End-of-Sample Instability Tests

Donald W. K. Andrews[1]
Cowles Foundation for Research in Economics
Yale University

January 2002

Abstract

This paper considers tests for structural instability of short duration, such as at the end of the sample. The key feature of the testing problem is that the number, m, of observations in the period of potential change is relatively small—possibly as small as one. The well-known F test of Chow (1960) for this problem only applies in a linear regression model with normally distributed iid errors and strictly exogenous regressors, even when the total number of observations, $n + m$, is large.

We generalize the F test to cover regression models with much more general error processes, regressors that are not strictly exogenous, and estimation by instrumental variables as well as least squares. In addition, we extend the F test to nonlinear models estimated by generalized method of moments and maximum likelihood.

Asymptotic critical values that are valid as $n \to \infty$ with m fixed are provided using a subsampling-like method. The results apply quite generally to processes that are strictly stationary and ergodic under the null hypothesis of no structural instability.

Keywords: Instrumental variables estimator, least squares estimator, parameter change, structural instability test, structural change.

JEL Classification Numbers: C12, C52.

1 Introduction

This paper considers the problem of testing for structural instability over a short time interval, such as at the end of a sample. Most tests in the literature are designed for detecting instability that lasts for a relatively long period of time starting somewhere in the middle of the sample, e.g., see Andrews and Fair (1988), Ghysels and Hall (1990), Hansen (1992), Andrews (1993), Ghysels, Guay, and Hall (1997), and other references in Stock (1994). These tests use asymptotics in which the number of observations before a potential changepoint, n, and the number after the potential changepoint, m, both go to infinity. Such tests are not valid in the case considered here in which the number of observations in the period of potential instability, m, is small—perhaps as small as one. In this paper we design tests that are asymptotically valid when $n \to \infty$ with m fixed, where $n + m$ is the total number of observations.

We start by considering the F test for parameter change in a linear regression model with iid normal errors and strictly exogenous regressors, as in Chow (1960). The F test is restrictive because it is asymptotically valid when m is small only under the stated conditions. Even normality of the errors is needed.

The main contribution of this paper is to introduce variants of the F test that are valid under weak assumptions and apply to a wide variety of models. We do so by constructing critical values using a subsampling-like method. In the linear regression model, the tests we propose are asymptotically valid with non-normal, heteroskedastic, conditionally heteroskedastic, and/or autocorrelated errors and with regressors that are not strictly exogenous. The observations and/or errors could even possess long memory. The main requirement is that the observations are strictly stationary and ergodic under the null hypothesis. Furthermore, the tests we propose apply to regression models estimated by instrumental variables (IV) and to nonlinear models estimated by generalized method of moments (GMM) and maximum likelihood (ML).

The bulk of this paper discusses tests for structural instability at the end of the sample. Extending such tests to the case of potential instability at the beginning, rather than the end, of the sample is trivial. Such tests can be used to determine the start of the sample period that is most appropriate for a given model. In addition, we show how end-of-sample tests can be used to test for structural instability that occurs over a small number of observations in the middle of the sample. For example, such tests can be used to test for instability during war years or during a short regime shift, such as the Federal Reserve Bank policy regime of 1979-82. Standard tests for structural instability are not valid in these situations because the number of observations in the period of change, m, is small and, hence, asymptotics that rely on $m \to \infty$ are inappropriate.

The form of the F test for parameter change over the last m observations in a sample depends on whether m is greater than or less than the number of regressors d, see Chow (1960). When $m \geq d$, the primary component of the F statistic is a quadratic form in the post-change residual vector projected onto the column space spanned by the post-change regressors with the residuals computed using the least squares (LS) estimator based on the pre-change observations. When $m \leq d$, the primary component of the F statistic can be written as the sum of squared post-change residuals computed using the full sample estimator. When $m = d$, the two forms of the F statistic are equal.

The tests that we consider here are of the same two types as the F test. For each type, we provide two different tests that differ mainly in the way their critical values are computed. In particular, we consider tests S_a and S_b for the case where $m \geq d$ and tests P_a and P_b for the case where $m \leq d$. (When $m = d$, the S_a and P_a tests are the same and similarly for S_b and P_b.) The S_b and P_b tests are simpler to compute than the S_a and P_a tests.

When the regression model is estimated by IV, rather than LS, the relevant value d for the form of the statistic is the number of IVs, rather than the number of regressors. In the GMM case, the relevant value d for the form of the statistic is the number of moment conditions, rather than the number of regressors.

Critical values are obtained by a subsampling-like method that we call *parametric sub-sampling*. Consider an end-of-sample test. One computes the $n - m + 1$ test statistics that are analogous to the test statistic of interest but are for testing for structural instability over the m observations that start at the j-th observation, rather than for instability starting at the $(n+1)$-th observation, for $j = 1, ..., n-m+1$. The $1-\alpha$ sample quantile of these statistics is the significance level α critical value for the end-of-sample instability test statistic. Computation of the critical value is relatively easy. It just requires calculation of $n - m + 1$ versions of the original statistic.

The critical values considered here are parametric subsampling critical values that rely on subsamples of length m. There is no arbitrary smoothing parameter or block length parameter to select. The appropriate subsample length is automatically specified by the number of post-change observations m. Also, no heteroskedasticity and autocorrelation consistent covariance matrix estimator is required. The critical value is not a pure subsampling critical value because the test statistic for a given value of j depends on observations other than those indexed by $j, ..., j + m - 1$ through the parameter estimator that is used to compute the residuals. See Politis, Romano, and Wolf (1999) for an in-depth treatment of, and references on, subsampling methods.

Given that m is fixed as $n \to \infty$, the tests considered here are not consistent tests. However, they typically are asymptotically unbiased. The power of the tests depends on the magnitude of the structural change relative to the error variance, as well as on the magnitude of m. The larger is m, the greater is the power *ceteris paribus*. For small m, the power may be low if the magnitude of structural change is not large. In consequence, failure to reject the null hypothesis should not necessarily be interpreted as strong evidence in favor of structural stability.

We provide some Monte Carlo results for the tests introduced in the paper. We consider end-of sample instability tests for linear regression models with first-order autoregressive (AR) errors and regressors with AR parameter $\rho = 0$, .4, and .8 and innovations that are normal, chi-square with two degrees of freedom, and uniform. We consider pre-change sample sizes $n = 100$, 250, and 500 and post-change sample sizes $m = 1$, 5, and 10. We find that the tests T_v for $T = S$ and P and $v = a$ and b have good size properties over the range of cases considered except when $\rho = .8$ and $n = 100$ or 250. The size properties of the F test are poor whenever $\rho > 0$ and

$m > 1$ and whenever $\rho = 0$ and errors and regressors are uniform. The size and power properties of the S_b test are quite similar to those of the S_a test. The same is true for P_b and P_a.

Given the simulation results and the relative ease of computation of the S_b and P_b tests, we recommend using the S_b test when $m \geq d$ and the P_b test when $m \leq d$. (The test are equivalent when $m = d$.) The recommended tests, S_b and P_b, are defined in Section 2 with as little preliminary notation and discussion as possible, for the ease of readers who wish to skip the technical details that are given in the main body of the paper.

The tests introduced here have been used effectively by Fair (2002). Fair (2002) finds evidence of structural change in a U.S. stock market equation in the late 1990s, but no structural change in most other U.S. macroeconometric equations that he considers.

We note that the tests considered here can be applied to p-th order autoregressive models that may have a unit root by differencing the observations and applying the tests to the differenced data.

The remainder of this paper is organized as follows. All sections of the paper except Section 6 discuss end-of-sample instability tests. Section 2 defines the P_b and S_b tests. Section 3 motivates the statistics considered in the paper using the F statistic for the normal linear regression model. Section 4 considers tests for the linear regression model estimated by IV. Section 5 considers models estimated by GMM. Section 6 discusses tests for structural instability that occurs at the beginning or in the middle of the sample for a small number of observations. Section 7 briefly discusses application of the tests to simple models with integrated variables. Section 8 introduces high-level assumptions, sufficient conditions for these assumptions for the LS, IV, and GMM cases, and the main asymptotic results. Section 9 provides some Monte Carlo results. An Appendix contains proofs.

2 Definitions of the P_b and S_b Tests

For readers who are primarily interested in the definitions of the proposed end-of-sample instability tests, we describe the recommended P_b and S_b tests here. For their extension to beginning-of-sample and middle-of-sample instability tests, see Section 6.

Consider a linear regression model estimated by LS. Let m denote the number of observations after the (potential) changepoint. Let d denote the number of regressors. If $m \leq d$, the P_b test is appropriate. The P_b test statistic is given by the average of the post-change squared residuals with the residuals computed using the LS estimator based on the observations indexed by $i = 1 + m, ..., n + m$. For $j = 1, ..., n - m + 1$, let $P_{b,j}$ denote the average of the m squared residuals for the observations indexed by $j, ..., j + m - 1$ with the residuals computed using the LS estimator based on the observations indexed by $i = 1, ..., n$. One rejects the null hypothesis at significance level α if P_b exceeds $100(1 - \alpha)\%$ of the values $\{P_{b,j} : j = 1, ..., n - m + 1\}$.[2]

For the case of a linear model estimated by IV or a model estimated by GMM with moments given by the product of a residual and a vector of IVs, the P_b test is the same as in the LS case, but with the IV or GMM estimator used place of the LS estimator.

Next, suppose $m \geq d$. In this case, the S_b test is appropriate. The S_b test statistic is given by m^{-1} times the quadratic form in the m-dimensional post-change residual vector projected onto the column space spanned by the post-change regressor vectors, with the residuals computed using the LS estimator based on the observations indexed by $i = 1 + m, ..., n + m$. Equivalently, the S_b test statistic equals the quadratic form in the d-vector average of the residual times the regressor vector for the post-change observations with $d \times d$ weight matrix given by the inverse of the average of the outer-product of the post-change regressor vectors.

For $j = 1, ..., n - m + 1$, define $S_{b,j}$ as S_b is defined except using the observations indexed by $j, ..., j + m - 1$ instead of the post-change observations and with the residuals computed using the LS estimator based on the observations indexed by $i = 1, ..., n$. One rejects the null hypothesis at significance level α if S_b exceeds $100(1 - \alpha)\%$ of the values $\{S_{b,j} : j = 1, ..., n - m + 1\}$.

For the case of a linear model estimated by IV, the S_b test is the same as in the LS case but with the IV vector in place of the regressor vector and the IV estimator in place of the LS estimator. For a model estimated by GMM, the test is the same as in the LS case but with the moment vector in place of the product of the residual and regressor vector, the GMM estimator in place of the LS estimator, and the weight matrix defined by the average of the outer-product of either the moment vectors or the IV vectors (if the moment vector is given by the product of a residual and an IV vector) in place of the outer product of regressor vectors.

3 Motivation Based on Linear Regression

In this section, we consider the linear regression model. We start by considering the standard F statistic for parameter change. This statistic motivates the form of the test statistics considered in the paper for linear regression models and more general models. The F statistic that we consider is based on a one-time shift in the parameters, but it has power against more general types of structural change. For this reason, we distinguish between the "test-generating" model and hypotheses and the more general model and hypotheses of interest.

The "test-generating" model is

$$Y_i = \begin{cases} X_i'\beta_0 + U_i & \text{for } i = 1, ..., n \\ X_i'\beta_1 + U_i & \text{for } i = n + 1, ..., n + m, \end{cases} \tag{3.1}$$

where $Y_i \in R$, $U_i \in R$, and $X_i, \beta_0, \beta_1 \in R^d$. The observations are $\{W_i : i = 1, ..., n + m\}$, where $W_i = (Y_i, X_i')'$. The errors and regressors of the "test-generating" model satisfy $U_i \sim N(0, \sigma_0^2)$, U_i is independent of X_i, and EX_iX_i' is positive definite. There are n observations before, and m observations after, the (possible) change in the parameter. In this paper, we focus on the case in which n is large and m is small. In the extremum, one could have $m = 1$.

In vector notation, the "test-generating" model is

$$\begin{aligned}\mathbf{Y}_0 &= \mathbf{X}_0'\beta_0 + \mathbf{U}_0 \text{ and} \\ \mathbf{Y}_1 &= \mathbf{X}_1'\beta_1 + \mathbf{U}_1,\end{aligned} \tag{3.2}$$

where $\mathbf{Y}_0, \mathbf{U}_0 \in R^n$, $\mathbf{X}_0 \in R^{n\times d}$, $\mathbf{Y}_1, \mathbf{U}_1 \in R^m$, and $\mathbf{X}_1 \in R^{m\times d}$.

The "test-generating" null and alternative hypotheses are

$$H_0^* : \beta_1 = \beta_0 \text{ and } H_1^* : \beta_1 \neq \beta_0. \tag{3.3}$$

The error variance, σ_0^2, is constant under the "test-generating" null and alternative hypotheses. The form of the F statistic for testing H_0^* versus H_1^* depends on whether $m \geq d$ or $m \leq d$. We treat these cases separately in two subsections below.

The more general model of interest is

$$Y_i = \begin{cases} X_i'\beta_0 + U_i & \text{for } i = 1, ..., n \\ X_i'\beta_{1i} + U_i & \text{for } i = n+1, ..., n+m, \end{cases} \tag{3.4}$$

where $EU_iX_i = 0$, EX_iX_i' is positive definite, and $\{(Y_i, X_i) : i \geq 1\}$ are stationary and ergodic under the null hypothesis (which implies that the error variance, σ_0^2, is constant under the null hypothesis).

The null and alternative hypotheses of interest are

$$\begin{aligned} H_0 &: \begin{cases} \beta_{1i} = \beta_0 \text{ for all } i = n+1, ..., n+m \text{ and} \\ \{(Y_i, X_i) : i \geq 1\} \text{ are stationary and ergodic} \end{cases} \\ H_1 &: \begin{cases} \beta_{1i} \neq \beta_0 \text{ for some } i = n+1, ..., n+m \text{ and/or} \\ \text{the distribution of } U_i \text{ for some } i = n+1, ..., n+m \\ \text{differs from the distribution of } U_i \text{ for } i = 1, ..., n. \end{cases} . \end{aligned} \tag{3.5}$$

Alternatively, these hypotheses can be expressed as

$$\begin{aligned} H_0 &: \begin{cases} E(Y_i - X_i'\beta_0)X_i = 0 \text{ for all } i = 1, ..., n+m \text{ and} \\ \{(Y_i, X_i) : i \geq 1\} \text{ are stationary and ergodic} \end{cases} \\ H_1 &: \begin{cases} E(Y_i - X_i'\beta_0)X_i = 0 \text{ for all } i = 1, ..., n, \text{ and} \\ E(Y_i - X_i'\beta_0)X_i \neq 0 \text{ for some } i = n+1, ..., n+m \text{ and/or} \\ \text{the distribution of } U_i \text{ for some } i = n+1, ..., n+m \\ \text{differs from the distribution of } U_i \text{ for } i = 1, ..., n. \end{cases} . \end{aligned} \tag{3.6}$$

For linear regression models estimated by LS, the hypotheses in (3.5) and (3.6) are equivalent. But, for linear regression models estimated by IV and for GMM models, which are considered below, hypotheses that are analogous to those in (3.6) allow for more general structural change than those in (3.5). In particular, in addition to parameter change and change in the error distribution, they allow for change in over-identifying restrictions. See Ghysels and Hall (1990). For IV and GMM cases, the hypotheses that we consider are analogues of (3.6), rather than (3.5) (although one could design tests for (3.5) if desired).

3.1 $\mathbf{m \geq d}$ Case

First, we consider the case where the number of observations after the structural change, m, is greater than or equal to the number of regressors, d. In this case, the F statistic for testing H_0^* against H_1^* can be written as

$$\begin{aligned} F &= (\widehat{\beta}_1 - \widehat{\beta}_0)' \left(\widehat{\sigma}^2 \left[(\mathbf{X}_0'\mathbf{X}_0)^{-1} + (\mathbf{X}_1'\mathbf{X}_1)^{-1} \right] \right)^{-1} (\widehat{\beta}_1 - \widehat{\beta}_0)/d \\ &= (\mathbf{Y}_1 - \mathbf{X}_1\widehat{\beta}_0)'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} \left[(\mathbf{X}_0'\mathbf{X}_0)^{-1} + (\mathbf{X}_1'\mathbf{X}_1)^{-1} \right]^{-1} \\ &\quad \times (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{Y}_1 - \mathbf{X}_1\widehat{\beta}_0)/(d\widehat{\sigma}^2), \end{aligned} \tag{3.7}$$

where $\widehat{\beta}_j = (\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathbf{Y}_j$ for $j = 0, 1$ are the LS estimators using the pre- and post-change observations and $\widehat{\sigma}^2$ is the sum of squares residuals for $i = 1, ..., n + m$ (computed using $\widehat{\beta}_0$ and $\widehat{\beta}_1$) divided by $n + m - 2d$. Note that $(\mathbf{X}_1'\mathbf{X}_1)^{-1}$ only exists if $m \geq d$, as is assumed here.

Because the number of post-change observations, m, is fixed as $n \to \infty$, the standard F test is asymptotically valid only if the errors are normal, iid, and homoskedastic. (Normality is required because $\widehat{\beta}_1$ is not asymptotically normal as $n \to \infty$ since it is determined by only m observations.) These conditions on the errors are very restrictive. There are few applications in economics in which a test of structural change is of interest and these conditions are satisfied. In consequence, we propose alternative tests to the F test that utilize critical values that allow for much more general error processes. We consider test statistics that are slight variants of the F statistic.

The variance estimator, $\widehat{\sigma}^2$, appears in the F statistic to yield invariance of the statistic to the error variance σ^2. It does not contribute to the power of the statistic and we can eliminate it without reducing the power of the test. The critical values that we employ yield invariance with respect to σ^2. Similarly, we can replace the constant d by m without affecting power. The form of the F test is simplified considerably if we replace $(\mathbf{X}_0'\mathbf{X}_0)^{-1} + (\mathbf{X}_1'\mathbf{X}_1)^{-1}$ by $(\mathbf{X}_1'\mathbf{X}_1)^{-1}$. This simplification is warranted because n is much larger than m and $\lim_{n \to \infty} \left((\mathbf{X}_0'\mathbf{X}_0)^{-1} + (\mathbf{X}_1'\mathbf{X}_1)^{-1} \right) = (\mathbf{X}_1'\mathbf{X}_1)^{-1}$. With these alterations, we obtain the following variant of the F statistic:

$$S_a = m^{-1}(\mathbf{Y}_1 - \mathbf{X}_1\widehat{\beta}_0)'P_{\mathbf{X}_1}(\mathbf{Y}_1 - \mathbf{X}_1\widehat{\beta}_0), \tag{3.8}$$

where $P_{\mathbf{X}_1} = \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$.

Note that a test based on S_a has power against alternatives of the form H_1, not just H_1^*, because any deviations of β_{1i} from β_0 for $i = n + 1, ..., n + m$ cause the distribution of S_a to be stochastically larger than its distribution under the null hypothesis.

We now specify critical values for the statistic S_a. These critical values allow for non-normal, dependent, heteroskedastic errors. The main assumptions are that $\{W_i : i \geq 1\}$ is stationary and ergodic under the null hypothesis, $EU_1X_1 = 0$, EX_1X_1' is positive definite, U_1 has an absolutely continuous distribution, and some moment conditions hold. (Assumptions are stated formally in Section 8 below.)

First, we introduce some additional notation. Let

$$
\begin{aligned}
S_j(\beta) &= A_j(\beta)' V_j^{-1} A_j(\beta), \text{ where} \\
A_j(\beta) &= m^{-1} \sum_{i=j}^{j+m-1} (Y_i - X_i'\beta) X_i, \text{ and} \\
V_j &= m^{-1} \sum_{i=j}^{j+m-1} X_i X_i'
\end{aligned} \tag{3.9}
$$

for $j \geq 1$. Let

$$
\begin{aligned}
\widehat{\beta}_{r,s} &= \text{LS estimator of } \beta \text{ using observations indexed by } i = r, ..., s \text{ and} \\
\widehat{\beta}_{(j)} &= \text{LS estimator of } \beta \text{ using observations indexed by } i = 1, ..., n \text{ with} \\
& \quad i \neq j, ..., j+m-1
\end{aligned} \tag{3.10}
$$

for $j = 1, ..., n - m + 1$.

The statistic S_a can be written as

$$S_a = S_{n+1}(\widehat{\beta}_{1,n}). \tag{3.11}$$

Under the null hypothesis, the distribution of $S_{n+1}(\beta)$ is same as that of $S_j(\beta)$ for all $j \geq 1$ for all β, because $\{W_i : i \geq 1\}$ are stationary. The estimator $\widehat{\beta}_{1,n}$, which appears in the statistic S_a, converges in probability to the true parameter, β_0, under the null hypothesis. Hence, the asymptotic distribution of S_a is the distribution of $S_1(\beta_0)$.

Note that $\{S_j(\beta) : j \geq 1\}$ are stationary and ergodic for all β. In consequence, the empirical distribution function (df) of $\{S_j(\beta) : j = 1, ..., n-m+1\}$ is a consistent estimator of the df of $S_1(\beta)$ for all β. Hence, we can consistently estimate the df of $S_1(\beta_0)$ by using the empirical df of $\{S_j(\beta) : j \geq 1\}$ evaluated at consistent estimators of β_0. The estimator $\widehat{\beta}_{1,n}$, which appears in the statistic S_a, does not depend on the observations that appear in $S_{n+1}(\beta)$. To mirror this property, we evaluate $S_j(\beta)$ at $\widehat{\beta}_{(j)}$, which does not depend on the observations that appear in $S_j(\beta)$.[3] The estimator $\widehat{\beta}_{(j)}$ is consistent for β_0 (uniformly over j, under suitable assumptions).

Define

$$S_{a,j} = S_j(\widehat{\beta}_{(j)}) \text{ for } j = 1, ..., n - m + 1. \tag{3.12}$$

The empirical df of $\{S_{a,j} : j = 1, ..., n - m + 1\}$ converges in probability (and almost surely) to the df of $S_1(\beta_0)$. In consequence, to obtain a test with asymptotic significance level α, we take the critical value for the test statistic S_a to be the $1-\alpha$ sample quantile, $\widehat{q}_{a,1-\alpha}$, of $\{S_{a,j} : j = 1, ..., n - m + 1\}$. That is,

$$\widehat{q}_{a,1-\alpha} = \inf\{x \in R : \widehat{F}_{a,n}(x) \geq 1 - \alpha\}. \tag{3.13}$$

One rejects H_0 if $S_a > \widehat{q}_{a,1-\alpha}$. Equivalently, one rejects H_0 if S_a exceeds $100(1-\alpha)\%$ of the values $\{S_{a,j} : j = 1, ..., n - m + 1\}$. That is, if

$$(n - m + 1)^{-1} \sum_{j=1}^{n-m+1} 1(S_a > S_{a,j}) \geq 1 - \alpha. \tag{3.14}$$

The p-value for the test is

$$p_a = 1 - (n-m+1)^{-1} \sum_{j=1}^{n-m+1} 1(S_a > S_{a,j}). \tag{3.15}$$

A test that is easier to compute (and to program) than that based on S_a is obtained by using the estimator $\widehat{\beta}_{1+m,n+m}$ in $S_{n+1}(\cdot)$. Because $\widehat{\beta}_{1+m,n+m}$ depends on the observations used by $S_{n+1}(\cdot)$ (as well as others), the estimator used with $S_j(\cdot)$ should depend on the observations used by $S_j(\cdot)$. In consequence, we can use the same estimator $\widehat{\beta}_{1,n}$ in $S_j(\cdot)$ for all $j = 1, ..., n-m+1$. This avoids having to compute the estimators $\{\widehat{\beta}_{(j)} : j = 1, ..., n-m+1\}$. The S_b and $S_{b,j}$ statistics are defined as:

$$\begin{aligned} S_b &= S_{n+1}(\widehat{\beta}_{1+m,n+m}) \text{ and} \\ S_{b,j} &= S_j(\widehat{\beta}_{1,n}) \text{ for } j = 1, ..., n-m+1. \end{aligned} \tag{3.16}$$

The S_b test rejects H_0 if $S_b > \widehat{q}_{b,1-\alpha}$, where $\widehat{q}_{b,1-\alpha}$ is the $1-\alpha$ quantile of $\{S_{b,j} : j = 1, ..., n-m+1\}$. The p-value of the S_b test is given by (3.15) with a replaced by b.

A possible drawback of the S_b test is that its power may be lower than that of the S_a test. The reason is that the estimator $\widehat{\beta}_{1+m,n+m}$ depends on the observations used by $S_{n+1}(\cdot)$ and, hence, $A_{n+1}(\widehat{\beta}_{1+m,n+m})$ may be closer to the zero vector under H_1 than $A_{n+1}(\widehat{\beta}_{1,n})$ is. We investigate the relative powers of the S_a and S_b tests in the Monte Carlo simulation section. We find that there is little or no sacrifice in power by using S_b and the size properties of S_b are slightly better than those of S_a. Hence, the S_b test is recommended, especially if computation time is an issue.

Next, we discuss the S_a and S_b statistics when $m = d$. In this case, $P_{\mathbf{X}_1}$ is the projection matrix onto the column space of the $m \times m$ matrix $\mathbf{X}_1'\mathbf{X}_1$, which is full rank with probability one (because its expectation is positive definite by assumption). In consequence, $P_{\mathbf{X}_1} = I_m$ and

$$\begin{aligned} S_a &= m^{-1}(\mathbf{Y}_1 - \mathbf{X}_1\widehat{\beta}_0)'(\mathbf{Y}_1 - \mathbf{X}_1\widehat{\beta}_0) \\ &= m^{-1} \sum_{i=n+1}^{n+m} (Y_i - X_i'\widehat{\beta}_{1,n})^2. \end{aligned} \tag{3.17}$$

The statistic S_b has an analogous expression when $m = d$ with $\widehat{\beta}_{1,n}$ replaced by $\widehat{\beta}_{1+m,n+m}$.

A natural extension of the definition of the statistic S_a to the case where $m < d$ is made using the definition in (3.8) with $P_{\mathbf{X}_1}$ denoting the projection matrix onto the column space of the $m \times d$ matrix $\mathbf{X}_1$, which has column rank m $(< d)$, and hence, $P_{X_1} = I_m$. That is, when $m < d$, the natural definition of S_a is that given in (3.17).

In fact, this definition is the same as the definition of the statistic P_a (defined below), that we obtain when we consider a variant of the F statistic for the case where $m \leq d$. A natural extension of the S_b statistic to the case where $m < d$ is given by (3.17) with $\widehat{\beta}_{1,n}$ replaced by $\widehat{\beta}_{1+m,n+m}$.

In some Monte Carlo simulations analogous to those reported in Section 9, we consider the F statistic as defined in (3.7), but with $\widehat{\sigma}^2$ deleted. Critical values are constructed analogously to those for S_a. Neither the size nor the power results differ much from those of the simpler test based on S_a. In most cases, the differences in rejection probabilities are .005 or less. In consequence, we focus on the S_a and S_b tests in this paper rather than the F test itself.

In some additional Monte Carlo simulations, we consider a variant of the S_a statistic that is a quadratic form in $A_j(\widehat{\beta}_{1,n})$ but with a weight matrix based on the full sample. In particular, we replace $V_j = m^{-1} \sum_{i=j}^{j+m-1} X_i X_i'$ by the matrix $V_{1,n+m} = (n+m)^{-1} \sum_{i=1}^{n+m} X_i X_i'$. The same weight matrix is used for $S_{a,j}$. The resulting test is inferior both in terms of size and power to the S_a test. For this reason, we do not discuss this test further.

3.2 m $\leq$ d Case

Next, we consider the case where $m \leq d$. In this case, the F statistic for testing H_0^* against H_1^* can be written as

$$F = \frac{\left(\sum_{i=1}^{n+m} (Y_i - X_i'\widehat{\beta}_{1,n+m})^2 - \sum_{i=1}^{n} (Y_i - X_i'\widehat{\beta}_{1,n})^2 \right) / m}{\sum_{i=1}^{n} (Y_i - X_i'\widehat{\beta}_{1,n})^2 / (n-d)}. \tag{3.18}$$

The second sum in the numerator is the unrestricted sum of squared residuals because the residuals from the last m observations equal zero due to the perfect fit that is possible when $m \leq d$.

As above, the denominator of the F statistic does not contribute to the power of the test just to its invariance with respect to the error variance σ^2. In addition, the estimator $\widehat{\beta}_{1,n+m}$ differs by little from $\widehat{\beta}_{1,n}$ under the null and alternative hypotheses because n is much larger than m. Hence, we evaluate both sums of squares at $\widehat{\beta}_{1,n}$, which simplifies the statistic considerably. This leads to the following variant of the F statistic:

$$P_a = m^{-1} \sum_{i=n+1}^{n+m} (Y_i - X_i'\widehat{\beta}_{1,n})^2. \tag{3.19}$$

(The statistic is denoted P_a because statistics of this form are sometimes called *predictive* statistics, e.g., see Chow (1960).)

Note that $P_a = S_a$ when $m = d$ by (3.17). The same sort of equivalence occurs for the F statistic when $m = d$. That is, the F statistics in (3.7) and (3.18) are equal when $m = d$. Given the above extension of the S_a statistic to the case where $m < d$, we find that $P_a = S_a$ whenever $m \leq d$. Of course, the statistic P_a can be defined even when $m > d$. In this case, the P_a and S_a statistics differ.

We now specify critical values for the statistic P_a that allow for non-normal, dependent, heteroskedastic errors. The main assumptions are as above, see Section 8 for details.

Define

$$P_j(\beta) = m^{-1} \sum_{i=j}^{j+m-1} (Y_i - X_i'\beta)^2 \tag{3.20}$$

for $j = 1, ..., n+1$ and $\beta \in R^d$. Note that

$$P_a = P_{n+1}(\widehat{\beta}_{1,n}). \tag{3.21}$$

Under the null hypothesis, the distribution of $P_{n+1}(\beta)$ is same as that of $P_j(\beta)$ for all $j \geq 1$ for all β, because $\{W_i : i \geq 1\}$ are stationary. The estimator $\widehat{\beta}_{1,n}$, which appears in the statistic P_a, converges in probability to the true parameter, β_0, under the null hypothesis. Hence, the asymptotic distribution of P_a is the distribution of $P_1(\beta_0)$.

As with $\{S_j(\beta) : j \geq 1\}$, $\{P_j(\beta) : j \geq 1\}$ are stationary and ergodic for all β. In consequence, the empirical df of $\{P_j(\beta) : j = 1, ..., n-m+1\}$ is a consistent estimator of the df of $P_1(\beta)$ for all β. We can consistently estimate the df of $P_1(\beta_0)$ by the empirical df of $\{P_j(\beta) : j \geq 1\}$ evaluated at consistent estimators of β_0.

Define

$$P_{a,j} = P_j(\widehat{\beta}_{(j)}) \text{ for } j = 1, ..., n-m+1, \tag{3.22}$$

where $\widehat{\beta}_{(j)}$ is as defined in the subsection above. Critical values and p-values for the P_a statistic are obtained using $\{P_{a,j} : j = 1, ..., n-m+1\}$ in the same manner as for S_a using $\{S_{a,j} : j = 1, ..., n-m+1\}$ as described in (3.13)-(3.15).

We define analogues $(P_b, P_{b,j})$ to $(P_a, P_{a,j})$ in the same way and for the same reason that $(S_b, S_{b,j})$ are defined relative to $(S_a, S_{a,j})$. That is,

$$P_b = P_{n+1}(\widehat{\beta}_{1+m,n+m}) \text{ and } P_{b,j} = P_j(\widehat{\beta}_{1,n}). \tag{3.23}$$

Critical values and p-values for P_b are constructed in the same way as for S_a, but using $P_{b,j}$ in place of $S_{a,j}$.

As in the $m \geq d$ case, we carry out some Monte Carlo simulations for the F test as defined in (3.18) but with $\widehat{\sigma}^2$ deleted. Critical values are constructed analogously to those for P_a. The size and power results are quite similar to those of P_a. Hence, we focus on the simpler statistics P_a and P_b in this paper, rather than the F statistic.

3.3 Changes in the Regressor and Error Distributions

The null hypothesis H_0 imposes stationarity of $\{(Y_i, X_i) : i \geq 1\}$. Hence, a change in the distribution of the regressors $\{X_i : i \geq 1\}$ is not part of H_0. In many cases, this is not desirable. One does not want to reject the null hypothesis due to just a change in the regressor distribution.

As it turns out, this is not a problem. The tests P_a and P_b have no power asymptotically against changes in the regressor distribution because the test statistics depend only on the squared residuals for $i = n+1, ..., n+m$. The tests S_a and S_b have no power against location and/or scale changes in the regressor distribution. Furthermore, Monte Carlo simulations show that changes in the shape of the regressor distribution beyond location and scale changes have very little effect on the rejection rates of the S_a and S_b tests when the parameters are constant and the error distribution is constant, see Section 9.2.3. Hence, the S_a and S_b tests appear to have little to no power against changes just in the regressor distribution.

The tests P_a and P_b have power against changes in the error distribution that increase the $1-\alpha$ quantile of the distribution of the average of the squared errors $\{U_i^2 : i = n+1, ...n+m\}$. Similarly, the tests S_a and S_b have power against changes in the error distribution that increase the $1-\alpha$ quantile of the distribution of the average of the squared error $\{U_i^2 : i = n+1, ..., n+m\}$ after projection onto the space spanned by the regressors $\{X_i : i = n+1, ..., n+m\}$. For example, a sufficiently large increase in the variance of the errors will cause these tests to reject the null hypothesis.

The tests T_v for $T = S$ and P and $v = a$ and b obviously have power against changes in the parameter vector β_0. Hence, rejection of the null hypothesis by one of these tests provides evidence that either the parameter vector changed or the error distributions became more variable (roughly speaking).

4 Linear Instrumental Variables

Extension of the tests based on T_v for $T = S$ and P and $v = a$ and b to the case of linear IV estimators of a linear regression model is fairly straightforward. We define the statistics $(T_v, T_{v,j})$ for the IV case here. Critical values and p-values are then constructed in the same way as in the previous subsection.

The model of interest is as in (3.4), but with regressors that may be correlated with the errors. Let Z_i denote a vector of IVs for $i = 1, ..., n+m$. The two cases distinguished in the previous section, namely, $m \geq d$ and $m \leq d$, also arise here, but the distinction depends on the dimension of the IV vector Z_i, rather than the dimension of the regressor vector X_i. Hence, in this section, we let d denote the dimension of the IV vector Z_i and we let d_X denote the dimension of the regressor vector X_i and the parameter β. We assume that $d \geq d_X$.

The null and alternative hypotheses of interest are as in (3.6) but with the LS moments, $E(Y_i - X_i'\beta_0)X_i$, replaced by the IV moments $E(Y_i - X_i'\beta_0)Z_i$. Thus,

$$H_0 : \begin{cases} E(Y_i - X_i'\beta_0)Z_i = 0 \text{ for all } i = 1, ..., n+m \text{ and} \\ \{(Y_i, X_i, Z_i) : i \geq 1\} \text{ are stationary and ergodic} \end{cases}$$

$$H_1 : \begin{cases} E(Y_i - X_i'\beta_0)Z_i = 0 \text{ for all } i = 1, ..., n \text{ and} \\ E(Y_i - X_i'\beta_0)Z_i \neq 0 \text{ for some } i = n+1, ..., n+m \text{ and/or} \\ \text{the distribution of } U_i \text{ for some } i = n+1, ..., n+m \\ \text{differs from the distribution of } U_i \text{ for } i = 1, ..., n. \end{cases} . \quad (4.1)$$

The alternative hypothesis H_1 covers parameter instability, i.e., $\beta_{1i} \neq \beta_0$ for some $i = n+1, ..., n+m$, instability in the validity of the IVs, i.e., $EU_iZ_i \neq 0$ for some $i = n+1, ..., n+m$, and/or instability in the distribution of the errors. Tests have power against instability in the validity of the IVs only if there are over-identifying restrictions, i.e., $d > d_X$. Hence, the alternative effectively encompasses parameter instability, instability in *over-identifying* restrictions, and instability in the error distribution. The tests considered below have power against changes in the error distribution that increase the variability of the errors, roughly speaking, as in the linear regression case. The tests have little to no power against changes in the regressor or IV distributions, as is desirable in most cases.

The main assumptions are that $\{W_i = (Y_i, X_i', Z_i')' : i \geq 1\}$ are stationary and ergodic under the null hypothesis, $EU_1Z_1 = 0$, EZ_1Z_1' is positive definite, EX_1Z_1' has full row rank, U_1 has an absolutely continuous distribution, and some moment conditions hold, see Section 8.

We now define the test statistics and critical values for the linear IV case. The IV estimator using the observations indexed by $i = r, ..., s$ is defined by

$$\widehat{\beta}_{r,s} = \left(\sum_{r=1}^{s} X_i Z_i' (\sum_{r=1}^{s} Z_i Z_i')^{-1} \sum_{r=1}^{s} Z_i X_i' \right)^{-1} \sum_{r=1}^{s} X_i Z_i' (\sum_{r=1}^{s} Z_i Z_i')^{-1} \sum_{r=1}^{s} Z_i Y_i. \quad (4.2)$$

For $j = 1, ..., n - m + 1$, the IV estimator $\widehat{\beta}_{(j)}$ is defined analogously using the observations indexed by $i = 1, ..., n$ with $i \neq j, ..., j + m - 1$.

For the case where $m \geq d$, the statistics S_a and $S_{a,j}$ are defined in the linear IV case to be

$$\begin{aligned} S_a &= S_{n+1}(\widehat{\beta}_{1,n}) \text{ and } S_{a,j} = S_j(\widehat{\beta}_{(j)}), \text{ where} \\ S_j(\beta) &= A_j(\beta)' V_j^{-1} A_j(\beta), \\ A_j(\beta) &= m^{-1} \sum_{i=j}^{j+m-1} (Y_i - X_i'\beta) Z_i, \text{ and} \\ V_j &= m^{-1} \sum_{i=j}^{j+m-1} Z_i Z_i'. \end{aligned} \quad (4.3)$$

The statistics $(S_b, S_{b,j})$ in the IV case are defined using (4.2) and (4.3) to be $S_b = S_{n+1}(\widehat{\beta}_{1+m,n+m})$ and $S_{b,j} = S_j(\widehat{\beta}_{1,n})$. (This is just as in the LS case, but using the IV definitions of $\widehat{\beta}_{r,s}$, $\widehat{\beta}_{(j)}$, $A_j(\beta)$, and V_j.)

The statistics $(P_v, P_{v,j})$ for $v = a$ and b are defined as in (3.19)-(3.23), but with $\widehat{\beta}_{r,s}$ and $\widehat{\beta}_{(j)}$ defined using the IV estimator as in (4.2).

5 Generalized Method of Moments

In this section, we extend the tests based on T_v for $T = S$ and P and $v = a$ and b to the case of models estimated by GMM. This extension covers tests of structural change for models estimated by ML by taking the GMM moment function $g(W_i, \beta)$ to be the ML score function for the i-th observation. We define the statistics $(T_v, T_{v,j})$ for the GMM case below. Accompanying critical values and p-values are constructed in the same way as above.

We consider GMM moment conditions given by

$$Eg(W_i, \beta_0) = 0, \quad (5.1)$$

where $g(\cdot,\cdot)$ is a vector-valued function. The two cases distinguished in the linear regression section, namely, $m \geq d$ and $m \leq d$, also arise here, but the distinction depends on the number of moments, rather than the dimension of X_i. Hence, in this section, we let d denote the dimension of the function $g(\cdot,\cdot)$ and we let d_β denote the dimension of the parameter β. We assume that $d \geq d_\beta$.

The null and alternative hypotheses of interest are

$$H_0 : \begin{cases} Eg(W_i, \beta_0) = 0 \text{ for all } i = 1, ..., n+m \text{ and} \\ \{W_i : i \geq 1\} \text{ are stationary and ergodic} \end{cases}$$

$$H_1 : \begin{cases} Eg(W_i, \beta_0) = 0 \text{ for all } i = 1, ..., n \text{ and} \\ Eg(W_i, \beta_0) \neq 0 \text{ for some } i = n+1, ..., n+m \text{ and/or} \\ \text{the distribution of } g(W_i, \beta_0) \text{ for some } i = n+1, ..., n+m \\ \text{differs from that of } g(W_i, \beta_0) \text{ for } i = 1, ..., n. \end{cases} \quad (5.2)$$

As in the linear IV testing case, the alternative hypothesis covers parameter instability, invalidity of over-identifying restrictions, and instability in the error distribution when the moments are the product of an error and a IV vector.

We consider one-step, two-step, and continuously updated (CU) GMM estimators. The GMM estimator using the observations indexed by $i = r, ..., s$, denoted $\widehat{\beta}_{r,s}$, is defined to minimize one of the following three criteria over the parameter space $\mathcal{B}$:

$$\begin{aligned} Q_{r,s}^{(1)}(\beta) &= (\sum_{i=r}^{s} g(W_i, \beta))' V^{-1} \sum_{i=r}^{s} g(W_i, \beta) \\ Q_{r,s}^{(2)}(\beta) &= (\sum_{i=r}^{s} g(W_i, \beta))' V_{r,s}^{-1}(\widetilde{\beta}_{r,s}) \sum_{i=r}^{s} g(W_i, \beta), \text{ and} \\ Q_{r,s}^{(CU)}(\beta) &= (\sum_{i=r}^{s} g(W_i, \beta))' V_{r,s}^{-1}(\beta) \sum_{i=r}^{s} g(W_i, \beta), \end{aligned} \quad (5.3)$$

where $Q_{r,s}^{(1)}(\beta)$, $Q_{r,s}^{(2)}(\beta)$, and $Q_{r,s}^{(CU)}(\beta)$ are the one-step, two-step, and CU GMM criterion functions, respectively; the one-step weight matrix V is some fixed non-stochastic matrix, such as I_d; the weight matrix $V_{r,s}(\beta)$ depends on the observations indexed by $i = r, ..., s$; and the estimator $\widetilde{\beta}_{r,s}$ that appears in the two-step weight matrix is the one-step GMM estimator based on the observations indexed by $i = r, ..., s$.

For $j = 1, ..., n-m+1$, the one-step, two-step, and CU GMM criterion functions, $Q_{(j)}^{(k)}(\beta)$ for $k = 1$, 2, and CU and estimators $\widehat{\beta}_{(j)}$ are defined analogously using the observations indexed by $i = 1, ..., n$ with $i \neq j, ..., j+m-1$. Let $V_{(j)}(\beta)$ denote the matrix-valued function that is used in the definitions of the two-step and CU GMM criterion functions $Q_{(j)}^{(2)}(\beta)$ and $Q_{(j)}^{(CU)}(\beta)$.

For the case where $m \geq d$, the statistics S_a and $S_{a,j}$ are defined in the GMM case to be

$$S_a = S_{n+1}(\widehat{\beta}_{1,n}) \text{ and } S_{a,j} = S_j(\widehat{\beta}_{(j)}), \text{ where}$$

$$
\begin{aligned}
S_j(\beta) &= A_j(\beta)' V_j^{-1}(\beta) A_j(\beta), \\
A_j(\beta) &= m^{-1} \sum_{i=j}^{j+m-1} g(W_i, \beta),
\end{aligned} \tag{5.4}
$$

and $V_j(\beta) = V(W_j, ..., W_{j+m-1}, \beta)$ is some positive definite weight matrix that is a function of the observations $W_j, ..., W_{j+m-1}$ and the parameter β for $j = 1, ..., n+1$.

For example, suppose $g(W_i, \beta)$ is of the form

$$
g(W_i, \beta) = U(W_i, \beta) Z(W_i, \beta), \tag{5.5}
$$

where $U(W_i, \beta)$ is real-valued, $U_i = U(W_i, \beta_0)$ is an error term, and $Z(W_i, \beta)$ is a d-vector of instruments. Then, one can take

$$
V_j(\beta) = m^{-1} \sum_{i=j}^{j+m-1} Z(W_i, \beta) Z(W_i, \beta)'. \tag{5.6}
$$

In the special case where the GMM estimator is the LS estimator, i.e., $g(W_i, \beta) = (Y_i - X_i'\beta)X_i$, or the linear IV estimator, i.e., $g(W_i, \beta) = (Y_i - X_i'\beta)Z_i$, this choice yields the same weight matrix as considered in the subsections above.

Alternatively, one could take

$$
V_j(\beta) = m^{-1} \sum_{i=j}^{j+m-1} g(W_i, \beta) g(W_i, \beta)'. \tag{5.7}
$$

The asymptotic results given below cover both choices and any other choice of $V_j(\beta)$ that satisfies the stated assumptions.

The statistics $(S_b, S_{b,j})$ in the GMM case are defined just as $(S_a, S_{a,j})$ are defined in (5.4) but with $(\widehat{\beta}_{1+m,n+m}, \widehat{\beta}_{1,n})$ in place of $(\widehat{\beta}_{1,n}, \widehat{\beta}_{(j)})$.

The statistics $(P_v, P_{v,j})$ for $v = a$ and b are only defined in the GMM case when (5.5) holds. In this case, we define

$$
\begin{aligned}
P_j(\beta) &= m^{-1} \sum_{i=j}^{j+m-1} U^2(W_i, \beta) \text{ for } j = 1, ..., n+1, \\
P_a &= P_{n+1}(\widehat{\beta}_{1,n}), \;\; P_{a,j} = P_j(\widehat{\beta}_{(j)}), \\
P_b &= P_{n+1}(\widehat{\beta}_{1+m,n+m}), \text{ and } P_{b,j} = P_j(\widehat{\beta}_{1,n})
\end{aligned} \tag{5.8}
$$

for $j = 1, ..., n-m+1$.

6 Tests for Instability at the Beginning, or in the Middle, of the Sample

The tests introduced above for detecting instability at the end of the sample can be altered to detect instability occurring at the beginning or in the middle of the sample. For example, one might be interested in determining the most suitable starting date for a given model. Alternatively, one might be interested in whether a model behaves differently during war years or during a policy regime shift compared to other years in the sample. Often such periods of potential instability are of relatively short duration, so that asymptotic tests that rely on their length going to infinity are not appropriate. In such cases, the testing approach introduced above is useful because the length, m, of the time period of potential instability is taken to be fixed and finite in the asymptotics.

We consider the case of testing for structural instability for the m observations indexed by $i = i_0, ..., i_0 + m - 1$ when the total number of observations is $n + m$. For example, for the GMM case, the null and alternative hypotheses are given by

$$H_0 : \begin{cases} Eg(W_i, \beta_0) = 0 \text{ for all } i = 1, ..., n+m \text{ and} \\ \{W_i : i \geq 1\} \text{ are stationary and ergodic} \end{cases}$$
$$H_1 : \begin{cases} Eg(W_i, \beta_0) = 0 \text{ for all } i = 1, ..., i_0 - 1,\ i_0 + m, ..., n+m \text{ and} \\ Eg(W_i, \beta_0) \neq 0 \text{ for some } i = i_0, ..., i_0 + m - 1 \text{ and/or} \\ \text{the distribution of } g(W_i, \beta_0) \text{ for some } i = i_0, ..., i_0 + m - 1 \\ \text{differs from that of } g(W_i, \beta_0) \text{ for } i = 1, ..., i_0 - 1,\ i_0 + m, ..., n+m. \end{cases} \quad (6.1)$$

One can construct tests for these hypotheses by taking the summands $\{g(W_i, \beta) : i = i_0, ..., i_0 + m - 1\}$ and switching them with the summands $\{g(W_i, \beta) : i = n+1, ..., n+m\}$ in the sums that appear in the components of the test statistics and estimator criterion functions considered in the sections above. After making this switch, the tests defined above can be used to test the hypotheses in (6.1).

For the LS and IV testing cases, analogous switches deliver tests for instability for the observations indexed by $i = i_0, ..., i_0 + m - 1$.

7 Application to Models with I(1) Variables

The tests introduced above can be applied to some models with integrated variables of order one (I(1)). For example, consider two common $(p+1)$-th order autoregressive models with possible unit roots written in Dickey-Fuller representation:

$$\begin{aligned} Y_i &= \mu + (\alpha - 1)Y_{i-1} + \gamma_1 \Delta Y_{i-1} + ... + \gamma_p \Delta Y_{i-p} + U_i \text{ and} \\ Y_i &= \mu + \beta i + (\alpha - 1)Y_{i-1} + \gamma_1 \Delta Y_{i-1} + ... + \gamma_p \Delta Y_{i-p} + U_i. \end{aligned} \quad (7.1)$$

The second model contains a time trend. If $\alpha = 1$, the models have unit roots and are non-stationary. However, differenced versions of these models are strictly stationary for $|\alpha| \leq 1$ under suitable conditions on the errors U_i :

$$\begin{aligned} \Delta Y_i &= (\alpha - 1)\Delta Y_{i-1} + \gamma_1 \Delta^2 Y_{i-1} + ... + \gamma_p \Delta^2 Y_{i-p} + \Delta U_i \text{ and} \\ \Delta Y_i &= \beta + (\alpha - 1)\Delta Y_{i-1} + \gamma_1 \Delta^2 Y_{i-1} + ... + \gamma_p \Delta^2 Y_{i-p} + \Delta U_i. \end{aligned} \quad (7.2)$$

In consequence, one can test for structural instability at the end of the sample using the tests above applied to the models written in differenced form (7.2).

8 Asymptotic Results

In this section, we show that the tests introduced above are asymptotically valid.

8.1 Assumptions

In order to determine the behavior of the random critical values defined above under both H_0 and H_1, it is convenient to consider a sequence of random variables $\{W_{0,i} : i \geq 1\}$ that are stationary and ergodic under both H_0 and H_1. Under H_0, the observations are $W_i = W_{0,i}$ for $i = 1, ..., n + m$. Under H_1, the observations are from a triangular array, rather than a sequence, because the changepoint n changes as $n \to \infty$. Under H_1, the observations are $W_i = W_{0,i}$ for $i = 1, ..., n$ and $W_i = W_{n,i}$ for $i = n+1, ..., n+m$, where $\{W_{n,i} : i = n+1, ..., n+m\}$ are some random variables whose joint distribution is different from that of $\{W_{0,i} : i = n + 1, ..., n + m\}$. We assume that the distribution of $\{W_{n,i} : i = n+1, ..., n+m\}$ is independent of n. That is, we consider fixed, not local, alternatives.

For brevity, we do not state separate assumptions for each of the test statistics and models considered. Rather, we state generic assumptions that cover the statistic T_v, where T denotes S or P and v denotes a or b. For any given statistic T_v, we do not state which model or estimator is considered. It could be the linear regression model estimated by LS, the linear regression model estimated by IV, or the GMM model estimated by the one-step, two-step, or CU GMM estimator. Provided the assumptions hold for the model/estimator of choice, the asymptotic results hold for this choice. In Assumption 3, the term $T_{n+1}(\beta)$ denotes $S_{n+1}(\beta)$ or $P_{n+1}(\beta)$ and the form of these statistics depends on the model/estimator under consideration as defined above. For example, the statistic $S_{n+1}(\beta)$ is defined in (3.9), (4.3), or (5.4) for the LS, IV, and GMM cases respectively.

Let $B(\beta_0, \varepsilon)$ denote a ball centered at β_0 with radius $\varepsilon > 0$.

Assumption 1. $\{W_{0,i} : i \geq 1\}$ are stationary and ergodic.

Assumption 2. $\sup_{j=1,...,n-m+1} ||\widehat{\beta}_{(j)} - \beta_0|| \to_p 0$ and $||\widehat{\beta}_{r,s} - \beta_0|| \to_p 0$ for $(r, s) = (1, n)$ and $(r, s) = (1 + m, n + m)$ as $n \to \infty$ with m fixed under H_0 and H_1.

Assumption 3. **(a)** $T_{n+1}(\beta)$ is continuously differentiable in a neighborhood of β_0 with probability one under H_0 and H_1 and either $E \sup_{\beta \in B(\beta_o,\varepsilon)} ||(\partial/\partial\beta)T_1(\beta)|| < \infty$ or $(n - m + 1)^{-1} \sum_{j=1}^{n-m+1} \sup_{\beta \in B(\beta_{0},\varepsilon)} ||(\partial/\partial\beta)T_j(\beta)|| = O_p(1)$ for some $\varepsilon > 0$.

(b) $P(T_1(\beta_0) = x) = 0$ for all x in a neighborhood of $q_{1-\alpha}$, where $q_{1-\alpha}$ is the $1 - \alpha$ quantile of $T_1(\beta_0)$.

Assumption 1 is fairly general compared to many results in the testing literature. It allows for both asymptotically weakly dependent processes, such as mixing and near epoch dependent processes, as well as long-memory processes. It allows for conditional variation in all moments, including conditional heteroskedasticity.

Assumptions 2 and 3 hold for LS, IV, and GMM estimators under appropriate regularity conditions. The following are sufficient:

Assumption LS. **(a)** $EU_1X_1 = 0$.

(b) $E||U_1X_1||^{1+\delta} < \infty$ and $E||X_1||^{2+\delta} < \infty$ for some $\delta > 0$.
(c) EX_1X_1' is positive definite.
(d) U_1 has an absolutely continuous distribution.

Assumption IV. (a) $EU_1Z_1 = 0$.
(b) $EU_1^2 < \infty$, $E||X_1||^2 < \infty$, and $E||Z_1||^{2+\delta} < \infty$ for some $\delta > 0$.
(c) EZ_1Z_1' is positive definite and EX_1Z_1' has full row rank.
(d) U_1 has an absolutely continuous distribution.

Assumption GMM. (a) $Eg(W_1, \beta) = 0$ for $\beta \in \mathcal{B}$ if and only if $\beta = \beta_0 \in \mathcal{B}$.
(b) $\mathcal{B}$ is compact.
(c) $g(W_1, \beta)$ is continuous on $\mathcal{B}$ almost surely and $Eg(W_1, \beta)$ is continuous on $\mathcal{B}$.
(d) $E \sup_{\beta \in \mathcal{B}} ||g(W_1, \beta)||^{1+\delta} < \infty$ for some $\delta > 0$.
(e) The one-step GMM weight matrix V is non-stochastic and positive definite; the two-step GMM weight matrix functions $V_{(j)}(\beta)$ and $V_{r,s}(\beta)$ satisfy $\sup_{j=1,...,n-m+1}$ $\sup_{\beta \in B(\beta_0, \varepsilon)} |V_{(j)}(\beta) - V_{00}(\beta)| \rightarrow_p 0$ and $\sup_{\beta \in B(\beta_0, \varepsilon)} |V_{r,s}(\beta) - V_{00}(\beta)| \rightarrow_p 0$ for some $\varepsilon > 0$ for $(r, s) = (1, n)$ and $(r, s) = (1 + m, n + m)$, for some symmetric positive definite non-stochastic function $V_{00}(\beta)$ defined on $B(\beta_0, \varepsilon)$ that is continuous at β_0; and the CU weight matrix functions $V_{(j)}(\beta)$ and $V_{r,s}(\beta)$ satisfy analogous convergence conditions but with $B(\beta_0, \varepsilon)$ replaced by $\mathcal{B}$ and with $V_{00}(\beta)$ being a symmetric non-stochastic function with eigenvalues bounded away from zero on $\mathcal{B}$.
(f) When $T = S$, $g(W_1, \beta)$ is continuously differentiable on a neighborhood of β_0 almost surely, $V_j(\beta) = V(W_j, ..., W_{j+m-1}, \beta)$ is a positive definite weight matrix that is a function of the observations $W_j, ..., W_{j+m-1}$ and the parameter β and is continuously differentiable in β on a neighborhood of β_0 almost surely for $j = 1, ..., n+1$, $E \sup_{\beta \in B(\beta_0, \varepsilon)} ||g(W_1, \beta)||^2 < \infty$, $E \sup_{\beta \in B(\beta_0, \varepsilon)} (||(\partial/\partial\beta')g(W_1, \beta)|| \cdot ||g(W_1, \beta)||) < \infty$, and $\sup_{j=1,...,n-m+1} \sup_{\beta \in B(\beta_0, \varepsilon)} (||V_j^{-1}(\beta)|| + ||(\partial/\partial\beta_r)V_j^{-1}(\beta)||) = O_p(1)$ for $r = 1, ..., d_\beta$ for some $\varepsilon > 0$. When $T = P$, $U(W_1, \beta)$ is continuously differentiable on a neighborhood of β_0 almost surely and $E \sup_{\beta \in B(\beta_0, \varepsilon)} ||(\partial/\partial\beta)U(W_1, \beta)|| < \infty$.
(g) $g(W_1, \beta_0)$ has an absolutely continuous distribution when $T = S$ and $U(W_1, \beta_0)$ has an absolutely continuous distribution when $T = P$.

Assumptions LS(d), IV(d), and GMM(g) provide simple sufficient conditions for Assumption 3(b). But, they are undoubtedly much stronger than is necessary for Assumption 3(b) to hold.

Assumptions GMM(a)-(e), (f), and (g) are used to verify Assumptions 2 and 3(c), 3(a), and 3(b) respectively. The conditions involving $V_{(j)}(\beta)$ in Assumption GMM are only needed for the statistics S_a and P_a. The conditions involving $V_{r,s}(\beta)$ for $(r, s) = (1+m, n+m)$ in Assumption GMM are only needed for the statistics S_b and P_b.

Lemma 1 (a) *Assumptions* 1 *and* LS *imply Assumptions* 2 *and* 3 *for the LS estimator for the statistic* T_v *when* $T = S$ *or* P *and* $v = a$ *or* b.
(b) *Assumptions* 1 *and* IV *imply Assumptions* 2 *and* 3 *for the IV estimator for the statistic* T_v *when* $T = S$ *or* P *and* $v = a$ *or* b.
(c) *Assumptions* 1 *and* GMM *imply Assumptions* 2 *and* 3 *for the GMM estimator for the statistic* T_v *when* $T = S$ *or* P *and* $v = a$ *or* b.

8.2 Results

In this subsection, we state the asymptotic results that justify the use of the data-dependent critical values that are introduced above.

For $T = S$ or P, let $\widehat{F}_{T,a,n}(x)$ denote the empirical df based on $\{T_{a,j} = T_j(\widehat{\beta}_{(j)}) : j = 1, ..., n-m+1\}$. That is,

$$\widehat{F}_{T,a,n}(x) = \frac{1}{n-m+1} \sum_{j=1}^{n-m+1} 1(T_j(\widehat{\beta}_{(j)}) \leq x). \tag{8.1}$$

Define $\widehat{F}_{T,b,n}(x)$ as $\widehat{F}_{T,a,n}(x)$ is defined, but with $\widehat{\beta}_{(j)}$ replaced by $\widehat{\beta}_{1,n}$.

Let $F_{T_1}(x)$ denote the df of $T_1(\beta_0)$ at x. Let $q_{1-\alpha}$ denote the $1-\alpha$ quantile of $T_1(\beta_0)$.

Theorem 1 *Suppose Assumptions* 1-3 *hold for $T = S$ or P, then*
(a) $T_a \rightarrow_d T_{n+1}(\beta_0)$ *as $n \rightarrow \infty$ under H_0 and H_1, where the distribution of $T_{n+1}(\beta_0)$ equals that of $T_1(\beta_0)$ under H_0 and the distribution of $T_{n+1}(\beta_0)$ does not depend on n under either H_0 or H_1,*
(b) $\widehat{F}_{T,a,n}(x) \rightarrow_p F_{T_1}(x)$ *for all x in a neighborhood of $q_{1-\alpha}$ under H_0 and H_1,*
(c) $\widehat{q}_{a,1-\alpha} \rightarrow_p q_{1-\alpha}$ *under H_0 and H_1,*
(d) $P(T_a > \widehat{q}_{a,1-\alpha}) \rightarrow \alpha$ *under H_0, and*
(e) *the results of parts* (a)-(d) *hold with $(T_a, \widehat{F}_{T,a,n}(x), \widehat{q}_{a,1-\alpha})$ replaced by $(T_b, \widehat{F}_{T,b,n}(x), \widehat{q}_{b,1-\alpha})$.*

Comments: 1. Part (a) gives the asymptotic distribution of T_a under the null hypothesis and fixed alternatives. The distribution of $T_{n+1}(\beta_0)$ does not depend on n under H_1 because we take the distribution of $\{W_{n,i} : i = n+1, ..., n+m\}$ to be independent of n, which is appropriate for fixed alternatives.

2. Part (c) of the Theorem shows that the random critical value $\widehat{q}_{a,1-\alpha}$ has the same asymptotic behavior under H_1 as under H_0. This is desirable for the power of the test.

3. The idea of the proof of part (b) of the Theorem is to show that (i) the difference between $\widehat{F}_{T,a,n}(x)$ and a smoothed version of it, say, $\widehat{F}_{T,a,n}(x, h_n)$ converges in probability to zero, where h_n indexes the amount of smoothing and $h_n \rightarrow 0$ as $n \rightarrow \infty$, (ii) the difference between $\widehat{F}_{T,a,n}(x, h_n)$ and an analogous df with $\widehat{\beta}_{(j)}$ replaced by β_0 converges in probability to zero, (iii) the difference between the latter and the empirical df of $\{T_j(\beta_0) : j = 1, ..., n-m+1)\}$ converges in probability to zero as $n \rightarrow \infty$, and (iv) the difference between the latter and its expectation, $F_{T_1}(x)$, is asymptotically negligible. The reason for considering a smoothed version of $\widehat{F}_{T,a,n}(x)$ is that it is a smooth function of $\widehat{\beta}_{(j)}$ and, hence, result (ii) can be established by taking a mean-value expansion about β_0. Result (iv) holds by the ergodic theorem because $\{T_j(\beta_0) : j = 1, ..., n-m+1)\}$ is a finite subset of stationary and ergodic random variables using Assumption 1.

4. Part (a) shows that T_a does not diverge to infinity as $n \rightarrow \infty$ under H_1. Hence, T_a is not a consistent test. However, if $T_{n+1}(\beta_0)$ is stochastically greater than $T_1(\beta_0)$

under H_1, then T_a is an asymptotically unbiased test. For example, this occurs in a normal linear regression model estimated by LS or IV by the fact that a noncentral χ^2 distribution is stochastically increasing in its noncentrality parameter.

5. Stationarity under H_0 is not essential for the tests considered in the Theorem to be asymptotically valid. For example, in a linear regression model what is essential is stationarity of the error but not stationarity of the regressor. Provided the regressor behaves in a way that yields consistent estimators of β_0, i.e., Assumption 2 holds, the P_a and P_b tests will have the correct size asymptotically. To verify Assumption 2, one could use near epoch dependence (NED) or mixing conditions in place of stationarity and ergodicity. We use the stationarity and ergodicity condition here because it allows for more general dependence, such as long-memory dependence, and is simpler and more elegant than NED or mixing conditions.

9 Monte Carlo Experiment

In this section, we describe some Monte Carlo results that are designed to assess the size and power properties of the T_v tests for $T = S$ and P and $v = a$ and b and to compare the size properties of the T_v tests to those of the F test.

9.1 Experimental Design

We consider linear regression models estimated by LS, as in (3.1). Three pre-change sample sizes, n, are considered: 100, 250, and 500. Three post-change sample sizes, m, are considered: 1, 5, and 10. The number of regressors, d, is taken to be five. One regressor is a constant; the other four are independent of each other. Each of the latter regressors and the error is generated by an autoregressive process of order one (AR(1)) with the same AR parameter ρ. We consider three values of ρ : 0, .4, and .8. The innovations for the AR(1) processes are iid. We consider three different distributions for the innovations: standard normal, chi-square with two degrees of freedom (recentered and rescaled to have mean zero and variance one), and uniform on $[-\sqrt{12}, \sqrt{12}]$ (which has mean zero and variance one). (Note that the test size results are invariant with respect to the error variance.) The initial observations used to start up the AR(1) processes are taken to have the same distribution as the innovations, but are scaled to yield variance stationary processes.

Under the null hypothesis, the sample of $n+m$ observations is computed using the regression parameter vectors $\beta_0 = 0$ and $\beta_1 = 0$. (The test size results are invariant with respect to the value of β_0 $(= \beta_1)$.) Under the alternative hypothesis, $\beta_0 = 0$ and $\beta_1 \propto (1, 1, 1, 1, 1)',$ where $\propto$ denotes "is proportional to". For most results, we take $||\beta_1|| = 2$, where $||\beta_1||$ denotes the Euclidean norm. For some results, we take $||\beta_1|| = 1.25$.

The number of simulation repetitions used is 40,000 for each case considered. This yields simulation standard errors of (approximately) .001 for the simulated null rejection rates of nominal .05 tests and simulated standard errors in the interval (.0020, .0025) for the simulated alternative hypothesis rejection rates when these rejection rates are in the interval (.20, .80).

9.2 Monte Carlo Results

9.2.1 Size

Table I provides true null rejection rates for nominal .05 tests for the case of normal errors and regressors. The rejection rates of the tests T_v for $T = S$ and P and $v = a$ and b are always greater than or equal to .05. The rejection rates of the S_a and S_b tests are almost always very similar. In the few cases where there is a noticeable difference, the S_b test is better. The same is true for the comparison of P_a and P_b.

The absolute performance of the tests T_v in Table I is quite good for $\rho = 0$ and .4. It is not as good when $\rho = .8$ and $n = 100$. When $\rho = .8$, performance improves considerably as n increases. This occurs because there are more subsamples of length m upon which to construct the critical values when n is larger. When $\rho = 0$ or .4, performance improves with n, but not as quickly as when $\rho = .8$.

The size performance of the tests T_v in Table I declines as m increases. Presumably this occurs because there is increased overlap across the subsamples of length m when m increases.

When $m = 10$, the statistics S_a and P_a differ and the size performance of S_a is slightly superior to that of P_a in Table I. The same is true of S_b relative to P_b.

In Table I, the true null rejection rates of the F test are good when $\rho = 0$ and when $m = 1$. This is expected because in these cases the usual F critical value is asymptotically correct. The size of the F test is poor when $m = 5$ or 10 and $\rho = .4$. It is very poor when $m = 5$ or 10 and $\rho = .8$. For example, even for $n = 500$, the size of the F test is .266 when $m = 10$ and $\rho = .8$. In these cases, the usual F critical value is not asymptotically correct because of serial correlation in the errors.

Tables II provides true null rejection rates for nominal .05 tests for the case of chi-square errors and regressors. In general, the results of Table II are similar to those of Table I. The main difference is that the F test over-rejects much more when $\rho = 0$ and $m = 5$ or 10 with chi-square than with normal errors and regressors. This reflects the fact that the usual F critical value is not valid asymptotically in these cases with chi-square errors, whereas it is with normal errors. The performance of the T_v tests is not effected very much by the change in error and regressor distribution from normal to chi-square.

Tables III provides true null rejection rates for nominal .05 tests for the case of uniform errors and regressors. The results of Table III for the T_v tests are quite similar to those of Tables I and II. The results for the F test, however, are quite different. The F test greatly under-rejects the null hypothesis whenever $\rho = 0$. For larger ρ values, the tendency of the F test to under-reject due to uniform errors is offset by its tendency to over-reject due to positive serial correlation. In consequence, the F test has pretty good size when $\rho = .4$ and $m = 1$ and 5, but over-rejects when $\rho = .8$ and when $\rho = .4$ and $m = 10$.

Based on the size results of Tables I-III, we make the following conclusions: (i) The overall performance of the T_v tests is pretty good. These tests are not sensitive to the error/regressor distribution and are not very sensitive to the amount of serial correlation present. The case in which these tests perform most poorly is when $\rho = .8$ and $n = 100$. In this case, their performance is not satisfactory. (ii) The T_v tests outperform the F test in an overall sense. The F test is sensitive to the error/regressor distribution and to the amount of serial correlation present when $m = 5$ or 10. (iii) The S_b test performs at least as well as the S_a test and the same is true for P_b versus P_a. (iv) When $m = 10$, the performance of S_b is slightly better than that of P_b and the same is true of S_a versus P_a.

9.2.2 Power

Next, we discuss the power of the T_v tests. Table IV gives results for the case of normal errors and regressors. The tests are not size-corrected because the critical values are data-dependent, which makes size-correction problematic. This does not have much of an effect on the comparison of the powers of the T_a and T_b tests because their rejection rates under the null are quite similar. It does have a misleading effect on the influence of n on power when $\rho = .8$ and on the absolute level of power when $\rho = .8$ and n is small. This should be taken into consideration when analyzing the results.

Table IV shows that the differences in rejection rates between T_a and T_b are small for $T = S$ and P. Hence, there is little or no sacrifice in power for using the computationally simpler T_b statistic rather than the T_a statistic.

Table IV also shows that the power of T_v is increasing in m, as expected. Power is not necessarily increasing in n. This is because a larger value of n only helps to estimate β_0, but does not provide additional information about the post-change distribution of the observations. For $m = 1$, power is not sensitive to ρ. This is because the distribution of a subsample of size one is independent of ρ. For $m = 5$ or 10, power decreases with ρ. This is because there is less information about post-change parameter values when ρ is increased. When $m = 10$, S_a and P_a differ, as do S_b and P_b. In this case, the S_v test is found to be more powerful than the corresponding P_v test, see the last six rows of Table IV. This is especially true when $\rho = 0$.

For brevity power results for the cases of chi-square and uniform errors and regressors are not reported. The results for these cases are broadly similar to those for the normal case. The main differences are that (i) power is noticeably lower for all cases considered for chi-square errors and regressors compared to normal errors and regressors and (ii) power is noticeably higher for all cases considered for uniform errors and regressors compared to normal errors and regressors. For example, for $m = 1$, $\rho = 0$, and $n = 500$, power is .384, .300, and .467 for normal, chi-square, and uniform errors and regressors, respectively.

Based on the size and power results discussed above, we recommend using the S_b test when $m \geq d$ and the P_b test when $m \leq d$. These tests' size and power performances are at least as good as those of the S_a and P_a tests. When $m > d$, the power performance of S_b is better than that of P_a or P_b. Not surprisingly, the size performances of the S_b and P_b tests are much better than that of the F test in an overall sense.

9.2.3 Change in Regressor Distribution

We carry out some simulations to see whether a change in the regressor distribution alone causes the T_v tests to reject the null hypothesis more frequently than when there is no change in the regressor distribution, the parameters, or the error distribution. Six cases are considered. In each case, the pre-change regressor innovation distribution is $N(0,1)$, recentered and rescaled χ^2_2, or $U[-\sqrt{12}, \sqrt{12}]$ and the post-change regressor innovation distribution is one of these three distributions but a different one. We consider the same values of n, m, and ρ as above.

The rejection rates of the T_v tests in the above cases are always within .006 of their rejection rates for the corresponding cases that have the same pre-change regressor innovation distributions and no change in this distribution after $i = n$. This indicates that the T_v tests do not reject the null with probability greater than α when the only instability present is instability in the regressor distribution. In most cases, this is a desirable feature.

10 Appendix of Proofs

Proof of Theorem 1. We prove parts (a)-(e) for $T = S$. The proof for $T = P$ is the same and, hence, is not given.

We prove part (a) first. By Assumption 2, $\widehat{\beta}_{1,n} \to_p \beta_0$. In consequence, there exists a sequence of non-negative constants $\{\varepsilon_n : n \geq 1\}$ for which $\varepsilon_n \to 0$ and $P(K_n) \to 1$, where $K_n = \{||\widehat{\beta}_{1,n} - \beta_0|| < \varepsilon_n\}$. Let $x \in R$ be a continuity point of the df of $S_{n+1}(\beta_0)$. Let K_n^c denote the complement of the set K_n. We have

$$\begin{aligned}
&P(S_{n+1}(\widehat{\beta}_{1,n}) \leq x) \\
&= P(\{S_{n+1}(\widehat{\beta}_{1,n}) \leq x\} \cap K_n) + P(\{S_{n+1}(\widehat{\beta}_{1,n}) \leq x\} \cap K_n^c) \\
&\leq P(\inf_{||\beta-\beta_0||\leq\varepsilon_n} S_{n+1}(\beta) \leq x) + o(1) \\
&= P(S_{n+1}(\beta_0) \leq x) + o(1), \qquad (10.1)
\end{aligned}$$

where the last equality holds because Assumption 3(a) implies that $\inf_{||\beta-\beta_0||\leq\varepsilon} S_{n+1}(\beta) \to S_{n+1}(\beta_0)$ as $\varepsilon \to 0$ a.s. and $S_{n+1}(\beta)$ has a distribution that does not depend on n. If the inf is replaced by sup, then the $\leq$ is replaced by $\geq$. In consequence, $P(S_{n+1}(\widehat{\beta}_{1,n}) \leq x) \to P(S_{n+1}(\beta_0) \leq x)$ and part (a) is proved.

Next, we prove part (b). We introduce the following notation. For some random or non-random vectors $\{\beta_j : j = 1, ..., n-m+1\}$, let $\widehat{F}_n(x, \{\beta_j\})$ denote the empirical df based on $\{S_j(\beta_j) : j = 1, ..., n-m+1\}$, where $S_j(\beta)$ is defined in (3.9), (4.3), or (5.4) for the LS, linear IV, and GMM cases respectively. That is,

$$\widehat{F}_n(x, \{\beta_j\}) = \frac{1}{n-m+1} \sum_{j=1}^{n-m+1} 1(S_j(\beta_j) \leq x) \qquad (10.2)$$

for $x \in R$. Note that $\widehat{F}_{S,a,n}(x) = \widehat{F}_n(x, \{\widehat{\beta}_{(j)}\})$ and $\widehat{F}_{S,b,n}(x) = \widehat{F}_n(x, \{\widehat{\beta}_{1,n}\})$.

We define a smoothed version of the df $\widehat{F}_n(x, \{\beta_j\})$ as follows. Let $k(\cdot)$ be a monotone decreasing, everywhere differentiable, real function on R with bounded derivative and such that $k(x) = 1$ for $x \in (-\infty, 0]$, $k(x) \in [0, 1]$ for $x \in (0, 1)$, and $k(x) = 0$ for $x \in [1, \infty)$. For example, one could take $k(x) = \cos(\pi x)/2 + 1/2$ for $x \in (0, 1)$. For some random or non-random vectors $\{\beta_j : j = 1, ..., n - m + 1\}$, we define the smoothed df

$$\widehat{F}_n(x, \{\beta_j\}, h_n) = \frac{1}{n-m+1} \sum_{j=1}^{n-m+1} k((S_j(\beta_j) - x)/h_n), \qquad (10.3)$$

where $\{h_n : n \geq 1\}$ is a sequence of positive constants such that $h_n \to 0$ and $\sup_{j=1,...,n-m+1} ||\widehat{\beta}_{(j)} - \beta_0||/h_n \to_p 0$ as $n \to \infty$. Such a sequence exists by Assumption 2.

We have

$$|\widehat{F}_{S,a,n}(x) - F_{S_1}(x)| \leq \sum_{\ell=1}^{4} D_{\ell,n}, \text{ where}$$

$$
\begin{aligned}
D_{1,n} &= |\widehat{F}_{S,a,n}(x) - \widehat{F}_n(x, \{\widehat{\beta}_{(j)}\}, h_n)|, \\
D_{2,n} &= |\widehat{F}_n(x, \{\widehat{\beta}_{(j)}\}, h_n) - \widehat{F}_n(x, \{\beta_0\}, h_n)|, \\
D_{3,n} &= |\widehat{F}_n(x, \{\beta_0\}, h_n) - \widehat{F}_n(x, \{\beta_0\})|, \text{ and} \\
D_{4,n} &= |\widehat{F}_n(x, \{\beta_0\}) - F_{S_1}(x)|. \qquad (10.4)
\end{aligned}
$$

We have $D_{4,n} \to_p 0$ under H_0 and H_1 by the ergodic theorem. This holds because $\{S_1(\beta_0), ..., S_{n-m+1}(\beta_0)\}$ only depend upon the observations $\{W_1, ..., W_n\}$, which come from the stationary and ergodic sequence $\{W_{0,i} : i \geq 1\}$, and not on the "post change" observations $\{S_{n+1}(\beta_0), ..., S_{n+m}(\beta_0)\}$. Each term $S_j(\beta_0)$ is the same measurable function of m observations $\{W_j, ..., W_{j+m-1}\}$ for $j = 1, ..., n-m+1$, where m is fixed and finite. Hence, $\{S_1(\beta_0), ..., S_{n-m+1}(\beta_0)\}$ is a finite subsequence of a stationary and ergodic sequence of random variables that depend on $\{W_{0,i} : i \geq 1\}$ and the ergodic theorem applies.

We have

$$
D_{1,n} \leq \frac{1}{n-m+1} \sum_{j=1}^{n-m+1} 1(S_j(\widehat{\beta}_{(j)}) - x \in (0, h_n)), \qquad (10.5)
$$

because $\widehat{F}_{S,a,n}(x)$ and $\widehat{F}_n(x, \{\widehat{\beta}_{(j)}\}, h_n)$ only differ when $(S_j(\widehat{\beta}_{(j)}) - x)/h_n \in (0, 1)$.

By Assumption 2, there exists a sequence of positive constants $\{\varepsilon_n : n \geq 1\}$ such that $\varepsilon_n \to 0$ and $P(L_n) \to 1$, where $L_n = \{||\widehat{\beta}_{(j)} - \beta_0|| \leq \varepsilon_n, \forall j = 1, ..., n-m+1\}$. Now, for all $\delta > 0$,

$$
\begin{aligned}
& P\left(D_{1,n} > \delta\right) \\
&\leq P((D_{1,n} > \delta) \cap L_n) + P(L_n^c) \\
&\leq P\left(\frac{1}{n-m+1} \sum_{j=1}^{n-m+1} \sup_{||\beta-\beta_0|| \leq \varepsilon_n} 1(S_j(\beta) - x \in (0, h_n)) > \delta\right) + o(1) \\
&\leq E \sup_{||\beta-\beta_0|| \leq \varepsilon_n} 1\left(S_1(\beta) - x \in (0, h_n)\right)/\delta + o(1) \\
&\leq E1\left(S_1(\beta_0) - x \in (-\varepsilon_n \sup_{||\beta-\beta_0|| \leq \varepsilon_n} ||\frac{\partial}{\partial\beta} S_1(\beta)||,\right. \\
&\qquad \left. h_n + \varepsilon_n \sup_{||\beta-\beta_0|| \leq \varepsilon_n} ||\frac{\partial}{\partial\beta} S_1(\beta)||)\right)/\delta + o(1), \qquad (10.6)
\end{aligned}
$$

where L_n^c denotes the complement of the set L_n, the third inequality uses Markov's inequality and the identical distributions of $S_j(\beta_0)$ for $j = 1, ..., n-m+1$, and the fourth inequality holds by a mean-value expansion of $S_1(\beta)$ about β_0 using Assumption 3(a). The right-hand side of (10.6) is $o(1)$ by the dominated convergence theorem using $f(\cdot) = 1$ as the dominating function, because $\varepsilon_n \to 0$, $h_n \to 0$, $\limsup_{n\to\infty} \sup_{||\beta-\beta_0|| \leq \varepsilon_n} ||(\partial/\partial\beta) S_1(\beta)|| < \infty$ a.s. by Assumption 3(a), and $S_1(\beta_0) \neq x$ a.s. by Assumption 3(b). Hence, $D_{1,n} \to_p 0$.

An analogous, but simpler, argument shows that $D_{3,n} \to_p 0$.

For part (b), it remains to show that $D_{2,n} \to_p 0$. By mean-value expansions about β_0, we have:

$$\begin{aligned} D_{2,n} &= \left| \frac{1}{n-m+1} \sum_{j=1}^{n-m+1} k'((S_j(\widetilde{\beta}_{(j)}) - x)/h_n) \frac{\partial}{\partial \beta'} S_j(\widetilde{\beta}_{(j)})(\widehat{\beta}_{(j)} - \beta_0)/h_n \right| \\ &\leq \left(\frac{1}{n-m+1} \sum_{j=1}^{n-m+1} B \sup_{||\beta - \beta_0|| \leq \delta} ||\frac{\partial}{\partial \beta} S_j(\beta)|| \right) \sup_{r=1,\ldots,n-m+1} ||\widehat{\beta}_{(r)} - \beta_0||/h_n \\ &= O_p(1) o_p(1), \qquad (10.7) \end{aligned}$$

where $k'(\cdot)$ denotes the derivative of $k(\cdot)$, $\widetilde{\beta}_{(j)}$ lies between $\widehat{\beta}_{(j)}$ and β_0, $B < \infty$ denotes the bound on the derivative of $k(\cdot)$, the inequality uses the fact that $\sup_{j=1,\ldots,n-m+1} ||\widetilde{\beta}_{(j)} - \beta_0|| < \delta$ for some $\delta > 0$ with probability that goes to one by Assumption 2, and the second equality holds by Assumption 3(a) (either directly by assumption or by Markov's inequality) and by the fact that h_n is defined such that $\sup_{r=1,\ldots,n-m+1} ||\widehat{\beta}_{(r)} - \beta_0||/h_n \to_p 0$. This completes the proof of part (b).

Part (c) is implied by part (b) using Assumption 3(b). This is a standard result. It follows from the fact that for all small $\varepsilon > 0$, $\widehat{F}_{S,a,n}(q_{1-\alpha} - \varepsilon) \to_p F_{S_1}(q_{1-\alpha} - \varepsilon) < 1 - \alpha$ and $\widehat{F}_{S,a,n}(q_{1-\alpha} + \varepsilon) \to_p F_{S_1}(q_{1-\alpha} + \varepsilon) > 1 - \alpha$.

Part (d) is implied by parts (a) and (c) using Assumption 3(b).

The proofs of parts (a)-(d) with $(S_a, \widehat{F}_{S,a,n}(x), \widehat{q}_{a,1-\alpha})$ replaced by $(S_b, \widehat{F}_{S,b,n}(x), \widehat{q}_{b,1-\alpha})$ are analogous to those given above because Assumption 2 implies that the estimator $\widehat{\beta}_{1+m,n+m}$ behaves like $\widehat{\beta}_{1,n}$ and $\widehat{\beta}_{(j)}$ asymptotically. Hence, part (e) holds. $\square$

Proof of Lemma 1. We start by showing that Assumptions 1 and LS imply that $\sup_{j=1,\ldots,n-m+1} ||\widehat{\beta}_{(j)} - \beta_0|| \to_p 0$ for the LS case and Assumptions 1 and IV imply the same result for the IV case. We use the following result. Suppose that $\{\xi_i : i \geq 1\}$ is a stationary and ergodic sequence of mean zero random variables and $E||\xi_i||^{1+\delta} < \infty$ for some $\delta > 0$. Then,

$$\begin{aligned} &\sup_{j=1,\ldots,n-m+1} ||(n-m)^{-1} \sum_{i=1,\ldots,n; i \neq j,\ldots,j+m-1} \xi_i|| \\ \leq &\sup_{j=1,\ldots,n-m+1} ||(n-m)^{-1} (\sum_{i=1,\ldots,n; i \neq j,\ldots,j+m-1} \xi_i - \sum_{i=1}^{n} \xi_i)|| + ||(n-m)^{-1} \sum_{i=1}^{n} \xi_i|| \\ = &\sup_{j=1,\ldots,n-m+1} ||(n-m)^{-1} \sum_{i=j,\ldots,j+m-1} \xi_i|| + o_p(1), \qquad (10.8) \end{aligned}$$

where the equality holds by the ergodic theorem. Let $\tau_j = \sum_{i=j,\ldots,j+m-1} \xi_i$. For all $\varepsilon > 0$,

$$P((n-m)^{-1} \sup_{j \leq n-m+1} ||\tau_j|| > \varepsilon) = P(\cup_{j=1}^{n-m+1} \{||\tau_j|| > (n-m)\varepsilon\})$$

$$
\begin{aligned}
&\leq \sum_{j=1}^{n-m+1} P(||\tau_j|| > (n-m)\varepsilon) \\
&\leq (n-m+1)E||\tau_j||^{1+\delta}(n-m)^{-(1+\delta)}\varepsilon^{-(1+\delta)} \\
&= o(1), \qquad (10.9)
\end{aligned}
$$

where the second inequality uses Markov's inequality. Hence, the right-hand side of (10.8) is $o_p(1)$.

The estimator $\widehat{\beta}_{(j)}$ in the LS case satisfies

$$
\begin{aligned}
&\sup_{j=1,...,n-m+1} ||\widehat{\beta}_{(j)} - \beta_0|| \\
&= \sup_{j=1,...,n-m+1} ||((n-m)^{-1} \sum_{i=1,...,n;i\neq j,...,j+m-1} X_i X_i')^{-1} \\
&\quad \times (n-m)^{-1} \sum_{i=1,...,n;i\neq j,...,j+m-1} X_i U_i|| \\
&\leq ||(EX_1X_1' + o_p(1))^{-1}(EX_1U_1 + o_p(1)|| \\
&= 0, \qquad (10.10)
\end{aligned}
$$

where the inequality holds by applying (10.8) and (10.9) twice with $\xi_i = X_i X_i'$ and $\xi_i = X_i U_i$ and, in consequence, the $o_p(1)$ terms hold uniformly over $j = 1, ..., n-m+1$.

The proof of the same result for the linear IV estimator is quite similar using the definition of the IV estimator in (4.2). In this case, (10.8) and (10.9) are applied with $\xi_i = X_i Z_i'$, $\xi_i = Z_i Z_i'$, and $\xi_i = Z_i U_i$. (Note that $EU_1^2 < \infty$ and $E||Z_1||^{2+\delta} < \infty$ imply that $E||U_1Z_1||^{1+\delta_1} < \infty$ for some $\delta_1 > 0$ by Holder's inequality.)

The proof of $||\widehat{\beta}_{r,s} - \beta_0|| \rightarrow_p 0$ under H_0 and H_1 for $(r, s) = (1, n)$ and $(r, s) = (1+m, n+m)$ for the LS and IV estimators is fairly standard and, hence, is not given. (Note that the proof for $\widehat{\beta}_{1+m,n+m}$ under H_1 uses the fact that the distribution of $\{W_{n,i} : i = n+1, ..., n+m\}$ is independent of n.) Thus, Assumption 2 holds for the LS and IV estimators.

Next, we consider the verification of Assumption 3(a) for the LS and IV estimators. First, we consider $T = S$. For the LS estimator,

$$
\frac{\partial}{\partial \beta} S_j(\beta) = -2m^{-1} \sum_{i=j}^{j+m-1} (Y_i - X_i'\beta)X_i. \qquad (10.11)
$$

We have $E \sup_{||\beta-\beta_0||<\delta} ||(\partial/\partial\beta)S_1(\beta)|| < \infty$ because $E||U_1X_1|| < \infty$ and $E||X_1||^2 < \infty$.

For the IV estimator,

$$
\frac{\partial}{\partial \beta} S_j(\beta) = -2(m^{-1} \sum_{i=j}^{j+m-1} X_i Z_i')(m^{-1} \sum_{i=j}^{j+m-1} Z_i Z_i)^{-1} m^{-1} \sum_{i=j}^{j+m-1} (Y_i - X_i'\beta)Z_i. \qquad (10.12)
$$

The rhs can be written as $-2m^{-1}X_{jj}'P_{Z_{jj}}(Y_{jj}-X_{jj}\beta)$, where $Y_{jj}=(Y_j,...,Y_{j+m-1})'$, $X_{jj}=(X_j,...,X_{j+m-1})'$, $Z_{jj}=(Z_j,...,Z_{j+m-1})'$, and $P_{Z_{jj}}$ denotes the projection matrix that projects onto the column space of Z_{jj}. For simplicity, suppose X_i is a scalar. Then, X_{jj} is a vector and by the Cauchy-Schwartz inequality,

$$\begin{aligned}|X_{jj}'P_{Z_{jj}}(Y_{jj}-X_{jj}\beta)| &\leq (X_{jj}'P_{Z_{jj}}X_{jj})^{1/2}((Y_{jj}-X_{jj}\beta)'P_{Z_{jj}}(Y_{jj}-X_{jj}\beta))^{1/2}\\ &\leq (X_{jj}'X_{jj})^{1/2}((Y_{jj}-X_{jj}\beta)'(Y_{jj}-X_{jj}\beta))^{1/2}. \qquad (10.13)\end{aligned}$$

The rhs has finite expectation because $E||X_1||^2<\infty$ and $E||U_1X_1||^2<\infty$. Hence, $E\sup_{||\beta-\beta_0||<\delta}||(\partial/\partial\beta)S_1(\beta)||<\infty$ holds for the IV estimator when $T=S$.

Now suppose $T=P$. In this case, for both the LS and IV estimators, we have

$$\frac{\partial}{\partial\beta}P_j(\beta)=-2m^{-1}\sum_{i=j}^{j+m-1}(Y_i-X_i'\beta)X_i \qquad (10.14)$$

and $E\sup_{||\beta-\beta_0||<\delta}||(\partial/\partial\beta)P_1(\beta)||<\infty$ because $E||U_1X_1||<\infty$ and $E||X_1||^2<\infty$.

Assumption 3(b) holds for the LS and IV estimators because U_i has an absolutely continuous distribution. This completes the proofs of parts (a) and (b) of the Lemma.

We now prove part (c) of the Lemma, which concerns the GMM estimator. To show that $\sup_{j=1,...,n-m+1}||\widehat{\beta}_{(j)}-\beta_0||\rightarrow_p 0$, we extend the standard proof of consistency for nonlinear extremum estimators. First, we verify that for $k=1$, 2, and CU and all $\varepsilon>0$,

$$\begin{aligned}&\sup_{\beta\in\mathcal{B}}\sup_{j=1,...,n-m+1}|Q_{(j)}^{(k)}(\beta)-Q^{(k)}(\beta)|\rightarrow_p 0 \text{ and} && (10.15)\\ &Q^{(k)}(\beta_0)<\inf_{\beta\notin B(\beta_0,\varepsilon)\cap\mathcal{B}}Q^{(k)}(\beta), \text{ where} && (10.16)\\ &Q^{(1)}(\beta)=Eg(W_1,\beta)'V^{-1}Eg(W_1,\beta), &&\\ &Q^{(2)}(\beta)=Eg(W_1,\beta)'V^{-1}(\beta_0)Eg(W_1,\beta), \text{ and} &&\\ &Q^{(CU)}(\beta)=Eg(W_1,\beta)'V^{-1}(\beta)Eg(W_1,\beta). && (10.17)\end{aligned}$$

Condition (10.15) holds provided

$$\sup_{\beta\in\mathcal{B}}\sup_{j=1,...,n-m+1}|(n-m)^{-1}\sum_{i=1,...,n;i\neq j,...,j+m-1}g(W_i,\beta)-Eg(W_1,\beta)|\rightarrow_p 0, \quad (10.18)$$

because Assumption GMM(e) insures that the weight matrices are well-behaved. Equation (10.18) holds pointwise in β for all $\beta\in\mathcal{B}$ by applying (10.8) and (10.9) with $\xi_i=g(W_i,\beta)-Eg(W_1,\beta)$ using Assumptions 1 and GMM(d). Then, a generic uniform convergence result strengthens pointwise convergence to uniform convergence over $\beta\in\mathcal{B}$. In particular, Theorem 5 of Andrews (1992) using Assumption TSE-1D gives the desired result under Assumptions 1 and GMM(b)-(d).

Condition (10.16) holds by Assumption GMM(a)-(c) and (e).

Next, we use (10.15) and (10.16) to show that $\sup_{j=1,...,n-m+1}||\widehat{\beta}_{(j)}-\beta_0||\rightarrow_p 0$. By (10.16), given $\delta>0$ there exists $\varepsilon>0$ such that $||\beta-\beta_0||>\varepsilon$ implies that

$Q^{(k)}(\beta) - Q^{(k)}(\beta_0) > \delta$ for $k = 1$, 2, and CU. Hence, we have

$$
\begin{aligned}
& P(\sup_{j=1,\ldots,n-m+1} ||\widehat{\beta}_{(j)} - \beta_0|| > \varepsilon) \\
& \leq P(\sup_{j=1,\ldots,n-m+1} Q^{(k)}(\widehat{\beta}_{(j)}) - Q^{(k)}(\beta_0) > \delta) \\
& = P(\sup_{j=1,\ldots,n-m+1} (Q^{(k)}(\widehat{\beta}_{(j)}) - Q^{(k)}_{(j)}(\widehat{\beta}_{(j)}) + Q^{(k)}_{(j)}(\widehat{\beta}_{(j)}) - Q^{(k)}(\beta_0)) > \delta) \\
& \leq P(\sup_{j=1,\ldots,n-m+1} (Q^{(k)}(\widehat{\beta}_{(j)}) - Q^{(k)}_{(j)}(\widehat{\beta}_{(j)}) + Q^{(k)}_{(j)}(\beta_0) - Q^{(k)}(\beta_0)) > \delta) \\
& \leq P(2 \sup_{j=1,\ldots,n-m+1} |Q^{(k)}_{(j)}(\beta) - Q^{(k)}(\beta)| > \delta) \\
& = o_p(1),
\end{aligned} \tag{10.19}
$$

where the second inequality holds because $\widehat{\beta}_{(j)}$ minimizes $Q^{(k)}_{(j)}(\beta)$ over $\beta \in \mathcal{B}$ and the second equality holds by (10.15).

The proof that $||\widehat{\beta}_{r,s} - \beta_0|| \rightarrow_p 0$ for (r, s) as in Assumption 2 is standard (and is a special case of the proof above) and, hence, is not given. This completes the verification of Assumption 2 for the GMM case.

To establish Assumption 3(a) for GMM estimators, we note that $T_{n+1}(\beta)$ is continuously differentiable on a neighborhood of β_0 by Assumption GMM(f). Next, we verify that $B_n := (n - m + 1)^{-1} \sum_{j=1}^{n-m+1} \sup_{\beta \in B(\beta_0, \varepsilon)} ||(\partial/\partial\beta)T_j(\beta)|| = O_p(1)$. For $T = S$, we have

$$
\begin{aligned}
\frac{\partial}{\partial \beta_r} S_j(\beta) = {} & 2 \left(m^{-1} \sum_{i=j}^{j+m-1} \frac{\partial}{\partial \beta_r} g(W_i, \beta) \right)' V_j^{-1}(\beta) m^{-1} \sum_{i=j}^{j+m-1} g(W_i, \beta) \\
& + m^{-1} \sum_{i=j}^{j+m-1} g(W_i, \beta)' \left(\frac{\partial}{\partial \beta_r} (V_j^{-1}(\beta)) \right) m^{-1} \sum_{i=j}^{j+m-1} g(W_i, \beta)
\end{aligned} \tag{10.20}
$$

for $r = 1, ..., d_\beta$. The matrices $V_j^{-1}(\beta)$ and $(\partial/\partial\beta_r)V_j^{-1}(\beta)$ have stochastically bounded Euclidean norms uniformly over β in a neighborhood of β_0 and over $j = 1, ..., n-m+1$ using Assumption GMM(f). In consequence, it suffices to show the desired result with $V_j^{-1}(\beta)$ and $(\partial/\partial\beta_r)V_j^{-1}(\beta)$ replaced by I_d. The latter holds by Markov's inequality given the moment conditions in Assumption GMM(f).

For $T = P$, $(\partial/\partial\beta)P_j(\beta) = 2 \sum_{i=j}^{j+m-1} (\partial/\partial\beta)U(W_i, \beta)$ and $B_n = O_p(1)$ by Markov's inequality and the moment condition in Assumption GMM(f).

Assumption 3(b) holds for GMM estimators because $g(W_1, \beta_0)$ or $U(W_1, \beta_0)$ has an absolutely continuous distribution by Assumption GMM(g). $\square$

Footnotes

[1] The author thanks Ray Fair for suggesting the testing problem considered in this paper and for helpful comments. The author gratefully acknowledges the research support of the National Science Foundation via grant number SBR-9730277.

[2] We use the observations indexed by $i = 1, ..., n$, rather than the full sample, to estimate the parameter when defining $P_{b,j}$ because we do not want the post-change observations to affect the critical value calculation, which could reduce power.

Working backwards, to ensure that the parameter estimator used by P_b uses the same number of observations, n, as $P_{b,j}$ does for each j, we use the n observations indexed by $i = 1+m, ..., n+m$, rather than the full sample, to estimate the parameter when defining P_b. Simulation results indicate that defining P_b using the full sample to estimate the parameter, rather than the observations indexed by $i = 1+m, ..., n+m$, coupled with $P_{b,j}$ defined as is, leads to a slight increase in over-rejection of the test.

[3] The estimator $\widehat{\beta}_{1,n}$, which appears in the statistic S_a, depends on n observations. In contrast, the estimators $\{\widehat{\beta}_{(j)} : j = 1, ..., n-m+1\}$, which appear in the statistics $\{S_j(\widehat{\beta}_{(j)}) : j = 1, ..., n-m+1\}$, only depend on $n-m$ observations. To see whether the use of the same number of observations by all estimates of β leads to better size results, we carried out some Monte Carlo simulations for the case where S_a is defined using the estimator $\widehat{\beta}_{1,n-m}$ instead of $\widehat{\beta}_{1,n}$. This had essentially no effect on the size of the test for the cases considered. (Why this occurs in the present case, but not in the case described in footnote 2, is unclear.)

References

Andrews, D. W. K. (1992): "Generic Uniform Convergence," *Econometric Theory,* 8, 241-257.

——— (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica,* 61, 821-856.

Andrews, D. W. K. and R. C. Fair (1988): "Inference in Nonlinear Econometric Models with Structural Change," *Review of Economic Studies*, 55, 615-640.

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

Fair, R. C. (2002): "Testing for a New Economy in the 1990s," unpublished manuscript, Cowles Foundation, Yale University.

Ghysels, E., A. Guay, and A. Hall (1997): "Predictive Tests for Structural Change with Unknown Breakpoint," *Journal of Econometrics,* 82, 209-233.

Ghysels, E. and A. Hall (1990): "A Test for Structural Stability of Euler Conditions Parameters Estimated Via the Generalized Method of Moments Estimator," *International Economic Review,* 31, 355-364.

Hansen, B. E. (1992): "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business and Economic Statistics,* 10, 321-336.

Politis, D. N., J. P. Romano, and M. Wolf (1999): *Subsampling.* Springer Series in Statistics. New York: Springer.

Stock, J. H. (1994): "Unit Roots, Structural Breaks and Trends," Chapter 46 in *Handbook of Econometrics, Volume IV,* ed. by R. F. Engle and D. L. McFadden. New York: North-Holland.

Table I

True Sizes of Tests with Nominal Size .05

Using Normally Distributed Regressors and Errors

Number of Post-change Observations m	Autoregressive Coefficient ρ	Test	Number of Pre-change Observations n 100	250	500
1	0	P_a	.060	.053	.052
		P_b	.059	.053	.052
		F	.050	.052	.050
1	.4	P_a	.064	.054	.053
		P_b	.063	.054	.053
		F	.053	.052	.052
1	.8	P_a	.087	.064	.057
		P_b	.085	.064	.057
		F	.074	.061	.054
5	0	S_a	.063	.059	.054
		S_b	.064	.059	.054
		F	.051	.051	.052
5	.4	S_a	.068	.059	.055
		S_b	.065	.059	.055
		F	.070	.068	.067
5	.8	S_a	.095	.070	.061
		S_b	.087	.070	.061
		F	.147	.124	.118

Table I (cont)

m	ρ	Test	$n = 100$	$n = 250$	$n = 500$
10	0	S_a	.076	.062	.056
		S_b	.075	.063	.055
		P_a	.085	.065	.057
		P_b	.088	.066	.057
		F	.049	.049	.050
10	.4	S_a	.083	.066	.055
		S_b	.086	.066	.055
		P_a	.089	.067	.057
		P_b	.090	.067	.057
		F	.120	.119	.116
10	.8	S_a	.115	.077	.061
		S_b	.113	.077	.060
		P_a	.116	.078	.061
		P_b	.114	.078	.061
		F	.333	.286	.266

Table II

True Sizes of Tests with Nominal Size .05

Using χ^2_2 Regressors and Errors

m	ρ	Test	$n = 100$	$n = 250$	$n = 500$
1	0	P_a	.059	.056	.052
		P_b	.059	.056	.052
		F	.055	.057	.052
1	.4	P_a	.062	.052	.051
		P_b	.060	.052	.051
		F	.052	.050	.050
1	.8	P_a	.085	.059	.056
		P_b	.083	.058	.056
		F	.069	.053	.049
5	0	S_a	.058	.059	.054
		S_b	.057	.059	.054
		F	.104	.102	.100
5	.4	S_a	.060	.057	.052
		S_b	.060	.057	.052
		F	.099	.094	.093
5	.8	S_a	.087	.067	.058
		S_b	.078	.065	.057
		F	.141	.117	.107
10	.0	S_a	.081	.063	.057
		S_b	.081	.063	.057
		F	.085	.087	.089
10	.4	S_a	.085	.064	.056
		S_b	.086	.065	.057
		F	.131	.121	.120
10	.8	S_a	.115	.074	.059
		S_b	.109	.072	.059
		F	.318	.267	.243

Table III
True Sizes of Tests with Nominal Size .05
Using Uniform Regressors and Errors

m	ρ	Test	$n = 100$	$n = 250$	$n = 500$
1	0	P_a	.058	.052	.053
		P_b	.057	.053	.053
		F	.009	.002	.001
1	.4	P_a	.065	.054	.054
		P_b	.065	.054	.055
		F	.032	.028	.026
1	.8	P_a	.090	.066	.059
		P_b	.087	.065	.059
		F	.071	.059	.054
5	0	S_a	.061	.057	.053
		S_b	.063	.057	.053
		F	.007	.004	.003
5	.4	S_a	.069	.060	.054
		S_b	.069	.059	.053
		F	.040	.035	.034
5	.8	S_a	.095	.067	.057
		S_b	.088	.067	.056
		F	.141	.119	.109
10	0	S_a	.070	.060	.055
		S_b	.073	.060	.055
		F	.029	.024	.023
10	.4	S_a	.083	.066	.057
		S_b	.084	.066	.058
		F	.105	.101	.100
10	.8	S_a	.115	.077	.060
		S_b	.111	.076	.060
		F	.327	.287	.270

Table IV
Power of Tests with Nominal Significance Level .05
Using Normally Distributed Regressors and Errors

$\|\|\beta_1\|\|$	m	ρ	Test	$n = 100$	$n = 250$	$n = 500$
2.0	1	0	P_a	.387	.379	.384
			P_b	.386	.379	.384
2.0	1	.4	P_a	.388	.375	.380
			P_b	.388	.376	.380
2.0	1	.8	P_a	.416	.375	.370
			P_b	.412	.374	.370
2.0	5	0	S_a	.796	.819	.817
			S_b	.787	.817	.816
2.0	5	.4	S_a	.705	.729	.730
			S_b	.707	.732	.731
2.0	5	.8	S_a	.508	.490	.478
			S_b	.522	.501	.485
2.0	10	0	S_a	.971	.978	.979
			S_b	.974	.978	.980
2.0	10	.4	S_a	.905	.923	.928
			S_b	.917	.928	.931
2.0	10	.8	S_a	.623	.612	.614
			S_b	.666	.640	.630
1.25	10	0	S_b	.755	.790	.799
			P_b	.635	.687	.695
1.25	10	.4	S_b	.614	.640	.641
			P_b	.542	.575	.576
1.25	10	.8	S_b	.375	.342	.321
			P_b	.363	.336	.317