END-OF-SAMPLE INSTABILITY TESTS

BY

DONALD W. K. ANDREWS

COWLES FOUNDATION PAPER NO. 1072

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2003

http://cowles.econ.yale.edu/

Econometrica, Vol. 71, No. 6 (November, 2003), 1661–1694

END-OF-SAMPLE INSTABILITY TESTS

BY D. W. K. ANDREWS[1]

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, generalized method of moments estimator, least squares estimator, parameter change, structural instability test, structural change.

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), Andrews and Ploberger (1994), Ghysels, Guay, and Hall (1997), and other references listed 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 appropriate 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 a test that is appropriate in this case because it is asymptotically valid when $n \to \infty$ with m fixed. The primary difficulty in constructing this test is in obtaining asymptotically valid critical values.

Dufour, Ghysels, and Hall (1994) (DGH) also consider the above testing problem. They specify three different methods of obtaining critical values. But, each of the three methods has some drawback. The first method requires

[1]The author thanks Ray Fair for suggesting the testing problem considered in this paper. He thanks Patrik Guggenberger, Ray Fair, three referees, and the co-editor for helpful comments. In particular, he is very grateful to a referee who suggested the simplified motivation for the test statistic presented in Section 2.2. The author thanks Alastair Hall for pointing out the closely related paper by Dufour, Ghysels, and Hall (1994). The author gratefully acknowledges the research support of the National Science Foundation via Grant Numbers SBR-9730277 and SES-0001706.

strong distributional assumptions, such as normality of the errors. The second method relies on a bound obtained using Markov's inequality and, hence, is not exact even in large samples. The third method utilizes semi-nonparametric density estimation methods, which involves the usual problems associated with choosing the most appropriate basis functions and truncation values.

The results of this paper differ from those of DGH primarily in the specification of the critical values. The critical values considered here apply under very weak distributional assumptions, do not involve any bounding, do not require the specification of any truncation or smoothing parameters, and are quick to compute. The test statistics that we consider are similar to those of DGH, but different in some cases. In particular, the tests that we propose are more powerful than those of DGH in the case where m exceeds the number of regressors in a linear regression testing problem.

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 a variant of the F test, called the S test, that is valid under weak assumptions and applies to a wide variety of models. We do so by constructing critical values using a subsampling-like method. In the linear regression model, the S test is asymptotically valid with nonnormal, 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 S test applies 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 the S test for structural instability at the *end* of the sample. Extending the S test to the case of potential instability at the beginning, rather than the end, of the sample is trivial. This test can be used to determine the start of the sample period that is most appropriate for a given model. In addition, we show how the S test can be used to test for structural instability that occurs over a small number of observations in the middle of the sample. For example, the S test 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 appropriate in these situations because the number of observations in the period of change, m, is small and, hence, asymptotics that rely on $m \to \infty$ often lead to distortions in the null rejection rates of the tests.

Note that the S test with $m = 1$ can be used to test for a single outlier in the errors at a known time. There is a considerable literature in statistics on outlier detection; e.g., see Belsley, Kuh, and Welsch (1980) and Cook and Weisberg (1982), to which the S test makes a contribution.

We now briefly describe the S test for a regression model. The S test is a variant of the F test that is obtained after transforming the model to account for serial correlation of the errors. In particular, the S test statistic is a quadratic form in a *transformed* post-change residual vector evaluated at the full sample least squares (LS) estimator. The transformation is by the square root of the inverse of an estimator of the $m \times m$ covariance matrix of the errors. Like the F statistic (see Chow (1960)), the weight matrix of the quadratic form depends on whether the number of post-change observations, m, is greater than or less than the number of regressors d. When $m > d$, the weight matrix is the projection matrix onto the column space of the transformed post-change regressor matrix. When $m \leq d$, the weight matrix is the inverse of the transformation matrix squared.

Critical values for the S test statistic are obtained by a subsampling-like method that we call *parametric subsampling*. One computes the $n - m + 1$ test statistics that are analogous to the S test statistic but are for testing for structural instability over the m observations that start at the jth observation, rather than for instability starting at the $(n+1)$th observation, for $j = 1, \ldots, n-m+1$. 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 S statistic. The critical value is asymptotically valid even in the presence of serial correlation of the errors because the critical value behaves like a sample quantile based on stationary and ergodic random variables when the sample size is large and such a sample quantile is consistent.

If the errors are believed to be uncorrelated or close to being uncorrelated, then the S test can be simplified by replacing the transformation matrix by the identity matrix. The resulting test is asymptotically valid whether or not the errors are correlated. But, it has lower power than the S test if there is significant correlation in the errors.

The parametric subsampling critical values rely on subsamples of length m, the number of post-change observations. There is no arbitrary smoothing parameter or block length parameter to select. Also, no heteroskedasticity and autocorrelation consistent covariance matrix estimator is required. These critical values are not pure subsampling critical values because the test statistic for a given value of j depends on observations other than those indexed by $j, \ldots, j + m - 1$ 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.

Part of the maintained hypothesis of the S test is structural stability over the first n observations. This can be tested using standard tests for structural stability over a relatively long period of time, such as those referenced above.

Given that m is fixed as $n \to \infty$, the S test is not a consistent test. However, it typically is asymptotically unbiased. The power of the S test 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.

The paper presents some Monte Carlo simulation results for end-of-sample instability tests for linear regression models with first-order autoregressive (AR) errors and regressors. The AR parameters considered are $\rho = 0$, .4, and .8. The AR innovations considered are normal, chi-square with two degrees of freedom, t_3, and uniform. The pre-change sample sizes are $n = 100$ and 250 and the post-change sample sizes are $m = 10, 5$, and 1. We find that the S test has null rejection probabilities that are quite close to the nominal size of the test over the range of cases considered. On the other hand, the null rejection probabilities of the F test are too large whenever $\rho > 0$ and $m > 1$ and too small whenever $\rho = 0$ and the errors and regressors are uniform. When the errors are uncorrelated, the power of the S test is comparable to that of the F test after both have been corrected to have the same null rejection probability. When the errors are correlated, the S test has higher power than the F test (because the S test utilizes a transformation to account for correlation of the errors).

The S test has been used effectively by Fair (2003). Fair (2003) finds evidence of structural change in a U.S. stock market equation in the late 1990's, but no structural change in most other U.S. macroeconometric equations that he considers.

We note that the S test can be applied to pth order autoregressive models that may have a unit root by differencing the observations and applying the tests to the differenced data. Andrews and Kim (2002) provide closely related tests for cointegration breakdown over short time intervals in linear cointegration models with nonstationary observations.

The remainder of this paper is organized as follows. All sections of the paper except Section 4 discuss tests for instability that occurs at the *end* of the sample. Section 2 considers the linear regression model estimated by LS and provides motivation for the statistic considered in this model and others. Section 3 considers moment condition models estimated by GMM. Section 4 discusses tests for structural instability that occurs at the beginning or in the middle of the sample. Section 5 briefly discusses application of the tests to simple models with integrated variables. Section 6 introduces high-level assumptions, provides sufficient conditions for these assumptions for LS, linear IV, and GMM cases, and states the main asymptotic results. Section 7 provides some Monte Carlo results. An Appendix contains proofs.

2. LINEAR REGRESSION MODEL

2.1. *Introduction*

In this section, we consider a linear regression model with d regressors, n observations before the potential changepoint, and m observations after the potential changepoint:

$$Y_i = \begin{cases} X_i'\beta_0 + U_i & \text{for } i = 1, \ldots, n, \\ X_i'\beta_{1i} + U_i & \text{for } i = n+1, \ldots, n+m. \end{cases} \tag{2.1}$$

We assume that $EU_iX_i = 0$, 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

$$\begin{aligned} &H_0 : \begin{cases} \beta_{1i} = \beta_0 \text{ for all } i = n+1, \ldots, n+m \text{ and} \\ \{(Y_i, X_i) : i \geq 1\} \text{ are stationary and ergodic,} \end{cases} \\ &H_1 : \begin{cases} \beta_{1i} \neq \beta_0 \text{ for some } i = n+1, \ldots, n+m \text{ and/or} \\ \text{the distribution of } (U_{n+1}, \ldots, U_{n+m}) \text{ differs from} \\ \text{that of } (U_i, \ldots, U_{i+m-1}) \text{ for } i = 1, \ldots, n-m+1. \end{cases} \end{aligned} \tag{2.2}$$

The hypotheses of interest also can be expressed as

$$\begin{aligned} &H_0 : \begin{cases} E(Y_i - X_i'\beta_0)X_i = 0 \text{ for all } i = 1, \ldots, n+m \text{ and} \\ \{(Y_i, X_i) : i \geq 1\} \text{ are stationary and ergodic,} \end{cases} \\ &H_1 : \begin{cases} E(Y_i - X_i'\beta_0)X_i = 0 \text{ for all } i = 1, \ldots, n, \text{ and} \\ E(Y_i - X_i'\beta_0)X_i \neq 0 \text{ for some } i = n+1, \ldots, n+m \text{ and/or} \\ \text{the distribution of } (U_{n+1}, \ldots, U_{n+m}) \text{ differs from} \\ \text{that of } (U_i, \ldots, U_{i+m-1}) \text{ for } i = 1, \ldots, n-m+1. \end{cases} \end{aligned} \tag{2.3}$$

For linear regression models estimated by LS, the hypotheses in (2.2) and (2.3) are equivalent. But, for overidentified moment condition models estimated by GMM, which are considered below, hypotheses that are analogous to those in (2.3) allow for more general structural change than those in (2.2). In particular, in addition to parameter change and change in the error distribution, they allow for change in overidentifying restrictions; see Ghysels and Hall (1990) and Hall and Sen (1999). For the GMM case, the hypotheses that we consider are analogues of (2.3), rather than (2.2) (although one could design tests for (2.2) if desired).

Let

$$\begin{aligned} &\mathbf{Y}_{r,s} = (Y_r, \dots, Y_s)', \\ &\mathbf{X}_{r,s} = (X_r, \dots, X_s)', \quad \text{and} \\ &\mathbf{U}_{r,s} = (U_r, \dots, U_s)' \end{aligned} \tag{2.4}$$

for $1 \le r \le s \le n + m$.

For notational simplicity, we abbreviate the subscript "$n+1, n+m$" by "$n+$." For example, $\mathbf{Y}_{n+} = \mathbf{Y}_{n+1,n+m}$.

2.2. *Motivation for the Test Statistic*

In this subsection, we consider the standard F statistic for a one-time shift in the parameters in the fixed-regressor normal linear regression model. This statistic motivates the form of the test statistics considered in the paper for more general models—both regression models and others. 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.

We note that the standard F test should not be used for the model and hypotheses in (2.1) and (2.3) because it is asymptotically valid only if the errors are normal, iid, and homoskedastic. (This occurs because the number of post-change observations, m, is fixed as $n \to \infty$.) 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, as we now describe.

We distinguish between the "test-generating" model and hypotheses considered in this subsection and the more general model and hypotheses of interest specified in (2.1) and (2.3). Let β_0 denote the parameter vector for the first n observations. Suppose β_0 is known. The test-generating model for the last m observations can be written in matrix notation as

$$\mathbf{Y}_{n+} - \mathbf{X}_{n+}\beta_0 = \mathbf{X}_{n+}\delta_0 + \mathbf{U}_{n+}. \tag{2.5}$$

The test-generating null and alternative hypotheses are $\delta_0 = 0$ and $\delta_0 \neq 0$, respectively, where δ_0 is a d-vector. Suppose the errors are iid normal and the regressors are fixed (i.e., exogenous). Then, the F test for testing $\delta_0 = 0$ has well-known optimality properties; e.g., see Scheffé (1959, Sec. 2.10). These optimality properties are the motivation for constructing test statistics below based on the F statistic.

The F statistic is based on the difference between the restricted and unrestricted sums of squares:

$$SSR_R = (\mathbf{Y}_{n+} - \mathbf{X}_{n+}\beta_0)'(\mathbf{Y}_{n+} - \mathbf{X}_{n+}\beta_0) \quad \text{and}$$
$$SSR_U = (\mathbf{Y}_{n+} - \mathbf{X}_{n+}\beta_0)'(I_m - \mathbf{P}_{\mathbf{X}_{n+}})(\mathbf{Y}_{n+} - \mathbf{X}_{n+}\beta_0), \tag{2.6}$$

where $\mathbf{P}_{\mathbf{X}_{n+}}$ is the projection matrix onto the column space of the post-change regressor matrix $\mathbf{X}_{n+}$. In particular, the numerator of the F statistic is

$$SSR_R - SSR_U = (\mathbf{Y}_{n+} - \mathbf{X}_{n+}\beta_0)'\mathbf{P}_{\mathbf{X}_{n+}}(\mathbf{Y}_{n+} - \mathbf{X}_{n+}\beta_0). \tag{2.7}$$

The denominator of the F statistic is given by an estimator of the error variance times a constant. The denominator is employed to yield a test that is invariant to the error variance. It does not add to the power of the test. For the tests introduced below, we use a subsampling critical value that does not require estimation of the error variance to achieve invariance. In consequence, the test statistic we consider is based on the numerator of the F statistic as given in (2.7).

If the errors are serially correlated, then a more powerful test can be obtained by transforming the model (2.5) using the $m \times m$ covariance matrix Σ_0 of $\mathbf{U}_{n+}$. Suppose Σ_0 is known and is positive definite. Then, the transformed test-generating model is

$$\mathbf{Y}^*_{n+} - \mathbf{X}^*_{n+}\beta_0 = \mathbf{X}^*_{n+}\delta_0 + \mathbf{U}^*_{n+}, \quad \text{where} \tag{2.8}$$
$$\mathbf{Y}^*_{n+} = \Sigma_0^{-1/2}\mathbf{Y}_{n+},$$
$$\mathbf{X}^*_{n+} = \Sigma_0^{-1/2}\mathbf{X}_{n+}, \quad \text{and}$$
$$\mathbf{U}^*_{n+} = \Sigma_0^{-1/2}\mathbf{U}_{n+}.$$

The numerator of the F statistic for the transformed model is

$$(\mathbf{Y}^*_{n+} - \mathbf{X}^*_{n+}\beta_0)'\mathbf{P}_{\mathbf{X}^*_{n+}}(\mathbf{Y}^*_{n+} - \mathbf{X}^*_{n+}\beta_0). \tag{2.9}$$

When the number of post-change observations, m, is greater than or equal to the dimension of the regression parameter vector, d (and $\mathbf{X}_{n+}$ has full column rank d), the numerator of the F statistic can be written as

$$(\mathbf{Y}_{n+} - \mathbf{X}_{n+}\beta_0)'\Sigma_0^{-1}\mathbf{X}_{n+}(\mathbf{X}'_{n+}\Sigma_0^{-1}\mathbf{X}_{n+})^{-1}\mathbf{X}'_{n+}\Sigma_0^{-1}(\mathbf{Y}_{n+} - \mathbf{X}_{n+}\beta_0). \tag{2.10}$$

When the number of post-change observations, m, is less than or equal to the dimension of the regression parameter vector, d (and $\mathbf{X}_{n+}$ has rank m), the projection matrix $\mathbf{P}_{\mathbf{X}^*_{n+}}$ equals I_m and the numerator of the F statistic reduces to

$$(\mathbf{Y}_{n+} - \mathbf{X}_{n+}\beta_0)'\Sigma_0^{-1}(\mathbf{Y}_{n+} - \mathbf{X}_{n+}\beta_0). \tag{2.11}$$

Note that when $m = d$, $\mathbf{X}_{n+}$ is a square invertible matrix and the two formulae coincide.

The test statistics that we consider below are based on the formulae of (2.10) and (2.11) for the cases where $m \geq d$ and $m \leq d$, respectively.

2.3. *Definition of the S Test Statistic*

In this subsection we define the recommended test statistic S for the model and hypotheses specified in (2.1) and (2.3). It is based on the formulae of (2.10) and (2.11) with β_0 and Σ_0 replaced by estimators. This replacement does not affect the asymptotic properties of the test because the estimators of β_0 and Σ_0 are based on $n+m$ observations, where $n+m \to \infty$, whereas the quadratic forms in (2.10) and (2.11) are based on only m observations, where $m \nrightarrow \infty$.

We now define S. Let

$$\widehat{\beta}_{n+m} = \text{LS estimator of } \beta \text{ using observations indexed by } i = 1, \ldots, n+m. \tag{2.12}$$

The estimator of the $m \times m$ covariance matrix of the errors, $\Sigma_0 = E\mathbf{U}_{1,m}\mathbf{U}'_{1,m}$, is

$$\widehat{\Sigma}_{n+m} = (n+1)^{-1}\sum_{j=1}^{n+1}\widehat{\mathbf{U}}_{j,j+m-1}\widehat{\mathbf{U}}'_{j,j+m-1}, \quad \text{where} \tag{2.13}$$
$$\widehat{\mathbf{U}}_{j,j+m-1} = \mathbf{Y}_{j,j+m-1} - \mathbf{X}_{j,j+m-1}\widehat{\beta}_{n+m}.$$

(Note that by stationarity and ergodicity under H_0, $\Sigma_0 = E\mathbf{U}_{j,j+m-1}\mathbf{U}'_{j,j+m-1}$ for any $j = 1, \ldots, n+1$.)

When $m \geq d$, the statistic S is defined as

$$S = S_{n+1}(\widehat{\beta}_{n+m}, \widehat{\Sigma}_{n+m}), \tag{2.14}$$

where

$$S_j(\beta, \Sigma) = A_j(\beta, \Sigma)'V_j^{-1}(\Sigma)A_j(\beta, \Sigma),$$
$$A_j(\beta, \Sigma) = \mathbf{X}'_{j,j+m-1}\Sigma^{-1}(\mathbf{Y}_{j,j+m-1} - \mathbf{X}_{j,j+m-1}\beta), \tag{2.15}$$
$$V_j(\Sigma) = \mathbf{X}'_{j,j+m-1}\Sigma^{-1}\mathbf{X}_{j,j+m-1},$$

$\beta \in R^d$, and Σ is a nonsingular $m \times m$ matrix, for $j = 1, \ldots, n+1$. (Note that we define $S_j(\beta, \Sigma)$, $A_j(\beta, \Sigma)$, and $V_j(\Sigma)$ for $j = n+1$ and $j \neq n+1$ in (2.15) because for $j \neq n+1$ these quantities are used in the subsample statistics introduced below.)

The statistic S is a positive definite quadratic form given by the projection of the transformed m-vector of post-change residuals, $\widehat{\Sigma}_{n+m}^{-1/2}(\mathbf{Y}_{j,j+m-1} -$

$\mathbf{X}_{j,j+m-1}\widehat{\beta}_{n+m}$), onto the column space of the post-change transformed regressor matrix $\widehat{\Sigma}_{n+m}^{-1/2}\mathbf{X}_{j,j+m-1}$. If the null hypothesis is true, the post-change residuals are centered around zero and the quadratic form has a distribution that is relatively close to zero. On the other hand, if the alternative hypothesis is true, the post-change residuals are not centered around zero, because the LS estimator $\widehat{\beta}_{n+m}$ is not a consistent estimator of the post-change β_{1i} vectors, and the quadratic form has a distribution that is farther from zero. Thus, a large value of S is evidence against the null hypothesis.

When $m \leq d$, the statistic S is defined as

$$S = P_{n+1}(\widehat{\beta}_{n+m}, \widehat{\Sigma}_{n+m}), \quad \text{where} \tag{2.16}$$

$$P_j(\beta, \Sigma) = (\mathbf{Y}_{j,j+m-1} - \mathbf{X}_{j,j+m-1}\beta)'\Sigma^{-1}(\mathbf{Y}_{j,j+m-1} - \mathbf{X}_{j,j+m-1}\beta) \tag{2.17}$$

and $\widehat{\beta}_{n+m}$ and $\widehat{\Sigma}_{n+m}$ are as defined above.

When $m \leq d$, S is the sum of squared transformed post-change residuals. Again, large values of S provide evidence against the null hypothesis.

The test statistic $P_{n+1}(\widehat{\beta}_{n+m}, \widehat{\Sigma}_{n+m})$ is well defined even if $m > d$. For convenience, for all m and d, we define

$$P = P_{n+1}(\widehat{\beta}_{n+m}, \widehat{\Sigma}_{n+m}). \tag{2.18}$$

(Hence, for $m \leq d$, $S = P$.) In Section 7, a test based on P is compared via simulation to one based on S when $m > d$ to see whether projection on the space spanned by the transformed post-change regressor matrix improves performance.

2.4. *Critical Values and p-Values*

Critical values for the statistic S are obtained as follows. Under the null, $\{S_j(\beta, \Sigma) : j \geq 1\}$ are stationary and ergodic for all β and Σ because $\{(Y_i, X_i) : i \geq 1\}$ are stationary and ergodic. In addition, $\widehat{\beta}_{n+m}$ and $\widehat{\Sigma}_{n+m}$ are consistent estimators of β_0 and Σ_0 respectively. In consequence, the asymptotic null distribution of $S_{n+1}(\widehat{\beta}_{n+m}, \widehat{\Sigma}_{n+m})$ is the distribution of $S_1(\beta_0, \Sigma_0)$ (see Theorem 1 below). The empirical distribution function (df) of $\{S_j(\beta, \Sigma) : j = 1, \ldots, n - m + 1\}$ is a consistent estimator of the df of $S_1(\beta, \Sigma)$ for all β and Σ. This holds under the null and the alternative because $\{S_j(\beta, \Sigma) : j = 1, \ldots, n - m + 1\}$ only depends on the stationary and ergodic pre-change observations $\{(Y_i, X_i) : i = 1, \ldots, n\}$. Hence, one can consistently estimate the df of $S_1(\beta_0, \Sigma_0)$ by using the empirical df of $\{S_j(\beta, \Sigma) : j = 1, \ldots, n - m + 1\}$ evaluated at consistent estimators of β_0 and Σ_0 (see Theorem 1 below).

We evaluate $S_j(\beta, \Sigma)$ at the following consistent estimator of β_0. Let

$$\widehat{\beta}_{2(j)} = \text{LS estimator of } \beta \text{ using observations indexed by } i = 1, \ldots, n \text{ with } i \neq j, \ldots, j + \lceil m/2 \rceil - 1, \tag{2.19}$$

where $\lceil m/2 \rceil$ denotes the smallest integer that is greater than or equal to $m/2$, for $j = 1, \dots, n-m+1$. Thus, $\widehat{\beta}_{2(j)}$ is a LS estimator that leaves out the $\lceil m/2 \rceil$ observations that start at observation j. The choice of leaving out $\lceil m/2 \rceil$ observations is based on small sample considerations discussed in detail in Andrews (2002) and discussed briefly below.

When $m \geq d$, the statistics $\{S_j : j = 1, \dots, n-m+1\}$ are defined as

$$S_j = S_j(\widehat{\beta}_{2(j)}, \widehat{\Sigma}_{n+m}), \tag{2.20}$$

where $\widehat{\Sigma}_{n+m}$ and $S_j(\cdot, \cdot)$ are defined in (2.13) and (2.15), respectively.

When $m \leq d$, the statistics $\{S_j : j = 1, \dots, n-m+1\}$ are defined as

$$S_j = P_j(\widehat{\beta}_{2(j)}, \widehat{\Sigma}_{n+m}), \tag{2.21}$$

where $\widehat{\Sigma}_{n+m}$ and $P_j(\cdot, \cdot)$ are defined in (2.13) and (2.17), respectively.

For a test with asymptotic significance level α, the critical value for S is the $1-\alpha$ sample quantile, $\widehat{q}_{S,1-\alpha}$, of $\{S_j : j = 1, \dots, n-m+1\}$. That is,

$$\widehat{q}_{S,1-\alpha} = \inf\{x \in R : \widehat{F}_{S,n}(x) \geq 1-\alpha\}, \tag{2.22}$$

where $\widehat{F}_{S,n}(x)$ denotes the empirical df of $\{S_j : j = 1, \dots, n-m+1\}$. One rejects H_0 if $S > \widehat{q}_{S,1-\alpha}$. Equivalently, one rejects H_0 if S exceeds $100(1-\alpha)\%$ of the values $\{S_j : j = 1, \dots, n-m+1\}$—that is, if

$$(n-m+1)^{-1} \sum_{j=1}^{n-m+1} 1(S > S_j) \geq 1-\alpha. \tag{2.23}$$

The p-value for the S test is

$$pv_S = (n-m+1)^{-1} \sum_{j=1}^{n-m+1} 1(S \leq S_j). \tag{2.24}$$

(Note that, although $S_j(\beta, \Sigma)$, $A_j(\beta, \Sigma)$, and $V_j(\Sigma)$ are defined for $j = 1, \dots, n+1$, the sums in (2.23) and (2.24) are only over $j = 1, \dots, n-m+1$.)

These critical values and p-values allow for nonnormal, dependent, heteroskedastic errors. The main assumptions for their asymptotic validity are that $\{(Y_i, X_i) : i \geq 1\}$ 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 6 below.)

The motivation for leaving out $\lceil m/2 \rceil$ observations in the definition of $\widehat{\beta}_{2(j)}$ is as follows. Suppose the estimator $\widehat{\beta}_{n+m}$ is used in the definition of the subsample statistics S_j. Then, the resulting test has the correct asymptotic null rejection rate, but simulations show that it rejects the null hypothesis too often in samples of sizes 100 and 250. That is, the subsample statistics are not variable enough relative to the statistic S to yield a test with the desired null rejection probability.

One can make the subsample statistics more variable by using the estimator, $\widehat{\beta}_{(j)}$, that leaves out the m observations starting at the jth observation because then $\widehat{\beta}_{(j)}$ does not depend on the observations used in the quadratic form. However, simulations show that the resulting test does not reject the null hypothesis often enough in samples of sizes 100 and 250. A compromise between the two estimators $\widehat{\beta}_{n+m}$ and $\widehat{\beta}_{(j)}$ for use in the jth subsample statistic is the estimator $\widehat{\beta}_{2(j)}$. Simulations show that this choice yields very good null rejection probabilities for a wide variety of different error processes and sample sizes 100 and 250; see Section 7 below. See Andrews (2002) for additional simulation results for several different statistics.

Simulation results reported in Andrews (2002) indicate that the estimation of Σ_0 via $\widehat{\Sigma}_{n+m}$, as compared to using I_m in place of $\widehat{\Sigma}_{n+m}$ in both S and S_j, does not distort null rejection probabilities and costs little or nothing in terms of power when the errors are uncorrelated. But, it yields considerable gains in power when the errors exhibit significant serial correlation. In consequence, our recommended test S employs estimation of Σ_0.

Critical values and p-values for the test based on P are defined analogously to those for S. Specifically, they are based on the subsample statistics

$$P_j = P_j(\widehat{\beta}_{2(j)}, \widehat{\Sigma}_{n+m}) \tag{2.25}$$

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

2.5. *Issues of Power*

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. When $m \leq d$, the S test has no power asymptotically against changes in the regressor distribution because the test statistic depends only on the residuals for $i = n + 1, \ldots, n + m$. When $m > d$, the S test has 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 test when the parameters are constant and the error distribution is constant; see Section 7.2.3. Hence, the S test appears to have little to no power against changes just in the regressor distribution.

On the other hand, when $m \leq d$ the S test has power against changes in the error distribution that increase the $1-\alpha$ quantile of the distribution of the quadratic form $\mathbf{U}_{n+}'\Sigma_0^{-1}\mathbf{U}_{n+}$. Similarly, when $m > d$, the S test has power against changes in the error distribution that increase the $1-\alpha$ quantile of the distribution of $\Sigma_0^{-1/2}\mathbf{U}_{n+}$ after projection onto the column space spanned by the transformed post-change regressors $\Sigma_0^{-1/2}\mathbf{X}_{n+}$. For example, a sufficiently large increase in the variance of the errors causes the test to reject the null hypothesis.

The S test obviously has power against changes in the parameter vector β_0. Hence, rejection of the null hypothesis by S provides evidence that either the parameter vector has changed or the error distributions have became more variable (roughly speaking).

3. GENERALIZED METHOD OF MOMENTS

3.1. *Introduction*

In this section, we extend the S test for the linear regression model to moment condition models estimated by GMM. This extension covers tests of structural change for linear regression models estimated by IV. It also covers models estimated by ML by taking the GMM moment function $g(W_i, \beta)$ to be the ML score function for the ith observation conditional on the previous observations.

We consider GMM moment conditions given by

$$(3.1) \qquad Eg(W_i, \beta_0) = 0,$$

where $g(\cdot, \cdot)$ is a vector-valued function. As in the linear regression case, the observations are indexed by $i = 1, \ldots, n+m$ and the potential changepoint is at $i = n$.

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 on the dimension of X_i. Hence, in the GMM case, 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

$$(3.2) \qquad \begin{aligned} &H_0 : \begin{cases} Eg(W_i, \beta_0) = 0 \text{ for all } i = 1, \ldots, n+m \text{ and} \\ \{W_i : i \geq 1\} \text{ are stationary and ergodic,} \end{cases} \\ &H_1 : \begin{cases} Eg(W_i, \beta_0) = 0 \text{ for all } i = 1, \ldots, n \text{ and} \\ Eg(W_i, \beta_0) \neq 0 \text{ for some } i = n+1, \ldots, n+m \text{ and/or} \\ \text{the distribution of } \{g(W_{n+1}, \beta_0), \ldots, g(W_{n+m}, \beta_0)\} \text{ differs from} \\ \text{that of } \{g(W_i, \beta_0), \ldots, g(W_{i+m-1}, \beta_0)\} \text{ for } i = 1, \ldots, n-m+1. \end{cases} \end{aligned}$$

The alternative hypothesis H_1 covers parameter instability, i.e., $\beta_{1i} \neq \beta_0$ for some $i = n+1, \ldots, n+m$ and instability in the validity of the moment conditions, i.e., $Eg(W_i, \beta_0) \neq 0$ for some $i = n+1, \ldots, n+m$. Tests have power against instability in the validity of the moment conditions only if there are overidentifying restrictions, i.e., $d > d_\beta$. Hence, the alternative effectively encompasses parameter instability and instability in *overidentifying* restrictions; see Hall and Sen (1999).

The main assumptions are that $\{W_i : i \geq 1\}$ are stationary and ergodic under the null hypothesis and $Eg(W_i, \beta_0) = 0$ for all $i = 1, \ldots, n+m$ for some $\beta_0 \in R^{d_\beta}$ under the null hypothesis. See Section 6 for additional regularity conditions.

A special case of the GMM moment condition model is the linear IV model. In this case, the model is as in (2.1), but with regressors that may be correlated with the errors, and one has a d-vector Z_i of IV's for $i = 1, \ldots, n+m$. For the linear IV model, the function $g(W_i, \beta)$ is

$$g(W_i, \beta) = (Y_i - X_i'\beta)Z_i. \tag{3.3}$$

In this model, the null and alternative hypotheses of (3.2) are the same as in (2.3) but with the LS moments, $E(Y_i - X_i'\beta_0)X_i$, replaced by the IV moments, $E(Y_i - X_i'\beta_0)Z_i$.

3.2. *GMM Estimators*

We consider one-step, two-step, and continuously updated (CU) GMM estimators. The GMM estimator using the observations indexed by $i = 1, \ldots, n+m$, denoted $\widehat{\beta}_{n+m}$, is defined to minimize one of the following three criteria over the parameter space $\mathcal{B}$:

$$\begin{aligned}
Q_{n+m}^{(1)}(\beta) &= \left(\sum_{i=1}^{n+m} g(W_i, \beta)\right)' \mathcal{V}^{-1} \sum_{i=1}^{n+m} g(W_i, \beta), \\
Q_{n+m}^{(2)}(\beta) &= \left(\sum_{i=1}^{n+m} g(W_i, \beta)\right)' \mathcal{V}_{n+m}^{-1}(\widetilde{\beta}_{n+m}) \sum_{i=1}^{n+m} g(W_i, \beta), \quad \text{and} \\
Q_{n+m}^{(CU)}(\beta) &= \left(\sum_{i=1}^{n+m} g(W_i, \beta)\right)' \mathcal{V}_{n+m}^{-1}(\beta) \sum_{i=1}^{n+m} g(W_i, \beta),
\end{aligned} \tag{3.4}$$

where $Q_{n+m}^{(1)}(\beta)$, $Q_{n+m}^{(2)}(\beta)$, and $Q_{n+m}^{(CU)}(\beta)$ are the one-step, two-step, and CU GMM criterion functions, respectively; the one-step weight matrix $\mathcal{V}$ is some fixed nonstochastic matrix, such as I_d; the weight matrix $\mathcal{V}_{n+m}(\beta)$ depends on the observations; and the estimator $\widetilde{\beta}_{n+m}$ that appears in the two-step weight matrix is the one-step GMM estimator.

For $j = 1, \ldots, n-m+1$, the one-step, two-step, and CU GMM criterion functions, $Q^{(k)}_{2(j)}(\beta)$ for $k = 1, 2, CU$ and estimators $\widehat{\beta}_{2(j)}$ are defined using the same weight matrix but taking the sum $\sum_{i=1}^{n+m} g(W_i, \beta)$ only over the observations indexed by $i = 1, \ldots, n$ with $i \neq j, \ldots, j + \lceil m/2 \rceil - 1$.

In the special case of the linear IV model, all three full-sample GMM estimators reduce to the linear IV estimator:

$$
\begin{aligned}
(3.5)\qquad & \widehat{\beta}_{n+m} = (\mathbf{X}'_{1,n+m}\mathbf{P}_{\mathbf{Z}_{1,n+m}}\mathbf{X}_{1,n+m})^{-1}\mathbf{X}'_{1,n+m}\mathbf{P}_{\mathbf{Z}_{1,n+m}}\mathbf{Y}_{1,n+m}, \quad \text{where} \\
& \mathbf{Z}_{1,n+m} = (Z_1, \ldots, Z_{n+m})' \quad \text{and} \\
& \mathbf{P}_{\mathbf{Z}_{1,n+m}} = \mathbf{Z}_{1,n+m}(\mathbf{Z}'_{1,n+m}\mathbf{Z}_{1,n+m})^{-1}\mathbf{Z}'_{1,n+m}.
\end{aligned}
$$

The estimator $\widehat{\beta}_{2(j)}$ in the linear IV model is defined as in (3.5), but with the rows of $\mathbf{Y}_{1,n+m}$, $\mathbf{X}_{1,n+m}$, and $\mathbf{Z}_{1,n+m}$ indexed by $i = j, \ldots, j + \lceil m/2 \rceil - 1$ eliminated.

3.3. *GMM: Case 1*

First, we consider the case where the moment conditions are of the form:

$$
(3.6)\qquad g(W_i, \beta) = U(W_i, \beta)Z(W_i, \beta),
$$

where $U(W_i, \beta) \in R$, $U_i = U(W_i, \beta_0)$ is an error term, and $Z(W_i, \beta)$ is a d-vector of instruments.

In this case, the (S, S_j) statistics are given by

$$
(3.7)\qquad
\begin{aligned}
& S = \begin{cases} S_{n+1}(\widehat{\beta}_{n+m}, \widehat{\Sigma}_{n+m}) & \text{when } m \geq d, \\ P_{n+1}(\widehat{\beta}_{n+m}, \widehat{\Sigma}_{n+m}) & \text{when } m \leq d, \end{cases} \quad \text{and} \\
& S_j = \begin{cases} S_j(\widehat{\beta}_{2(j)}, \widehat{\Sigma}_{n+m}) & \text{when } m \geq d, \\ P_j(\widehat{\beta}_{2(j)}, \widehat{\Sigma}_{n+m}) & \text{when } m \leq d, \end{cases}
\end{aligned}
$$

where $S_j(\beta, \Sigma)$ and $P_j(\beta, \Sigma)$ are defined by

$$
(3.8)\qquad
\begin{aligned}
& S_j(\beta, \Sigma) = A_j(\beta, \Sigma)' V_j^{-1}(\Sigma) A_j(\beta, \Sigma), \\
& P_j(\beta, \Sigma) = \mathbf{U}_{j,j+m-1}(\beta)' \Sigma^{-1} \mathbf{U}_{j,j+m-1}(\beta),
\end{aligned}
$$

$$
(3.9)\qquad \mathbf{U}_{j,j+m-1}(\beta) = (U(W_j, \beta), \ldots, U(W_{j+m-1}, \beta))',
$$

$$
(3.10)\qquad
\begin{aligned}
& A_j(\beta, \Sigma) = \mathbf{Z}_{j,j+m-1}(\beta)' \Sigma^{-1} \mathbf{U}_{j,j+m-1}(\beta), \\
& \mathbf{Z}_{j,j+m-1}(\beta) = (Z(W_j, \beta), \ldots, Z(W_{j+m-1}, \beta))',
\end{aligned}
$$

$$V_j(\Sigma) = V_j(\widehat{\beta}_{n+m}, \Sigma),$$
(3.11)
$$V_j(\beta, \Sigma) = \mathbf{Z}_{j,j+m-1}(\beta)'\Sigma^{-1}\mathbf{Z}_{j,j+m-1}(\beta),$$

and

$$\widehat{\Sigma}_{n+m} = \widehat{\Sigma}_{n+m}(\widehat{\beta}_{n+m}),$$
(3.12)
$$\widehat{\Sigma}_{n+m}(\beta) = (n+1)^{-1}\sum_{j=1}^{n+1}\mathbf{U}_{j,j+m-1}(\beta)\mathbf{U}_{j,j+m-1}(\beta)'.$$

Note that the (S, S_j) statistics have the same definition as in the LS case, but with $\mathbf{U}_{j,j+m-1}(\beta)$ in place of $\mathbf{Y}_{j,j+m-1} - \mathbf{X}_{j,j+m-1}\beta$ and $\mathbf{Z}_{j,j+m-1}(\beta)$ in place of $\mathbf{X}_{j,j+m-1}$.

Critical values and p-values for the S test in the GMM context are constructed in the same way as in Section 2.4 for the linear regression model.

As stated above, the S test has power against parameter instability and invalidity of overidentifying restrictions. In addition, when the moment conditions are of the form (3.6), the S test has power against changes in the error distribution that increase the variability of the errors, as in the linear regression case.

The GMM version of the test based on P is defined as in (2.18) and (2.25), but with $P_j(\beta, \Sigma)$, $\widehat{\beta}_{n+m}$, and $\widehat{\Sigma}_{n+m}$ as defined above.

3.4. *GMM: Case 2*

When the moment conditions are not of the form (3.6), S cannot be defined as in (3.6) because the matrices Σ_0 and $\widehat{\Sigma}_{n+m}$ are not defined. For example, for models estimated by ML, the score function is not of the form (3.6) (except for normal linear or nonlinear regression models). In this case, we define the statistics S and S_j as follows:

(3.13) $$S = S_{n+1}(\widehat{\beta}_{n+m}) \quad \text{and} \quad S_j = S_j(\widehat{\beta}_{2(j)}), \quad \text{where}$$
$$S_j(\beta) = A_j(\beta)'V_j^{-1}(\beta)A_j(\beta),$$
$$A_j(\beta) = \sum_{i=j}^{j+m-1} g(W_i, \beta),$$

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

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

When $m < d$, one can take

$$(3.15) \qquad V_j(\beta) = \frac{1}{n+m} \sum_{i=1}^{n+m} g(W_i, \beta) g(W_i, \beta)'$$

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

Note that the above definition of S is essentially the same as that given in the previous subsection but with $\widehat{\Sigma}_{n+m} = I_m$ and with a slightly different weight matrix.

There is no GMM version of P in case 2 because there is no scalar error $U(W_i, \beta)$ upon which to base it.

4. TESTS FOR INSTABILITY AT THE BEGINNING OR IN THE MIDDLE OF THE SAMPLE

The tests introduced above can be altered to detect instability occurring at the beginning or in the middle of the sample. We consider the case of testing for structural instability for the m observations indexed by $i = i_0, \dots, i_0 + m - 1$ when the total number of observations is $n + m$.

For the GMM case, the null and alternative hypotheses are given by

$$(4.1) \qquad H_0 : \begin{cases} Eg(W_i, \beta_0) = 0 \text{ for all } i = 1, \dots, n+m \text{ and} \\ \{W_i : i \geq 1\} \text{ are stationary and ergodic,} \end{cases}$$

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

One can construct tests for these hypotheses by first computing $\widehat{\beta}_{n+m}$ and $\widehat{\Sigma}_{n+m}$ using the formulae given above. Then, one moves the summands $\{g(W_i, \beta) : i = i_0, \dots, i_0 + m - 1\}$ to the end of the sample and shifts the observations following $\{g(W_i, \beta) : i = i_0, \dots, i_0 + m - 1\}$ up to eliminate the gap. In this way, the original observations $\{g(W_i, \beta) : i = i_0, \dots, i_0 + m - 1\}$ become indexed by $i = n + 1, \dots, n + m$ and the original observations $\{g(W_i, \beta) : i = i_0 + m, \dots, n + m\}$ become indexed by $i = i_0, \dots, n$. After making this shift, one computes the statistics S and S_j using the formulae given above. (The weight matrix used to compute the GMM estimator $\widehat{\beta}_{2(j)}$ is the same as that used to compute $\widehat{\beta}_{n+m}$ above.) One constructs critical values and/or p-values as above, but only using the subsample statistics $\{S_j : j = 1, \dots, n + m - 1$ with $j \neq \max(i_0 - m + 1, 1), \dots, \min(i_0 - 1, n - m + 1)\}$. Thus, in (2.23) and

(2.24), the sum is over fewer than $n+m-1$ terms (unless $i_0=1$ or $n+1$) and the multiplicand $(n+m-1)^{-1}$ is replaced by the reciprocal of the number of subsample values being employed.

The reason that some subsample statistics are dropped is that there is a join point between the observations indexed by $i=i_0-1$ and $i=i_0$ in the shifted sample. These adjacent observations really occur m time periods apart. After dropping the subsample statistics indicated above, none of the subsample statistics are based on observations indexed by both $i=i_0-1$ and $i=i_0$.

At most $m-1$ subsample statistics are dropped due to the join point issue. If $i_0<m$ or $i_0>n-m+2$, then fewer than $m-1$ are dropped. If $i_0=1$ or $n+1$, none are dropped.

For the linear regression testing problem, the same procedure as above is followed taking $g(W_i,\beta)=(Y_i-X_i'\beta)X_i$.

5. 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:

$$(5.1)\qquad \begin{aligned} &Y_i=\mu+\alpha Y_{i-1}+\gamma_1\Delta Y_{i-1}+\cdots+\gamma_p\Delta Y_{i-p}+U_i \quad \text{and}\\ &Y_i=\mu+\beta i+\alpha Y_{i-1}+\gamma_1\Delta Y_{i-1}+\cdots+\gamma_p\Delta Y_{i-p}+U_i. \end{aligned}$$

The second model contains a time trend. If $\alpha=1$, the models have unit roots and are nonstationary. However, if the characteristic polynomial associated with the parameters $(\alpha,\gamma_1,\dots,\gamma_p)$ has at most one unit root and all other roots lie outside the unit circle, then differenced versions of these models are strictly stationary for $|\alpha|\le 1$ under suitable conditions on the errors U_i:

$$(5.2)\qquad \begin{aligned} \Delta Y_i &= (\alpha+\gamma_1)\Delta Y_{i-1}+(\gamma_2-\gamma_1)\Delta Y_{i-2}+\cdots\\ &\quad +(\gamma_p-\gamma_{p-1})\Delta Y_{i-p-2}-\gamma_p\Delta Y_{i-p-1}+\Delta U_i \quad \text{and}\\ \Delta Y_i &= \beta+(\alpha+\gamma_1)\Delta Y_{i-1}+(\gamma_2-\gamma_1)\Delta Y_{i-2}+\cdots\\ &\quad +(\gamma_p-\gamma_{p-1})\Delta Y_{i-p-2}-\gamma_p\Delta Y_{i-p-1}+\Delta U_i. \end{aligned}$$

In consequence, in the unit root case, one can test for structural instability at the end of the sample by applying the tests above to the models written in differenced form (5.2).

6. ASYMPTOTIC RESULTS

In this section, we show that the S test introduced above is asymptotically valid under suitable conditions.

6.1. *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 $i = 1, \ldots, n+m$. 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 $i = 1, \ldots, n$ and $W_i = W_{n,i}$ for $i = n+1, \ldots, n+m$, where $\{W_{n,i} : i = n+1, \ldots, n+m\}$ are some random variables whose joint distribution is different from that of $\{W_{0,i} : i = n+1, \ldots, n+m\}$. We assume that the distribution of $\{W_{n,i} : i = n+1, \ldots, n+m\}$ is independent of n. That is, we consider fixed, not local, alternatives.

For simplicity and generality, we state one set of high-level conditions that applies to linear regression and moment condition models. Then, we provide primitive sufficient conditions for the linear regression model estimated by LS, the linear regression model estimated by IV, and the moment condition model estimated by GMM.

Let $B(\beta_0, \varepsilon)$ denote a ball centered at β_0 with radius $\varepsilon > 0$. Let $\partial/\partial(\beta, \Sigma^{-1})$ denote partial differentiation with respect to β and the nonredundant elements of Σ^{-1}.

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

ASSUMPTION 2: (a) $\|\widehat{\beta}_{n+m} - \beta_0\| \to_p 0$ *and* $\sup_{j=1,\ldots,n-m+1} \|\widehat{\beta}_{2(j)} - \beta_0\| \to_p 0$ $n \to \infty$ *with* m *fixed under* H_0 *and* H_1.

(b) $\sup_{\beta \in B(\beta_0, \varepsilon_n)} \|\widehat{\Sigma}_{n+m}(\beta) - \Sigma_0\| \to_p 0$ *as* $n \to \infty$ *for some nonsingular matrix* Σ_0, *for all sequences of constants* $\{\varepsilon_n : n \geq 1\}$ *for which* $\varepsilon_n \to 0$ *as* $n \to \infty$.

ASSUMPTION 3: (a) $S_{n+1}(\beta, \Sigma)$ *is continuously differentiable in a neighborhood of* (β_0, Σ_0) *with probability one under* H_0 *and* H_1, *where* Σ_0 *is as in Assumption* 2(b), *when* $m > d$ *and likewise for* $P_{n+1}(\beta, \Sigma)$ *when* $m \leq d$.

(b) *Either*

$$E \sup_{\beta \in B(\beta_0, \varepsilon), \Sigma \in N(\Sigma_0)} \|(\partial/\partial(\beta, \Sigma^{-1}))S_1(\beta, \Sigma)\| < \infty$$

or

$$(n-m+1)^{-1} \sum_{j=1}^{n-m+1} \sup_{\beta \in B(\beta_0, \varepsilon), \Sigma \in N(\Sigma_0)} \|(\partial/\partial(\beta, \Sigma^{-1}))S_j(\beta, \Sigma)\| = O_p(1)$$

for some $\varepsilon > 0$, *where* Σ_0 *is as in Assumption* 2(b) *and* $N(\Sigma_0)$ *denotes some neighborhood of* Σ_0, *when* $m > d$ *and likewise for* $P_j(\beta, \Sigma)$ *when* $m \leq d$.

(c) *The distribution function of* $S_1(\beta_0, \Sigma_0)$ *is continuous and increasing at its* $1-\alpha$ *quantile*, *where* Σ_0 *is as in Assumption* 2(b), *when* $m > d$ *and likewise for* $P_1(\beta_0, \Sigma_0)$ *when* $m \leq d$.

Assumption 1 is fairly general compared to many assumptions 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, linear IV, and GMM estimators under appropriate regularity conditions. The following are sufficient:

ASSUMPTION LS:
(a) $EU_1X_1 = 0$.
(b) $EU_1^2 < \infty$ *and* $E\|X_1\|^{2+\delta} < \infty$ *for some* $\delta > 0$.
(c) EX_1X_1' *and* $\Sigma_0 = E\mathbf{U}_{1,m}\mathbf{U}_{1,m}'$ *are positive definite*.
(d) *The df of* $S_1(\beta_0, \Sigma_0)$ *is continuous and increasing at its* $1-\alpha$ *quantile when* $m > d$ *and likewise for* $P_1(\beta_0, \Sigma_0)$ *when* $m \leq d$.

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' *and* $\Sigma_0 = E\mathbf{U}_{1,m}\mathbf{U}_{1,m}'$ *are positive definite and* EX_1Z_1' *has full row rank*.
(d) *The df of* $S_1(\beta_0, \Sigma_0)$ *is continuous and increasing at its* $1-\alpha$ *quantile when* $m > d$ *and likewise for* $P_1(\beta_0, \Sigma_0)$ *when* $m \leq d$.

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* $\mathcal{V}$ *is nonstochastic and positive definite*; *the two-step GMM weight matrix function* $\mathcal{V}_{n+m}(\beta)$ *satisfies* $\sup_{\beta\in B(\beta_0,\varepsilon)} |\mathcal{V}_{n+m}(\beta) - \mathcal{V}(\beta)| \to_p 0$ *for some* $\varepsilon > 0$, *for some symmetric positive definite nonstochastic function* $\mathcal{V}(\beta)$ *defined on* $B(\beta_0, \varepsilon)$ *that is continuous at* β_0; *and the CU weight matrix function* $\mathcal{V}_{n+m}(\beta)$ *satisfies analogous convergence conditions on* $\mathcal{B}$.
(f) *When* S *is defined as in case* 2, (i) $g(W_1, \beta)$ *is continuously differentiable on a neighborhood of* β_0 *almost surely*, (ii) $E\sup_{\beta\in B(\beta_0,\varepsilon)} \|g(W_1, \beta)\|^2 < \infty$, (iii) $E\sup_{\beta\in B(\beta_0,\varepsilon)}(\|(\partial/\partial\beta')g(W_1, \beta)\| \cdot \|g(W_1, \beta)\|) < \infty$, (iv) $V_j(\beta) = V(W_j, \ldots, W_{j+m-1}, \beta)$ *is a positive definite weight matrix that is a function of the observations* $W_j, \ldots, W_{j+m-1}$ *and the parameter* β *and is continuously differentiable in* β *on a neighborhood of* β_0 *almost surely for* $j = 1, \ldots,$

$n+1$, (v) $\sup_{j=1,\ldots,n-m+1} \sup_{\beta \in B(\beta_0,\varepsilon)} (\|V_j^{-1}(\beta)\| + \|(\partial/\partial\beta_r)V_j^{-1}(\beta)\|) = O_p(1)$ *for* $r = 1, \ldots, d_\beta$ *for some* $\varepsilon > 0$. *When* S *is defined as in case* 1, $E\mathbf{U}_{1,m}\mathbf{U}'_{1,m}$ *is positive definite and* $E\sup_{\beta \in B(\beta_0,\varepsilon)} U^2(W_1, \beta) < \infty$. *When* S *is defined as in case* 1 *and* $m > d$, $U(W_1, \beta)$ *and* $Z(W_1, \beta)$ *are continuously differentiable on a neighborhood of* β_0 *almost surely*, $E\sup_{\beta \in B(\beta_0,\varepsilon)} \|Z(W_1, \beta)\|^2 < \infty$, $E\sup_{\beta \in B(\beta_0,\varepsilon)} \|(\partial/\partial\beta')U(W_1, \beta)\|^2 < \infty$, $E\sup_{\beta \in B(\beta_0,\varepsilon)} \|(\partial/\partial\beta')Z(W_1, \beta)\|^2 < \infty$, *conditions* (iv) *and* (v) *hold with* $V_j(\beta)$ *replaced by* $V_j(\beta, \Sigma)$, *the suprema also taken over* Σ *in some neighborhood* $N(\Sigma_0)$ *of* Σ_0, *and* $\partial/\partial\beta_r$ *replaced by* $\partial/\partial(\beta, \Sigma^{-1})_r$, *where the latter denotes partial differentiation with respect to the rth element of the vector comprised of* β *and the nonredundant elements of* Σ^{-1}. *When* S *is defined as in case* 1 *and* $m \le d$, $U(W_1, \beta)$ *is continuously differentiable on a neighborhood of* β_0 *almost surely and* $E\sup_{\beta \in B(\beta_0,\varepsilon)} \|U(W_1, \beta)(\partial/\partial\beta)U(W_1, \beta)\| < \infty$.

(g) *The df of* $S_1(\beta_0, \Sigma_0)$ *is continuous and increasing at its* $1-\alpha$ *quantile when* $m > d$ *and likewise for* $P_1(\beta_0, \Sigma_0)$ *when* $m \le d$.

Assumptions LS, IV, and GMM(a)–(d) only place restrictions on the distribution of the first observation. By stationarity, this has implications for the distributions of all of the observations under H_0 and for the "pre-change" observations under H_1. Assumptions LS, IV, and GMM(a)–(d) place no restrictions on the distributions of the "post-change" observations even though Assumptions 2 and 3 are required to hold under H_0 and H_1. Nevertheless, Assumptions LS, IV, and GMM are each sufficient for Assumptions 2 and 3. This is possible because in Assumption 2 the post-change observations only affect the estimators $\widehat{\beta}_{n+m}$ and $\widehat{\Sigma}_{n+m}$ and their behavior is dominated by the pre-change observations and in Assumption 3 the post-change observations only affect $S_{n+1}(\beta_0, \Sigma_0)$ or $P_{n+1}(\beta_0, \Sigma_0)$ and whether they are continuously differentiable does not depend on the distribution of the observations.

A simple sufficient condition for Assumptions LS(d) and IV(d) is that U_1 has an absolutely continuous distribution. A simple sufficient condition for Assumption GMM(g) is that $U(W_1, \beta_0)$ has an absolutely continuous distribution when S is defined as in case 1 and that $g(W_1, \beta_0)$ has an absolutely continuous distribution when S is defined as in case 2.

Assumptions GMM(a)–(e) are used to verify Assumption 2(a). Assumption GMM(f) is used to verify Assumptions 2(b), 3(a), and 3(b). Assumption GMM(g) is equivalent to Assumption 3(c).

LEMMA 1: (a) *Assumptions* 1 *and LS imply that Assumptions* 2 *and* 3 *hold for the linear regression model estimated using the LS estimator*.

(b) *Assumptions* 1 *and IV imply that Assumptions* 2 *and* 3 *hold for the IV regression model estimated using the IV estimator*.

(c) *Assumptions* 1 *and GMM imply Assumptions* 2 *and* 3 *hold for the moment condition model estimated using a GMM estimator with* S *defined as in case* 1 *or* 2.

6.2. *Results*

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

Let $\widehat{F}_{S,n}(x)$ denote the empirical df based on $\{S_j : j = 1, \ldots, n-m+1\}$. That is,

$$\widehat{F}_{S,n}(x) = \frac{1}{n-m+1} \sum_{j=1}^{n-m+1} 1(S_j \leq x). \tag{6.1}$$

Let $F_S(x)$ denote the df at x of $S_1(\beta_0, \Sigma_0)$ when $m > d$ and of $P_1(\beta_0, \Sigma_0)$ when $m \leq d$. (Note that the distributions of $S_1(\beta_0, \Sigma_0)$ and $P_1(\beta_0, \Sigma_0)$ are the same as those of $S_j(\beta_0, \Sigma_0)$ and $P_j(\beta_0, \Sigma_0)$ for all $j = 1, \ldots, n$ under H_0 and H_1 by stationarity.) Let $q_{S,1-\alpha}$ denote the $1-\alpha$ quantile of $S_1(\beta_0, \Sigma_0)$ when $m > d$ and of $P_1(\beta_0, \Sigma_0)$ when $m \leq d$. Let $\widehat{q}_{S,1-\alpha}$ denote the $1-\alpha$ sample quantile of $\{S_j : j = 1, \ldots, n-m+1\}$, as defined in (2.22).

Let S_∞ be a random variable with the same distribution as $S_{n+1}(\beta_0, \Sigma_0)$ when $m > d$ and as $P_{n+1}(\beta_0, \Sigma_0)$ when $m \leq d$. Under Assumptions 1–3 and H_0, the distribution of $S_{n+1}(\beta_0, \Sigma_0)$ equals that of $S_1(\beta_0, \Sigma_0)$ and the distribution of $P_{n+1}(\beta_0, \Sigma_0)$ equals that of $P_1(\beta_0, \Sigma_0)$. Also, the distributions of $S_{n+1}(\beta_0, \Sigma_0)$ and $P_{n+1}(\beta_0, \Sigma_0)$ do not depend on n under either H_0 or H_1. Under H_0, this holds by stationarity. Under H_1, this holds because we take the distribution of $\{W_{n,i} : i = n+1, \ldots, n+m\}$ to be independent of n, which is appropriate for fixed alternatives.

THEOREM 1: *Suppose Assumptions* 1–3 *hold. Then, as* $n \to \infty$,
(a) $S \to_d S_\infty$ *under* H_0 *and* H_1,
(b) $\widehat{F}_{S,n}(x) \to_p F_S(x)$ *for all* x *in a neighborhood of* $q_{S,1-\alpha}$ *under* H_0 *and* H_1,
(c) $\widehat{q}_{S,1-\alpha} \to_p q_{S,1-\alpha}$ *under* H_0 *and* H_1, *and*
(d) $P(S > \widehat{q}_{S,1-\alpha}) \to \alpha$ *under* H_0.

COMMENTS: 1. Part (a) gives the asymptotic distribution of S under the null hypothesis and fixed alternatives.

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

3. Part (a) shows that S does not diverge to infinity as $n \to \infty$ under H_1. Hence, S is not a consistent test. However, if $S_{n+1}(\beta_0, \Sigma_0)$ is stochastically greater than $S_1(\beta_0, \Sigma_0)$ under H_1 when $m > d$ and $P_{n+1}(\beta_0, \Sigma_0)$ is stochastically greater than $P_1(\beta_0, \Sigma_0)$ under H_1 when $m \leq d$, then S is an asymptotically unbiased test.

4. 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 and Σ_0, i.e., Assumption 2 holds, the S test for $m \leq d$ has 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.

5. The idea of the proof of part (b) of the Theorem is to show that (i) the difference between $\widehat{F}_{S,n}(x)$ and a smoothed version of it, say $\widehat{F}_{S,n}(x, h_n)$, converges in probability to zero, where h_n indexes the amount of smoothing and $h_n \to 0$ as $n \to \infty$, (ii) the difference between $\widehat{F}_{S,n}(x, h_n)$ and an analogous df with $\widehat{\beta}_{2(j)}$ replaced by β_0 converges in probability to zero, (iii) the difference between the latter and the empirical df of $\{S_j(\beta_0, \Sigma_0) : j = 1, \ldots, n - m + 1)\}$ converges in probability to zero as $n \to \infty$ (when $m \geq d$), and (iv) the difference between the latter and its expectation, $F_S(x)$, is asymptotically negligible. The reason for considering a smoothed version of $\widehat{F}_{S,n}(x)$ is that it is a smooth function of $\widehat{\beta}_{2(j)}$ and, hence, result (ii) can be established by taking a mean-value expansion about β_0. Result (iv) holds by the ergodic theorem (which states that the sample average of mean zero stationary and ergodic random variables converges in probability to zero as $n \to \infty$; e.g., see Hannan (1970, Ch. IV, Sec. 2)) because $\{S_j(\beta_0, \Sigma_0) : j = 1, \ldots, n - m + 1)\}$ is a finite subset of stationary and ergodic random variables using Assumption 1.

7. MONTE CARLO EXPERIMENT

In this section, we describe some Monte Carlo results that are designed to assess the null rejection probabilities (NRP's) and the power properties of the S and P tests and to compare them to the F test.

7.1. *Experimental Design*

We consider linear regression models estimated by LS, as in (2.1). Two pre-change sample sizes, n, are considered: 100 and 250. Three post-change sample sizes, m, are considered: 10, 5, and 1. 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 four different distributions for the innovations: standard normal, chi-square with two degrees of freedom (recentered and rescaled to have mean zero and variance one), t_3 (rescaled to have variance one), and uniform on $[-\sqrt{12}/2, \sqrt{12}/2]$ (which has mean zero and variance one). (Note that the NRP results, but not the power 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. The χ^2, t_3, and uniform distributions are chosen because they exhibit asymmetry, thick tails, and thin tails, respectively.

Under the null hypothesis, the sample of $n+m$ observations is computed using the regression parameter vectors $\beta_0 = \beta_{1i} = 0$ for $i = n+1, \dots, n+m$. (The NRP results are invariant with respect to the value of β_0 $(= \beta_{1i})$.) Under the alternative hypothesis, $\beta_0 = 0$ and $\beta_{1i} = \beta_1 \propto (1,1,1,1,1)'$ for $i = n+1, \dots, n+m$, where $\propto$ denotes "is proportional to." For most results, we take $\|\beta_1\| = 1.75$, where $\|\beta_1\|$ denotes the Euclidean norm. For some results, we take $\|\beta_1\| = 7.0$.

Results are reported for tests with nominal size .05.

The power results that we report are for NRP-corrected tests because we do not want to confound power differences with NRP distortions (which are quite large for the F test in some scenarios). For the F test, NRP correction is straightforward. By simulation we determine critical values that yield the desired NRP, .05, for each distribution and each n, m, and ρ value when $\beta_0 = \beta_{1i} = 0$ (which are the pseudo-true values of the parameters under the alternative hypothesis). These critical values are employed when computing the power of the F test.

For the S and P tests, NRP correction is not as straightforward because their critical values are sample quantiles, not constants. For each of these tests, we determine by simulation the nominal significance levels that yield the finite sample NRP to be as close to the desired NRP, .05, as possible for each distribution and each n, m, and ρ value when $\beta_0 = \beta_{1i} = 0$. (The NRP's cannot be made exactly equal to .05 because the sample quantile functions are not continuous. But, the differences are fairly small.) These significance levels are employed when computing the power of the tests. Note that this method of NRP correction is equivalent to the method in which the critical value is adjusted for any test that has a nonrandom critical value.

The number of simulation repetitions used is 40,000 for each case considered. This yields simulation standard errors of (approximately) .001 for the simulated NRP's of nominal .05 tests and simulated standard errors in the interval (.0020, .0025) for the simulated power when power is in the interval (.20, .80).

7.2. *Monte Carlo Results*

7.2.1. *Null Rejection Probabilities*

Table I presents the NRP's for nominal .05 tests. When $m = 5$ or 1, separate results are not given for P because $P = S$ by definition.

TABLE I

TRUE NULL REJECTION PROBABILITIES OF TESTS WITH NOMINAL SIGNIFICANCE LEVEL .05 USING NORMAL, χ^2_2, t_3, AND UNIFORM REGRESSORS AND ERRORS

			Normal n		χ^2_2 n		t^3 n		Uniform n	
m	ρ	Test	100	250	100	250	100	250	100	250
10	0	S	.046	.052	.056	.055	.058	.055	.043	.048
		P	.063	.060	.072	.060	.075	.065	.048	.052
		F	.051	.050	.088	.089	.090	.085	.028	.024
10	.4	S	.047	.050	.056	.058	.058	.057	.042	.049
		P	.062	.056	.072	.065	.078	.065	.048	.050
		F	.123	.115	.131	.125	.140	.129	.105	.101
10	.8	S	.053	.055	.061	.058	.064	.060	.049	.054
		P	.062	.058	.076	.063	.080	.066	.050	.055
		F	.329	.286	.318	.270	.314	.255	.334	.288
5	0	$S\ (=P)$	.047	.052	.049	.056	.050	.055	.040	.045
		F	.050	.049	.103	.102	.099	.089	.007	.004
5	.4	$S\ (=P)$	.050	.053	.050	.054	.052	.056	.041	.048
		F	.071	.068	.099	.094	.102	.093	.040	.036
5	.8	$S\ (=P)$	.056	.055	.060	.057	.061	.058	.049	.055
		F	.146	.125	.142	.118	.147	.125	.141	.123
1	0	$S\ (=P)$	.048	.048	.053	.053	.053	.049	.034	.039
		F	.050	.052	.055	.057	.054	.049	.008	.002
1	.4	$S\ (=P)$	.051	.049	.053	.050	.053	.050	.046	.046
		F	.054	.052	.052	.050	.055	.052	.030	.025
1	.8	$S\ (=P)$	.072	.058	.068	.056	.069	.058	.073	.059
		F	.074	.059	.068	.054	.075	.061	.072	.059

The main results are as follows:

1. The F test has (exactly) correct NRP for all values of (n, m) when the distribution is normal and $\rho = 0$. The F test also has fairly good NRP when $m = 1$ and the distribution is normal, χ^2, or t_3.

2. In most other cases, the NRP of the F test is poor and, in some cases, it is very poor. Across all of the cases considered, the NRP of the F test varies between .002 and .329. The standard deviation of the NRP of the F test from the desired value .05 across the 72 cases in Table I is .095, which is very high. When $\rho = 0$, the F test over-rejects when the distribution is χ^2 and t_3 and under-rejects when the distribution is uniform. For example, for $\rho = 0$, $m = 5$, $n = 250$, and χ^2 distribution, the NRP is .102; while for $\rho = 0$, $m = 5$, $n = 250$, and uniform distribution, the NRP is .004. When $\rho = .4$ or .8, the F test over-rejects for all distributions, including the normal, except for one case with the t_3 distribution. For example, for $\rho = .4$, $m = 10$, $n = 250$, and normal distribution, the NRP is .115. For the same case except with $\rho = .8$, the NRP is .286. The reason for the poor performance of the F test when either $\rho \neq 0$ or the

distribution is not normal is that the F test does not have correct NRP asymptotically in these cases.

3. The S test performs very much better than the F test in terms of NRP. In addition, the NRP performance of S is quite good in an absolute sense. Except for the case of $m = 1$ and uniform distribution, the NRP's of the S test vary between .040 and .058 when $\rho = 0$ or .4 and between .049 and .073 when $\rho = .8$. Across all cases, the standard deviation of the NRP's for the S test from the desired value .05 is only .008. The S test has better NRP when $m = 10$ or 5 than when $m = 1$. When $m = 10$ or 5, its NRP varies between .040 and .064.

4. The P test also performs very much better than the F test in terms of NRP's. For the case in which the P and S tests differ, i.e., when $m = 10$, the P test over-rejects the null hypothesis somewhat. In consequence, the S test outperforms the P test in terms of its NRP performance.

5. In general, for all tests, the rejection rates tend to be somewhat lower for the uniform and normal distributions and higher for the χ^2 and t_3 distributions, although the differences are not great except for the F test. For all tests, the rejection rates are higher for $n = 100$ than for $n = 250$. This is because the estimator of β is a constant in the asymptotic approximations and this is closer to being true when $n = 250$ than when $n = 100$. For all tests, the rejection rates increase as ρ increases, but the extent of the increase varies dramatically across different tests. For the F test, the increase is very large. For the S and P tests, however, the increase is slight. For all tests, the rejection rates do not vary much with m when $\rho = 0$. When $\rho = .4$ or .8, the rejection rates of the F test increase in m. For the S and P tests, the rejection rates do not vary much with m even when $\rho = .4$ or .8, although the rates tend to be highest for $m = 1$ and $\rho = .8$.

To conclude, the NRP results of Table I show that the F test performs poorly in many of the cases considered. The S and P tests greatly outperform the F test. But, the P test over-rejects the null hypothesis somewhat when $m = 10$. The S test performs best and its performance in an absolute sense is quite good.

7.2.2. *Power*

Next, we consider Table II which provides the NRP-corrected power results. The principle findings are as follows:

1. The S test is more powerful than the P test by 35.5% on average when $m = 10$ (which is the only case in which the two tests differ). Hence, there is a substantial gain in power by using a weight matrix that projects onto the column space of the transformed post-change regressors rather than using an identity weight matrix.

2. The F test is 4.1% more powerful than the S test on average when $\rho = 0$. The F test is 1.8% more powerful on average when $\rho = .4$. The F and S tests have essentially the same power when $\rho = .8$ and $m = 1$. The S test is 51.3%

TABLE II
POWER OF SIGNIFICANCE LEVEL .05 NRP-CORRECTED TESTS USING NORMAL, χ_2^2, t_3, AND UNIFORM REGRESSORS AND ERRORS

				Normal		χ_2^2		t^3		Uniform	
				n		n		n		n	
$\Vert\beta_1\Vert$	m	ρ	Test	100	250	100	250	100	250	100	250
1.75	10	0	S	.90	.94	.67	.78	.67	.78	.93	.96
			P	.80	.88	.39	.57	.39	.41	.92	.95
			F	.94	.95	.79	.82	.81	.83	.96	.97
1.75	10	.4	S	.85	.91	.62	.72	.61	.70	.89	.94
			P	.74	.85	.36	.45	.36	.37	.88	.93
			F	.83	.86	.67	.71	.69	.71	.87	.89
1.75	10	.8	S	.76	.87	.56	.71	.53	.64	.90	.80
			P	.68	.81	.35	.48	.33	.41	.90	.79
			F	.42	.47	.36	.44	.40	.40	.49	.44
1.75	5	0	$S\ (=P)$	.66	.72	.33	.42	.34	.43	.79	.83
			F	.70	.73	.41	.44	.44	.46	.81	.83
1.75	5	.4	$S\ (=P)$	.61	.67	.30	.39	.32	.39	.74	.79
			F	.60	.63	.36	.41	.40	.43	.68	.71
1.75	5	.8	$S\ (=P)$	.54	.63	.31	.38	.30	.37	.65	.73
			F	.36	.38	.34	.35	.30	.32	.39	.40
1.75	1	0	$S\ (=P)$	.31	.32	.25	.25	.29	.29	.41	.42
			F	.32	.32	.24	.25	.28	.29	.40	.41
1.75	1	.4	$S\ (=P)$	.31	.32	.24	.27	.29	.29	.36	.36
			F	.32	.32	.26	.27	.29	.29	.36	.37
1.75	1	.8	$S\ (=P)$	.30	.30	.31	.29	.27	.27	.31	.31
			F	.30	.30	.29	.29	.28	.28	.31	.31
7	1	0	$S\ (=P)$	.77	.78	.72	.72	.76	.77	.81	.82
7	1	.4	$S\ (=P)$	.77	.78	.72	.74	.76	.76	.79	.80
7	1	.8	$S\ (=P)$	.76	.76	.76	.75	.74	.75	.77	.77

more powerful than the F test when $\rho = .8$ and $m = 10$ or 5. Hence, the S test has power close to or equal to that of the F test when $\rho = 0$ and .4 and when $m = 1$ and noticeably higher power when $\rho = .8$ and $m = 10$ or 5. (Of course, the results of Table I indicate that the F test is not a viable competitor to the S test because of its NRP distortions.)

3. For all tests, power increases greatly in m and only marginally in n. This is because m indexes the number of residuals upon which the tests depend. An increase in n provides a less variable estimator of β, which improves power, but not by nearly as much as an increase in m. For all tests, power is highest for the uniform distribution and lowest for the χ^2 and t_3 distributions. For all tests, power decreases sharply as ρ increases when $m = 10$ or 5, but is more or less independent of ρ when $m = 1$. This occurs because increasing ρ increases the correlation between the post-change residuals when $m > 1$, which can be

viewed as reducing the effective post-change sample size. When $m = 1$, there is only one post-change observation, so increasing ρ only reduces the precision with which β can be estimated, but does not affect the effective post-change sample size.

Based on the NRP and power results discussed above, we recommend using the S test. The S test has the best NRP performance. Its NRP performance is far superior to that of the F test. In addition, the NRP-corrected power of the S test is close to that of the F test when the errors are uncorrelated and better when the errors have noticeable correlation. The NRP-corrected power of the S test is noticeably superior to that of the P test when $m = 10$.

7.2.3. *Change in Regressor Distribution*

We carry out some simulations to see whether a change in the regressor distribution alone causes the S test 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 NRP's of the S test 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 S test does not reject the null with probability greater than .05 when the only instability present is instability in the regressor distribution. This is a desirable feature of the tests.

Cowles Foundation for Research in Economics, Yale University, 30 Hillhouse Ave., Box 208281, New Haven, CT 06520-8281, U.S.A.; donald.andrews@yale.edu.

Manuscript received April, 2002; final revision received January, 2003.

APPENDIX: PROOFS

PROOF OF THEOREM 1: For notational simplicity, we start by proving parts (a)–(d) for the case in which S and S_j are defined with $\widehat{\Sigma}_{n+m}$ replaced by Σ_0 and $m \geq d$. Let $S_j(\beta) = S_j(\beta, \Sigma_0)$.

We prove part (a) first. By Assumption 2(a), $\widehat{\beta}_{n+m} \to_p \beta_0$. In consequence, there exists a sequence of nonnegative constants $\{\varepsilon_n : n \geq 1\}$ for which $\varepsilon_n \to 0$ and $P(K_n) \to 1$, where $K_n = \{\|\widehat{\beta}_{n+m} - \beta_0\| < \varepsilon_n\}$. 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

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

$$= P(S_{n+1}(\beta_0) \le x) + o(1)$$
$$= P(S_\infty \le x) + o(1),$$

where the second equality holds because Assumptions 3(a) and (b) imply that $\inf_{\|\beta-\beta_0\|\le\varepsilon} S_{n+1}(\beta) \to S_{n+1}(\beta_0)$ as $\varepsilon \to 0$ a.s. and $S_{n+1}(\beta)$ has a distribution that does not depend on n and the last equality holds by definition of S_∞. If the inf is replaced by sup, then the first $\le$ in the third line is replaced by $\ge$. In consequence, $P(S_{n+1}(\widehat{\beta}_{n+m}) \le x) \to P(S_\infty \le x)$ and part (a) is proved.

Next, we prove part (b). We introduce the following notation. For some random or nonrandom vectors $\{\beta_j : j = 1, \ldots, n-m+1\}$, let $\widehat{F}_n(x, \{\beta_j\})$ denote the empirical df based on $\{S_j(\beta_j) : j = 1, \ldots, n-m+1\}$. That is,

$$\widehat{F}_n(x, \{\beta_j\}) = \frac{1}{n-m+1} \sum_{j=1}^{n-m+1} 1(S_j(\beta_j) \le x) \tag{A.2}$$

for $x \in R$. Note that $\widehat{F}_{S,n}(x) = \widehat{F}_n(x, \{\widehat{\beta}_{2(j)}\})$.

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 nonrandom vectors $\{\beta_j : j = 1, \ldots, n-m+1\}$, we define the smoothed df

$$\widehat{F}_n(x, \{\beta_j\}, h_n) = \frac{1}{n-m+1} \sum_{j=1}^{n-m+1} k((S_j(\beta_j) - x)/h_n), \tag{A.3}$$

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

We have

$$\big|\widehat{F}_{S,n}(x) - F_S(x)\big| \le \sum_{\ell=1}^{4} D_{\ell,n}, \quad \text{where} \tag{A.4}$$

$$D_{1,n} = \big|\widehat{F}_{S,n}(x) - \widehat{F}_n(x, \{\widehat{\beta}_{2(j)}\}, h_n)\big|,$$

$$D_{2,n} = \big|\widehat{F}_n(x, \{\widehat{\beta}_{2(j)}\}, h_n) - \widehat{F}_n(x, \{\beta_0\}, h_n)\big|,$$

$$D_{3,n} = \big|\widehat{F}_n(x, \{\beta_0\}, h_n) - \widehat{F}_n(x, \{\beta_0\})\big|, \quad \text{and}$$

$$D_{4,n} = \big|\widehat{F}_n(x, \{\beta_0\}) - F_S(x)\big|.$$

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

We have

$$D_{1,n} \le \frac{1}{n-m+1} \sum_{j=1}^{n-m+1} 1\big(S_j(\widehat{\beta}_{2(j)}) - x \in (0, h_n)\big), \tag{A.5}$$

because $\widehat{F}_{S,n}(x)$ and $\widehat{F}_n(x, \{\widehat{\beta}_{2(j)}\}, h_n)$ only differ when $(S_j(\widehat{\beta}_{2(j)}) - x)/h_n \in (0, 1)$.

By Assumption 2(a), 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 $L_n = \{\|\widehat{\beta}_{2(j)} - \beta_0\| \leq \varepsilon_n, \forall j = 1, \ldots, n-m+1\}$. Now, for all $\delta > 0$,

$$\begin{aligned}
(\text{A.6}) \qquad & P(D_{1,n} > \delta) \\
& \leq P((D_{1,n} > \delta) \cap L_n) + P(L_n^c) \\
& \leq P\left(\frac{1}{n-m+1} \sum_{j=1}^{n-m+1} \sup_{\|\beta-\beta_0\| \leq \varepsilon_n} 1(S_j(\beta) - x \in (0, h_n)) > \delta\right) + o(1) \\
& \leq E \sup_{\|\beta-\beta_0\| \leq \varepsilon_n} 1\big(S_1(\beta) - x \in (0, h_n)\big)/\delta + o(1) \\
& \leq E1\left(S_1(\beta_0) - x \in \left(-\varepsilon_n \sup_{\|\beta-\beta_0\| \leq \varepsilon_n} \left\|\frac{\partial}{\partial\beta} S_1(\beta)\right\|, h_n + \varepsilon_n \sup_{\|\beta-\beta_0\| \leq \varepsilon_n} \left\|\frac{\partial}{\partial\beta} S_1(\beta)\right\|\right)\right)\Big/\delta \\
& \quad + o(1),
\end{aligned}$$

where L_n^c denotes the complement of the set L_n, the third inequality uses Markov's inequality and the identical distributions of $S_j(\cdot)$ for $j = 1, \ldots, n-m+1$, 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 (A.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(b), and $S_1(\beta_0) \neq x$ a.s. by Assumption 3(c). 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

$$\begin{aligned}
(\text{A.7}) \qquad D_{2,n} &= \left|\frac{1}{n-m+1} \sum_{j=1}^{n-m+1} k'((S_j(\widetilde{\beta}_{2(j)}) - x)/h_n) \frac{\partial}{\partial\beta'} S_j(\widetilde{\beta}_{2(j)})(\widehat{\beta}_{2(j)} - \beta_0)/h_n\right| \\
&\leq \left(\frac{1}{n-m+1} \sum_{j=1}^{n-m+1} B \sup_{\|\beta-\beta_0\| \leq \varepsilon} \left\|\frac{\partial}{\partial\beta} S_j(\beta)\right\|\right) \sup_{r=1,\ldots,n-m+1} \|\widehat{\beta}_{2(r)} - \beta_0\|/h_n \\
&= O_p(1) o_p(1),
\end{aligned}$$

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

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

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

This completes the proof for the case in which S and S_j are defined with $\widehat{\Sigma}_{n+m}$ replaced by Σ_0 and $m \geq d$. The corresponding proof when $m < d$ is the same but with $S_j(\cdot)$ replaced by $P_j(\cdot) = P_j(\cdot, \Sigma_0)$ for $j = 1, \ldots, n+1$.

The proof for the case in which S and S_j are defined with $\widehat{\Sigma}_{n+m}$, rather than Σ_0, is essentially the same as that given above, but with $S_j(\beta)$ replaced by $S_j(\beta, \Sigma)$ when $m \geq d$ and by $P_j(\beta, \Sigma)$ when $m < d$ and with β, β_0, $\widehat{\beta}_{n+m}$, $\widehat{\beta}_{2(j)}$, and β_j replaced by the vectors comprised of the nonredundant elements of (β, Σ), (β_0, Σ_0), $(\widehat{\beta}_{n+m}, \widehat{\Sigma}_{n+m})$, $(\widehat{\beta}_{2(j)}, \widehat{\Sigma}_{n+m})$, and (β_j, Σ_j), respectively, where Σ_j is some random or nonrandom $m \times m$ matrix. In addition, the mean-value expansions in β around β_0 in the proof are replaced by expansions in (β, Σ^{-1}) around (β_0, Σ_0^{-1}). *Q.E.D.*

PROOF OF LEMMA 1: We start by showing that Assumptions 1 and LS imply that $\sup_{j=1,\dots,n-m+1}\|\widehat{\beta}_{2(j)}-\beta_0\|\to_p 0$ 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 $\delta>0$. Let m be fixed and suppose $n\to\infty$. For notational simplicity, let $\ell=\lceil m/2\rceil$. Then,

$$\text{(A.8)}\qquad \begin{aligned}&\sup_{j=1,\dots,n-\ell+1}\left\|(n-\ell)^{-1}\sum_{i=1,\dots,n;i\neq j,\dots,j+\ell-1}\xi_i\right\|\\ &\leq \sup_{j=1,\dots,n-m+1}\left\|(n-\ell)^{-1}\left(\sum_{i=1,\dots,n;i\neq j,\dots,j+\ell-1}\xi_i-\sum_{i=1}^n\xi_i\right)\right\|+\left\|(n-\ell)^{-1}\sum_{i=1}^n\xi_i\right\|\\ &=\sup_{j=1,\dots,n-m+1}\left\|(n-\ell)^{-1}\sum_{i=j,\dots,j+\ell-1}\xi_i\right\|+o_p(1),\end{aligned}$$

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

$$\text{(A.9)}\qquad \begin{aligned}P\Big((n-\ell)^{-1}\sup_{j\leq n-m+1}\|\tau_j\|>\varepsilon\Big)&=P\left(\bigcup_{j=1}^{n-m+1}\{\|\tau_j\|>(n-\ell)\varepsilon\}\right)\\ &\leq\sum_{j=1}^{n-m+1}P(\|\tau_j\|>(n-\ell)\varepsilon)\\ &\leq(n-m+1)E\|\tau_j\|^{1+\delta}(n-\ell)^{-(1+\delta)}\varepsilon^{-(1+\delta)}\\ &=o(1),\end{aligned}$$

where the second inequality uses Markov's inequality. Hence, the right-hand side of (A.8) is $o_p(1)$.

The estimator $\widehat{\beta}_{2(j)}$ in the LS case satisfies

$$\text{(A.10)}\qquad \begin{aligned}&\sup_{j=1,\dots,n-m+1}\|\widehat{\beta}_{2(j)}-\beta_0\|\\ &=\sup_{j=1,\dots,n-m+1}\left\|\left((n-\ell)^{-1}\sum_{i=1,\dots,n;i\neq j,\dots,j+\ell-1}X_iX_i'\right)^{-1}(n-\ell)^{-1}\sum_{i=1,\dots,n;i\neq j,\dots,j+\ell-1}X_iU_i\right\|\\ &\leq\left\|(EX_1X_1'+o_p(1))^{-1}(EX_1U_1+o_p(1)\right\|\\ &=O_p(1),\end{aligned}$$

where the inequality holds by applying (A.8) and (A.9) twice with $\xi_i=X_iX_i'-EX_iX_i'$ and $\xi_i=X_iU_i$ and, in consequence, the $o_p(1)$ terms hold uniformly over $j=1,\dots,n-m+1$.

The proof of the same result for the linear IV estimator is quite similar using the definition of the IV estimator in (3.5). In this case, (A.8) and (A.9) are applied with $\xi_i=X_iZ_i'$, $\xi_i=Z_iZ_i'$, and $\xi_i=Z_iU_i$. (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 Hölder's inequality.)

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

It is straightforward to verify Assumption 2(b) for the LS and IV estimators.

Assumption 3(a) holds for the LS and IV estimators because $S_j(\beta,\Sigma)$ and $P_j(\beta,\Sigma)$ are quadratic functions of β and quite simple functions of Σ^{-1}.

Next, we verify Assumption 3(b) for the LS and IV estimators. Let $\mathbf{Y}$, $\mathbf{X}$, and $\mathbf{Z}$ abbreviate $\mathbf{Y}_{j,j+m-1}$, $\mathbf{X}_{j,j+m-1}$, and $\mathbf{Z}_{j,j+m-1}$, respectively. First, we suppose $m \geq d$. For the LS estimator,

$$\frac{\partial}{\partial\beta} S_j(\beta, \Sigma) = -2\mathbf{X}'\Sigma^{-1}(\mathbf{Y} - \mathbf{X}\beta). \tag{A.11}$$

We have $E \sup_{\beta \in B(\beta_0,\varepsilon), \Sigma \in N(\Sigma_0)} \|(\partial/\partial\beta) S_1(\beta, \Sigma)\| < \infty$ because $EU_1^2 < \infty$, $E\|X_1\|^2 < \infty$, and $\inf_{\Sigma \in N(\Sigma_0)} \lambda_{\min}(\Sigma) > 0$ for some neighborhood $N(\Sigma_0)$ of Σ_0, where $\lambda_{\min}(\Sigma)$ denotes the smallest eigenvalue of Σ. An analogous result holds when $\Sigma = I_m$.

Let $\omega_{k,\ell}$ denote the (k, ℓ) element of Σ^{-1}. For the LS estimator,

$$\begin{aligned} \frac{\partial}{\partial\omega_{k,\ell}} S_j(\beta, \Sigma) &= 2(\mathbf{Y} - \mathbf{X}\beta)'\Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\frac{\partial}{\partial\omega_{k,\ell}}(\Sigma^{-1})(\mathbf{Y} - \mathbf{X}\beta) \\ &\quad + (\mathbf{Y} - \mathbf{X}\beta)'\Sigma^{-1}\mathbf{X}\frac{\partial}{\partial\omega_{k,\ell}}[(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}]\mathbf{X}'\Sigma^{-1}(\mathbf{Y} - \mathbf{X}\beta). \end{aligned} \tag{A.12}$$

We find that $E \sup_{\beta \in B(\beta_0,\varepsilon), \Sigma \in N(\Sigma_0)} \|(\partial/\partial\omega_{k,\ell}) S_1(\beta, \Sigma)\| < \infty$ because $EU_1^2 < \infty$, $E\|X_1\|^2 < \infty$, $\inf_{\Sigma \in N(\Sigma_0)} \lambda_{\min}(\Sigma) > 0$, and $(\partial/\partial\alpha)(A^{-1}) = -A^{-1}((\partial/\partial\alpha)A)A^{-1}$, where A is a nonsingular matrix that depends on α.

For the IV estimator,

$$\begin{aligned} \frac{\partial}{\partial\beta} S_j(\beta, \Sigma) &= -2\mathbf{X}'\Sigma^{-1}\mathbf{Z}(\mathbf{Z}'\Sigma^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\Sigma^{-1}(\mathbf{Y} - \mathbf{X}\beta) \\ &= -2\mathbf{X}'\Sigma^{-1/2}P_{\Sigma^{-1/2}\mathbf{Z}}\Sigma^{-1/2}(\mathbf{Y} - \mathbf{X}\beta), \end{aligned} \tag{A.13}$$

where $P_{\Sigma^{-1/2}\mathbf{Z}}$ is the projection matrix that projects onto the column space of $\Sigma^{-1/2}\mathbf{Z}$. Let $\widetilde{\mathbf{Y}}$, $\widetilde{\mathbf{X}}$, and $\widetilde{\mathbf{Z}}$ denote $\Sigma^{-1/2}\mathbf{Y}$, $\Sigma^{-1/2}\mathbf{X}$, and $\Sigma^{-1/2}\mathbf{Z}$, respectively. We can consider the columns of $\widetilde{\mathbf{X}}$ one at a time. So, for notational simplicity, we just suppose $\widetilde{\mathbf{X}}$ is a vector. Then, by the Cauchy–Schwarz inequality,

$$\begin{aligned} \big|\widetilde{\mathbf{X}}'P_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\beta)\big| &\leq \big(\widetilde{\mathbf{X}}'P_{\widetilde{\mathbf{Z}}}\widetilde{\mathbf{X}}\big)^{1/2}\big((\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\beta)'P_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\beta)\big)^{1/2} \\ &\leq \big(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\big)^{1/2}\big((\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\beta)'(\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\beta)\big)^{1/2}. \end{aligned} \tag{A.14}$$

The supremum of the right-hand side over $\beta \in B(\beta_0, \delta)$ and $\Sigma \in N(\Sigma_0)$ has finite expectation because $EU_1^2 < \infty$, $E\|X_1\|^2 < \infty$, $\inf_{\Sigma \in N(\Sigma_0)} \lambda_{\min}(\Sigma) > 0$. Hence,

$$E \sup_{\beta \in B(\beta_0,\varepsilon), \Sigma \in N(\Sigma_0)} \left\| \frac{\partial}{\partial\beta} S_1(\beta, \Sigma) \right\| < \infty \tag{A.15}$$

for the IV estimator when $m \geq d$. Note that we apply the Cauchy-Schwarz inequality in (A.14) in order to eliminate the $(\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}})^{-1}$ term that appears in the left-hand side, which cannot be bounded on its own.

An analogous result to (A.15) holds with $\partial/\partial\beta$ replaced by $\partial/\partial\omega_{k,l}$ by combining the calculations in (A.12) and (A.14). To bound the second term of (A.12), which involves $(\partial/\partial\omega_{k,l})[(\mathbf{Z}'\Sigma^{-1}\mathbf{Z})^{-1}]$ in the IV case, we make use of the formula for the derivative of the inverse of a matrix given above and the inequality

$$|x'\widetilde{P}A\widetilde{P}x| \leq \sum_{r=1}^{m}\sum_{s=1}^{m} |a_{r,s}| \cdot \|\widetilde{P}x\| \leq \sum_{r=1}^{m}\sum_{s=1}^{m} |a_{r,s}| \cdot \|x\|, \tag{A.16}$$

where x is a vector, $\widetilde{P}$ is a projection matrix, and A is a matrix with (r, s) element $a_{r,s}$.

Now, suppose $m < d$. In this case, for both the LS and IV estimators, we have

$$\frac{\partial}{\partial\beta}P_j(\beta,\Sigma) = -2\mathbf{X}'\Sigma^{-1}(\mathbf{Y}-\mathbf{X}\beta) \quad \text{and}$$
$$\frac{\partial}{\partial\omega_{k,\ell}}P_j(\beta,\Sigma) = (\mathbf{Y}-\mathbf{X}\beta)'\frac{\partial}{\partial\omega_{k,\ell}}(\Sigma^{-1})(\mathbf{Y}-\mathbf{X}\beta). \tag{A.17}$$

The expectation of the supremum over $\beta \in B(\beta_0, \varepsilon)$ and $\Sigma \in N(\Sigma_0)$ of the right-hand side of the second equation in (A.17) is finite because $EU_1^2 < \infty$ and $E\|X_1\|^2 < \infty$. Also, the right-hand side of the first equation in (A.17) is the same as in (A.11). Hence, Assumption 3(b) holds by the same argument as above.

Assumption 3(c) holds for the LS and IV estimators by Assumptions LS(d) and IV(d).

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

$$\sup_{\beta\in\mathcal{B}}\sup_{j=1,\dots,n-m+1}\left|Q_{2(j)}^{(k)}(\beta) - Q^{(k)}(\beta)\right| \to_p 0 \quad \text{and} \tag{A.18}$$

$$Q^{(k)}(\beta_0) < \inf_{\beta\notin B(\beta_0,\varepsilon)\cap\mathcal{B}} Q^{(k)}(\beta), \quad \text{where} \tag{A.19}$$

$$Q^{(1)}(\beta) = Eg(W_1,\beta)'\mathcal{V}^{-1}Eg(W_1,\beta),$$
$$Q^{(2)}(\beta) = Eg(W_1,\beta)'\mathcal{V}^{-1}(\beta_0)Eg(W_1,\beta), \quad \text{and} \tag{A.20}$$
$$Q^{(CU)}(\beta) = Eg(W_1,\beta)'\mathcal{V}^{-1}(\beta)Eg(W_1,\beta).$$

Condition (A.18) holds provided

$$\sup_{\beta\in\mathcal{B}}\sup_{j=1,\dots,n-m+1}\left|(n-\ell)^{-1}\sum_{i=1,\dots,n;i\neq j,\dots,j+\ell-1} g(W_i,\beta) - Eg(W_1,\beta)\right| \to_p 0, \tag{A.21}$$

where $\ell = \lceil m/2 \rceil$ as above, because Assumption GMM(e) insures that the weight matrices are well-behaved. Equation (A.21) holds pointwise in β for all $\beta \in \mathcal{B}$ by applying (A.8) and (A.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 (A.19) holds by Assumption GMM(a)–(c) and (e).

Next, we use (A.18) and (A.19) to show that $\sup_{j=1,\dots,n-m+1}\|\widehat{\beta}_{2(j)} - \beta_0\| \to_p 0$. By (A.19), 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

$$\begin{aligned}
&P\Big(\sup_{j=1,\dots,n-m+1}\|\widehat{\beta}_{2(j)} - \beta_0\| > \varepsilon\Big)\\
&\leq P\Big(\sup_{j=1,\dots,n-m+1} Q^{(k)}(\widehat{\beta}_{2(j)}) - Q^{(k)}(\beta_0) > \delta\Big)\\
&= P\Big(\sup_{j=1,\dots,n-m+1}\big(Q^{(k)}(\widehat{\beta}_{2(j)}) - Q_{2(j)}^{(k)}(\widehat{\beta}_{2(j)}) + Q_{2(j)}^{(k)}(\widehat{\beta}_{2(j)}) - Q^{(k)}(\beta_0)\big) > \delta\Big)\\
&\leq P\Big(\sup_{j=1,\dots,n-m+1}\big(Q^{(k)}(\widehat{\beta}_{2(j)}) - Q_{2(j)}^{(k)}(\widehat{\beta}_{2(j)}) + Q_{2(j)}^{(k)}(\beta_0) - Q^{(k)}(\beta_0)\big) > \delta\Big)\\
&\leq P\Big(2\sup_{j=1,\dots,n-m+1}\left|Q_{2(j)}^{(k)}(\beta) - Q^{(k)}(\beta)\right| > \delta\Big)\\
&= o_p(1),
\end{aligned} \tag{A.22}$$

where the second inequality holds because $\widehat{\beta}_{2(j)}$ minimizes $Q^{(k)}_{2(j)}(\beta)$ over $\beta \in \mathcal{B}$ and the second equality holds by (A.18).

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

To verify Assumption 2(b) (which only applies to case 1 GMM estimators), we write

$$\text{(A.23)} \qquad \sup_{\beta \in B(\beta_0, \varepsilon_n)} \|\widehat{\Sigma}_{n+m}(\beta) - \Sigma_0\| \le \sup_{\beta \in B(\beta_0, \varepsilon_n)} \|\widehat{\Sigma}_{n+m}(\beta) - \Sigma(\beta)\| + \sup_{\beta \in B(\beta_0, \varepsilon_n)} \|\Sigma(\beta) - \Sigma_0\|,$$

where $\Sigma(\beta) = E\mathbf{U}_{1,m}(\beta)\mathbf{U}_{1,m}(\beta)'$. The second term on the right-hand side is $o(1)$ by the dominated convergence theorem because $U(W_i, \beta)$ is continuous at β_0 almost surely and $E\sup_{\beta \in B(\beta_0, \varepsilon)} U^2(W_i, \beta) < \infty$ by Assumption GMM(f). For any fixed β, the first term on the right-hand side is $o_p(1)$ by the ergodic theorem. A generic uniform convergence result strengthens pointwise convergence to uniform convergence over $\beta \in B(\beta_0, \varepsilon)$ for some $\varepsilon > 0$. For example, Theorem 5 of Andrews (1992) using Assumption TSE-1D gives the desired result under Assumptions 1 and GMM(f).

Assumption 3(a) holds for GMM estimators by Assumption GMM(f). Next, to establish Assumption 3(b) for case 2 GMM estimators, we verify that

$$B_n := (n-m+1)^{-1} \sum_{j=1}^{n-m+1} \sup_{\beta \in B(\beta_0, \varepsilon)} \left\| \frac{\partial}{\partial \beta} S_j(\beta) \right\| = O_p(1).$$

We have

$$\text{(A.24)} \qquad \frac{\partial}{\partial \beta_r} S_j(\beta, \Sigma) = 2\left(\sum_{i=j}^{j+\ell-1} \frac{\partial}{\partial \beta_r} g(W_i, \beta)\right)' V_j^{-1}(\beta) \sum_{i=j}^{j+\ell-1} g(W_i, \beta) + \sum_{i=j}^{j+\ell-1} g(W_i, \beta)' \left(\frac{\partial}{\partial \beta_r}(V_j^{-1}(\beta))\right) \sum_{i=j}^{j+\ell-1} g(W_i, \beta)$$

for $r = 1, \ldots, d_\beta$. 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 $j = 1, \ldots, n-m+1$ 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 case 1 GMM estimators, the verification of Assumption 3(b) for GMM estimators is essentially the same as that for IV estimators with $Y_{j,j+m-1} - X_{j,j+m-1}\beta$ and $Z_{j,j+m-1}$ replaced by $U_{j,j+m-1}(\beta)$ and $Z_{j,j+m-1}(\beta)$, respectively, using Assumption GMM(f).

Assumption 3(c) holds for GMM estimators by Assumption GMM(g). *Q.E.D.*

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.

——— (2002): "End-of-sample Instability Tests," Cowles Foundation Discussion Paper No. 1369, Yale University. Available at http://cowles.econ.yale.edu.

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

ANDREWS, D. W. K., AND J.-Y. KIM (2002): "End-of-sample Cointegration Breakdown Tests," Cowles Foundation Discussion Paper No. 1404, Yale University. Available at http://cowles.econ.yale.edu.

ANDREWS, D. W. K., AND W. PLOBERGER (1994): "Optimal Tests when a Nuisance Parameter Is Present only under the Alternative," *Econometrica*, 62, 1383–1414.

BELSLEY, D. A., E. KUH, AND R. E. WELSCH (1980): *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

CHOW, G. C. (1960): "Tests of Equality between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 28, 591–605.

COOK, R. D., AND S. WEISBERG (1982): *Residuals and Influence in Regression*. New York: Chapman Hall.

DUFOUR, J.-M., E. GHYSELS, AND A. HALL (1994): "Generalized Predictive Tests and Structural Change Analysis in Econometrics," *International Economic Review*, 35, 199–229.

FAIR, R. C. (2003): "Testing for a New Economy in the 1990s," *Business Economics*, 38, forthcoming.

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.

HALL, A. R., AND A. SEN (1999): "Structural Stability Testing in Models Estimated by Generalized Method of Moments," *Journal of Business and Economic Statistics*, 17, 335–348.

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

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.

SCHEFFÉ, H. (1959): *Analysis of Variance*. New York: Wiley.

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.