The case of a known restricted interval Π arises when one wants to test for structural change that is initiated by some political or institutional change that has occurred in a known time period. For example, in a model estimated using annual data from 1920 to 1989, say, one might want to test for structural change occurring sometime in the "war period" 1939–1949. In this case, one would specify $\Pi = [20/70, 31/70]$. Analogously, for some models of post-World War

II U.S. economic behavior, the Viet Nam war period might be of interest as a potential time of structural change. Alternatively, one could test for a presidential administration effect (or a chairman of the Fed effect) on certain parameters by letting Π correspond to a president's (chairman's) term of office.

A second set of cases where one can specify a restricted, but nondegenerate, interval Π includes those in which a specific exogenous event is the potential cause of structural change, but change occurs only after a lag of unknown length or before the event due to anticipation of the event. For example, in a model of aggregate or disaggregate productivity, one might want to test for structural change that occurs some time around the 1973 oil price shock. Or, in a model of the communications industry, one might want to test for structural change that occurs some time close to the court decision to break up AT&T. Or, in a small open economy, one might want to test for structural change that occurs some time close to a significant change in tariff or exchange rate policy. As a last example, in an industry study, one might want to test for structural change that occurs some time close to the introduction of a new product or technological process (which may take some time to diffuse), such as a new drug, a new chemical, or new computer equipment.

For any of the above examples, the tests considered in this paper can be applied using the critical values provided below for a very broad range of different Π intervals. The only requirement is that Π be bounded away from zero and one for reasons discussed below.

Note that the structural changes associated with the above events may be more complicated than an abrupt change. For example, there may be a movement from one regime to another with a transition period in between. It is shown below that the tests considered here have power against alternatives of this sort even though they are not the alternatives for which the tests are designed.

Next, we consider the case where no information is available regarding the time of structural change. This case arises, for example, when one wants to apply a test of structural change as a general diagnostic test of model adequacy. The usefulness of such tests is well recognized in the literature, as shown by the widespread use of the CUSUM test of Brown, Durbin, and Evans (1975) for linear regression models and by the inclusion of rolling change point tests (even without accompanying distributional theory) in popular econometric packages such as PC-GIVE and Datafit; see Hendry (1989, pp. 44, 49).

When a structural change test is employed as a general diagnostic test of model adequacy, the range of alternatives of interest is usually broader than $\bigcup_{\pi \in \Pi} H_{1T}(\pi)$ for some $\Pi \subset (0,1)$. In such cases, the alternative hypothesis may be

$$(2.4) \qquad H_1\colon \beta_s \neq \beta_t \quad \text{for some } s, t \geqslant 1.$$

Although the tests $\sup_{\pi \in \Pi} W_T(\pi), \ldots, \sup_{\pi \in \Pi} LR_T(\pi)$ are constructed with the more restricted alternatives $\bigcup_{\pi \in \Pi} H_{1T}(\pi)$ in mind, we show below that they have power against more general alternatives in H_1. In particular, we consider local alternatives of the form $\beta_t = \beta_0 + \eta(t/T)/\sqrt{T}$ for some bounded function $\eta(\cdot)$ on [0, 1] (as do Ploberger, Krämer, and Kontrus (1989)) and show that the tests of (2.3) have power against all alternatives for which $\eta(\cdot)$ is not almost everywhere on Π equal to a constant. In addition, these tests even have power against many alternatives for which $\eta(\cdot)$ is constant on Π and nonconstant elsewhere. Thus, as tests of parameter instability, the tests of (2.3) have some desirable properties.

On the other hand, if the model is stationary under the null and parameter instability can be characterized by the omission of some relevant, but unobserved, stationary variables (which can be viewed as causing structural change with an infinite number of regime changes as $T \to \infty$), then tests of the form (2.3) will not detect it asymptotically. The reason is that the model is stationary under the alternative, and so, the nonrandom parameters β_t (however defined) are constant across all $t \geqslant 1$.

A natural choice of the set of change points Π for use with the statistics of (2.3) is (0, 1) when one has no information regarding the change point. This choice, however, is not desirable. When $\Pi = (0, 1)$, the statistics $\sup_{\pi \in \Pi} W_T(\pi), \ldots, \sup_{\pi \in \Pi} LR_T(\pi)$ are shown below to diverge to infinity in probability, whereas when Π is bounded away from zero and one the statistics converge in distribution. In consequence, the use of the full interval (0, 1) results in a test whose concern for power against alternatives with a change point near zero or one leads to much reduced power against alternatives with change points anywhere else in (0, 1). Thus, when no knowledge of the change point is available, we suggest using a restricted interval Π, such as $\Pi = [.15, .85]$.

As tests of general parameter instability, the tests of (2.3) can be compared with several other tests in the literature, such as the CUSUM test of Brown, Durbin, and Evans (1975) and the fluctuation test of Sen (1980) and Ploberger, Krämer, and Kontrus (1989). These tests are all designed for the linear regression model whereas the tests of (2.3) apply more generally. A drawback of the CUSUM test is that it exhibits only trivial power against alternatives in certain directions, as shown by Krämer, Ploberger, and Alt (1988) using asymptotic local power and by Garbade (1977) and others using simulations. The tests of (2.3) do not exhibit these local power problems. The fluctuation test is similar to the $\sup_{\pi \in \Pi} W_T(\pi)$ test considered here, but the latter possesses large sample optimality properties for each fixed π, whereas the former does not.

Monte Carlo comparisons of the CUSUM, fluctuation, and $\sup_{\pi \in \Pi} W_T(\pi)$ tests reported in Andrews (1989c) show that $\sup_{\pi \in \Pi} W_T(\pi)$ is superior to the CUSUM test in terms of closeness of true and nominal size and very much superior in terms of power (both size-corrected and uncorrected) for almost all scenarios considered. In addition, $\sup_{\pi \in \Pi} W_T(\pi)$ is clearly preferable to the fluctuation test in terms of the difference between true and nominal size and in terms of uncorrected power and more marginally preferable in terms of size-corrected power.

Several additional tests in the literature for testing for parameter instability are the tests of Leybourne and McCabe (1989), Nyblom (1989), and B. E.

Hansen (1992). (Also, see the references in Krämer and Sonnberger (1986, pp. 56–59).) These tests are designed for alternatives with stochastic trends and, hence, have a different focus than the tests considered here. On the other hand, they also have a number of similarities.

3. PARTIAL-SAMPLE GMM ESTIMATORS

In this section we analyze partial-sample GMM (PS-GMM) estimators. PS-GMM estimators are GMM estimators that primarily use the pre-$T\pi$ or the post-$T\pi$ data in estimating a parameter β for variable values of π in Π and use all the data in estimating an additional parameter δ. These estimators are the basic components of the sup Wald test. Furthermore, the properties of PS-GMM estimators are used to obtain the asymptotic distribution of the corresponding sup LM and LR-like statistics.

The first subsection below defines the class of estimators to be considered. The second subsection establishes the weak convergence of PS-GMM estimators to a function of a vector Brownian motion process on $[0,1]$ restricted to Π. The third subsection considers the estimation of unknown matrices that arise in the limiting Brownian motion process. Estimators of these matrices are needed to construct the Wald and LM statistics.

3.1. *Definition of Partial-Sample GMM Estimators*

First we define the standard GMM estimator, which we call the *full-sample GMM estimator*. Under the null hypothesis of parameter stability, the unknown parameter to be estimated is a $p+q$-vector $(\beta', \delta')'$. Let $B\ (\subset R^p)$ and $\Delta\ (\subset R^q)$ denote the parameter spaces of β and δ respectively. We assume the data are given by a triangular array of rv's $\{W_{Tt}: 1 \leqslant t \leqslant T,\ T \geqslant 1\}$ defined on a probability space $(\Omega, \mathscr{F}, \boldsymbol{P})$. (By definition, a rv is Borel measurable.) Triangular arrays are considered because they are required for the local power results below.

The observed sample is $\{W_t: 1 \leqslant t \leqslant T\}$, where W_t is used here and elsewhere below to denote W_{Tt} for notational simplicity. The population orthogonality conditions that are used by the GMM estimator to estimate the true parameter $(\beta_0', \delta_0')'$ are $(1/T)\Sigma_1^T Em(W_t, \beta_0, \delta_0) = 0$ for a specified R^v-valued function $m(\cdot,\cdot,\cdot)$.

DEFINITION: A sequence of *full-sample GMM estimators* $\{(\tilde{\beta}, \tilde{\delta}): T \geqslant 1\}$ is any sequence of (Borel measurable) estimators that satisfies

$$(3.1)\qquad \frac{1}{T}\sum_1^T m\left(W_t, \tilde{\beta}, \tilde{\delta}\right)'\hat{\gamma}\frac{1}{T}\sum_1^T m\left(W_t, \tilde{\beta}, \tilde{\delta}\right) = \inf_{(\beta,\delta)\in B\times\Delta} \frac{1}{T}\sum_1^T m(W_t, \beta, \delta)'\hat{\gamma}\frac{1}{T}\sum_1^T m(W_t, \beta, \delta)$$

with probability $\rightarrow 1$, where $(\beta', \delta')' \in B \times \Delta \subset R^p \times R^q$, $m(\cdot,\cdot,\cdot)$ is a function from $W \times B \times \Delta$ to R^{ν}, $W \subset R^k$, and $\hat{\gamma}$ is a random (Borel measurable) symmetric $\nu \times \nu$ matrix (which depends on T in general). As is well-known (e.g., see L. P. Hansen (1982)), the class of GMM estimators is quite broad. Among others, it includes least squares, nonlinear instrumental variables, ML, and pseudo-ML estimators.

Next consider the case where the sample is broken into two parts, viz., $t = 1, \ldots, T\pi$, and $t = T\pi + 1, \ldots, T$, the parameter β takes the value β_1 for the first part of the sample and another value β_2 for the second part, and the parameter δ is constant across the whole sample. In this case, the unknown parameter of interest is $\theta = (\beta_1', \beta_2', \delta')' \in \Theta = B \times B \times \Delta \subset R^p \times R^p \times R^q$.

Let $\tilde{\theta} = (\tilde{\beta}', \tilde{\beta}', \tilde{\delta}')'$. We call $\tilde{\theta}$ the full-sample GMM estimator of θ. It is a restricted estimator that is consistent only under the null hypothesis that $\beta_1 = \beta_2$.

We now define an unrestricted GMM estimator of θ that allows the estimates of β_1 and β_2 to differ. Suppose the true value of θ is $(\beta_{10}', \beta_{20}', \delta_0')'$. For the observations $t = 1, \ldots, T\pi$, we have the population orthogonality conditions $(1/T)\Sigma_1^{T\pi} Em(W_t, \beta_{10}, \delta_0) = 0$, and for the observations $t = T\pi + 1, \ldots, T$, we have a second set of orthogonality conditions $(1/T)\Sigma_{T\pi+1}^{T} Em(W_t, \beta_{20}, \delta_0) = 0$. For each potential change point $\pi \in \Pi \subset (0,1)$, we can define an estimator that is based on the sample analogues of these orthogonality conditions. The collection of such estimators for $\pi \in \Pi$ is called the partial-sample GMM estimator of θ.

DEFINITION: A sequence of *partial-sample GMM estimators* $\{\hat{\theta}(\cdot): T \geqslant 1\} = \{(\hat{\theta}(\pi): \pi \in \Pi): T \geqslant 1\}$ is any sequence of estimators that satisfies

$$(3.2) \qquad \overline{m}_T(\hat{\theta}(\pi), \pi)' \hat{\gamma}(\pi) \overline{m}_T(\hat{\theta}(\pi), \pi) = \inf_{\theta \in \Theta} \overline{m}_T(\theta, \pi)' \hat{\gamma}(\pi) \overline{m}_T(\theta, \pi) \quad \text{for all } \pi \in \Pi$$

with probability $\to 1$ and $\hat{\theta}(\cdot)$ is a random element, where $\theta = (\beta_1', \beta_2', \delta')' \in \Theta = B \times B \times \Delta \subset R^p \times R^p \times R^q$,

$$\overline{m}_T(\theta, \pi) = \frac{1}{T} \sum_1^{T\pi} \begin{pmatrix} m(W_t, \beta_1, \delta) \\ 0 \end{pmatrix} + \frac{1}{T} \sum_{T\pi+1}^{T} \begin{pmatrix} 0 \\ m(W_t, \beta_2, \delta) \end{pmatrix} \in R^{2\nu},$$

$m(\cdot, \cdot, \cdot)$ is a function from $W \times B \times \Delta$ to R^{ν}, $W \subset R^k, \hat{\gamma}(\pi)$ is a random symmetric $2\nu \times 2\nu$ matrix (which depends on T in general), and $\hat{\gamma}(\cdot)$ is a random element.

Existence of partial-sample GMM estimators can be established under standard conditions. For example, compactness of Θ and continuity of the criterion function above are sufficient.

(By definition, a *random element* is a measurable function from $(\Omega, \mathscr{F}, \boldsymbol{P})$ to a space of bounded Euclidean-valued cadlag functions on $[0,1]$ or on $\Pi \subset (0,1)$ equipped with the σ-field $\boldsymbol{A}$ generated by the closed balls under the uniform metric. Note that such a function, say $\hat{g}$, is necessarily $\mathscr{F}/A$-measurable, and hence is a random element, if $\hat{g}(\pi)$ is $\mathscr{F}$/Borel-measurable for all $\pi \in \Pi$; see Pollard (1984, Problems 2 and 4(b), pp. 80–81). Thus, measurability of $\hat{\theta}(\cdot)$ can be established under standard assumptions. For brevity, we do not give sufficient conditions here.)

As the definition indicates, $\hat{\theta}(\pi) = (\hat{\beta}_1(\pi)', \hat{\beta}_2(\pi)', \hat{\delta}(\pi)')'$ is a $2p+q$-vector comprised of an estimator $\hat{\beta}_1(\pi) \in R^p$ that primarily uses the pre-$T\pi$ data, an estimator $\hat{\beta}_2(\pi) \in R^p$ that primarily uses the post-$T\pi$ data, and an estimator $\hat{\delta}(\pi) \in R^q$ that uses all of the data. For a fixed value of π, the PS-GMM estimators defined above are a special case of the extremum estimators of Andrews and Fair (1988).

3.2. *Weak Convergence of Partial-Sample GMM Estimators*

In this subsection we establish the asymptotic distribution of the PS-GMM estimator $\hat{\theta}(\cdot)$ for the case of no structural change. To do so, we need to introduce some additional notation and definitions. The asymptotic distribution of $\hat{\theta}(\cdot)$ depends on the following matrices:

$$(3.3) \qquad S = \lim_{T\to\infty} \operatorname{Var}\left(\frac{1}{\sqrt{T}} \sum_1^T m(W_t, \beta_0, \delta_0)\right) \in R^{\nu\times\nu},$$

$$M = \lim_{T\to\infty} \frac{1}{T} \sum_1^T E \frac{\partial m(W_t, \beta_0, \delta_0)}{\partial \beta'} \in R^{\nu\times p},$$

$$M_\delta = \lim_{T\to\infty} \frac{1}{T} \sum_1^T E \frac{\partial m(W_t, \beta_0, \delta_0)}{\partial \delta'} \in R^{\nu\times q}, \quad \text{and}$$

$$M(\pi) = \begin{bmatrix} \pi M & 0 & \pi M_\delta \\ 0 & (1-\pi)M & (1-\pi)M_\delta \end{bmatrix} \in R^{2\nu\times(2p+q)}.$$

For simplicity, let m_t or m_{Tt} denote $m(W_t, \beta_0, \delta_0)$. Let the domain $\boldsymbol{W}$ of $m(\cdot, \beta, \delta)$ be chosen to include the support of W_t $\forall t, \forall T$. Let B_0 and Δ_0 denote some compact subsets of R^p and R^q that contain neighborhoods of β_0 and δ_0 and are contained in the parameter spaces B and Δ respectively (where $\theta_0 = (\beta_0', \beta_0', \delta_0')'$ when no structural change occurs). Let μ_{Tt} denote the distribution of W_{Tt} and let $\bar{\mu}_T = (1/T)\Sigma_1^T \mu_{Tt}$. We say that $\{\bar{\mu}_T\colon T \geqslant 1\}$ is tight on $\boldsymbol{W}$ if $\lim_{j\to\infty} \sup_{T\geqslant 1} (1/T)\Sigma_1^T P(W_{Tt} \notin C_j) = 0$ for some sequence of compact sets $C_j \subset \boldsymbol{W}$ for $j \geqslant 1$. (A sufficient condition for tightness of $\{\bar{\mu}_T\colon T \geqslant 1\}$ when $\boldsymbol{W}$ is closed is the weak moment condition $\lim_{T\to\infty} (1/T)\Sigma_1^T E\|W_{Tt}\|^\varepsilon < \infty$ for some $\varepsilon > 0$.) For a sequence of nonrandom matrices $\{A_T(\lambda)\colon T \geqslant 1\}$ indexed by a parameter $\lambda \in \Lambda$, we say that $\lim_{T\to\infty} A_T(\lambda)$ exists uniformly over $\lambda \in \Lambda$, if there exist matrices $\{A(\lambda)\colon \lambda \in \Lambda\}$ such that $\lim_{T\to\infty} \sup_{\lambda\in\Lambda} |A_T(\lambda) - A(\lambda)| = 0$. Following McLeish (1975a), for a constant $q > 0$ and a sequence of nonnegative constants $\{\nu_m\colon m \geqslant 1\}$, we say that $\{\nu_m\colon m \geqslant 1\}$ is of size $-q$ if ν_m converges to zero at a fast enough rate. A sufficient condition is that $\nu_m = O(m^{-\lambda})$ for some $\lambda > q$. The precise definition of size is given in McLeish (1975a).

Next, we define a type of asymptotically weak temporal dependence called near epoch dependence (NED). This concept has origins as far back (at least) as Ibragimov (1962). It has appeared in various forms in the work of Billingsley (1968), McLeish (1975a, b), Bierens (1981), Gallant (1987), Gallant and White (1988), Andrews (1988), Wooldridge and White (1988), B. E. Hansen (1991), and Pötscher and Prucha (1991) among others. The NED condition is used to obtain laws of large numbers, CLTs, and invariance principles for triangular arrays of temporally dependent rv's. It is one of the most general concepts of weak temporal dependence for nonlinear models that is available. See Bierens (1981), Gallant (1987), Gallant and White (1988), and Pötscher and Prucha (1991) for examples of its application to particular econometric models.

DEFINITION: For $p \geqslant 0$, a triangular array of rv's $\{X_{Tt}: t = 1, \dots, T, T \geqslant 1\}$ is said to be *L^p-NED on the strong mixing base* $\{Y_{Tt}: t = \dots, 0, 1, \dots; T \geqslant 1\}$ if $\{Y_{Tt}: t = \dots, 0, 1, \dots; T \geqslant 1\}$ is a strong mixing (i.e., α-mixing) array of rv's and

$$\overline{\lim_{T\to\infty}} \frac{1}{T}\sum_{1}^{T} \| X_{Tt} - E(X_{Tt}|Y_{Tt-m}, \dots, Y_{Tt+m})\|_p \to 0$$

as $m \to \infty$ when $p > 0$ or

$$\overline{\lim_{T\to\infty}} \frac{1}{T}\sum_{1}^{T} P(\| X_{Tt} - E(X_{Tt}|Y_{Tt-m}, \dots, Y_{Tt+m})\| > \varepsilon) \to 0$$

as $m \to \infty$ $\forall \varepsilon > 0$ when $p = 0$. For $p > 0$, $q > 0$, and $r > 0$, $\{X_{Tt}: t = 1, \dots, T, T \geqslant 1\}$ is said to be *L^p-NED of size $-q$ on a strong mixing base* $\{Y_{Tt}: t = \dots, 0, 1, \dots; T \geqslant 1\}$ *of size* $-r$ if $\{\nu_m: m \geqslant 1\}$ is of size $-q$, where $\nu_m = \sup_{t \leqslant T, T \geqslant 1} \|X_{Tt} - E(X_{Tt}|Y_{Tt-m}, \dots, Y_{Tt+m})\|_p$ and $\{\alpha_m: m \geqslant 1\}$ is of size $-r$, where $\{\alpha_m: m \geqslant 1\}$ are the strong mixing numbers of $\{Y_{Tt}\}$.

The following assumption is sufficient to obtain the weak convergence under the null hypothesis of the PS-GMM estimator $\hat{\theta}(\cdot)$ as a process indexed by $\pi \in \Pi$.

ASSUMPTION 1: (a) $\{W_{Tt}: t \leqslant T, T \geqslant 1\}$ *is a triangular array of* $\mathbf{W}$*-valued rv's that is L^0-NED on a strong mixing base* $\{Y_{Tt}: t = \dots, 0, 1, \dots; T \geqslant 1\}$, *where* $\mathbf{W}$ *is a Borel subset of* R^k, *and* $\{\bar{\mu}_T: T \geqslant 1\}$ *is tight on* $\mathbf{W}$.

(b) *For some* $r > 2$, $\{m_{Tt}: t \leqslant T, T \geqslant 1\}$ *is a triangular array of mean zero R^v-valued rv's that is L^2-NED of size* $-1/2$ *on a strong mixing base* $\{Y_{Tt}: t = \dots, 0, 1, \dots; T \geqslant 1\}$ *of size* $-r/(r-2)$ *and* $\sup_{t \leqslant T, T \geqslant 1} E\|m_{Tt}\|^r < \infty$.

(c) $\operatorname{Var}((1/\sqrt{T})\Sigma_1^{T\pi} m_{Tt}) \to \pi S$ $\forall \pi \in [0,1]$ *for some positive definite* $v \times v$ *matrix* S.

(d) $\sup_{\pi \in \Pi} \|\hat{\theta}(\pi) - \theta_0\| \to_p 0$ *and* $\tilde{\theta} \to_p \theta_0$ *for some* $\theta_0 = (\beta_0', \beta_0', \delta_0')'$ *in the interior of* $\Theta = B \times B \times \Delta$.

(e) $\sup_{\pi \in \Pi} \|\hat{\gamma}(\pi) - \gamma(\pi)\| \to_p 0$ *for some symmetric* $2\nu \times 2\nu$ *matrices* $\gamma(\pi)$ *for which* $\sup_{\pi \in \Pi} \|\gamma(\pi)\| < \infty$.

(f) $m(w, \beta, \delta)$ *is partially differentiable in* (β, δ) $\forall(\beta, \delta) \in B_0 \times \Delta_0$ $\forall w \in W_0 \subset W$ *for a Borel measurable set* W_0 *that satisfies* $P(W_{Tt} \in W_0) = 1$ $\forall t \leqslant T$, $T \geqslant 1$, $m(w, \beta, \delta)$ *is Borel measurable in* w $\forall(\beta, \delta) \in B_0 \times \Delta_0$, $\partial m(w, \beta, \delta)/\partial(\beta', \delta')$ *is continuous in* (w, β, δ) *on* $W \times B_0 \times \Delta_0$, *and*

$$\sup_{t \leqslant T, T \geqslant 1} E \sup_{(\beta, \delta) \in B_0 \times \Delta_0} \|\partial m(W_{Tt}, \beta, \delta)/\partial(\beta', \delta')\|^{1+\varepsilon} < \infty$$

for some $\varepsilon > 0$.

(g) $\lim_{T \to \infty} (1/T) \Sigma_1^{T\pi} E \partial m(W_{Tt}, \beta_0, \delta_0)/\partial(\beta', \delta')$ *exists uniformly over* $\pi \in \Pi$ *and equals* πM $\forall \pi \in \Pi$.

(h) $M(\pi)'\gamma(\pi)M(\pi)$ *is nonsingular* $\forall \pi \in \Pi$ *and has eigenvalues bounded away from zero.*

We now discuss Assumption 1. Assumptions 1(a) and (b) are typical of asymptotic weak dependence conditions found in the literature on nonlinear dynamic models; see the references above. They are closest to conditions given by Pötscher and Prucha (1991). Assumptions 1(c) and (g) are asymptotic covariance stationarity conditions that are used for the results of the present paper, but are not needed for results in the literature that deal only with the estimation of nonlinear dynamic models. Assumptions 1(d) and (e) are used to show that various remainder terms in the proof of weak convergence of $\hat{\theta}(\cdot)$ are negligible. Sufficient conditions for Assumption 1(d) are provided in the Appendix. The verification of Assumption 1(e) depends, of course, on the choice of the weight matrix $\hat{\gamma}(\pi)$. Often $\hat{\gamma}(\pi)$ is of the form $\text{Diag}\{\hat{S}_1^{-1}(\pi)/\pi, \hat{S}_2^{-1}(\pi)/(1-\pi)\}$, where $\hat{S}_r(\pi)$ is an estimator of S for $r = 1, 2$. The definition of, motivation for, and uniform consistency of the estimators $\hat{S}_r(\pi)$ are discussed in the Comment following Theorem 1 and in Section 3.3 below. Assumption 1(f) is a standard smoothness condition on the function $m(w, \beta, \delta)$. It could be relaxed at the expense of greater complexity by using the approaches taken in Huber (1967), in Andrews (1989a, b), or elsewhere in the literature. Assumption 1(h) ensures that the estimator $\hat{\theta}(\pi)$ has a well-defined asymptotic variance $\forall \pi \in \Pi$. In the common case that $\gamma(\pi)$ is of the form $\text{Diag}\{\gamma/\pi, \gamma/(1-\pi)\}$, Assumption 1(h) holds if γ, M, and M_δ are full rank ν, p, and q respectively and Π has closure in $(0, 1)$. For the ML estimator, for instance, this requires that the information matrix for (β, δ) in the case of parameter constancy is nonsingular, as usually occurs.

Before stating the main result of this section, we introduce some additional notation. Let $\{B(\pi): \pi \in [0, 1]\}$ denote a ν-vector of independent Brownian motions on $[0, 1]$. Let

$$G(\pi) = \begin{pmatrix} S^{1/2}B(\pi) \\ S^{1/2}(B(1) - B(\pi)) \end{pmatrix}. \tag{3.4}$$

THEOREM 1: *Under Assumption* 1, *every sequence of PS-GMM estimators* $\{\hat{\theta}(\cdot): T \geqslant 1\}$ *satisfies*

$$\sqrt{T}\left(\hat{\theta}(\cdot) - \theta_0\right) \Rightarrow \left(M(\cdot)'\gamma(\cdot)M(\cdot)\right)^{-1} M(\cdot)'\gamma(\cdot)G(\cdot)$$

as a process indexed by $\pi \in \Pi$, *provided* Π *has closure in* $(0,1)$.

COMMENT: For any fixed value of π, an optimal choice of the asymptotic weight matrix $\gamma(\pi)$ is

$$\gamma(\pi) = \operatorname{Diag}\left\{S^{-1}/\pi, S^{-1}/(1-\pi)\right\}. \tag{3.5}$$

This matrix is asymptotically optimal in the sense of minimizing the covariance matrix of the asymptotic normal distribution of the normalized estimator $\sqrt{T}(\hat{\theta}(\pi) - \theta_0)$.[3] For $\gamma(\pi)$ as in (3.5), the limit process of $\sqrt{T}(\hat{\theta}(\pi) - \theta_0)$ evaluated at π can be written as

$$\begin{bmatrix} \pi M'S^{-1}M & 0 & \pi M'S^{-1}M_\delta \\ 0 & (1-\pi)M'S^{-1}M & (1-\pi)M'S^{-1}M_\delta \\ \pi M_\delta' S^{-1}M & (1-\pi)M_\delta' S^{-1}M & M_\delta' S^{-1}M_\delta \end{bmatrix}^{-1} \times \begin{pmatrix} M'S^{-1/2}B(\pi) \\ M'S^{-1/2}(B(1) - B(\pi)) \\ M_\delta' S^{-1/2}B(1) \end{pmatrix}. \tag{3.6}$$

3.3. *Covariance Matrix Estimation for Partial-Sample GMM Estimators*

The Wald statistic defined in Section 4 below is based on the vector $\sqrt{T}(\hat{\beta}_1(\cdot) - \hat{\beta}_2(\cdot))$. Here we introduce estimators of the unknown matrices that appear in the limit distribution of $\sqrt{T}(\hat{\beta}_1(\cdot) - \hat{\beta}_2(\cdot))$. These estimators are needed to construct the weight matrices of the Wald and LM test statistics. For brevity, we consider the standard case where the weight matrix $\hat{\gamma}(\pi)$ is chosen to be asymptotically optimal:

ASSUMPTION 2: $\gamma(\pi) = \operatorname{Diag}\{S^{-1}/\pi, S^{-1}/(1-\pi)\}$ *for*

$$S = \lim_{T\to\infty} \operatorname{Var}\left(\frac{1}{\sqrt{T}} \sum\nolimits_1^T m_{Tt}\right)$$

(*as in Assumption* 1).

Let

$$H = \left[I_p \vdots -I_p \vdots 0\right] \in R^{p\times(2p+q)}. \tag{3.7}$$

[3] This result can be proved using standard arguments; e.g., see the proof of Theorem 3.2 of L. P. Hansen (1982).

By Theorem 1, $\sqrt{T}(\hat{\beta}_1(\pi) - \hat{\beta}_2(\pi)) = H\sqrt{T}(\hat{\theta}(\pi) - \theta_0)$ converges in distribution to $H(M(\pi)'\gamma(\pi)M(\pi))^{-1}M(\pi)'\gamma(\pi)G(\pi)$. By (3.6) and Lemma A5 of the Appendix, the latter simplifies to

$$(3.8) \qquad [I_p \vdots -I_p]\begin{pmatrix} \frac{1}{\pi}(M'S^{-1}M)^{-1}M'S^{-1/2}B(\pi) \\ \frac{1}{1-\pi}(M'S^{-1}M)^{-1}M'S^{-1/2}(B(1)-B(\pi)) \end{pmatrix}$$

$$\sim N\left(0, \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)V\right), \quad \text{where}$$

$$V = (M'S^{-1}M)^{-1}.$$

Consider the following estimators of V:

$$(3.9) \qquad \hat{V}_r(\pi) = \left(\hat{M}_r(\pi)'\hat{S}_r^{-1}(\pi)\hat{M}_r(\pi)\right)^{-1} \quad \text{for } r = 1, 2.^4$$

The estimators $\hat{M}_r(\pi)$ and $\hat{S}_r(\pi)$ can be defined in two ways. The first way uses only the data for $t = 1, \ldots, T\pi$ for the case $r = 1$ and only the data for $t = T\pi + 1, \ldots, T$ for the case $r = 2$:

$$(3.10) \qquad \hat{M}_1(\pi) = \frac{1}{T\pi}\sum_1^{T\pi}\frac{\partial m\left(W_t, \hat{\beta}_1(\pi), \hat{\delta}(\pi)\right)}{\partial \beta_1'},$$

$$\hat{M}_2(\pi) = \frac{1}{T - T\pi}\sum_{T\pi+1}^{T}\frac{\partial m\left(W_t, \hat{\beta}_2(\pi), \hat{\delta}(\pi)\right)}{\partial \beta_2'}.$$

(Corresponding estimators $\hat{S}_r(\pi)$ are defined below.) The second way uses all of the data for $r = 1$ and $r = 2$:

$$(3.11) \qquad \hat{M}_r(\pi) = \hat{M} = \frac{1}{T}\sum_1^T \frac{\partial m\left(W_t, \tilde{\beta}, \tilde{\delta}\right)}{\partial \beta'},$$

where $(\tilde{\beta}, \tilde{\delta})$ are the full-sample GMM estimators of (β, δ).

The estimators defined in (3.10) and (3.11) have the same probability limits under the null hypothesis and under sequences of local alternatives (see Section 5). They do not necessarily have the same probability limits, however, under sequences of fixed alternatives. Typically, unrestricted estimators, such as those of (3.10), are used to construct weight matrices for Wald statistics whereas restricted estimators, such as those of (3.11), are used for LM statistics. The weight matrix is taken to equal $(\hat{V}_1(\pi)/\pi + \hat{V}_2(\pi)/(1-\pi))^{-1}$ in either case, but in the latter case $\hat{V}_1(\pi) = \hat{V}_2(\pi) = \hat{V}$. As with the choice of weight matrices for classical Wald and LM tests, one cannot distinguish between the two methods based on local power.

[4] Under the assumptions, $(\hat{M}_r(\pi)'\hat{S}_r^{-1}(\pi)\hat{M}_r(\pi))^{-1}$ may exist only with probability $\to 1$. When $(\hat{M}_r(\pi)'\hat{S}_r^{-1}(\pi)\hat{M}_r(\pi))^{-1}$ is singular, a g-inverse can be used in place of the inverse. Similar comments apply elsewhere below.

Next we consider the definition of the estimators $\hat{S}_r(\pi)$ of $S(=\lim_{T\to\infty}\operatorname{Var}((1/\sqrt{T})\Sigma_1^T m_t))$. If $\{m_t\colon t \geqslant 1\}$ consists of mean zero uncorrelated rv's, then $S = \lim_{T\to\infty}(1/T)\Sigma_1^T Em_t m_t'$ and we define either

$$(3.12)\quad \hat{S}_1(\pi) = \frac{1}{T\pi}\sum_1^{T\pi}\left(m\left(W_t, \hat{\beta}_1(\pi), \hat{\delta}(\pi)\right) - \overline{m}_{1T}(\pi)\right) \times\left(m\left(W_t, \hat{\beta}_1(\pi), \hat{\delta}(\pi)\right) - \overline{m}_{1T}(\pi)\right)' \quad \text{and}$$

$$\hat{S}_2(\pi) = \frac{1}{T - T\pi}\sum_{T\pi+1}^{T}\left(m\left(W_t, \hat{\beta}_2(\pi), \hat{\delta}(\pi)\right) - \overline{m}_{2T}(\pi)\right) \times\left(m\left(W_t, \hat{\beta}_2(\pi), \hat{\delta}(\pi)\right) - \overline{m}_{2T}(\pi)\right)', \quad \text{or}$$

$$(3.13)\quad \hat{S}_r(\pi) = \hat{S} = \frac{1}{T}\sum_1^T\left(m\left(W_t, \tilde{\beta}, \tilde{\delta}\right) - \overline{m}_T\right)\left(m\left(W_t, \tilde{\beta}, \tilde{\delta}\right) - \overline{m}_T\right)' \quad \text{for } r = 1, 2,$$

where $\overline{m}_{1T}(\pi) = (1/T\pi)\Sigma_1^{T\pi} m(W_t, \hat{\beta}_1(\pi), \hat{\delta}(\pi))$, $\overline{m}_{2T}(\pi) = (1/(T - T\pi))\Sigma_{T\pi+1}^T m(W_t, \hat{\beta}_2(\pi), \hat{\delta}(\pi))$, and $\overline{m}_T = (1/T)\Sigma_1^T m(W_t, \tilde{\beta}, \tilde{\delta})$.

Alternatively, if $\{m_t\colon t \geqslant 1\}$ consists of mean zero temporally dependent rv's, then $S = \Sigma_{\nu=0}^{\infty}\Gamma_\nu + \Sigma_{\nu=1}^{\infty}\Gamma_\nu'$, where $\Gamma_\nu = \lim_{T\to\infty}(1/T)\Sigma_{\nu+1}^T Em_t m_{t-\nu}'$. In this case, the estimator $\hat{S}_r(\pi)$ corresponding to (3.11) can be taken to be

$$(3.14)\quad \hat{S}_r(\pi) = \hat{S} = \sum_{\nu=0}^{T-1} k(\nu/l(T)) \times\frac{1}{T}\sum_{\nu+1}^{T}\left(m\left(W_t, \tilde{\beta}, \tilde{\delta}\right) - \overline{m}_T\right)\left(m\left(W_{t-\nu}, \tilde{\beta}, \tilde{\delta}\right) - \overline{m}_T\right)' + \sum_{\nu=1}^{T-1} k(\nu/l(T)) \times\frac{1}{T}\sum_{\nu+1}^{T}\left(m\left(W_{t-\nu}, \tilde{\beta}, \tilde{\delta}\right) - \overline{m}_T\right)\left(m\left(W_t, \tilde{\beta}, \tilde{\delta}\right) - \overline{m}_T\right)'$$

for $r = 1, 2$, where $k(\cdot)$ is a kernel and $l(T)$ is a (possibly data-dependent) bandwidth parameter. The estimator $\hat{S}$ is a kernel estimator of the spectral density matrix at frequency zero of the sequence of rv's $\{m(W_t, \delta_0, \beta_0)\colon t \leqslant T\}$; e.g., see Hannan (1970). For a suitable choice of kernel, $\hat{S}$ is necessarily positive semi-definite. See Andrews (1991) regarding the choice of kernel and bandwidth parameter.[5] Unrestricted kernel estimators $\hat{S}_1(\pi)$ and $\hat{S}_2(\pi)$ that correspond to

[5] An attractive alternative to the kernel estimator of (3.16) is a prewhitened kernel estimator described in Andrews and Monahan (1992) (and for brevity not defined here). This estimator has proved to work well in simulation studies in terms of minimizing the discrepancy between the nominal and true size of test statistics constructed using a nonparametric covariance matrix estimator.

(3.10) can be defined analogously to (3.14) using the data from the time periods $1,\dots,T\pi$ and $T\pi+1,\dots,T$, respectively, and using the estimators $(\hat{\beta}_1(\pi),\hat{\delta}(\pi))$ and $(\hat{\beta}_2(\pi),\hat{\delta}(\pi))$ respectively.

Under Assumptions 1 and 2 and the following assumption, the estimators $\hat{V}_r(\pi)$ defined above are consistent for V uniformly over $\pi \in \Pi$:

ASSUMPTION 3: *$\hat{V}_r(\pi)$ is constructed using an estimator $\hat{S}_r(\pi)$ that satisfies $\sup_{\pi\in\Pi}\|\hat{S}_r(\pi)-S\|\to_p 0$ and $\hat{V}_r(\cdot)$ is a random element for $r=1,2$.*

Assumption 3 holds for $\hat{S}_r(\pi)$ as defined in (3.13) under Assumption 1 plus

$$E\sup_{\beta\in B_0,\,\delta\in\Delta_0}\left\|m(W_t,\beta,\delta)\frac{\partial m(W_t,\beta,\delta)}{\partial(\beta',\delta')}\right\|<\infty.$$

Assumption 3 holds for $\hat{S}_r(\pi)$ as defined in (3.12) under the same conditions provided Π has closure in (0,1) (using Lemmas A3 and A4 of the Appendix in the proof). Assumption 3 holds for $\hat{S}_r(\pi)$ as in (3.14) under the conditions given in Andrews (1991).

THEOREM 2: *Under Assumptions* 1–3,

$$\sup_{\pi\in\Pi}\left\|\hat{V}_r(\pi)-V\right\|\to_p 0\quad\text{for } r=1,2,$$

provided Π has closure in (0,1).

4. DEFINITIONS OF THE TEST STATISTICS

4.1. *The Wald Statistic*

The Wald statistic for testing H_0 against $H_{1T}(\pi)$ is given by

$$\begin{aligned}(4.1)\qquad W_T(\pi)&=T\big(\hat{\beta}_1(\pi)-\hat{\beta}_2(\pi)\big)'\\&\quad\times\big(\hat{V}_1(\pi)/\pi+\hat{V}_2(\pi)/(1-\pi)\big)^{-1}\big(\hat{\beta}_1(\pi)-\hat{\beta}_2(\pi)\big),\end{aligned}$$

where $\hat{V}_1(\pi)$ and $\hat{V}_2(\pi)$ are as in (3.9) plus either (3.10) or (3.11), etc. Based on $W_T(\pi)$, the following statistic can be used for testing H_0 versus $\bigcup_{\pi\in\Pi}H_{1T}(\pi)$ or H_0 versus H_1:

$$(4.2)\qquad \sup_{\pi\in\Pi}W_T(\pi),$$

where Π is a set with closure in (0,1). One rejects H_0 for large values of $\sup_{\pi\in\Pi}W_T(\pi)$.

Note that the asymptotic variance of $\sqrt{T}(\hat{\beta}_1(\pi)-\hat{\beta}_2(\pi))$ takes on the additive form $V/\pi+V/(1-\pi)$ even though Assumption 1 allows for temporal dependence. This occurs because of the assumption of asymptotically weak temporal dependence plus the fact that the fraction of observations that are close to the change point, say within R time periods, goes to zero as $T \to \infty$ and this holds for all R.

The sup Wald test of (4.2) has been considered previously by others in less general contexts. For example, D. L. Hawkins (1987) considers it in the context of tests of pure structural change based on ML estimators for models with iid observations. Hawkins takes $\Pi = [\varepsilon, 1-\varepsilon]$ for small $\varepsilon > 0$, whereas we consider more flexible choices of Π.

One can compute $W_T(\pi)$ using a standard GMM computer routine as follows. For given $\pi \in \Pi$, form the vector of orthogonality conditions $\overline{m}_T(\theta, \pi)$ and the weight matrix $\hat{\gamma}(\pi) = \mathrm{Diag}\{\hat{S}_1^{-1}(\pi)/\pi, \hat{S}_2^{-1}(\pi)/(1-\pi)\}$. Let $\hat{\theta}(\pi)$ and $\hat{\Omega}(\pi)$ denote the parameter vector and its estimated covariance matrix that are produced by the GMM computer routine. Then, $W_T(\pi)$ equals $\hat{\theta}(\pi)'H'(H\hat{\Omega}(\pi)H')^{-1}H\hat{\theta}(\pi)$, where $H = [I_p \vdots -I_p \vdots 0]$.[6]

4.2. *The LM Statistic*

Next we define the $LM_T(\pi)$ statistic. It makes use of the full-sample GMM estimator $\tilde{\theta} = (\tilde{\beta}', \tilde{\beta}', \tilde{\delta}')'$. For fixed change point π, the LM statistic is a quadratic form based on the vector of first-order conditions from the minimization of the PS-GMM criterion function evaluated at the restricted estimator $\tilde{\theta}$ (i.e., $[\partial \overline{m}_T(\tilde{\theta}, \pi)/\partial\theta']'\hat{\gamma}(\pi)\overline{m}_T(\tilde{\theta}, \pi)$). The weight matrix of the quadratic form is chosen such that the statistic has a χ_p^2 distribution under the null for each fixed π. The LM statistic can be written as

$$(4.3) \qquad LM_T(\pi) = c_T(\pi)'\left(\hat{V}_1(\pi)/\pi + \hat{V}_2(\pi)/(1-\pi)\right)^{-1} c_T(\pi), \quad \text{where}$$

$$c_T(\pi) = \left[I_p \vdots -I_p\right] \times \begin{bmatrix} \frac{1}{\pi}\left(\hat{M}_1'\hat{S}_1^{-1}\hat{M}_1\right)^{-1}\hat{M}_1'\hat{S}_1^{-1} & 0 \\ 0 & \frac{1}{1-\pi}\left(\hat{M}_2'\hat{S}_2^{-1}\hat{M}_2\right)^{-1}\hat{M}_2'\hat{S}_2^{-1} \end{bmatrix} \times \sqrt{T}\,\overline{m}_T(\tilde{\theta}, \pi)$$

and $\hat{M}_r = \hat{M}_r(\pi)$, $\hat{S}_r = \hat{S}_r(\pi)$, and $\hat{V}_r(\pi)$ are as in Section 3.3 for $r = 1, 2$.

Typically one uses "restricted" estimators $\hat{M}_r(\pi)$, $\hat{S}_r(\pi)$, and $\hat{V}_r(\pi)$, when constructing the LM statistic. In this case, $LM_T(\pi)$ simplifies. In particular, suppose $\hat{M}_r(\pi) = \hat{M}$ is as in (3.11) and $\hat{S}_r(\pi) = \hat{S}$ is as in (3.13) or (3.14). Then,

[6] As defined, the weight matrix $(H\hat{\Omega}(\pi)H')^{-1}$ will not necessarily be identical to that specified in (4.1), but it will at least be asymptotically equivalent to it.

$LM_T(\pi)$ simplifies to

$$(4.4) \qquad LM_T(\pi) \cong \frac{T}{\pi(1-\pi)} \overline{m}_{1T}(\tilde{\theta},\pi)' \hat{S}^{-1} \hat{M}\left(\hat{M}' \hat{S}^{-1} \hat{M}\right)^{-1} \hat{M}' \hat{S}^{-1} \overline{m}_{1T}(\tilde{\theta},\pi),$$

where

$$\overline{m}_{1T}(\tilde{\theta},\pi) = \frac{1}{T} \sum_{1}^{T\pi} m\left(W_t, \tilde{\beta}, \tilde{\delta}\right)$$

and $\cong$ denotes equality that holds with probability $\to 1$.[7] The LM statistic is particularly easy to compute because the only estimate of θ that is required is the full-sample GMM estimate.

4.3. *The LR-like Statistic*

Lastly, we define the LR-like test statistic. For fixed change point π, it is given by the difference between the PS-GMM objective function evaluated at the full sample GMM and the PS-GMM estimators:

$$(4.5) \qquad LR_T(\pi) = T\overline{m}_T(\tilde{\theta},\pi)' \hat{\gamma}(\pi) \overline{m}_T(\tilde{\theta},\pi) - T\overline{m}_T\left(\hat{\theta}(\pi),\pi\right)' \hat{\gamma}(\pi) \overline{m}_T\left(\hat{\theta}(\pi),\pi\right).$$

As in (4.2), for testing H_0 versus $\bigcup_{\pi \in \Pi} H_{1T}(\pi)$ or H_0 versus H_1 based on $LM_T(\cdot)$ or $LR_T(\cdot)$, we consider

$$(4.6) \qquad \sup_{\pi \in \Pi} LM_T(\pi) \quad \text{and} \quad \sup_{\pi \in \Pi} LR_T(\pi).$$

The null hypothesis H_0 is rejected for large values of these statistics.

5. ASYMPTOTIC PROPERTIES OF THE TEST STATISTICS

5.1. *Asymptotic Distributions under the Null Hypothesis*

This subsection provides the asymptotic null distributions of the test statistics introduced in Section 4.

[7] The simplification of $LM_T(\pi)$ from (4.3) to (4.4) occurs because the first-order conditions of the full-sample GMM estimator are $[\hat{M} \vdots \hat{M}_\delta]' \hat{S}^{-1}(1/T)\Sigma_1^T m(W_t, \tilde{\beta}, \tilde{\delta}) \cong 0$, where $\hat{M}_\delta = (1/T)\Sigma_1^T \partial m(W_t, \tilde{\beta}, \tilde{\delta})/\partial \delta'$. In consequence,

$$\begin{bmatrix} \hat{M}'\hat{S}^{-1} & 0 \\ 0 & \hat{M}'\hat{S}^{-1} \end{bmatrix} \overline{m}_T(\tilde{\theta},\pi) \cong \begin{pmatrix} \hat{M}'\hat{S}^{-1}\overline{m}_{1T}(\tilde{\theta},\pi) \\ -\hat{M}'\hat{S}^{-1}\overline{m}_{1T}(\tilde{\theta},\pi) \end{pmatrix}.$$

THEOREM 3: *Suppose Assumptions* 1–3 *hold. Given any set* Π *whose closure lies in* (0, 1), *the following processes indexed by* $\pi \in \Pi$ *satisfy*:

(a) $W_T(\cdot) \Rightarrow Q_p(\cdot)$ *and* $\sup_{\pi \in \Pi} W_T(\pi) \to_d \sup_{\pi \in \Pi} Q_p(\pi)$, *where* $Q_p(\pi) = (B_p(\pi) - \pi B_p(1))'(B_p(\pi) - \pi B_p(1))/[\pi(1-\pi)]$;

(b) $LM_T(\cdot) \Rightarrow Q_p(\cdot)$ *and* $\sup_{\pi \in \Pi} LM_T(\pi) \to_d \sup_{\pi \in \Pi} Q_p(\pi)$;

(c) $LR_T(\cdot) \Rightarrow Q_p(\cdot)$ *and* $\sup_{\pi \in \Pi} LR_T(\pi) \to_d \sup_{\pi \in \Pi} Q_p(\pi)$;

where $B_p(\cdot)$ *is a p-vector of independent Brownian motions on* [0, 1] *restricted to* Π. *The convergence in* (a)–(c) *holds jointly.*

COMMENTS: 1. The limit process $Q_p(\cdot)$ is referred to in the literature as the square of a standardized tied-down Bessel process of order p; see Sen (1981, p. 46). For any fixed $\pi \in (0,1)$, $Q_p(\pi)$ has a chi-square distribution with p degrees of freedom. Under the assumptions, the asymptotic null distribution of $\sup_{\pi \in \Pi} W_T(\pi), \ldots, \sup_{\pi \in \Pi} LR_T(\pi)$ is free of nuisance parameters except for the dimension p of β. Thus, critical values for the test statistics can be tabulated; see Section 5.3 below.

2. The requirement that Π is bounded away from zero and one is made to ensure that the estimators upon which the test statistics are based are uniformly consistent for $\pi \in \Pi$ and to ensure that the function mapping $B_p(\cdot)$ into $Q_p(\cdot)$ is continuous. For example, if $\Pi = [0,1]$, the functions $\pi \to 1/\pi$ and $\pi \to 1/(1-\pi)$ are not continuous. In fact, if $\Pi = [0,1]$, the test statistics $\sup_{\pi \in \Pi} W_T(\pi), \ldots, \sup_{\pi \in \Pi} LR_T(\pi)$ do not converge in distribution; see Corollary 1 below.

3. Theorem 3 establishes the asymptotic distributions of test statistics of the form $g(\{W_T(\pi)\colon \pi \in \Pi\})$ for arbitrary continuous functions g (using the uniform metric on the space of bounded cadlag Euclidean-valued functions on Π). In particular, $g(\{W_T(\pi)\colon \pi \in \Pi\}) \Rightarrow g(\{Q_p(\pi)\colon \pi \in \Pi\})$ under the assumptions and likewise for $LM_T(\cdot)$ and $LR_T(\cdot)$.

5.2. *Asymptotic Behavior of the Test Statistics When* $\Pi = [0,1]$

Next, we consider the limiting behavior under the null of the statistics $\sup_{\pi \in \Pi} W_T(\pi), \ldots, \sup_{\pi \in \Pi} LR_T(\pi)$ when $\Pi = [0,1]$. For the location model with iid $N(0,1)$ errors, D. M. Hawkins (1977) has already investigated this behavior (heuristically). In the general model scenario considered here, this behavior is determined using the results of Theorem 3. Note that Anderson and Darling (1952, Sec. 5) have considered a similar problem.

COROLLARY 1: *Suppose the conditions of Theorem* 3 *and the null hypothesis* H_0 *hold. Then,*

$$\sup_{\pi \in [0,1]} W_T(\pi) \to_p \infty, \qquad \sup_{\pi \in [0,1]} LM_T(\pi) \to_p \infty, \quad \textit{and}$$

$$\sup_{\pi \in [0,1]} LR_T(\pi) \to_p \infty.$$

COMMENTS: 1. The Corollary shows that the restriction in Theorem 3 to sets Π whose closure is in $(0,1)$ is not made only for technical convenience. Unless Π is bounded away from zero and one, critical values for the test statistics $\sup_{\pi \in \Pi} W_T(\pi), \ldots, \sup_{\pi \in \Pi} LR_T(\pi)$ must diverge to infinity as $T \to \infty$ to obtain a sequence of level α tests. By bounding π away from zero and one, however, a fixed critical value suffices for all T large. This suggests that the restriction of Π to a set whose closure is in $(0,1)$ yields significant power gains if the change point is in Π or is close to Π. Some Monte Carlo results of Talwar (1983) and James, James, and Siegmund (1987) for the location model substantiate this result. Furthermore, the Monte Carlo results of Talwar (1983) show that the test statistic $\sup_{\pi \in \Pi} W_T(\pi)$ has much closer true and nominal sizes in the location model under nonnormal errors when Π is restricted than when $\Pi = [0,1]$.

2. Suppose $\hat{\pi}$ maximizes $W_T(\pi)$, $LM_T(\pi)$, or $LR_T(\pi)$ over $[0,1]$. By Theorem 3 and Corollary 1, $\sup_{\pi \in [\varepsilon, 1-\varepsilon]} W_T(\pi) = O_p(1)$ $\forall \varepsilon > 0$, $\sup_{\pi \in [0,1]} W_T(\pi) \to_p \infty$ under the null hypothesis, and analogous results hold for $LM_T(\pi)$ and $LR_T(\pi)$. In consequence, $\hat{\pi} \to_p \{0,1\}$ under the null hypothesis. By symmetry, presumably, $\hat{\pi} \to_d \mathrm{Bern}(1/2)$, where $\mathrm{Bern}(1/2)$ denotes a Bernoulli distribution with parameter $1/2$. In contrast, if Π has closure in $(0,1)$ and $Q_p(\cdot)$ has a unique maximum on Π with probability one, then $\hat{\pi} \to_d \operatorname{argmax}\{Q_p(\pi)\colon \pi \in \Pi\}$ by the continuous mapping theorem. The latter distribution has support equal to Π.

5.3. *Asymptotic Critical Values*

Critical values c_α for the test statistics $\sup_{\pi \in \Pi} W_T(\pi), \ldots, \sup_{\pi \in \Pi} LR_T(\pi)$ are provided in Table I based on their asymptotic null distribution $\sup_{\pi \in \Pi} Q_p(\pi)$. By definition, c_α satisfies $P(\sup_{\pi \in \Pi} Q_p(\pi) > c_\alpha) = \alpha$. The table covers $\alpha = .01, .05$, and $.10$, $p = 1, 2, \ldots, 20$, and $\Pi = [\pi_0, 1 - \pi_0]$ for an array of π_0 values between .05 and .50.

Table I covers a much wider range of intervals Π, however, than just the symmetric intervals $[\pi_0, 1-\pi_0]$. If $\Pi = [\pi_1, \pi_2]$ for $0 < \pi_1 \leqslant \pi_2 < 1$, then it can be shown (see the proof of Corollary 1 in the Appendix) that

$$(5.1) \qquad P\Big(\sup_{\pi \in \Pi} Q_p(\pi) > c_\alpha\Big) = P\Big(\sup_{s \in [1, \pi_2(1-\pi_1)/(\pi_1(1-\pi_2))]} BM(s)' BM(s)/s > c_\alpha\Big),$$

where $BM(\cdot)$ denotes a p-vector of independent Brownian motion processes on $[0, \infty)$. In consequence, critical values based on the distribution of $\sup_{\pi \in [\pi_1, \pi_2]} Q_p(\pi)$ depend on π_1 and π_2 only through the parameter $\lambda = \pi_2(1 - \pi_1)/(\pi_1(1-\pi_2))$. Table I provides the value of λ corresponding to each value of π_0 considered (viz., $\lambda = (1-\pi_0)^2/\pi_0^2$). This allows one to obtain critical values for all intervals $\Pi = [\pi_1, \pi_2]$ whose corresponding value of $\lambda = \pi_2(1-\pi_1)/(\pi_1(1-\pi_2))$ either is tabulated or can be interpolated from the

TABLE I
Asymptotic Critical Values

		$p=1$			$p=2$			$p=3$			$p=4$			$p=5$		
π_0	λ	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
.50	1.00	2.71	3.84	6.63	4.61	5.99	9.21	6.25	7.81	11.34	7.78	9.49	13.28	9.24	11.07	15.09
.49	1.08	3.47	4.73	7.82	5.42	6.86	10.30	7.19	8.83	12.58	8.93	10.63	14.64	10.39	12.28	16.34
.48	1.17	3.79	5.10	8.26	5.80	7.31	10.71	7.64	9.29	13.05	9.42	11.17	15.17	10.96	12.88	16.83
.47	1.27	4.02	5.38	8.65	6.12	7.67	11.01	7.98	9.62	13.39	9.82	11.63	15.91	11.40	13.27	17.32
.45	1.49	4.38	5.91	9.00	6.60	8.11	11.77	8.50	10.15	14.23	10.35	12.27	16.64	12.05	14.00	18.06
.40	2.25	5.10	6.57	9.82	7.45	9.02	12.91	9.46	11.17	14.88	11.39	13.32	17.66	13.09	15.16	19.23
.35	3.45	5.59	7.05	10.53	8.06	9.67	13.53	10.16	12.05	15.71	12.10	14.12	18.54	13.86	15.93	19.99
.30	5.44	6.05	7.51	10.91	8.57	10.19	14.16	10.76	12.58	16.24	12.80	14.79	19.10	14.58	16.48	20.67
.25	9.00	6.46	7.93	11.48	9.10	10.75	14.47	11.29	13.16	16.60	13.36	15.34	19.78	15.17	17.25	21.39
.20	16.00	6.80	8.45	11.69	9.59	11.26	15.09	11.80	13.69	17.28	13.82	15.84	20.24	15.63	17.88	21.90
.15	32.11	7.17	8.85	12.35	10.01	11.79	15.51	12.27	14.15	17.68	14.31	16.45	20.71	16.20	18.35	22.49
.10	81.00	7.63	9.31	12.69	10.50	12.27	16.04	12.81	14.62	18.28	14.94	16.98	21.04	16.87	18.93	23.34
.05	361.00	8.19	9.84	13.01	11.20	12.93	16.44	13.47	15.15	19.06	15.62	17.56	21.54	17.69	19.61	24.18

		$p=6$			$p=7$			$p=8$			$p=9$			$p=10$		
π_0	λ	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
.50	1.00	10.64	12.59	16.81	12.02	14.07	18.48	13.36	15.51	20.09	14.68	16.92	21.67	15.99	18.31	23.21
.49	1.08	11.81	13.74	18.32	13.27	15.52	19.93	13.29	15.63	20.53	16.17	18.56	23.05	17.35	19.79	24.62
.48	1.17	12.42	14.45	19.12	13.92	16.14	20.64	13.89	16.31	21.14	16.82	19.25	23.83	18.08	20.35	25.75
.47	1.27	12.90	14.86	19.64	14.32	16.63	21.14	14.43	16.74	21.72	17.26	19.74	24.80	18.67	20.92	26.43
.45	1.49	13.53	15.59	20.45	14.97	17.38	22.32	15.05	17.53	22.28	18.10	20.59	25.52	19.39	21.78	27.30
.40	2.25	14.71	16.91	21.60	16.23	18.41	23.35	16.26	18.73	23.63	19.56	22.12	26.86	20.74	23.15	28.86
.35	3.45	15.56	17.75	22.33	17.09	19.34	24.10	17.06	19.46	24.64	20.49	22.93	27.77	21.87	24.17	29.76
.30	5.44	16.32	18.46	23.06	17.74	20.01	24.86	17.90	20.36	25.64	21.27	23.65	28.50	22.73	25.05	30.74
.25	9.00	17.00	19.07	23.65	18.38	20.63	25.11	18.61	20.95	26.10	21.93	24.31	29.23	23.32	25.80	31.32
.20	16.00	17.56	19.64	24.27	19.04	21.07	25.72	19.17	21.47	26.76	22.54	24.91	29.92	24.00	26.42	31.98
.15	32.11	18.12	20.26	24.79	19.69	21.84	26.23	19.82	22.13	27.25	23.15	25.47	30.52	24.62	27.03	32.33
.10	81.00	18.78	20.82	25.21	20.32	22.51	26.91	20.45	22.87	27.69	23.77	26.16	31.15	25.39	27.87	32.95
.05	361.00	19.49	21.56	25.96	21.02	23.22	27.53	21.23	23.60	28.77	24.64	26.94	31.61	26.24	28.63	33.86

		$p=11$			$p=12$			$p=13$			$p=14$			$p=15$		
π_0	λ	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
.50	1.00	17.28	19.68	24.73	18.55	21.03	26.22	19.81	22.36	27.69	21.06	23.68	29.14	22.31	25.00	30.58
.49	1.08	18.65	21.02	26.43	20.27	22.70	28.04	21.05	23.73	29.67	22.52	25.17	31.23	24.30	26.93	32.73
.48	1.17	19.40	21.81	27.04	21.01	23.54	28.64	21.87	24.53	30.23	23.45	26.11	32.11	25.13	27.82	33.55
.47	1.27	19.94	22.36	27.84	21.54	24.08	29.14	22.46	25.24	30.87	24.09	26.75	32.57	25.65	28.42	34.08
.45	1.49	20.70	23.27	28.47	22.42	24.99	30.03	23.31	25.95	31.62	24.98	27.65	33.39	26.59	29.25	35.16
.40	2.25	22.24	24.95	30.20	23.81	26.26	31.52	24.91	27.56	33.38	26.46	29.20	34.97	28.21	31.07	36.59
.35	3.45	23.25	25.93	31.14	24.80	27.45	32.46	26.00	28.77	34.65	27.52	30.30	36.10	29.23	32.11	37.32
.30	5.44	24.19	26.71	31.65	25.63	28.18	33.27	26.95	29.70	35.26	28.41	31.08	37.02	30.21	32.80	38.28
.25	9.00	24.82	27.43	32.39	26.37	28.98	33.77	27.64	30.48	35.85	29.31	31.83	37.25	30.93	33.65	38.93
.20	16.00	25.41	27.93	33.42	27.08	29.61	34.56	28.36	31.10	36.47	29.98	32.65	38.41	31.68	34.41	40.07
.15	32.11	26.14	28.55	33.75	27.74	30.16	35.07	29.13	31.80	37.04	30.67	33.45	39.22	32.51	35.06	40.56
.10	81.00	26.86	29.21	34.46	28.47	30.88	35.75	30.01	32.62	37.64	31.48	34.22	39.60	33.28	35.76	41.28
.05	361.00	27.91	30.15	35.14	29.31	31.76	36.70	30.90	33.42	38.44	32.48	35.00	40.10	34.25	36.74	42.05

		$p=16$			$p=17$			$p=18$			$p=19$			$p=20$		
π_0	λ	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
.50	1.00	23.54	26.30	32.00	24.76	27.58	33.43	25.99	28.87	34.81	27.19	30.14	36.21	28.41	31.41	37.57
.49	1.08	25.32	28.17	34.11	26.77	29.72	33.57	27.85	30.79	36.82	29.36	32.40	38.97	30.44	33.56	40.04
.48	1.17	26.22	29.05	35.09	27.65	30.58	36.78	28.76	31.57	37.70	30.18	33.20	39.93	31.39	34.62	40.88
.47	1.27	26.83	29.73	35.87	28.16	31.25	37.38	29.32	32.24	38.65	30.87	33.86	40.28	32.01	35.12	41.72
.45	1.49	27.75	30.59	36.88	29.14	32.18	38.25	30.38	33.17	39.31	31.95	34.82	41.40	33.09	36.41	42.97
.40	2.25	29.34	32.28	38.56	30.82	33.74	39.66	32.11	35.10	40.99	33.54	36.53	43.34	34.78	37.93	44.20
.35	3.45	30.56	33.40	39.18	31.96	34.86	40.81	33.40	36.23	41.84	34.71	37.79	44.30	36.04	39.21	45.34
.30	5.44	31.56	34.41	40.29	32.99	35.83	41.73	34.33	37.09	42.69	35.72	38.77	44.76	37.04	40.25	46.14
.25	9.00	32.33	35.19	41.07	33.85	36.71	42.46	35.27	37.94	43.68	36.60	39.75	45.66	37.92	40.98	46.79
.20	16.00	33.21	35.95	41.78	34.53	37.49	43.27	36.07	38.77	44.09	37.47	40.43	46.68	38.97	41.90	47.75
.15	32.11	33.90	36.66	42.40	35.39	38.12	43.95	36.85	39.55	44.84	38.36	41.25	47.25	39.96	43.00	48.40
.10	81.00	34.72	37.48	43.10	36.19	39.05	44.52	37.64	40.38	45.89	39.16	42.01	47.95	40.74	43.76	49.58
.05	361.00	35.81	38.51	44.20	37.24	40.05	45.39	38.68	41.36	46.69	40.27	43.05	49.02	41.83	44.52	50.76

table. The table covers values of λ between 1 and 361, so almost any interval of interest can be considered.

Note that $\|BM(\cdot)\|$ is a Bessel process of order p. In consequence, the probability given in (5.1) is the probability that a Bessel process exceeds a square root boundary somewhere in the given interval. Such probabilities and corresponding critical values for given significance levels have been computed numerically for $p \leqslant 4$ for a variety of λ values by DeLong (1981). In contrast, the critical values given here have been computed by simulation. They cover a considerably wider range of values of p and λ than those considered by DeLong.

The values reported in Table I are estimates of the critical values c_α obtained by (i) approximating the distribution of the supremum of $Q_p(\pi)$ over $\pi \in [\pi_0, 1-\pi_0]$ by its maximum over a fine grid of points $\Pi(N)$ and (ii) simulating the distribution of $\max_{\pi \in \Pi(N)} Q_p(\pi)$ by Monte Carlo. The grid $\Pi(N)$ is defined by

$$\Pi(N) = [\pi_0, 1-\pi_0] \cap \{\pi = j/N: j = 0, 1, \ldots, N\}. \tag{5.2}$$

The value of N was chosen to be 3,600 based on a comparison of the approximations obtained here with the numerical results of DeLong (1981), which are available for $p \leqslant 4$. A single realization from the distribution of $\max_{\pi \in \Pi(N)} Q_p(\pi)$ was obtained by simulating a p-vector $B_p(\cdot)$ of independent Brownian motions at the discrete points in $\Pi(N)$ and computing $\max_{\pi \in \Pi(N)} (B_p(\pi) - \pi B_p(1))'(B_p(\pi) - \pi B_p(1))/[\pi(1-\pi)]$. The number of repetitions R used was 10,000. The accuracy of the simulated critical values for approximating the critical values based on the statistic $\max_{\pi \in \Pi(N)} Q_p(\pi)$ can be determined by noting that the rejection probability of the statistic $\max_{\pi \in \Pi(N)} Q_p(\pi)$ using the simulated critical value has mean α and standard error approximately equal to $(\alpha(1-\alpha)/R)^{1/2}$. For $\alpha = .01$, .05, and .10, the standard errors due to simulation are .001, .002, and .003 respectively.

5.4. *Asymptotic Local Power*

In this section, we consider the behavior of $\hat{\theta}(\cdot), W_T(\cdot)$, etc. under sequences of local alternatives. We introduce the following assumption:

ASSUMPTION $1-LP$: *Assumption* 1 *holds but with the assumption in part* (*b*) *that* $Em_{Tt} = 0\ \forall t \leqslant T, T \geqslant 1$ *replaced by* $\sup_{\pi \in \Pi} \|\sqrt{T}\,\overline{m}_T(\theta_0, \pi) - \mu(\pi)\| = o_p(1)$ *for some nonrandom bounded* R^{2v}-*valued function* μ *on* Π.

We write $\mu(\pi) = (\mu_1(\pi)', \mu_2(\pi)')'$ for $\mu_1(\pi), \mu_2(\pi) \in R^v$.

In many cases, $\mu(\pi)$ can be expressed in more primitive terms. For example, suppose (i) Assumption $1-LP$ holds, (ii) $\{W_{Tt}: t \leqslant T, T \geqslant 1\}$ is such that $Em(W_{Tt}, \beta_0 + \eta(t/T)/\sqrt{T}, \delta_0) = 0\ \forall t \leqslant T, T \geqslant 1$, for some bounded R^p-valued function $\eta(\cdot)$ on $[0,1]$ that is Riemann integrable on $[0, \pi]$ uniformly over

$\pi \in \Pi \cup \{1\}$,[8] and (iii) $\max_{t \leqslant T} \sup_{\beta:\, \|\beta-\beta_0\| \leqslant K/\sqrt{T}} \|E(\partial/\partial\beta')m(W_{Tt}, \beta, \delta_0) - M\| \to 0$ as $T \to \infty$, where $K = \sup_{\pi \in \Pi} \|\eta(\pi)\|$. In this case,

$$(5.3)\qquad \mu(\pi) = \begin{pmatrix} \mu_1(\pi) \\ \mu_2(\pi) \end{pmatrix} = \begin{pmatrix} -M\int_0^{\pi} \eta(s)\, ds \\ -M\int_{\pi}^{1} \eta(s)\, ds \end{pmatrix}.$$

THEOREM 4: *Suppose Assumption* $1-LP$ *holds, Assumption* 2 *holds except in part* (a) *below, and Assumption* 3 *holds except in parts* (a) *and* (e) *below. Given any set* Π *whose closure lies in* $(0,1)$, *the following processes indexed by* $\pi \in \Pi$ *satisfy*:

(a) $\sqrt{T}(\hat{\theta}(\cdot) - \theta_0) \Rightarrow (M(\cdot)'\gamma(\cdot)M(\cdot))^{-1}M(\cdot)'\gamma(\cdot)(G(\cdot) - \mu(\cdot))$;

(b) $\sup_{\pi \in \Pi} \|\hat{V}_r(\pi) - V\| \to_p 0$ *for* $r = 1, 2$;

(c) $W_T(\cdot) \Rightarrow Q_p^*(\cdot) = J_p^*(\cdot)'J_p^*(\cdot)$ *and* $\sup_{\pi \in \Pi} W_T(\pi) \to_d \sup_{\pi \in \Pi} Q_p^*(\pi)$, *where*

$$J_p^*(\pi) = \frac{B_p(\pi) - \pi B_p(1)}{[\pi(1-\pi)]^{1/2}} - AS^{-1/2}\left(\left(\frac{1-\pi}{\pi}\right)^{1/2}\mu_1(\pi) - \left(\frac{\pi}{1-\pi}\right)^{1/2}\mu_2(\pi)\right);$$

(d) $LM_T(\cdot) \Rightarrow Q_p^*(\cdot)$ *and* $\sup_{\pi \in \Pi} LM_T(\pi) \to_d \sup_{\pi \in \Pi} Q_p^*(\pi)$;

(e) $LR_T(\cdot) \Rightarrow Q_p^*(\cdot)$ *and* $\sup_{\pi \in \Pi} LR_T(\pi) \to_d \sup_{\pi \in \Pi} Q_p^*(\pi)$;

where $B_p(\cdot)$ *is a p-vector of independent Brownian motions on* $[0,1]$ *restricted to* Π, $A = (CC')^{-1/2}C \in R^{p \times \nu}$, *and* $C = (M'S^{-1}M)^{-1}M'S^{-1/2} \in R^{p \times \nu}$. *If* $p = \nu$, *one can take* $A = I_p$. *The convergence in* (b)–(e) *holds jointly.*

COMMENTS: 1. The local power results of Theorem 4 are similar to those obtained by D. L. Hawkins (1987), but are more general. Hawkins' results cover the particular case of one-time structural change in which $\mu(\cdot)$ is as in (5.3) with $\eta(\cdot)$ of the form $\eta(\pi) = b1(\pi \leqslant \pi_0)$ for some fixed $\pi_0 \in (0,1)$ and some constant b. His results apply to ML estimators in iid contexts.

2. When $\mu(\cdot)$ satisfies (5.3), $Q_p^*(\cdot)$ depends on $\eta(\cdot)$ in the following way: $Q_p^*(\pi) = J_p^*(\pi)'J_p^*(\pi)$ and

$$(5.4)\qquad J_p^*(\pi) = \frac{B_p(\pi) - \pi B_p(1)}{[\pi(1-\pi)]^{1/2}} + AS^{-1/2}M\left(\left(\frac{1-\pi}{\pi}\right)^{1/2}\int_0^{\pi}\eta(s)\, ds - \left(\frac{\pi}{1-\pi}\right)^{1/2}\int_{\pi}^{1}\eta(s)\, ds\right).$$

[8] By definition, this means that η is Riemann integrable on $[0,\pi]$ $\forall \pi \in \Pi \cup \{1\}$ and $(1/T)\sum_1^{T\pi}\eta(t/T) \to \int_0^{\pi}\eta(s)\, ds$ uniformly over $\pi \in \Pi \cup \{1\}$ as $T \to \infty$.

3. For fixed $\pi \in (0,1)$, $Q_p^*(\pi)$ has a noncentral chi-square distribution with p degrees of freedom and noncentrality parameter given by the squared length of the second summand in the definition of $J_p^*(\pi)$.

4. By simulating the distribution of $\sup_{\pi \in \Pi} Q_p^*(\pi)$, the sensitivity of the power of the test considered here to the form of the alternative, as specified by $\mu(\cdot)$ or $\eta(\cdot)$, can be determined and the results hold asymptotically for a wide variety of models and estimators. For example, one can determine the effect of the location of the change point on the tests' power by simulating $\sup_{\pi \in \Pi} Q_p^*(\pi)$ with $\eta(\pi) = 1(\pi \leqslant \pi_0)$ for a variety of values of π_0. The results of Theorem 4 also can be used to compare the asymptotic power of the tests considered here for a wide variety of models with that of other tests in the literature, again by simulation.

5. The local power of the tests considered in Theorem 4 is the same whether δ_0 is estimated or is known.

The local power results of Theorem 4 can be used to show that the tests based on $\sup_{\pi \in \Pi} W_T(\pi), \ldots, \sup_{\pi \in \Pi} LR_T(\pi)$ each have nontrivial power against alternatives for which the parameter β_t is nonconstant on Π. These results are analogous to results obtained by Ploberger et al. (1989, Cor. 1) for the fluctuation test in the more restrictive context of testing for pure structural change in an iid linear regression model.

COROLLARY 2: *Suppose the assumptions of Theorem* 4(c) (*resp*. 4(d), 4(e)) *hold with* $\mu(\cdot)$ *as in* (5.3) *but with* $\eta(\cdot)$ *replaced by* $\xi\eta(\cdot)$. *Suppose* Π *is an interval whose closure lies in* (0, 1). *If* η *is not almost everywhere* (*Lebesgue*) *equal to a constant vector on* Π, *then*

$$\lim_{\xi \to \infty} \lim_{T \to \infty} P\Big(\sup_{\pi \in \Pi} W_T(\pi) > c_\alpha\Big) = 1$$

(*resp*.

$$\lim_{\xi \to \infty} \lim_{T \to \infty} P\Big(\sup_{\pi \in \Pi} LM_T(\pi) > c_\alpha\Big) = 1,$$

$$\lim_{\xi \to \infty} \lim_{T \to \infty} P\Big(\sup_{\pi \in \Pi} LR_T(\pi) > c_\alpha\Big) = 1),$$

where c_α *is as defined above and* $\alpha \in (0,1)$.

Next, using Theorem 4, we can establish a weak optimality result for the test statistics $\sup_{\pi \in \Pi} W_T(\pi), \ldots, \sup_{\pi \in \Pi} LR_T(\pi)$ for testing against the alternatives in $\bigcup_{\pi \in \Pi} H_{1T}(\pi)$. This result is a generalization to multiparameter two-sided tests of a result of Davies (1977, Thm. 4.2) for scalar parameter one-sided tests. The result shows that as the significance level α goes to zero, the power against all local alternatives of the level α test based on $\sup_{\pi \in \Pi} W_T(\pi)$ is at least as large as that of the level α test based on $W_T(\tilde{\pi})$ for any fixed $\tilde{\pi} \in \Pi$. Thus, if $W_T(\tilde{\pi})$ possesses asymptotic local power optimality properties against certain alternatives, e.g., as it does in the ML case against one-time structural changes (i.e., for $\eta(s)=0$ for $s<\tilde{\pi}$, $\eta(s)=\delta$ for $s \geqslant \tilde{\pi}$), then $\sup_{\pi \in \Pi} W_T(\pi)$ inherits these properties as $\alpha \to 0$. The same also holds for $\sup_{\pi \in \Pi} LM_T(\pi)$ and $\sup_{\pi \in \Pi} LR_T(\pi)$.

THEOREM 5: *Let η denote a bounded R^p-valued function on* $[0,1]$ *that is Riemann integrable on* $[0,\pi]$ *uniformly over* $\pi \in \Pi \cup \{1\}$. *Let Ξ denote the set of all such functions η for which there exists a distribution P_η of the triangular array* $\{W_{Tt}\colon t \leqslant T,\ T \geqslant 1\}$ *such that* $1-LP$, 2, *and* 3 *hold with* $\mu(\cdot)$ *as in* (5.3). *Then,*

$$\lim_{\alpha \to 0} \inf_{\eta \in \Xi} \inf_{\tilde{\pi} \in \Pi} \lim_{T \to \infty} \left[P_\eta\Big(\sup_{\pi \in \Pi} W_T(\pi) > c_\alpha \Big) - P_\eta\big(W_T(\tilde{\pi}) > \tilde{c}_\alpha\big)\right] \geqslant 0, \tag{5.5}$$

where c_α and $\tilde{c}_\alpha$ are such that the tests based on $\sup_{\pi \in \Pi} W_T(\pi)$ and $W_T(\tilde{\pi})$ have asymptotic level $\alpha \in (0,1)$. The result (5.5) *also holds with $W_T(\cdot)$ replaced by $LM_T(\cdot)$ or $LR_T(\cdot)$.*

COMMENT: The optimality result (5.5) is referred to above as a *weak* result because it appears that α must be quite small before the result is illustrative of finite sample behavior of the test statistics $\sup_{\pi \in \Pi} W_T(\pi)$ and $W_T(\tilde{\pi})$. Nevertheless, the result does serve to indicate that as α decreases the difference decreases between the power function of the level α test based on $\sup_{\pi \in \Pi} W_T(\pi)$ and the envelope of the power functions of the level α tests based on $W_T(\tilde{\pi})$ for fixed $\tilde{\pi} \in \Pi$.

6. CONCLUDING COMMENTS

1. The tests discussed in this paper are asymptotic in general. Nevertheless, exact versions of them can be obtained in some cases. In particular, consider a linear regression model with fixed regressors and iid normal errors. In this case, the sup Wald test statistic based on the least squares estimator has null distribution that is invariant with respect to the regression and variance parameter values. In consequence, one can set the regression parameters equal to zero and the error variance equal to one and generate exact critical values by simulating the resultant model. Since least squares regressions are very quick to compute, this procedure is not very burdensome computationally. See Andrews, Lee, and Ploberger (1992) for further details.

2. The basic Assumption 1 employed above utilizes the concept of near epoch dependence. This assumption can be simplified if the underlying random variables $\{W_t\colon t = \ldots, 0, 1, \ldots\}$ are stationary and ergodic. In particular, it can be shown that Assumption 1 can be replaced by the following assumption and Theorems 1–5 still hold.[9] Let $\mathscr{F}_s$ denote the σ-field generated by $\{W_t\colon -\infty < t \leqslant s\}$. Let m_t, B_0, and Δ_0 be as in Section 3.2.

[9] The invariance principle used to show this is given by Hall and Heyde (1980, Cor. 5.4, p. 145).

ASSUMPTION 1*: (a) $\{W_t: t = \ldots, 0, 1, \ldots\}$ *is stationary and ergodic*;

(b) $Em_t = 0, Em_t' m_t < \infty$, $\Sigma_{s=1}^{\infty}(E\|E(m_t|\mathscr{F}_{t-s})\|^2)^{1/2} < \infty$, *and* S *is positive definite*;

(c) $m(w, \beta, \delta)$ *is continuously partially differentiable in* (β, δ) $\forall(\beta, \delta) \in B_0 \times \Delta_0$, $\forall w \in W_0 \subset W$ *for a Borel measurable set* W_0 *that satisfies* $P(W_t \in W_0) = 1$, $m(w, \beta, \delta)$ *and* $\partial m(w, \beta, \delta)/\partial(\beta', \delta')$ *are Borel measurable functions of* w *for each* $(\beta, \delta) \in B_0 \times \Delta_0$, *and* $E \sup_{(\beta, \delta) \in B_0 \times \Delta_0} \|\partial m(W_t, \beta, \delta)/\partial(\beta', \delta')\| < \infty$;

(d) *Assumptions* 1(d), (e), *and* (h) *hold*.

3. In the event that a test for structural change rejects the null hypothesis, it may be of interest to estimate the parametric model defined by the restricted alternative $\bigcup_{\pi \in \Pi} H_{1T}(\pi)$. This involves estimating the time of change parameter π. Properties of the maximum likelihood (ML) estimator of π have been considered by Hinkley (1970), Picard (1983, 1985), Deshayes (1983), Bai (1991), and Bai, Lumsdaine, and Stock (1991) for a variety of models. No optimality properties are known for the ML estimator of π.

Cowles Foundation for Research in Economics, Yale University, P.O. Box 2125 Yale Station, New Haven, CT 06520, U.S.A.

Manuscript received June, 1991; final revision received November, 1992.

APPENDIX

For notational simplicity, we say $X_T(\pi) = o_{p\pi}(1)$ if $\sup_{\pi \in \Pi} \|X_T(\pi)\| = o_p(1)$ and we say $X_T(\pi) = O_{p\pi}(1)$ if $\sup_{\pi \in \Pi} \|X_T(\pi)\| = O_p(1)$.

First we provide conditions under which the PS-GMM estimator $\hat{\theta}(\cdot)$ is consistent for θ_0 uniformly over $\pi \in \Pi$ under the null hypothesis.

ASSUMPTION A: (a) *Assumption* 1(a) *holds*.

(b) $\sup_{\pi \in \Pi} \|\hat{\gamma}(\pi) - \gamma(\pi)\| \to_p 0$ *for some symmetric* $2\nu \times 2\nu$ *matrices* $\gamma(\pi)$ *for which* $\sup_{\pi \in \Pi} \|\gamma(\pi)\| < \infty$.

(c) B *and* Δ *are bounded subsets of* R^p *and* R^q *respectively*.

(d) $m(w, \beta, \delta)$ *is continuous in* w *for all* $(\beta, \delta) \in B \times \Delta$ *and is continuous in* (β, δ) *uniformly over* $(\beta, \delta, w) \in B \times \Delta \times C$ *for all compact sets* $C \subset W$.

(e) $\overline{\lim}_{T \to \infty}(1/T)\Sigma_1^T E \sup_{(\beta, \delta) \in B \times \Delta} |m(W_{Tt}, \beta, \delta)|^{1+\varepsilon} < \infty$ *for some* $\varepsilon > 0$.

(f) $\lim_{T \to \infty}(1/T)\Sigma_1^{T\pi} Em(W_{Tt}, \beta, \delta)$ *exists uniformly over* $(\beta, \delta, \pi) \in B \times \Delta \times \Pi$ *and equals* $\pi \lim_{T \to \infty}(1/T)\Sigma_1^T Em(W_{Tt}, \beta, \delta)$.

(g) $\bar{m}(\beta_0, \delta_0) = 0$, *where* $\bar{m}(\beta, \delta) = \lim_{T \to \infty}(1/T)\Sigma_1^T Em(W_{Tt}, \beta, \delta)$, *and for every neighborhood* Θ_0 $(\subset \Theta)$ *of* θ_0, $\inf_{\pi \in \Pi} \inf_{\theta \in \Theta/\Theta_0} m(\theta, \pi)' \gamma(\pi) m(\theta, \pi) > 0$, *where* $m(\theta, \pi) = (\pi \bar{m}(\beta_1, \delta)', (1 - \pi)\bar{m}(\beta_2, \delta)')'$.

When Assumptions 2 and 3 hold, Assumption A(b) automatically holds and A(g) simplifies to: $\bar{m}(\beta_0, \delta_0) = 0$ and for all neighborhoods B_0 and Δ_0 of β_0 and Δ_0, respectively, $\inf_{(\beta, \delta) \in B \times \Delta / B_0 \times \Delta_0} \bar{m}(\beta, \delta)' S^{-1} \bar{m}(\beta, \delta) > 0$.

THEOREM A1: *Under Assumption A, the PS-GMM estimator* $\hat{\theta}(\cdot)$ *satisfies* $\sup_{\pi \in \Pi} \|\hat{\theta}(\pi) - \theta_0\| \to_p 0$ *for any set* Π *whose closure lies in* $(0, 1)$.

COMMENT: To obtain consistency of the full-sample GMM estimator $\tilde{\theta}$, Assumption A only needs to be altered slightly. Consider the condition $A(b')$: $\hat{\gamma} \to_p \gamma$ for some nonsingular symmetric $\nu \times \nu$ matrix γ, where $\hat{\gamma}$ is the weight matrix of the full-sample GMM estimator. We have $\tilde{\theta} \to_p \theta_0$ if Assumption A holds with A(b) replaced by A(b′) and with A(g) replaced by the simplified version of A(g) given above except with γ in place of S^{-1}.

The proof of Theorem A1 uses the following three lemmas (the latter two of which are also used in the proofs of other results below).

LEMMA A1: *Suppose* $\hat{\theta}(\pi)$ *minimizes a random real function* $Q_T(\theta,\pi)$ *over* $\theta \in \Theta \subset R^{2p+q}$ *for each* $\pi \in \Pi$ *with probability* $\to 1$. *If* (a) $\sup_{\pi \in \Pi,\, \theta \in \Theta} |Q_T(\theta,\pi) - Q(\theta,\pi)| \to_p 0$ *for some real function* Q *on* $\Theta \times \Pi$ *and* (b) *for every neighborhood* Θ_0 $(\subset \Theta)$ *of* θ_0, $\inf_{\pi \in \Pi}(\inf_{\theta \in \Theta/\Theta_0} Q(\theta,\pi) - Q(\theta_0,\pi)) > 0$, *then* $\sup_{\pi \in \Pi} \|\hat{\theta}(\pi) - \theta_0\| \to_p 0$.

LEMMA A2: *Suppose* $\{X_{Tt}\colon t \leqslant T,\ T \geqslant 1\}$ *is a triangular array of mean zero real-valued rv's that is* L^0*-NED with respect to a strong mixing base* $\{Y_{Tt}\colon t = \ldots,0,1,\ldots;\ T \geqslant 1\}$ *and* $\overline{\lim}_{T\to\infty}(1/T)\Sigma_1^T E|X_{Tt}|^{1+\varepsilon} < \infty$ *for some* $\varepsilon > 0$. *Then*, $E\sup_{S \leqslant T}|(1/T)\Sigma_1^S X_{Tt}| \to 0$ *as* $T \to \infty$.

LEMMA A3: *Suppose* (a) *Assumption* 1(a) *holds*, (b) Λ *is a bounded subset of* R^s, (c) $f(w,\lambda)$ *is an* R^c*-valued function on* $W \times \Lambda$ *that is continuous in* w *for all* $\lambda \in \Lambda$ *and is continuous in* λ *uniformly over* $(\lambda, w) \in \Lambda \times C$ *for all compact sets* $C \subset W$, *and* (d) $\overline{\lim}_{T\to\infty}(1/T)\Sigma_1^T E\sup_{\lambda\in\Lambda}|f(W_{Tt},\lambda)|^{1+\varepsilon} < \infty$ *for some* $\varepsilon > 0$. *Then*,

$$\sup_{\lambda\in\Lambda}\sup_{S\leqslant T}\left|\frac{1}{T}\sum_1^S\left[f(W_{Tt},\lambda) - Ef(W_{Tt},\lambda)\right]\right| \to_p 0.$$

PROOF OF THEOREM A1: We apply Lemma A1 with $Q_T(\theta,\pi) = \overline{m}_T(\theta,\pi)'\hat{\gamma}(\pi)\overline{m}_T(\theta,\pi)$ and $Q(\theta,\pi) = m(\theta,\pi)'\gamma(\pi)m(\theta,\pi)$. Condition (b) of Lemma A1 holds by (Assumption) A(g). Given A(b), condition (a) of Lemma A1 holds if

$$\text{(A.1)} \qquad \sup_{\pi\in\Pi}\sup_{\theta\in\Theta}\|\overline{m}_T(\theta,\pi) - m(\theta,\pi)\| \to_p 0.$$

Using $\Sigma_{T\pi+1}^T = \Sigma_1^T - \Sigma_1^{T\pi}$, the latter holds if

$$\text{(A.2)} \qquad \sup_{(\beta,\delta)\in B\times\Delta}\ \sup_{T\pi_1\leqslant S\leqslant T}\left|\frac{1}{T}\sum_1^S\left[m(W_t,\beta,\delta) - Em(W_t,\beta,\delta)\right]\right| \to_p 0 \quad \text{and}$$

$$\text{(A.3)} \qquad \sup_{(\beta,\delta)\in B\times\Delta}\ \sup_{\pi\in\Pi}\left|\frac{1}{T}\sum_1^{T\pi}\left[Em(W_t,\beta,\delta) - \tilde{m}(\beta,\delta)\right]\right| \to 0,$$

where $\pi_1 = \inf\{\pi\colon \pi \in \Pi\} > 0$. (A.2) holds by Lemma A3 under A(a)–(e). (A.3) holds by A(f). *Q.E.D.*

PROOF OF LEMMA A1: By Assumption (b), given any neighborhood Θ_0 of θ_0, there exists a constant $\varepsilon > 0$ such that $\inf_{\pi\in\Pi}[\inf_{\theta\in\Theta/\Theta_0} Q(\theta,\tilde{\pi}) - Q(\theta_0,\tilde{\pi})] \geqslant \varepsilon > 0$. Thus,

$$\begin{aligned}\text{(A.4)} \qquad & P\big(\hat{\theta}(\pi) \in \Theta/\Theta_0 \quad \text{for some } \pi \in \Pi\big)\\ &\leqslant P\Big(\inf_{\tilde{\pi}\in\Pi}\left[Q\big(\hat{\theta}(\pi),\tilde{\pi}\big) - Q(\theta_0,\tilde{\pi})\right] \geqslant \varepsilon \quad \text{for some } \pi \in \Pi\Big)\\ &\leqslant P\big(Q\big(\hat{\theta}(\pi),\pi\big) - Q(\theta_0,\pi) \geqslant \varepsilon \quad \text{for some } \pi \in \Pi\big) \to 0,\end{aligned}$$

where "$\to 0$" holds provided $\sup_{\pi\in\Pi}|Q(\hat{\theta}(\pi),\pi) - Q(\theta_0,\pi)| \to_p 0$. Using Assumptions (a) and (b), the latter follows from

$$
\begin{aligned}
\text{(A.5)}\qquad 0 &\leqslant \inf_{\pi\in\Pi}\left[Q(\hat{\theta}(\pi),\pi)-Q(\theta_0,\pi)\right] \leqslant \sup_{\pi\in\Pi}\left[Q(\hat{\theta}(\pi),\pi)-Q(\theta_0,\pi)\right]\\
&\leqslant \sup_{\pi\in\Pi}\left[Q(\hat{\theta}(\pi),\pi)-Q_T(\hat{\theta}(\pi),\pi)\right]+\sup_{\pi\in\Pi}\left[Q_T(\hat{\theta}(\pi),\pi)-Q(\theta_0,\pi)\right]\\
&\leqslant \sup_{\pi\in\Pi}\left[Q(\hat{\theta}(\pi),\pi)-Q_T(\hat{\theta}(\pi),\pi)\right]+\sup_{\pi\in\Pi}\left[Q_T(\theta_0,\pi)-Q(\theta_0,\pi)\right]\\
&\leqslant 2\sup_{\pi\in\Pi,\,\theta\in\Theta}\left|Q_T(\theta,\pi)-Q(\theta,\pi)\right| \to_p 0. \qquad\qquad \textit{Q.E.D.}
\end{aligned}
$$

Proof of Lemma A2: Under the moment conditions, $\{X_{Tt}\}$ is L^1-NED by Theorem 6.1 of Pötscher and Prucha (1991). This property and inequalities (2) and (3) of Andrews (1988) show that (i) $\overline{\lim}_{T\to\infty}(1/T)\Sigma_1^T|E(X_{Tt}|Y_{T1},\ldots,Y_{Tt-m})| \to 0$ as $m\to\infty$ and (ii) $\overline{\lim}_{T\to\infty}(1/T)\Sigma_1^T|X_{Tt}-E(X_{Tt}|Y_{Tt+m},\ldots,Y_{TT})| \to 0$ as $m\to\infty$. Conditions (i) and (ii) are a slightly weaker version of the L^1-mixingale condition of Andrews (1988) with constants $c_t = 1$ and σ-fields F_{Tt} given by those generated by $\{Y_{Tt}\}$. Theorem 1 of Andrews (1988) holds with the L^1-mixingale condition replaced by (i) and (ii). In fact, the conclusion of Theorem 1 can be strengthened from $E|(1/T)\Sigma_1^T X_{Tt}| \to 0$ to $E\sup_{S\leqslant T}|(1/T)\Sigma_1^S X_{Tt}| \to 0$ as $T\to\infty$ with some alterations in its proof. This gives the result of Lemma A2. The alterations in the proof of Theorem 1 include changing $(1/n)\Sigma_{i=1}^n\cdot$ to $\sup_{j\leqslant n}|(1/n)\Sigma_{i=1}^j\cdot|$ and $E|\overline{Y}_{mn}|$ to $E\sup_{j\leqslant n}|\overline{Y}_{mj}|$ in equation (7) and strengthening the result of the Lemma in Andrews (1988) from $\|\overline{Y}_n\|_p \to 0$ as $n\to\infty$ to $\|\sup_{j\leqslant n}|(j/n)\overline{Y}_j|\,\|_p \to 0$ as $n\to\infty$. To achieve the latter, the proof of the Lemma needs to be changed by replacing $(1/n)\Sigma_{i=1}^n\cdot$ by $\sup_{j\leqslant n}|(1/n)\Sigma_{i=1}^j\cdot|$ throughout and by using the martingale inequality $\|\sup_{j\leqslant n}|(1/n)\Sigma_{i=1}^j(W_i - E(W_i|\mathscr{F}_{i-1}))|\,\|_2 \leqslant 2\|(1/n)\Sigma_{i=1}^n(W_i-E(W_i|\mathscr{F}_{i-1}))\|_2$ due to Doob (see Theorem 2.2 of Hall and Heyde (1980, p. 15)) in equation (4). *Q.E.D.*

Proof of Lemma A3: The desired result follows from Theorem 1 of Andrews (1992) with $G_n(\theta)$ set equal to $\sup_{S\leqslant T}|(1/T)\Sigma_1^S[f(W_{Tt},\lambda)-Ef(W_{Tt},\lambda)]|$. Hence, it suffices to verify the conditions BD, P-WCON, and SE of Theorem 1. BD holds by Assumption (b). For $G_n(\theta)$ as above, the proof of Lemma 3 of Andrews (1992) shows that DM and TSE imply SE. DM holds by Assumption (d). By Lemma 4(b) of Andrews (1992), TSE-2 implies TSE. TSE-2 holds by Assumptions (a) and (c). It remains to show P-WCON. Under Assumptions (a), (c), and (d), Theorem 6.5 of Pötcher and Prucha (1991) implies that $\{f(W_{Tt},\lambda)\colon t\leqslant T,\ T\geqslant 1\}$ is L^0-approximable by the base $\{Y_{Tt}\}$ for all $\lambda\in\Lambda$. By Assumption (d) and Theorem 6.1 of Pötscher and Prucha (1991), the approximators can be taken to be the conditional means $\{E(f(W_{Tt},\lambda)|Y_{Tt-m},\ldots,Y_{Tt+m})\colon t\leqslant T,\ T\geqslant 1,\ m\geqslant 1\}$. In consequence, $\{f(W_{Tt},\lambda)\colon t\leqslant T,\ T\geqslant 1\}$ is L^0-NED on the strong mixing base $\{Y_{Tt}\}$ for all $\lambda\in\Lambda$. We now apply Lemma A2 with X_{Tt} equal to an element of the c-vector $f(W_{Tt},\lambda)-Ef(W_{Tt},\lambda)$ to obtain P-WCON. *Q.E.D.*

The following lemma is used in the proof of Theorem 1.

Lemma A4: *Let $\{\eta_T(\cdot)\colon T\geqslant 0\}$ be a sequence of random elements of the space of bounded R^v-valued cadlag functions on a set $\Lambda\subset[0,1]$. If* (i) $\alpha'\eta_T(\cdot)\Rightarrow\alpha'\eta_0(\cdot)\ \forall\alpha\in R^v$ *and* (ii) $\{\eta_T(\cdot)\colon T\geqslant 1\}$ *has asymptotically independent increments* (*as defined, e.g., by Billingsley* (1968, *p.* 157)), *then* $\eta_T(\cdot)\Rightarrow\eta_0(\cdot)$.

(Note that Prop. 4.1 of Wooldridge and White (1988), which claims that condition (i) alone is sufficient for a multivariate invariance principle, is not correct. Their proposition cannot be derived in the manner they suggest.)

Proof of Theorem 1: Since $\hat{\theta}(\pi)$ minimizes $\overline{m}_T(\theta,\pi)'\hat{\gamma}(\pi)\overline{m}_T(\theta,\pi)$ and $\hat{\theta}(\pi)$ is in the interior of Θ $\forall\pi\in\Pi$ with probability $\to 1$ by (Assumption) 1(d), we have

$$
\text{(A.6)}\qquad \left[\frac{\partial\overline{m}_T(\hat{\theta}(\pi),\pi)}{\partial\theta'}\right]'\hat{\gamma}(\pi)\sqrt{T}\,\overline{m}_T(\hat{\theta}(\pi),\pi) = o_{p\pi}(1).
$$

Let $\overline{m}_{Tj}(\theta,\pi)$ denote the jth element of $\overline{m}_T(\theta,\pi)$. A mean value expansion of $\sqrt{T}\overline{m}_{Tj}(\hat{\theta}(\pi),\pi)$ about θ_0 gives: $\forall j = 1,\ldots,2\nu$,

$$\text{(A.7)} \qquad \sqrt{T}\overline{m}_{Tj}(\hat{\theta}(\pi),\pi) = \sqrt{T}\overline{m}_{Tj}(\theta_0,\pi) + \frac{\partial \overline{m}_{Tj}(\theta^*,\pi)\sqrt{T}(\hat{\theta}(\pi)-\theta_0)}{\partial\theta'},$$

where $\theta^*(=\theta_j^*(\pi))$ is a rv on the line segment joining $\hat{\theta}(\pi)$ and θ_0 (see Jennrich (1969, Lemma 3) for the mean value theorem for random functions). The latter property and 1(d) imply that $\theta^* = \theta_0 + o_{p\pi}(1)$.

Below we show that

$$\text{(A.8)} \qquad \sup_{\pi\in\Pi}\left\|\frac{\partial\overline{m}_T(\theta^*(\pi),\pi)}{\partial\theta'} - M(\pi)\right\| \to_p 0$$

whenever $\theta^*(\pi)$ satisfies $\sup_{\pi\in\Pi}\|\theta^*(\pi)-\theta_0\| \to_p 0$. We also show that

$$\text{(A.9)} \qquad \sqrt{T}\overline{m}_T(\theta_0,\cdot) \Rightarrow G(\cdot)$$

as a process indexed by $\pi\in\Pi$. Equations (A.6)–(A.9), 1(e) and (h), and the continuous mapping theorem (see Pollard (1984, Thm. IV.12, p. 70)) combine to give the desired result:

$$\text{(A.10)} \qquad \begin{aligned}\sqrt{T}(\hat{\theta}(\cdot)-\theta_0) &= -(M(\cdot)'\gamma(\cdot)M(\cdot))^{-1}M(\cdot)'\gamma(\cdot)\sqrt{T}\overline{m}_T(\theta_0,\cdot) + o_{p\pi}(1)\\ &\Rightarrow -(M(\cdot)'\gamma(\cdot)M(\cdot))^{-1}M(\cdot)'\gamma(\cdot)G(\cdot).\end{aligned}$$

To establish (A.8), we write

$$\text{(A.11)} \qquad \begin{aligned}&\sup_{\pi\in\Pi}\left\|\frac{\partial\overline{m}_T(\theta^*(\pi),\pi)}{\partial\theta'} - M(\pi)\right\|\\ &\leqslant \sup_{\pi\in\Pi}\left\|\frac{\partial\overline{m}_T(\theta^*(\pi),\pi)}{\partial\theta'} - E\frac{\partial\overline{m}_T(\theta,\pi)}{\partial\theta'}\bigg|_{\theta=\theta^*(\pi)}\right\|\\ &\quad + \sup_{T\geqslant 1}\sup_{\pi\in\Pi}\left\|E\frac{\partial\overline{m}_T(\theta,\pi)}{\partial\theta'}\bigg|_{\theta=\theta^*(\pi)} - E\frac{\partial\overline{m}_T(\theta_0,\pi)}{\partial\theta'}\right\|\\ &\quad + \sup_{\pi\in\Pi}\left\|E\frac{\partial\overline{m}_T(\theta_0,\pi)}{\partial\theta'} - M(\pi)\right\|.\end{aligned}$$

The third summand on the right-hand side of (A.11) $\to_p 0$ by 1(g). The first summand $\to_p 0$ because Assumption 1 and Lemma A4 yield

$$\text{(A.12)} \qquad \sup_{\pi\in\Pi}\sup_{\theta\in\Theta_0}\left\|\frac{\partial\overline{m}_T(\theta,\pi)}{\partial\theta'} - E\frac{\partial\overline{m}_T(\theta,\pi)}{\partial\theta'}\right\| \to_p 0.$$

Finally, the second summand on the right-hand side of (A.11) $\to_p 0$, because (i) by the tightness of $\{\overline{\mu}_T\colon T\geqslant 1\}$, $\sup_{T\geqslant 1}(1/T)\Sigma_1^T P(W_{Tt}\notin C_j)\to 0$ as $j\to\infty$ for some sequence of compact sets $\{C_j\colon j\geqslant 1\}$ in W; (ii) for all $j\geqslant 1$, we have

$$\text{(A.13)} \qquad \begin{aligned}&\sup_{T\geqslant 1}\sup_{\pi\in\Pi}\left\|\frac{1}{T}\sum_1^{T\pi}E\left[\frac{\partial m(W_{Tt},\beta,\delta)}{\partial(\beta',\delta')} - \frac{\partial m(W_{Tt},\beta_0,\delta_0)}{\partial(\beta',\delta')}\right]1(W_{Tt}\in C_j)\right\|\\ &\leqslant \sup_{w\in C_j}\left\|\frac{\partial m(w,\beta,\delta)}{\partial(\beta',\delta')} - \frac{\partial m(w,\beta_0,\delta_0)}{\partial(\beta',\delta')}\right\| \to 0\end{aligned}$$

as $(\beta,\delta)\to(\beta_0,\delta_0)$ using 1(f); and (iii) results (i) and (ii) combine to give

$$\text{(A.14)} \qquad \sup_{T\geqslant 1}\sup_{\pi\in\Pi}\left\|\frac{1}{T}\sum_1^{T\pi}E\left[\frac{\partial m(W_{Tt},\beta,\delta)}{\partial(\beta',\delta')} - \frac{\partial m(W_{Tt},\beta_0,\delta_0)}{\partial(\beta',\delta')}\right]\right\| \to 0$$

as $(\beta,\delta)\to(\beta_0,\delta_0)$. Thus, the right-hand side of (A.11) $\to_p 0$ and (A.8) is established.

Next, to show (A.9), let $\nu_T(\pi) = (1/\sqrt{T})\Sigma_1^{T\pi} m_t$. Since $\sqrt{T} m_T(\theta_0, \pi) = (\nu_T(\pi)', \nu_T(1)' - \nu_T(\pi)')'$, (A.9) follows from $\nu_T(\cdot) \Rightarrow S^{1/2}B(\cdot)$. To obtain the latter, we apply Lemma A4. Under 1(a)–(c), we obtain condition (i) of Lemma A4 by Cor. 3.1 of Wooldridge and White (1988), which utilizes results of McLeish (1977). (Note that Cor. 3.1 yields weak convergence of the standard partial sum process in $D[0,1]$ with the Skorokhod metric and the Borel σ-field generated by it. This can be converted into weak convergence in $D[0,1]$ with the uniform metric and the σ-field generated by the closed balls under the uniform metric as follows. The result of Cor. 3.1 implies weak convergence of the smoothed partial sum process (i.e., $\alpha'\nu_T(\pi) + (T\pi - [T\pi])\alpha' m_{[T\pi]+1}/\sqrt{T}$) using the Skorokhod metric on $D[0,1]$, because the difference between the standard and the smoothed partial sum processes is $\leqslant \sup_{t \leqslant T} |\alpha' m_t| / \sqrt{T}$ and the latter is $o_p(1)$ by the Lindeberg condition; see Hall and Hyde (1980, p. 53). Since the smoothed process is in $C[0,1]$, the Skorokhod and uniform metrics are equivalent for $C[0,1]$, and the Borel σ-field and the σ-field generated by the closed balls under the uniform metric are equivalent for $C[0,1]$, the smoothed partial sum process converges weakly as a sequence of random elements of $C[0,1]$ with the uniform metric and its Borel σ-field. This yields the desired univariate invariance principle for the standard partial sum process, $\alpha'\nu_T(\pi)$, in $D[0,1]$ with the uniform metric and the σ-field generated by the closed balls under the uniform metric, because the difference between these two processes is less than or equal to $\sup_{t \leqslant T} |\alpha' m_t| / \sqrt{T} = o_p(1)$.

To obtain condition (ii) of Lemma A4, it suffices to show that

$$\text{(A.15)} \qquad \begin{pmatrix} \nu_T(\pi_2) - \nu_T(\pi_1) \\ \nu_T(\pi_0) \end{pmatrix} \to_d N\left(0, \begin{pmatrix} (\pi_2 - \pi_1)S & 0 \\ 0 & \pi_0 S \end{pmatrix}\right) \quad \forall 0 \leqslant \pi_0 < \pi_1 \leqslant \pi_2 \leqslant 1.$$

By the Cramér-Wold device, the latter holds if

$$\text{(A.16)} \qquad \alpha_1'\nu_T(\pi_2) - \alpha_1'\nu_T(\pi_1) + \alpha_2'\nu_T(\pi_0) \to_d N(0, (\pi_2 - \pi_1)\alpha_1' S \alpha_1 + \pi_0 \alpha_2' S \alpha_2) \qquad \forall \alpha_1, \alpha_2 \in R^{\nu}.$$

(Note that this result is not implied by $\alpha'\nu_T(\cdot) \Rightarrow \alpha'\nu(\cdot)\ \forall \alpha \in R^{\nu}$.) To obtain (A.16), the same central limit theorem as used above, viz. Cor. 3.1 of Wooldridge and White (1988), can be employed. *Q.E.D.*

PROOF OF LEMMA A4: Conditions (i) and (ii) are sufficient because (a) tightness of $\{\alpha'\eta_T(\cdot): T \geqslant 1\}\ \forall \alpha \in R^{\nu}$ implies tightness of $\{\eta_T(\cdot): T \geqslant 1\}$ on the ν-dimensional product space, (b) asymptotically independent increments plus weak convergence of $\eta_T(\pi_2) - \eta_T(\pi_1)\ \forall 0 \leqslant \pi_1 < \pi_2 \leqslant 1$ is sufficient for joint convergence of all the finite dimensional distributions of $\{\eta_T(\cdot): T \geqslant 1\}$, and (c) weak convergence of $\alpha'\eta_T(\cdot)$ to $\alpha'\eta_0(\cdot)\ \forall \alpha \in R^{\nu}$ implies weak convergence of $\alpha'(\eta_T(\pi_2) - \eta_T(\pi_1))$ to $\alpha'(\eta_0(\pi_2) - \eta_0(\pi_1))\ \forall 0 \leqslant \pi_1 < \pi_2 \leqslant 1$ which, in turn, implies weak convergence of $\eta_T(\pi_2) - \eta_T(\pi_1)$ to $\eta_0(\pi_2) - \eta_0(\pi_1)$ using the Cramér-Wold device. *Q.E.D.*

PROOF OF THEOREM 2: Assumptions 1(h) and 2 imply that $M'S^{-1}M$ is nonsingular and hence that V is well defined. By the argument of (A.11)–(A.14) and Assumption 1(d), $\sup_{\pi \in \Pi} \|\hat{M}_r(\pi) - M\| \to_p 0$. By Assumption 3, $\sup_{\pi \in \Pi} \|\hat{S}_r(\pi) - S\| \to_p 0$. Using Assumption 2, this gives the desired result. *Q.E.D.*

The following Lemma is used in the proof of Theorem 3.

LEMMA A5: *Let $J(\pi)$ be a nonsingular $(2p+q) \times (2p+q)$ matrix of the form*

$$\begin{bmatrix} \pi_1 J & 0 & \pi_1 J_2 \\ 0 & \pi_2 J & \pi_2 J_2 \\ J_3 & J_4 & J_5 \end{bmatrix},$$

where π_1 and π_2 are nonzero scalar constants, $J \in R^{p \times p}$, $J_2 \in R^{p \times q}$, $J_3 \in R^{q \times p}$, $J_4 \in R^{q \times p}$, and

$J_5 \in R^{q\times q}$. Let $g=(g_1', g_2', g_3')'$ be any vector in R^{2p+q} and let $H=[I_p \vdots -I_p \vdots 0]\in R^{p\times(2p+q)}$. Let

$$J_*(\pi)=\begin{bmatrix}\pi_1 J & 0\\ 0 & \pi_2 J\end{bmatrix}, \qquad g_*=\begin{pmatrix}g_1\\ g_2\end{pmatrix}\in R^{2p},$$

and $H_*=[I_p \vdots -I_p]$. Then, $HJ^{-1}(\pi)g=H_*J_*^{-1}(\pi)g_*$.

PROOF OF LEMMA A5: *Let $\nu=(\nu_1', \nu_2', \nu_3')=J^{-1}(\pi)g$ and $\tilde{\nu}=(\tilde{\nu}_1', \tilde{\nu}_2')'=J_*^{-1}(\pi)g*$. Since $J(\pi)\nu=g$, we have*

$$\text{(A.17)}\qquad \begin{bmatrix}\pi_1 J & 0\\ 0 & \pi_2 J\end{bmatrix}\begin{pmatrix}\nu_1\\ \nu_2\end{pmatrix}=\begin{pmatrix}g_1\\ g_2\end{pmatrix}-\begin{pmatrix}\pi_1 J_2\nu_3\\ \pi_2 J_2 \nu_3\end{pmatrix}, \quad \begin{pmatrix}\nu_1\\ \nu_2\end{pmatrix}=\begin{pmatrix}\tilde{\nu}_1\\ \tilde{\nu}_2\end{pmatrix}-\begin{pmatrix}J^{-1}J_2\nu_3\\ J^{-1}J_2\nu_3\end{pmatrix}, \quad \text{and}$$

$$HJ^{-1}(\pi)g=\nu_1-\nu_2=\tilde{\nu}_1-\tilde{\nu}_2=H_*J^{-1}(\pi)g_*. \qquad Q.E.D.$$

PROOF OF THEOREM 3: Let the subscript $*$ be a deletion operator that deletes the last q rows and columns of $(2p+q)\times(2p+q)$ matrices, the last q rows of $2p+q$-vectors and $(2p+q)\times p$ matrices, and the last q columns of $p\times(2p+q)$ matrices. Let

$$\text{(A.18)}\qquad J(\pi)=M(\pi)'\gamma(\pi)M(\pi)=\begin{bmatrix}\pi M'S^{-1}M & 0 & \pi M'S^{-1}M_\delta\\ 0 & (1-\pi)M'S^{-1}M & (1-\pi)M'S^{-1}M_\delta\\ \pi M_\delta' S^{-1}M & (1-\pi)M_\delta' S^{-1}M & M_\delta' S^{-1}M_\delta\end{bmatrix},$$

where the second equality holds by Assumption 2. By Lemma A5, we have $HJ^{-1}(\pi)g=H_*J_*^{-1}(\pi)g_*$ for all $g\in R^{2p+q}$ (where $J_*^{-1}(\pi)=[J_*(\pi)]^{-1}$).

First we establish part (a) of Theorem 3. Let $C=(M'S^{-1}M)^{-1}M'S^{-1/2}$. By Theorem 1 and Lemma A5, we have

$$\begin{aligned}\text{(A.19)}\qquad &\sqrt{T}\left(\hat{\beta}_1(\cdot)-\hat{\beta}_2(\cdot)\right)\\ &=H\sqrt{T}\left(\hat{\theta}(\cdot)-\theta_0\right)\Rightarrow HJ^{-1}(\cdot)M(\cdot)'\gamma(\cdot)G(\cdot)\\ &\qquad=H_*J_*^{-1}(\cdot)[M(\cdot)'\gamma(\cdot)G(\cdot)]_*\\ &\qquad=[I_p \vdots -I_p]\begin{bmatrix}\iota(\cdot)M'S^{-1}M & 0\\ 0 & (1-\iota(\cdot))M'S^{-1}M\end{bmatrix}^{-1}\\ &\qquad\quad\times\begin{bmatrix}M'S^{-1/2}B(\cdot)\\ M'S^{-1/2}(B(1)-B(\cdot))\end{bmatrix}\\ &\qquad=C[B(\cdot)/\iota(\cdot)-(B(1)-B(\cdot))/(1-\iota(\cdot))],\end{aligned}$$

where $\iota(\pi)=\pi$. By Theorem 2,

$$\text{(A.20)}\qquad \hat{V}_1(\cdot)/\iota(\cdot)+\hat{V}_2(\cdot)/(1-\iota(\cdot))\Rightarrow\left(\frac{1}{\iota(\cdot)}+\frac{1}{1-\iota(\cdot)}\right)V=CC'/[\iota(\cdot)(1-\iota(\cdot))].$$

(A.19) and (A.20) and the continuous mapping theorem (see Pollard (1984, Thm. IV.12, p. 70)) give

$$\text{(A.21)}\qquad W_T(\cdot)\Rightarrow\left(B_p(\cdot)-\iota(\cdot)B_p(1)\right)'\left(B_p(\cdot)-\iota(\cdot)B_p(1)\right)/[\iota(\cdot)(1-\iota(\cdot))],$$

where $B_p(\cdot)=(CC')^{-1/2}CB(\cdot)$. $B_p(\cdot)$ is a p-vector of independent Brownian motions because $(CC')^{-1/2}C[(CC')^{-1/2}C]'=I_p$. This establishes the first result of part (a). The second and third results of part (a) follow from the first using the continuous mapping theorem. The same is true for parts (b) and (c), so it suffices to establish the first result in parts (b) and (c).

Next we establish part (b). By Theorem 2, $\hat{V}_r(\pi) = V + o_{p\pi}(1)$. In addition, we show below that

$$\text{(A.22)} \qquad \sup_{\pi \in \Pi} \left\| \sqrt{T}\,\overline{m}_T(\tilde{\theta}(\pi), \pi) \right\| = O_p(1).$$

Hence, it suffices to show that $LM_T^0(\cdot) \Rightarrow Q_p(\cdot)$, where

$$\text{(A.23)} \qquad LM_T^0(\pi) = c_T^0(\pi)'(V/\pi + V/(1-\pi))^{-1} c_T^0(\pi) \quad \text{and}$$

$$c_T^0(\pi) = \left[I_p \vdots -I_p \right] \begin{bmatrix} \frac{1}{\pi}(M'S^{-1}M)^{-1}M'S^{-1} & 0 \\ 0 & \frac{1}{1-\pi}(M'S^{-1}M)^{-1}M'S^{-1} \end{bmatrix}$$

$$\times \sqrt{T}\,\overline{m}_T(\tilde{\theta}(\pi), \pi)$$

$$= H_* J_*^{-1}(\pi)\left[M(\pi)'\gamma(\pi)\sqrt{T}\,\overline{m}_T(\tilde{\theta}(\pi), \pi) \right]_*.$$

Using Lemma A5, we obtain

$$\text{(A.24)} \qquad c_T^0(\pi) = HJ^{-1}(\pi)M(\pi)'\gamma(\pi)\sqrt{T}\,\overline{m}_T(\tilde{\theta}(\pi), \pi).$$

Equations (A.7)–(A.9) with $\hat{\theta}(\cdot)$ replaced by $\tilde{\theta}(\cdot)$ yield

$$\text{(A.25)} \qquad \sqrt{T}\,\overline{m}_T(\tilde{\theta}(\pi), \pi) = \sqrt{T}\,\overline{m}_T(\theta_0, \pi) + M(\pi)\sqrt{T}\left(\tilde{\theta}(\pi) - \theta_0\right) + o_{p\pi}(1).$$

This result and (A.24) give

$$\text{(A.26)} \qquad \begin{aligned} c_T^0(\cdot) &= HJ^{-1}(\cdot)M(\cdot)'\gamma(\cdot)\sqrt{T}\,\overline{m}_T(\theta_0, \cdot) + H\sqrt{T}\left(\tilde{\theta}(\cdot) - \theta_0\right) + o_{p\pi}(1) \\ &\Rightarrow HJ^{-1}(\cdot)M(\cdot)'\gamma(\cdot)G(\cdot) \\ &= C[B(\cdot)/\iota(\cdot) - (B(1) - B(\cdot))/(1 - \iota(\cdot))] \end{aligned}$$

using the last three equalities of (A.19) and the fact that $H\tilde{\theta}(\pi) = H\theta_0 = 0$. Equations (A.23) and (A.26) combine to give the desired result $LM_T^0(\cdot) \Rightarrow Q_p(\cdot)$ in the same way that (A.19) and (A.20) yield (A.21).

For part (b), it remains to show (A.22). This follows from (A.25), since (i) $\sup_{\pi \in \Pi} \|\sqrt{T}\,\overline{m}_T(\theta_0, \pi)\| \to_d \sup_{\pi \in \Pi} \|G(\pi)\|$ ($< \infty$ a.s.) by (A.9) and the continuous mapping theorem, (ii) $\sup_{\pi \in \Pi} \|M(\pi)\| < \infty$, and (iii) $\sqrt{T}(\tilde{\theta} - \theta_0) = O_p(1)$ because $\tilde{\theta}$ is consistent by 1(d) and, given consistency, is asymptotically normal by standard arguments given the remainder of Assumption 1.

Next, we consider part (c). For brevity, we only give a sketch of its proof. First, by element by element mean value expansions, one obtains

$$\text{(A.27)} \qquad \sqrt{T}\,\overline{m}_T(\tilde{\theta}, \pi) = \sqrt{T}\,\overline{m}_T(\hat{\theta}, \pi) + \frac{\partial \overline{m}_T(\hat{\theta}, \pi)\sqrt{T}(\tilde{\theta} - \hat{\theta})}{\partial \theta'} + o_{p\pi}(1) \quad \text{and}$$

$$\text{(A.28)} \qquad \begin{aligned} LR_T(\pi) &= 2T\overline{m}_T(\hat{\theta}, \pi)'\hat{\gamma}(\pi)\frac{\partial \overline{m}_T(\hat{\theta}, \pi)\sqrt{T}(\tilde{\theta} - \hat{\theta})}{\partial \theta'} \\ &\quad + T(\tilde{\theta} - \hat{\theta})'\left[\frac{\partial \overline{m}_T(\hat{\theta}, \pi)}{\partial \theta'}\right]'\hat{\gamma}(\pi)\frac{\partial \overline{m}_T(\hat{\theta}, \pi)(\tilde{\theta} - \hat{\theta})}{\partial \theta'} + o_{p\pi}(1) \\ &= \sqrt{T}(\tilde{\theta} - \hat{\theta})'M(\pi)'\gamma(\pi)M(\pi)\sqrt{T}(\tilde{\theta} - \hat{\theta}) + o_{p\pi}(1), \end{aligned}$$

where the second equality uses (A.6). Let $\tilde{\theta}(\pi)$ be the restricted PS-GMM estimator that minimizes $\overline{m}_T(\theta, \pi)'\hat{\gamma}(\pi)\overline{m}_T(\theta, \pi)$ over $\Theta_0 = \{\theta \in \Theta: \theta = (\beta', \beta', \delta')'\}$. The first-order conditions for $\tilde{\theta}(\pi)$ yield $[\partial \overline{m}_T(\tilde{\theta}(\pi), \pi)/\partial \theta']'\hat{\gamma}(\pi)\overline{m}_T(\tilde{\theta}(\pi), \pi) \cong H'\lambda$ for some p-vector of Lagrange multipliers λ ($= \lambda(\pi)$). Under Assumption 2, the full sample GMM estimator $\tilde{\theta}$ can be shown to satisfy the same first-order conditions up to $o_{p\pi}(1/\sqrt{T})$. This result, premultiplication of (A.27) by $[\partial \overline{m}_T(\hat{\theta}, \pi)/\partial \theta']'\hat{\gamma}(\pi)$, rearrangement of (A.27), and application of (A.6) give

$$\text{(A.29)} \qquad \sqrt{T}(\tilde{\theta}-\hat{\theta}) = J^{-1}(\pi)\left[\frac{\partial \overline{m}_T(\hat{\theta},\pi)}{\partial \theta'}\right]' \hat{\gamma}(\pi)\sqrt{T}\overline{m}_T(\tilde{\theta},\pi) + o_{p\pi}(1)$$

$$= \sqrt{T}J^{-1}(\pi)H'\tilde{\lambda} + o_{p\pi}(1) \quad \text{and}$$

$$\text{(A.30)} \qquad LR_T(\pi) = \sqrt{T}\tilde{\lambda}'HJ^{-1}(\pi)H'\left(HJ^{-1}(\pi)H'\right)^{-1}\sqrt{T}HJ^{-1}(\pi)H'\tilde{\lambda} + o_{p\pi}(1).$$

By Lemma A5 and the definitions of $\tilde{\lambda}$ and $c_T^0(\pi)$, we have

$$\text{(A.31)} \qquad HJ^{-1}(\pi)H' = H_* J_*^{-1}(\pi)H_*' = \left(\frac{1}{\pi}+\frac{1}{1-\pi}\right)(M'S^{-1}M)^{-1} = \left(\frac{1}{\pi}+\frac{1}{1-\pi}\right)V \quad \text{and}$$

$$\text{(A.32)} \qquad \sqrt{T}HJ^{-1}(\pi)H'\tilde{\lambda} = HJ^{-1}(\pi)M(\pi)'\gamma(\pi)\sqrt{T}\overline{m}_T(\tilde{\theta},\pi) + o_{p\pi}(1) = c_T^0(\pi) + o_{p\pi}(1).$$

Combining (A.30)–(A.32) gives $LR_T(\pi) = LM_T^0(\pi) + o_{p\pi}(1)$ and the desired result follows from the proof above that $LM_T^0(\cdot) \Rightarrow Q_p(\cdot)$. *Q.E.D.*

Proof of Corollary 1: The process $BB(\cdot) = B_p(\cdot) - \iota(\cdot)B_p(\cdot)$, which appears in the definition of $Q_p(\cdot)$, is a p-vector of independent Brownian bridge processes on $[0,1]$. An alternative method of defining such a process is via a p-vector $BM(\cdot)$ of independent Brownian motion processes on $[0,\infty)$. In particular, we have

$$\text{(A.33)} \qquad \{BB(\pi): \pi \in [0,1]\} \approx \{(1-\pi)BM(\pi/(1-\pi)): \pi \in [0,1]\},$$

where $\approx$ denotes equality in distribution. Hence, we have

$$\text{(A.34)} \qquad P\Big(\sup_{\pi\in[\pi_1,\pi_2]} Q_p(\pi) < c\Big)$$

$$= P\left(\sup_{\pi\in[\pi_1,\pi_2]} BM\left(\frac{\pi}{1-\pi}\right)' BM\left(\frac{\pi}{1-\pi}\right)\Big/\left(\frac{\pi}{1-\pi}\right) < c\right)$$

$$= P\left(\sup_{s\in[1,\pi_2(1-\pi_1)/(\pi_1(1-\pi_2))]} BM\left(\frac{\pi_1 s}{1-\pi_1}\right)' BM\left(\frac{\pi_1 s}{1-\pi_1}\right)\Big/\left(\frac{\pi_1 s}{1-\pi_1}\right) < c\right)$$

$$= P\Big(\sup_{s\in[1,\pi_2(1-\pi_1)/(\pi_1(1-\pi_2))]} \overline{BM}(s)'\overline{BM}(s)/s < c\Big)$$

for all $0 < \pi_1 \leqslant \pi_2 < 1$ and $c \geqslant 0$, where the second equality holds by change of variables with

$$s = \left(\frac{1-\pi_1}{\pi_1}\right)\frac{\pi}{1-\pi}, \qquad \overline{BM}(s) = BM\left(\frac{\pi_1 s}{1-\pi_1}\right)\Big/\left(\frac{\pi_1}{1-\pi_1}\right)^{1/2}$$

by definition, and $\overline{BM}(\cdot)$ is also a Brownian motion on $[0,\infty)$ (by direct verification).

The result of Corollary 1 is now obtained as follows:

$$\text{(A.35)} \qquad \lim_{T\to\infty} P\Big(\sup_{\pi\in[0,1]} W_T(\pi) < c\Big) \leqslant \overline{\lim_{\varepsilon\to 0}} \lim_{T\to\infty} P\Big(\sup_{\pi\in[\varepsilon,1-\varepsilon]} W_T(\pi) < c\Big)$$

$$= \overline{\lim_{\varepsilon\to 0}} P\Big(\sup_{\pi\in[\varepsilon,1-\varepsilon]} Q_p(\pi) < c\Big)$$

$$= \overline{\lim_{\varepsilon\to 0}} P\Big(\sup_{s\in[1,(1-\varepsilon)^2/\varepsilon^2]} BM(s)'BM(s)/s < c\Big)$$

$$= P\Big(\sup_{s\in[1,\infty)} BM(s)'BM(s)/s < c\Big) = 0,$$

where the first equality holds by Theorem 3, the second by (A.34), and the last by well known properties of Brownian motion (i.e., the law of the iterated logarithm). The proof is identical for $LM_T(\pi)$ and $LR_T(\pi)$. *Q.E.D.*

PROOF OF THEOREM 4: Part (a) holds by the proof of Theorem 1, noting that (A.9) holds with $G(\cdot)$ replaced by $G(\cdot)+\mu(\cdot)$, since $\sqrt{T}\overline{m}_T(\theta_0,\cdot)=\sqrt{T}(\overline{m}_T(\theta_0,\cdot)-E\overline{m}_T(\theta_0,\cdot))+\sqrt{T}E\overline{m}_T(\theta_0,\cdot)$ $\Rightarrow G(\cdot)+\mu(\cdot)$ under Assumption 1-LP. Part (b) holds by the proof of Theorem 2.

Parts (c)–(e) hold using the proof of Theorem 3 with references to Theorems 1 and 2 replaced by references to Theorem 4(a) and (b), respectively, with $G(\cdot)$ replaced by $G(\cdot)+\mu(\cdot)$, with the right-hand side of (A.19) replaced by

$$\text{(A.36)} \qquad C\big[B(\cdot)/\iota(\cdot)-(B(1)-B(\cdot))/(1-\iota(\cdot)) \\ -S^{-1/2}\mu_1(\cdot)/\iota(\cdot)+S^{-1/2}\mu_2(\cdot)/(1-\iota(\cdot))\big],$$

and with the right-hand side of (A.21) and (A.26) changed accordingly. *Q.E.D.*

PROOF OF COROLLARY 2: By Theorem 4(c)–(e) and the nonsingularity of $AS^{-1/2}M$ in (5.4), it suffices for Corollary 2 to show that

$$\text{(A.37)} \qquad \left(\frac{1-\pi}{\pi}\right)^{1/2}\int_0^{\pi}\eta(s)\,ds=\left(\frac{\pi}{1-\pi}\right)^{1/2}\int_{\pi}^{1}\eta(s)\,ds \qquad \forall\pi\in\Pi$$

does not hold. Note that (A.37) holds if and only if

$$\text{(A.38)} \qquad \int_0^{\pi}\eta_j(s)\,ds=\pi\int_0^{1}\eta_j(s)\,ds \qquad \forall\pi\in\Pi,\quad \forall j=1,\dots,p,$$

where $\eta(\pi)=(\eta_1(\pi),\dots,\eta_p(\pi))'$. Thus, it suffices to show that (A.38) does not hold.

Suppose (A.38) holds. Then, since $\pi\int_0^1\eta_j(s)\,ds$ is twice differentiable in π $\forall\pi\in\operatorname{int}(\Pi)$, $\forall j=1,\dots,p$, so must be $\int_0^{\pi}\eta_j(s)\,ds$, where $\operatorname{int}(\Pi)$ denotes the interior of Π. In particular, we must have

$$\text{(A.39)} \qquad \frac{d}{d\pi}\int_0^{\pi}\eta_j(s)\,ds=\int_0^{1}\eta_j(s)\,ds \quad\text{and}\quad \frac{d^2}{d\pi^2}\int_0^{\pi}\eta_j(s)\,ds=0 \\ \forall\pi\in\operatorname{int}(\Pi),\quad \forall j=1,\dots,p.$$

This implies that $\eta_j=c_j$ almost everywhere (Lebesgue) on Π for some constant c_j $\forall j=1,\dots,p$, which is a contradiction. *Q.E.D.*

PROOF OF THEOREM 5: Let $u_\alpha=c_\alpha^{1/2}$ and $t_\alpha=\tilde{c}_\alpha^{1/2}$. We will show that

$$\text{(A.40)} \qquad u_\alpha-t_\alpha\to 0 \quad\text{as}\quad \alpha\to 0.$$

Then, using Theorem 4, we have

$$\text{(A.41)} \qquad \lim_{\alpha\to 0}\inf_{\eta\in\Xi}\inf_{\tilde{\pi}\in\Pi}\lim_{T\to\infty}\left[P_\eta\Big(\sup_{\pi\in\Pi}W_T(\pi)>c_\alpha\Big)-P_\eta\big(W_T(\tilde{\pi})>\tilde{c}_\alpha\big)\right] \\ =\lim_{\alpha\to 0}\inf_{\eta\in\Xi}\sup_{\tilde{\pi}\in\Pi}\left[P_\eta\Big(\sup_{\pi\in\Pi}Q_p^*(\pi)^{1/2}>u_\alpha\Big)-P_\eta\Big(Q_p^*(\tilde{\pi})^{1/2}>t_\alpha\Big)\right] \\ \geqslant\lim_{\alpha\to 0}\inf_{\eta\in\Xi}\sup_{\tilde{\pi}\in\Pi}\left[P_\eta\Big(Q_p^*(\tilde{\pi})^{1/2}>u_\alpha\Big)-P_\eta\Big(Q_p^*(\tilde{\pi})^{1/2}>t_\alpha\Big)\right] \\ =0,$$

where the last equality uses (A.40) and the fact that $Q_p^*(\tilde{\pi})$ is a noncentral chi-square rv and the density of the square root of a noncentral chi-square rv is bounded above uniformly over all possible values of its noncentrality parameter.

To show (A.40) we use an argument similar to that of van Zwet and Oosterhoff (1967, p. 675). Let $\pi_1=\inf\{\pi\in\Pi\}>0$, let $\pi_2=\sup\{\pi\in\Pi\}<1$, and let ν_α be such that $P(\sup_{\pi\in[\pi_1,\pi_2]}Q_p^*(\pi)^{1/2}>\nu_\alpha)=\alpha$. Since $t_\alpha\leqslant u_\alpha\leqslant\nu_\alpha$, to establish (A.40) it suffices to show that $\nu_\alpha-t_\alpha\to 0$ as $\alpha\to 0$.

By a result of James, James, and Siegmund (1987, eqn. (26), p. 78), we have

$$P\Big(\sup_{\pi\in[\pi_1,\pi_2]} Q_p(\pi)^{1/2} > \nu_\alpha\Big) = K_p \nu_\alpha^{p-2} \exp\big(-\nu_\alpha^2/2\big)\big\{\big(\nu_\alpha^2 - p\big)\log\lambda + 4 + o(1)\big\} \tag{A.42}$$

as $\alpha \to 0$, where $Q_p(\cdot)$ is as in Theorem 3, K_p is a constant that depends only on the dimension p of the Brownian bridge vector that underlies $Q_p(\cdot)$, and $\lambda = \pi_2(1-\pi_1)/[\pi_1(1-\pi_2)]$. Taking $\pi_1 = \pi_2 = \tilde{\pi}$ in (A.42) yields $\log\lambda = 0$ and

$$P\big(Q_p(\tilde{\pi})^{1/2} > t_\alpha\big) = K_p t_\alpha^{p-2} \exp\big(-t_\alpha^2/2\big)\{4 + o(1)\} \quad \text{as } \alpha \to 0. \tag{A.43}$$

The left-hand side of (A.42) and (A.43) each equal α. Thus, the logs of the right-hand side of (A.42) and (A.43) can be equated to yield

$$\begin{aligned}(p-2)\log\nu_\alpha - \nu_\alpha^2/2 + \log\big\{\big(\nu_\alpha^2 - p\big)\log\lambda + 4 + o(1)\big\} \\ = (p-2)\log t_\alpha - t_\alpha^2/2 + \log\{4 + o(1)\} \quad \text{and}\end{aligned} \tag{A.44}$$

$$\begin{aligned}\nu_\alpha - \frac{t_\alpha^2}{\nu_\alpha} &= \frac{2}{\nu_\alpha}\big[(p-2)\log\nu_\alpha - (p-2)\log t_\alpha \\ &\quad + \log\big\{\big(\nu_\alpha^2 - p\big)\log\lambda + 4 + o(1)\big\} - \log\{4 + o(1)\}\big] \\ &= o(1)\end{aligned} \tag{A.45}$$

as $\alpha \to 0$, using the fact that $t_\alpha \to \infty$ as $\alpha \to 0$ and $t_\alpha \leqslant \nu_\alpha$. Since $|\nu_\alpha - t_\alpha| \leqslant \nu_\alpha - t_\alpha^2/\nu_\alpha$, (A.45) implies that $\nu_\alpha - t_\alpha \to 0$ as $\alpha \to 0$. *Q.E.D.*

REFERENCES

ANDERSON, T. W., AND D. A. DARLING (1952): "Asymptotic Theory of Certain 'Goodness of Fit' Criteria Based on Stochastic Processes," *Annals of Mathematical Statistics*, 23, 193–212.

ANDREWS, D. W. K. (1988): "Laws of Large Numbers for Dependent Non-identically Distributed Random Variables," *Econometric Theory*, 4, 458–467.

——— (1989a): "Asymptotics for Semiparametric Econometric Models: I. Estimation and Testing," Cowles Foundation Discussion Paper No. 908R, Yale University.

——— (1989b): "Asymptotics for Semiparametric Econometric Models: II. Stochastic Equicontinuity and Nonparametric Kernel Estimation," Cowles Foundation Discussion Paper No. 909R, Yale University.

——— (1989c): "Tests for Parameter Instability and Structural Change with Unknown Change Point," Cowles Foundation Discussion Paper No. 943, Yale University.

——— (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

——— (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, 241–257.

ANDREWS, D. W. K., AND R. C. FAIR (1988): "Inference in Nonlinear Econometric Models with Structural Change," *Review of Economic Studies*, 55, 615–640.

ANDREWS, D. W. K., I. LEE, AND W. PLOBERGER (1992): "Optimal Changepoint Tests for Linear Regression," Cowles Foundation Discussion Paper No. 1016, Yale University.

ANDREWS, D. W. K., AND J. C. MONAHAN (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953–966.

ANDREWS, D. W. K., AND W. PLOBERGER (1992): "Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative," Cowles Foundation Discussion Paper No. 1015, Yale University.

BAI, J. (1991): "Estimation of Structural Change in Econometric Models," unpublished manuscript, Department of Economics, University of California, Berkeley.

BAI, J., R. LUMSDAINE, AND J. H. STOCK (1991): "Testing for and Dating Breaks in Integrated and Cointegrated Time Series," unpublished manuscript, Kennedy School of Government, Harvard University.

Banerjee, A., R. L. Lumsdaine, and J. H. Stock (1992): "Recursive and Sequential Tests for a Unit Root: Theory and International Evidence," *Journal of Business and Economic Statistics*, 10, 271–287.

Bierens, H. J. (1981): *Robust Methods and Asymptotic Theory*, Lecture Notes in Economics and Mathematical Systems No. 192. Berlin: Springer Verlag.

Billingsley, P. (1968): *Convergence of Probability Measures*. New York: Wiley.

Brown, R. L., J. Durbin, and J. M. Evans (1975): "Techniques for Testing the Constancy of Regression Relationships over Time," *Journal of the Royal Statistical Society, Series B*, 37, 149–192.

Chow, Y. S., and H. Teicher (1978): *Probability Theory: Independence, Interchangeability, Martingales*. New York: Springer-Verlag.

Chu, C.-S. J. (1989): "New Tests for Parameter Constancy in Stationary and Nonstationary Regression Models," unpublished manuscript, Department of Economics, University of California-San Diego.

Davies, R. B. (1977): "Hypothesis Testing When a Nuisance Parameter Is Present Only under the Alternative," *Biometrika*, 64, 247–254.

——— (1987): "Hypothesis Testing When a Nuisance Parameter Is Present Only under the Alternative," *Biometrika*, 74, 33–43.

DeLong, D. M. (1981): "Crossing Probabilities for a Square Root Boundary by a Bessel Process," *Communications in Statistics—Theory and Methods*, A10 (21), 2197–2213.

Deshayes, J. (1983): *Rupture de Modèles en Statistique*. Ph.D. Thesis, Department of Mathematics, University of Paris, Orsay.

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

Gallant, A. R., and H. White (1988): *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Oxford: Basil Blackwell.

Garbade, K. (1977): "Two Methods of Examining the Stability of Regression Coefficients," *Journal of the American Statistical Association*, 72, 54–63.

Gordin, M. I. (1969): "The Central Limit Theorem for Stationary Processes," *Soviet Mathematics-Doklady*, 10, 1174–1176.

Hall, P., and C. C. Heyde (1980): *Martingale Limit Theory and Its Applications*. New York: Academic Press.

Hannan, E. J. (1970): *Multiple Time Series*. New York: Wiley.

——— (1973): "Central Limit Theorems for Time Series Regressions," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebeite*, 26, 157–170.

Hansen, B. E. (1990): "Lagrange Multiplier Tests for Parameter Instability in Nonlinear Models," unpublished manuscript, Department of Economics, University of Rochester.

——— (1991): "Strong Laws for Dependent Heterogeneous Processes," *Econometric Theory*, 7, 213–221.

——— (1992): "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business and Economic Statistics*, 10, 321–335.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Hawkins, D. L. (1987): "A Test for a Change Point in a Parametric Model Based on a Maximal Wald-type Statistic," *Sankhya*, 49, 368–376.

Hawkins, D. M. (1977): "Testing a Sequence of Observations for a Shift in Location," *Journal of the American Statistical Association*, 72, 180–186.

Hendry, D. F. (1989): *PC-GIVE: An Interactive Econometric Modelling System*, Version 6.0. Oxford: Oxford University Press.

Hinkley, D. V. (1970): "Inference about the Change-point in a Sequence of Random Variables," *Biometrika*, 57, 1–17.

Huber, P. J. (1967): "The Behaviour of Maximum Likelihood Estimates under Nonstandard Conditions," in *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*, Vol. 1, 221–233. Berkeley: University of California.

Ibragimov, I. A. (1962): "Some Limit Theorems for Stationary Processes," *Theory of Probability and its Applications*, 7, 349–382.

James, B., K. L. James, and D. Siegmund (1987): "Tests for a Change-point," *Biometrika*, 74, 71–83.

Jennrich, R. I. (1969): "Asymptotic Properties of Nonlinear Least Squares Estimators," *Annals of Mathematical Statistics*, 40, 633–643.

Kim, H. J., and D. Siegmund (1989): "The Likelihood Ratio Test for a Change-point in Simple Linear Regression," *Biometrika*, 76, 409–423.

Krämer, W., W. Ploberger, and R. Alt (1988): "Testing for Structural Change in Dynamic Models," *Econometrica*, 56, 1355–1369.

Krämer, W., and H. Sonnberger (1986): *The Linear Regression Model under Test*. Heidelberg: Physica-Verlag.

Krishnaiah, P. R., and B. Q. Miao (1988): "Review about Estimation of Change Points," in *Handbook of Statistics*, Vol. 7, ed. by P. R. Krishnaiah and C. R. Rao. New York: Elsevier.

Leybourne, S. J., and B. P. M. McCabe (1989): "On the Distribution of Some Test Statistics for Coefficient Constancy," *Biometrika*, 76, 169–177.

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

——— (1975b): "Invariance Principles for Dependent Variables," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebeite*, 32, 165–178.

——— (1977): "On the Invariance Principle for Nonstationary Mixingales," *Annals of Probability*, 5, 616–621.

Nyblom, J. (1989): "Testing for the Constancy of Parameters over Time," *Journal of the American Statistical Association*, 84, 223–230.

Picard, D. (1983): *Rupture de Modèles en Statistique*. Ph.D. Thesis, Department of Mathematics, University of Paris, Orsay.

——— (1985): "Testing and Estimating Change-points in Time Series," *Advances in Applied Probability*, 17, 841–867.

Ploberger, W., W. Krämer, and K. Kontrus (1989): "A New Test for Structural Stability in the Linear Regression Model," *Journal of Econometrics*, 40, 307–318.

Pollard, D. (1984): *Convergence of Stochastic Processes*. New York: Springer-Verlag.

Pötscher, P. M., and I. R. Prucha (1991): "Basic Structure of the Asymptotic Theory in Dynamic Nonlinear Econometric Models, Part I: Consistency and Approximation Concepts," *Econometric Reviews*, 10, 125–216.

Roy, S. N. (1953): "On a Heuristic Method of Test Construction and Its Use in Multivariate Analysis," *Annals of Mathematical Statistics*, 24, 220–238.

Roy, S. N., R. Gnanadesikan, and J. N. Srivastava (1971): *Analysis and Design of Certain Quantitative Multiresponse Experiments*. Oxford: Pergamon Press.

Sen, P. K. (1980): "Asymptotic Theory of Some Tests for a Possible Change in Regression Slope Occurring at an Unknown Time Point," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebeite*, 52, 203–218.

——— (1981): *Sequential Nonparametrics: Invariance Principles and Statistical Inference*. New York: Wiley.

Talwar, P. P. (1983): "Detecting a Shift in Location," *Journal of Econometrics*, 23, 353–367.

van Zwet, W. R., and J. Oosterhoff (1967): "On the Combination of Independent Test Statistics," *Annals of Mathematical Statistics*, 38, 659–680.

Wooldridge, J. M., and H. White (1988): "Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes," *Econometric Theory*, 4, 210–230.

Zacks, S. (1983): "Survey of Classical and Bayesian Approaches to the Change-point Problem: Fixed Sample and Sequential Procedures of Testing and Estimation," in *Recent Advances in Statistics*, ed. by M. H. Rivzi, J. S. Rustagi, and D. Siegmund. New York: Academic Press.

Zivot, E., and D. W. K. Andrews (1992): "Further Evidence on the Great Crash, the Oil Price Shock and the Unit Root Hypothesis," *Journal of Business and Economic Statistics*, 10, 251–270.