

End-of-Sample Instability Tests

Donald W. K. Andrews¹

*Cowles Foundation for Research in Economics
Yale University*

January 2002

Revised: April 2002

Abstract

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

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

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

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

JEL Classification Numbers: C12, C52.

1 Introduction

This paper considers the problem of testing for structural instability over a short time interval, such as at the end of a sample. Most tests in the literature are designed for detecting instability that lasts for a relatively long period of time starting somewhere in the middle of the sample, e.g., see Andrews and Fair (1988), Ghysels and Hall (1990), Hansen (1992), Andrews (1993), Ghysels, Guay, and Hall (1997), and other references in listed Stock (1994). These tests use asymptotics in which the number of observations before a potential changepoint, n , and the number after the potential changepoint, m , both go to infinity. Such tests are not valid in the case considered here in which the number of observations in the period of potential instability, m , is small—perhaps as small as one. In this paper, we design tests that are asymptotically valid when $n \rightarrow \infty$ with m fixed. The primary difficulty in doing so 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 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 variants of the F test that are valid under weak assumptions and apply to a wide variety of models. We do so by constructing critical values using a subsampling-like method. In the linear regression model, the tests we propose are asymptotically valid with non-normal, heteroskedastic, conditionally heteroskedastic, and/or autocorrelated errors and with regressors that are not strictly exogenous. The observations and/or errors could even possess long memory. The main requirement is that the observations are strictly stationary and ergodic under the null hypothesis. Furthermore, the tests we propose apply to regression models estimated by instrumental variables (IV) and to nonlinear models estimated by generalized method of moments (GMM) and maximum likelihood (ML).

The bulk of this paper discusses tests for structural instability at the end of the sample. Extending such tests to the case of potential instability at the beginning,

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

We now discuss the end-of-sample tests for a regression model. The first test statistic that we consider, S_a , is a quadratic form in the post-change residual vector evaluated at the pre-change least squares (LS) estimator. Like the F statistic, see Chow (1960), the form of S_a depends on whether the number of post-change observations, m , is greater than or equal to the number of regressors d . When $m \geq d$, the weight matrix of the quadratic form is the projection matrix onto the column space of the post-change regressor matrix. When $m \leq d$, the weight matrix is the m dimensional identity matrix I_m . Critical values for S_a are obtained using a subsampling-like method described below.

Simulations indicate that the S_a test over-rejects the null hypothesis somewhat when the null is true in many cases. In consequence, we make finite sample adjustments to both the test statistic and the critical value that lead to better finite sample size properties. The resulting test is called the S_b test.

The S_a and S_b test statistics are variants of the F test that are designed for the case of uncorrelated errors. These tests have correct size asymptotically whether or not the errors are correlated. However, if the errors are correlated, they sacrifice power. To address this issue, we introduce a third statistic, S_c , that is similar to S_a but is a quadratic form in *transformed* post-change residuals. The transformation matrix is the square root of the inverse of an estimator of the $m \times m$ covariance matrix of the errors. Correspondingly, when $m \geq d$, the weight matrix is the projection matrix onto the column space of *transformed* post-change regressors.

As with the S_a test, simulations show that the S_c test tends to over-reject under the null hypothesis. In consequence, we introduce finite sample adjustments to the test statistic and critical value to obtain a test S_d that has improved finite sample size properties. The S_d test is our recommended test. Its size and power properties are found to be quite good over a range of regression scenarios.

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

The 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 $j, \dots, j + m - 1$ through the parameter estimator that is used to compute the residuals. See Politis, Romano, and Wolf (1999) for an in-depth treatment of, and references on, subsampling methods.

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

The Monte Carlo simulations mentioned above are 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 . As noted above, we find that the S_b and S_d tests have quite good size properties over the range of cases considered. On the other hand, the size properties of the F test are poor whenever $\rho > 0$ and $m > 1$ and whenever $\rho = 0$ and the errors and regressors are uniform. The power of the S_d test is superior to that of the S_b test when $\rho > 0$ and similar when $\rho = 0$. In consequence, the S_d test is the recommended test.

The recommended test, S_d , is defined in Section 2 for the ease of readers who wish to skip the details of the test's motivation and the details of its asymptotic properties. The latter are given in subsequent sections of the paper.

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

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

The remainder of this paper is organized as follows. All sections of the paper except Section 6 discuss end-of-sample instability tests. Section 2 defines the recommended test S_d . Section 3 motivates the statistics considered in the paper using the F statistic for the normal linear regression model. Section 4 considers tests for the linear regression model estimated by IV. Section 5 considers moment condition models estimated by GMM. Section 6 discusses tests for structural instability that occurs at the beginning or in the middle of the sample for a small number of observations. Section 7 briefly discusses application of the tests to simple models with integrated variables. Section 8 introduces high-level assumptions, provides sufficient conditions

for these assumptions for the LS, IV, and GMM cases, and states the main asymptotic results. Section 9 provides some Monte Carlo results. An Appendix contains proofs.

2 Definition of the Recommended Test S_d

In this section, we define the recommended end-of-sample instability test.

Consider a linear regression model with d regressors, n observations before the potential changepoint, and m observations after the potential changepoint:

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

We assume that $EU_i X_i = 0$, $EX_i X_i'$ 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

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

The recommended test statistic is S_d , defined below. The critical value for S_d is determined by certain *parametric subsample* statistics $\{S_{d,j} : j = 1, \dots, n - m + 1\}$, defined below. For a test with asymptotic significance level α , the critical value for S_d is the $1 - \alpha$ sample quantile, $\hat{q}_{S_d, 1-\alpha}$, of $\{S_{d,j} : j = 1, \dots, n - m + 1\}$. That is,

$$\hat{q}_{S_d, 1-\alpha} = \inf\{x \in R : \hat{F}_{S_d, n}(x) \geq 1 - \alpha\}, \quad (2.3)$$

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

$$(n - m + 1)^{-1} \sum_{j=1}^{n-m+1} 1(S_d > S_{d,j}) \geq 1 - \alpha. \quad (2.4)$$

The p -value for the S_d test is

$$pv_{S_d} = (n - m + 1)^{-1} \sum_{j=1}^{n-m+1} 1(S_d \leq S_{d,j}). \quad (2.5)$$

The test statistic S_d is a positive definite quadratic form in a transformed m -vector of post-change residuals evaluated at the LS estimator based on the whole sample. The transformation matrix is the square root of the inverse of an estimator of the

$m \times m$ error covariance matrix. 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 full sample LS estimator 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_d is evidence against the null hypothesis.

The weight matrix for the quadratic form depends on whether $m \geq d$ or $m \leq d$, where m is the number of post-change observations and d is the number of regressors. When $m \geq d$, the weight matrix is the projection matrix onto the column space of the $m \times d$ matrix of transformed post-change regressors, where the same transformation is applied to the regressors as to the residuals. When $m \leq d$, the column space of the transformed post-change regressors is R^m and the weight matrix for S_d is the m -dimensional identity matrix, I_m . Note that when $m = d$, the two definitions of the weight matrix coincide.

We now define S_d . Let

$$\begin{aligned}\mathbf{Y}_{r,s} &= (Y_r, \dots, Y_s)', \\ \mathbf{X}_{r,s} &= (X_r, \dots, X_s)', \text{ and} \\ \mathbf{U}_{r,s} &= (U_r, \dots, U_s)'\end{aligned}\tag{2.6}$$

for $1 \leq r \leq s \leq n + m$. Let

$$\begin{aligned}\widehat{\beta}_{r,s} &= \text{LS estimator of } \beta \text{ using observations indexed by } i = r, \dots, s \text{ and} \\ \widehat{\beta}_{2(j)} &= \text{LS estimator of } \beta \text{ using observations indexed by } i = 1, \dots, n \text{ with} \\ & i \neq j, \dots, j + \lceil m/2 \rceil - 1,\end{aligned}\tag{2.7}$$

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

$$\begin{aligned}\widehat{\Sigma}_{1,n+m} &= (n+1)^{-1} \sum_{j=1}^{n+1} \widehat{\mathbf{U}}_{j,j+m-1} \widehat{\mathbf{U}}'_{j,j+m-1}, \text{ where} \\ \widehat{\mathbf{U}}_{j,j+m-1} &= \mathbf{Y}_{j,j+m-1} - \mathbf{X}_{j,j+m-1} \widehat{\beta}_{1,n+m}.\end{aligned}\tag{2.8}$$

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

$$\begin{aligned}S_d &= S_{n+1}(\widehat{\beta}_{1,n+m}, \widehat{\Sigma}_{1,n+m}) \text{ and} \\ S_{d,j} &= S_j(\widehat{\beta}_{2(j)}, \widehat{\Sigma}_{1,n+m}),\end{aligned}\tag{2.9}$$

where

$$\begin{aligned}S_j(\beta, \Sigma) &= A_j(\beta, \Sigma)' V_j^{-1}(\Sigma) A_j(\beta, \Sigma), \\ A_j(\beta, \Sigma) &= \mathbf{X}'_{j,j+m-1} \Sigma^{-1} (\mathbf{Y}_{j,j+m-1} - \mathbf{X}_{j,j+m-1} \beta), \\ V_j(\Sigma) &= \mathbf{X}'_{j,j+m-1} \Sigma^{-1} \mathbf{X}_{j,j+m-1},\end{aligned}\tag{2.10}$$

$\beta \in R^d$, and Σ is a nonsingular $m \times m$ matrix, for $j = 1, \dots, n+1$. (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 (2.4) and (2.5) are only over $j = 1, \dots, n-m+1$.)

In terms of the description of the statistic S_d given above, the transformed residual vector is $\widehat{\Sigma}_{1,n+m}^{-1/2}(\mathbf{Y}_{j,j+m-1} - \mathbf{X}_{j,j+m-1}\widehat{\beta}_{1,n+m})$, the transformed regressor matrix is $\widehat{\Sigma}_{1,n+m}^{-1/2}\mathbf{X}_{j,j+m-1}$, and the projection matrix is $\widehat{\Sigma}_{1,n+m}^{-1/2}\mathbf{X}_{j,j+m-1} \times (\mathbf{X}'_{j,j+m-1}\widehat{\Sigma}_{1,n+m}^{-1}\mathbf{X}_{j,j+m-1})^{-1}\mathbf{X}'_{j,j+m-1}\widehat{\Sigma}_{1,n+m}^{-1/2}$.

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

$$\begin{aligned} S_d &= P_{n+1}(\widehat{\beta}_{1,n+m}, \widehat{\Sigma}_{1,n+m}) \text{ and} \\ S_{d,j} &= P_j(\widehat{\beta}_{2(j)}, \widehat{\Sigma}_{1,n+m}), \end{aligned} \quad (2.11)$$

where

$$P_j(\beta, \Sigma) = (\mathbf{Y}_{j,j+m-1} - \mathbf{X}_{j,j+m-1}\beta)' \Sigma^{-1} (\mathbf{Y}_{j,j+m-1} - \mathbf{X}_{j,j+m-1}\beta) \quad (2.12)$$

and $\widehat{\beta}_{1,n+m}$, $\widehat{\beta}_{2(j)}$, and $\widehat{\Sigma}_{1,n+m}$ are as defined above.

Next, we define the S_d test for the case of a linear regression model estimated by IV with IV vectors Z_i for $i = 1, \dots, n+m$. In this case, the S_d test is the same as in the LS case, but with $\widehat{\beta}_{1,n+m}$ and $\widehat{\beta}_{2(j)}$ being IV estimators, rather than LS estimators (but still defined as in (2.7)), and in the case where $m \geq d$ with $A_j(\beta, \Sigma)$ and $V_j(\Sigma)$ defined by

$$\begin{aligned} A_j(\beta, \Sigma) &= \mathbf{Z}'_{j,j+m-1} \Sigma^{-1} (\mathbf{Y}_{j,j+m-1} - \mathbf{X}_{j,j+m-1}\beta) \text{ and} \\ V_j(\Sigma) &= \mathbf{Z}'_{j,j+m-1} \Sigma^{-1} \mathbf{Z}_{j,j+m-1}, \text{ where} \\ \mathbf{Z}_{j,j+m-1} &= (Z_j, \dots, Z_{j+m-1})'. \end{aligned} \quad (2.13)$$

In the IV case, d denotes the number of IVs rather than the number of regressors. Thus, the form of the weight matrix depends on the number of IVs, not on the number of regressors.

Lastly, we define the S_d test for the case of a moment condition model estimated by GMM. Suppose the moment conditions are

$$\begin{aligned} Eg(W_i, \beta_0) &= 0 \text{ and} \\ g(W_i, \beta) &= U(W_i, \beta)Z(W_i, \beta) \end{aligned} \quad (2.14)$$

for $i = 1, \dots, n+m$, where the potential changepoint is at $i = n$, $U(W_i, \beta) \in R$, $U_i = U(W_i, \beta_0)$ is an error term, and $Z(W_i, \beta)$ is a d -vector of instruments. In this case, the S_d test is the same as in the LS case, but with (i) $\widehat{\beta}_{1,n+m}$ and $\widehat{\beta}_{2(j)}$ being one-step, two-step, or continuously updated GMM estimators, rather than LS estimators (defined using the same observation indices i as in (2.7)), (ii) $\widehat{\Sigma}_{1,n+m}$ defined by

$$\begin{aligned} \widehat{\Sigma}_{1,n+m} &= \widehat{\Sigma}_{1,n+m}(\widehat{\beta}_{1,n+m}), \text{ where} \\ \widehat{\Sigma}_{1,n+m}(\beta) &= (n+1)^{-1} \sum_{j=1}^{n+1} \mathbf{U}_{j,j+m-1}(\beta) \mathbf{U}_{j,j+m-1}(\beta)' \text{ and} \\ \mathbf{U}_{j,j+m-1}(\beta) &= (U(W_j, \beta), \dots, U(W_{j+m-1}, \beta))', \end{aligned} \quad (2.15)$$

(iii) $A_j(\beta, \Sigma)$ and $V_j(\Sigma)$ defined by

$$\begin{aligned} A_j(\beta, \Sigma) &= \mathbf{Z}_{j,j+m-1}(\beta)' \Sigma^{-1} \mathbf{U}_{j,j+m-1}(\beta), \text{ and} \\ V_j(\Sigma) &= V_j(\widehat{\beta}_{1,n+m}, \Sigma), \text{ where} \\ \mathbf{Z}_{j,j+m-1}(\beta) &= (Z(W_j, \beta), \dots, Z(W_{j+m-1}, \beta))' \text{ and} \\ V_j(\beta, \Sigma) &= \mathbf{Z}_{j,j+m-1}(\beta)' \Sigma^{-1} \mathbf{Z}_{j,j+m-1}(\beta), \end{aligned} \quad (2.16)$$

and (iv) $P_j(\beta, \Sigma)$ defined by

$$P_j(\beta, \Sigma) = \mathbf{U}_{j,j+m-1}(\beta)' \Sigma^{-1} \mathbf{U}_{j,j+m-1}(\beta). \quad (2.17)$$

As in the linear IV case, the form of the weight matrix depends on the number of IVs d compared to the number of post-change observations m . For details of the definition of the GMM estimators, see Section 5.

In the special case where the GMM moment conditions are LS normal equations, i.e., $g(W_i, \beta) = (Y_i - X_i' \beta) X_i$, or linear IV normal equations, i.e., $g(W_i, \beta) = (Y_i - X_i' \beta) Z_i$, the GMM S_d test reduces to the S_d tests described above for the LS and linear IV cases.

3 Motivation Based on Linear Regression

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

The “test-generating” model is

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

where $Y_i \in R$, $U_i \in R$, and $X_i, \beta_0, \beta_1 \in R^d$. The observations are $\{W_i : i = 1, \dots, n + m\}$, where $W_i = (Y_i, X_i)'$. The errors and regressors of the “test-generating” model satisfy $U_i \sim N(0, \sigma_0^2)$, U_i is independent of X_i , and $EX_i X_i'$ is positive definite.

For notational simplicity, we abbreviate the subscript “ $n + 1, n + m$ ” by “ $n +$ ”. For example, $\mathbf{Y}_{n+} = \mathbf{Y}_{n+1, n+m}$. In vector notation, the “test-generating” model is

$$\begin{aligned} \mathbf{Y}_{1,n} &= \mathbf{X}_{1,n} \beta_0 + \mathbf{U}_{1,n} \text{ and} \\ \mathbf{Y}_{n+} &= \mathbf{X}_{n+} \beta_1 + \mathbf{U}_{n+}. \end{aligned} \quad (3.2)$$

The “test-generating” null and alternative hypotheses are

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

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

The more general model and hypotheses of interest are given in (2.1) and (2.2) above. The hypotheses of interest also can be expressed as

$$\begin{aligned} H_0 : & \begin{cases} E(Y_i - X_i' \beta_0) X_i = 0 \text{ for all } i = 1, \dots, n + m \text{ and} \\ \{(Y_i, X_i) : i \geq 1\} \text{ are stationary and ergodic} \end{cases} \\ H_1 : & \begin{cases} E(Y_i - X_i' \beta_0) X_i = 0 \text{ for all } i = 1, \dots, n, \text{ and} \\ E(Y_i - X_i' \beta_0) X_i \neq 0 \text{ for some } i = n + 1, \dots, n + m \text{ and/or} \\ \text{the distribution of } U_i \text{ for some } i = n + 1, \dots, n + m \\ \text{differs from the distribution of } U_i \text{ for } i = 1, \dots, n. \end{cases} \end{aligned} \quad (3.4)$$

For linear regression models estimated by LS, the hypotheses in (2.2) and (3.4) are equivalent. But, for over-identified linear regression models estimated by IV and for over-identified moment condition models estimated by GMM, which are considered below, hypotheses that are analogous to those in (3.4) 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 over-identifying restrictions, see Ghysels and Hall (1990). For IV and GMM cases, the hypotheses that we consider are analogues of (3.4), rather than (2.2) (although one could design tests for (2.2) if desired).

3.1 $m \geq d$ Case

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

$$\begin{aligned} F &= (\widehat{\beta}_{n+} - \widehat{\beta}_{1,n})' \left(\widehat{\sigma}^2 \left[(\mathbf{X}'_{1,n} \mathbf{X}_{1,n})^{-1} + (\mathbf{X}'_{n+} \mathbf{X}_{n+})^{-1} \right] \right)^{-1} (\widehat{\beta}_{n+} - \widehat{\beta}_{1,n}) / d \\ &= (\mathbf{Y}_{n+} - \mathbf{X}_{n+} \widehat{\beta}_{1,n})' \mathbf{X}_{n+} (\mathbf{X}'_{n+} \mathbf{X}_{n+})^{-1} \left[(\mathbf{X}'_{1,n} \mathbf{X}_{1,n})^{-1} + (\mathbf{X}'_{n+} \mathbf{X}_{n+})^{-1} \right]^{-1} \times \\ &\quad (\mathbf{X}'_{n+} \mathbf{X}_{n+})^{-1} \mathbf{X}'_{n+} (\mathbf{Y}_{n+} - \mathbf{X}_{n+} \widehat{\beta}_{1,n}) / (d \widehat{\sigma}^2), \end{aligned} \quad (3.5)$$

where $\widehat{\sigma}^2$ is the sum of squares residuals for $i = 1, \dots, n + m$ (computed using $\widehat{\beta}_{1,n}$ and $\widehat{\beta}_{n+}$) divided by $n + m - 2d$. Note that $(\mathbf{X}'_{n+} \mathbf{X}_{n+})^{-1}$ only exists if $m \geq d$, as is assumed here.

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

more general error processes. We consider test statistics that are slight variants of the F statistic.

The variance estimator, $\hat{\sigma}^2$, appears in the F statistic to yield invariance of the statistic to the error variance σ_0^2 . It does not contribute to the power of the statistic and we can eliminate it without reducing the power of the test. The critical values that we employ yield invariance with respect to σ_0^2 . Similarly, we can delete the constant d without affecting power. The form of the F test is simplified considerably if we replace $(\mathbf{X}'_{1,n}\mathbf{X}_{1,n})^{-1}+(\mathbf{X}'_{n+}\mathbf{X}_{n+})^{-1}$ by $(\mathbf{X}'_{n+}\mathbf{X}_{n+})^{-1}$. This simplification is warranted because n is much larger than m , $\lim_{n \rightarrow \infty}(\mathbf{X}'_{1,n}\mathbf{X}_{1,n})^{-1} = 0$, and $(\mathbf{X}'_{n+}\mathbf{X}_{n+})^{-1}$ does not depend on n under the assumptions introduced below. With these alterations, we obtain the following variant of the F statistic:

$$S_a = (\mathbf{Y}_{n+} - \mathbf{X}_{n+}\hat{\beta}_{1,n})'P_{\mathbf{X}_{n+}}(\mathbf{Y}_{n+} - \mathbf{X}_{n+}\hat{\beta}_{1,n}), \quad (3.6)$$

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

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

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

For $j \geq 1$, let

$$S_j(\beta) = S_j(\beta, I_m), \quad (3.7)$$

where I_m is the m dimensional identity matrix and $S_j(\beta, \Sigma)$ is defined in (2.10).

The statistic S_a can be written as

$$S_a = S_{n+1}(\hat{\beta}_{1,n}). \quad (3.8)$$

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

Note that $\{S_j(\beta) : j \geq 1\}$ are stationary and ergodic for all β . In consequence, the empirical distribution function (df) of $\{S_j(\beta) : j = 1, \dots, n-m+1\}$ is a consistent estimator of the df of $S_1(\beta)$ for all β . Hence, we can consistently estimate the df of $S_1(\beta_0)$ by using the empirical df of $\{S_j(\beta) : j \geq 1\}$ evaluated at consistent estimators of β_0 (see Theorem 1 below). The estimator $\hat{\beta}_{1,n}$, which appears in the statistic S_a , does not depend on the observations that appear in $S_{n+1}(\beta)$. To mirror this property in the subsample statistics, we evaluate $S_j(\beta)$ at an estimator $\hat{\beta}_{(j)}$ that does not depend on the observations that appear in $S_j(\beta)$.² By definition, for $j =$

$1, \dots, n - m + 1,$

$$\begin{aligned} \widehat{\beta}_{(j)} &= \text{LS estimator of } \beta \text{ using observations indexed by } i = 1, \dots, n \text{ with} \\ & i \neq j, \dots, j + m - 1. \end{aligned} \quad (3.9)$$

The estimator $\widehat{\beta}_{(j)}$ is consistent for β_0 (uniformly over j , under suitable assumptions).

Define

$$S_{a,j} = S_j(\widehat{\beta}_{(j)}) \text{ for } j = 1, \dots, n - m + 1. \quad (3.10)$$

The empirical df of $\{S_{a,j} : j = 1, \dots, n - m + 1\}$, which we denote by $\widehat{F}_{S_a, n}(x)$, converges in probability (and almost surely) to the df of $S_1(\beta_0)$. In consequence, to obtain a test with asymptotic significance level α , we take the critical value for the test statistic S_a to be the $1 - \alpha$ sample quantile, $\widehat{q}_{S_a, 1-\alpha}$, of $\{S_{a,j} : j = 1, \dots, n - m + 1\}$, defined as in (2.3). One rejects H_0 if $S_a > \widehat{q}_{S_a, 1-\alpha}$. Equivalently, one rejects H_0 if (2.4) holds with d replaced by a . The p -value for S_a is defined in (2.5) with d replaced by a .

Next, we consider a variant of the S_a test that is designed to have better finite-sample properties. We define the S_b and $S_{b,j}$ statistics as follows:

$$\begin{aligned} S_b &= S_{n+1}(\widehat{\beta}_{1, n+m}) \text{ and} \\ S_{b,j} &= S_j(\widehat{\beta}_{2(j)}) \text{ for } j = 1, \dots, n - m + 1, \end{aligned} \quad (3.11)$$

where $\widehat{\beta}_{2(j)}$ is defined in (2.7). Critical values and p -values for S_b are obtained using $\{S_{b,j} : j = 1, \dots, n - m + 1\}$ as in (2.3)-(2.5) with d replaced by b .

The motivation for the S_b test is as follows. Simulations indicate that the S_a test tends to over-reject the null hypothesis somewhat in finite samples. The S_b statistic is slightly less variable than the S_a statistic because the estimator $\widehat{\beta}_{1, n+m}$ makes use of the observations indexed by $i = n + 1, \dots, n + m$ and, hence, the residuals indexed by $i = n + 1, \dots, n + m$ (upon which $S_{n+1}(\cdot)$ depends) are less variable when computed using $\widehat{\beta}_{1, n+m}$ than when computed using $\widehat{\beta}_{1, n}$. It is natural to consider basing the critical values for S_b on $S_j(\widehat{\beta}_{1, n})$ for $j = 1, \dots, n - m + 1$, because $\widehat{\beta}_{1, n}$ correspondingly makes use of the observations that appear in the residuals upon which $S_j(\cdot)$ depends. Simulation results show, however, that this yields a test that has very similar size and power properties to those of the S_a test. In particular, the test tends to over-reject the null hypothesis, just as the S_a test does.

Alternatively, one could consider basing the critical values for the statistic S_b on $S_j(\widehat{\beta}_{(j)})$ for $j = 1, \dots, n - m + 1$. This yields a test whose rejection rate under the null hypothesis is smaller than when using $S_j(\widehat{\beta}_{1, n})$, because $\widehat{\beta}_{(j)}$ does not make use of the observations that appear in the residuals upon which $S_j(\cdot)$ depends and, hence, $S_j(\widehat{\beta}_{(j)})$ is more variable than $S_j(\widehat{\beta}_{1, n})$. In fact, simulations show that $S_j(\widehat{\beta}_{(j)})$ is too variable and the resulting test tends to under-reject the null hypothesis when using $S_j(\widehat{\beta}_{(j)})$ to compute the critical values.

We do not use either of these ways of computing the critical values. Instead, we base the critical values for S_b on $S_j(\widehat{\beta}_{2(j)})$ for $j = 1, \dots, n - m + 1$, as defined in (3.11). This is a compromise between using $S_j(\widehat{\beta}_{1, n})$ and $S_j(\widehat{\beta}_{(j)})$. The estimator $\widehat{\beta}_{2(j)}$ uses

half the observations that appear in the residuals upon which $S_j(\cdot)$ depends. Hence, its variability lies between that of $S_j(\widehat{\beta}_{1,n})$ and $S_j(\widehat{\beta}_{(j)})$. This is an *ad hoc*, but quite natural, finite sample adjustment. Simulations show that the use of $S_j(\widehat{\beta}_{2(j)})$ works quite well in finite samples over a wide range of parameter combinations, see Section 9.

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

$$\begin{aligned} S_a &= (\mathbf{Y}_{n+} - \mathbf{X}_{n+}\widehat{\beta}_{1,n})'(\mathbf{Y}_{n+} - \mathbf{X}_{n+}\widehat{\beta}_{1,n}) \\ &= \sum_{i=n+1}^{n+m} (Y_i - X_i'\widehat{\beta}_{1,n})^2. \end{aligned} \quad (3.12)$$

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

A natural extension of the definition of the statistic S_a to the case where $m < d$ is made using its definition in (3.6), in which $P_{\mathbf{X}_{n+}} = I_m$ when $m < d$ because \mathbf{X}_{n+} has column rank m . That is, when $m < d$, the natural definition of S_a is that given in (3.12). In fact, as shown in the next subsection, this definition is the definition that we obtain when we consider a variant of the F statistic for the case where $m \leq d$. Correspondingly, a natural extension of the S_b statistic to the case where $m < d$ is given by (3.12) with $\widehat{\beta}_{1,n}$ replaced by $\widehat{\beta}_{1,n+m}$.

In some Monte Carlo simulations analogous to those reported in Section 9, we consider the F statistic as defined in (3.5), but with $\widehat{\sigma}^2$ deleted. Critical values are constructed analogously to those for S_a . Neither the size nor the power results differ much from those of the simpler test based on S_a . In most cases, the differences in rejection probabilities are .005 or less. In consequence, we focus on the S_a and S_b tests in this paper rather than the F test itself.

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

3.2 $m \leq d$ Case

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

$$F = \frac{\left(\sum_{i=1}^{n+m} (Y_i - X_i'\widehat{\beta}_{1,n+m})^2 - \sum_{i=1}^n (Y_i - X_i'\widehat{\beta}_{1,n})^2\right) / m}{\sum_{i=1}^n (Y_i - X_i'\widehat{\beta}_{1,n})^2 / (n-d)}. \quad (3.13)$$

The second sum in the numerator is the unrestricted sum of squared residuals because the residuals from the last m observations equal zero due to the perfect fit that is possible when $m \leq d$.³

As above, the denominator of the F statistic does not contribute to the power of the test. It just contributes to the test's invariance with respect to the error variance σ_0^2 . In addition, the estimator $\widehat{\beta}_{1,n+m}$ differs by little from $\widehat{\beta}_{1,n}$ under the null and alternative hypotheses because n is much larger than m . Hence, we evaluate both sums of squares at $\widehat{\beta}_{1,n}$, which simplifies the statistic considerably. This leads to the following variant of the F statistic:

$$\begin{aligned} P_a &= \sum_{i=n+1}^{n+m} (Y_i - X_i' \widehat{\beta}_{1,n})^2 \\ &= P_{n+1}(\widehat{\beta}_{1,n}, I_m), \end{aligned} \tag{3.14}$$

where $P_j(\beta, \Sigma)$ is defined in (2.12). (The statistic is denoted P_a because statistics of this form are sometimes called *predictive* statistics, e.g., see Chow (1960).)

We define S_a to equal P_a when $m \leq d$. Of course, P_a is defined even when $m > d$. In this case, P_a differs from S_a and P_a is not a variant of the F statistic. Note that when $m = d$ the two definitions of S_a for $m \leq d$ and $m \geq d$ coincide by (3.12). Similarly, the two definitions of the F statistic for $m \geq d$ and $m \leq d$, given in (3.5) and (3.13), coincide when $m = d$.

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

$$P_{a,j} = P_j(\widehat{\beta}_{(j)}, I_m) \text{ for } j = 1, \dots, n - m + 1, \tag{3.15}$$

where $\widehat{\beta}_{(j)}$ is defined in (3.9). Critical values and p -values for the P_a statistic are obtained using $\{P_{a,j} : j = 1, \dots, n - m + 1\}$ in the same manner as in (2.3)-(2.5) with $(S_d, S_{d,j})$ replaced by $(P_a, P_{a,j})$.

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

$$P_b = P_{n+1}(\widehat{\beta}_{1,n+m}, I_m) \text{ and } P_{b,j} = P_j(\widehat{\beta}_{2(j)}, I_m). \tag{3.16}$$

Critical values and p -values for P_b are constructed in the same way as for S_d in (2.3)-(2.5).

When $m \leq d$, the S_b and $S_{b,j}$ statistics are defined to equal the P_b and $P_{b,j}$ statistics.

3.3 Tests Adjusted for Autocorrelation

The S_a , S_b , P_a , and P_b tests are designed for the case where the errors in the regression model are uncorrelated—although the tests have the correct size asymptotically whether or not the errors are correlated. If the errors are correlated, it

is advantageous in terms of power to include weights in the statistics based on an estimator of the error covariance matrix. We introduce statistics $(S_c, S_{c,j})$ that are analogues of $(S_a, S_{a,j})$ that do so. We define

$$\begin{aligned} S_c &= S_{n+1}(\widehat{\beta}_{1,n}, \widehat{\Sigma}_{1,n+m}) \text{ and} \\ S_{c,j} &= S_j(\widehat{\beta}_{(j)}, \widehat{\Sigma}_{1,n+m}) \end{aligned} \quad (3.17)$$

for $j = 1, \dots, n+1$, where $\widehat{\Sigma}_{1,n+m}$ and $S_j(\beta, \Sigma)$ are defined in (2.8) and (2.10), respectively. Critical values and p -values for S_c are obtained using $\{S_{c,j} : j = 1, \dots, n-m+1\}$ as in (2.3)-(2.5).

As with the S_a test, the S_c test tends to reject the null hypothesis too often when the null is true. Hence, we introduce an analogue of the S_c test that rejects less frequently under the null hypothesis. This is the S_d test defined in Section 2. The finite sample adjustment that is employed by the S_d test is the same as that for S_b , which is described above. The S_d test is the recommended test.

Next, for the case where $m \leq d$, we introduce statistics $(P_c, P_{c,j})$ that are analogous to $(P_a, P_{a,j})$, but with weights that take account of possible correlation in the errors:

$$\begin{aligned} P_c &= P_{n+1}(\widehat{\beta}_{1,n}, \widehat{\Sigma}_{1,n+m}) \text{ and} \\ P_{c,j} &= P_j(\widehat{\beta}_{(j)}, \widehat{\Sigma}_{1,n+m}) \end{aligned} \quad (3.18)$$

for $j = 1, \dots, n+1$, where $\widehat{\Sigma}_{1,n+m}$ and $P_j(\beta, \Sigma)$ are defined in (2.8) and (2.12), respectively. Critical values and p -values for P_c are obtained using $\{P_{c,j} : j = 1, \dots, n-m+1\}$ as in (2.3)-(2.5).

As with the S_c test, the P_c test tends to reject the null hypothesis too often. In consequence, we introduce analogues of the $(P_c, P_{c,j})$ statistics that reject the null hypothesis less frequently. Let

$$\begin{aligned} P_d &= P_{n+1}(\widehat{\beta}_{1,n+m}, \widehat{\Sigma}_{1,n+m}) \text{ and} \\ P_{d,j} &= P_j(\widehat{\beta}_{2(j)}, \widehat{\Sigma}_{1,n+m}). \end{aligned} \quad (3.19)$$

Critical values and p -values for the P_d statistic are obtained using $\{P_{d,j} : j = 1, \dots, n-m+1\}$ as in (2.3)-(2.5).

The S_d test when $m \leq d$, defined in Section 2, is the same as the P_d test. When $m > d$, the S_d and P_d tests differ.

3.4 Changes in the Regressor and Error Distributions

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

As it turns out, this is not a problem. The P_v tests for $v = a, b, c$, and d have no power asymptotically against changes in the regressor distribution because the test

statistics depend only on the residuals for $i = n + 1, \dots, n + m$. By definition, the same is true for the S_v tests when $m \leq d$. When $m \geq d$, the S_v tests for $v = a, b, c$, and d have no power against location and/or scale changes in the regressor distribution. Furthermore, Monte Carlo simulations show that changes in the shape of the regressor distribution beyond location and scale changes have very little effect on the rejection rates of the S_v tests when the parameters are constant and the error distribution is constant, see Section 9.2.3. Hence, the S_v tests appear to have little to no power against changes just in the regressor distribution.

On the other hand, the tests P_a and P_b have power against changes in the error distribution that increase the $1 - \alpha$ quantile of the distribution of the average of the squared errors $\{U_i^2 : i = n + 1, \dots, n + m\}$. Similarly, when $m \geq d$, the tests S_a and S_b have power against changes in the error distribution that increase the $1 - \alpha$ quantile of the distribution of the average of the squared errors after projection onto the column space spanned by the post-change regressors \mathbf{X}_{n+} . For example, a sufficiently large increase in the variance of the errors causes these tests to reject the null hypothesis.

The tests P_c and P_d have power against changes in the error distribution that increase the $1 - \alpha$ quantile of the distribution of the quadratic form $\mathbf{U}_{n+}' \Sigma_0^{-1} \mathbf{U}_{n+}$ in the errors, where $\Sigma_0 = E\mathbf{U}_{1,m} \mathbf{U}_{1,m}'$. The same is true of S_c and S_d when $m \leq d$. When $m \geq d$, the tests S_c and S_d have power against changes in the error distribution that increase the $1 - \alpha$ quantile of the distribution of the quadratic form in $\Sigma_0^{-1/2} \mathbf{U}_{n+}$ after projection onto the column space spanned by the transformed post-change regressors $\Sigma_0^{-1/2} \mathbf{X}_{n+}$.

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

4 Linear Instrumental Variables

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

The model of interest is as in (2.1), but with regressors that may be correlated with the errors. Let Z_i denote a vector of IVs for $i = 1, \dots, n + m$. The two cases distinguished in the previous section, namely, $m \geq d$ and $m \leq d$, also arise here, but the distinction depends on the dimension of the IV vector Z_i , rather than the dimension of the regressor vector X_i . Hence, in the linear IV case, we let d denote the dimension of the IV vector Z_i and we let d_X denote the dimension of the regressor vector X_i and the parameter β . We assume that $d \geq d_X$.

The null and alternative hypotheses of interest are as in (3.4) but with the LS moments, $E(Y_i - X_i' \beta_0) X_i$, replaced by the IV moments $E(Y_i - X_i' \beta_0) Z_i$. Thus,

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

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

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

The main assumptions are that $\{W_i = (Y_i, X_i', Z_i')' : i \geq 1\}$ are stationary and ergodic under the null hypothesis, $EU_1 Z_1 = 0$, $EZ_1 Z_1'$ is positive definite, $EX_1 Z_1'$ has full row rank, U_1 has an absolutely continuous distribution, and some moment conditions hold, see Section 8.

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

$$\begin{aligned} \hat{\beta}_{r,s} &= (\mathbf{X}'_{r,s} \mathbf{P}_{\mathbf{Z}_{r,s}} \mathbf{X}_{r,s})^{-1} \mathbf{X}'_{r,s} \mathbf{P}_{\mathbf{Z}_{r,s}} \mathbf{Y}_{r,s}, \text{ where} \\ \mathbf{Z}_{r,s} &= (Z_r, \dots, Z_s)' \text{ and} \\ \mathbf{P}_{\mathbf{Z}_{r,s}} &= \mathbf{Z}_{r,s} (\mathbf{Z}'_{r,s} \mathbf{Z}_{r,s})^{-1} \mathbf{Z}'_{r,s}. \end{aligned} \quad (4.2)$$

For $j = 1, \dots, n - m + 1$, the IV estimators $\hat{\beta}_{(j)}$ and $\hat{\beta}_{2(j)}$ are defined analogously using the observations indexed by $i = 1, \dots, n$ with $i \neq j, \dots, j + m - 1$ and by $i = 1, \dots, n$ with $i \neq j, \dots, j + \lceil m/2 \rceil - 1$, respectively.

In the IV case, the function $S_j(\beta, \Sigma)$ is as defined in (2.10), but with $A_j(\beta, \Sigma)$ and $V_j(\Sigma)$ defined in (2.13).

Apart from the changes in the definitions of the estimators and the function $S_j(\beta, \Sigma)$, the statistics $(T_v, T_{v,j})$ in the IV case are the same as in the LS regression case.

5 Generalized Method of Moments

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

We consider GMM moment conditions given by

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

where $g(\cdot, \cdot)$ is a vector-valued function that may be, but is not necessarily, of the form given in (2.14). The two cases distinguished in the linear regression section, namely, $m \geq d$ and $m \leq d$, also arise here, but the distinction depends on the number of moments, rather than the dimension of X_i . Hence, in 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

$$\begin{aligned} H_0 : & \begin{cases} Eg(W_i, \beta_0) = 0 \text{ for all } i = 1, \dots, n+m \text{ and} \\ \{W_i : i \geq 1\} \text{ are stationary and ergodic} \end{cases} \\ H_1 : & \begin{cases} Eg(W_i, \beta_0) = 0 \text{ for all } i = 1, \dots, n \text{ and} \\ Eg(W_i, \beta_0) \neq 0 \text{ for some } i = n+1, \dots, n+m \text{ and/or} \\ \text{the distribution of } g(W_i, \beta_0) \text{ for some } i = n+1, \dots, n+m \\ \text{differs from that of } g(W_i, \beta_0) \text{ for } i = 1, \dots, n. \end{cases} \end{aligned} \quad (5.2)$$

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

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

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

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

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

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

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

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

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

For example, suppose $g(W_i, \beta)$ is of the form in (2.14). Then, one can take

$$\begin{aligned}
V_j(\beta) &= \mathbf{Z}_{j,j+m-1}(\beta)' \mathbf{Z}_{j,j+m-1}(\beta), \text{ where} \\
\mathbf{Z}_{j,j+m-1}(\beta) &= (Z(W_j, \beta), \dots, Z(W_{j+m-1}, \beta))'.
\end{aligned} \tag{5.5}$$

Alternatively, one could take

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

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

The statistics $(S_b, S_{b,j})$ are defined just as $(S_a, S_{a,j})$ are defined in (5.4) when $m \geq d$, but with $(\widehat{\beta}_{1,n+m}, \widehat{\beta}_{2(j)})$ in place of $(\widehat{\beta}_{1,n}, \widehat{\beta}_{(j)})$.

The remaining statistics for the GMM case are only defined when $g(\cdot, \cdot)$ is of the form in (2.14). The $(S_a, S_{a,j})$ and $(S_b, S_{b,j})$ statistics when $m \leq d$ are defined to equal $(P_a, P_{a,j})$ and $(P_b, P_{b,j})$, respectively. The latter are defined (whether or not $m \leq d$) by

$$\begin{aligned}
P_a &= P_{n+1}(\widehat{\beta}_{1,n}, I_m), \quad P_{a,j} = P_j(\widehat{\beta}_{(j)}, I_m), \\
P_b &= P_{n+1}(\widehat{\beta}_{1,n+m}, I_m), \quad P_{b,j} = P_j(\widehat{\beta}_{2(j)}, I_m),
\end{aligned} \tag{5.7}$$

where $P_j(\beta, \Sigma)$ is defined in (2.17).

Again assuming that $g(\cdot, \cdot)$ is of the form in (2.14), we define $(S_c, S_{c,j})$ in the same way that $(S_d, S_{d,j})$ is defined in (2.14)-(2.17), but with $(\widehat{\beta}_{1,n}, \widehat{\beta}_{(j)})$ in place of $(\widehat{\beta}_{1,n+m}, \widehat{\beta}_{2(j)})$.

Lastly, $(P_c, P_{c,j})$ and $(P_d, P_{d,j})$ are defined by

$$\begin{aligned}
P_c &= P_{n+1}(\widehat{\beta}_{1,n}, \widehat{\Sigma}_{1,n+m}(\widehat{\beta}_{1,n})), \quad P_{c,j} = P_j(\widehat{\beta}_{(j)}, \widehat{\Sigma}_{1,n+m}(\widehat{\beta}_{(j)})), \\
P_d &= P_{n+1}(\widehat{\beta}_{1,n+m}, \widehat{\Sigma}_{1,n+m}(\widehat{\beta}_{1,n+m})), \quad P_{d,j} = P_j(\widehat{\beta}_{2(j)}, \widehat{\Sigma}_{1,n+m}(\widehat{\beta}_{2(j)})),
\end{aligned} \tag{5.8}$$

where $\widehat{\Sigma}_{1,n+m}(\beta)$ and $P_j(\beta, \Sigma)$ are defined in (2.15) and (2.17), respectively.

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

The tests introduced above for detecting instability at the end of the sample can be altered to detect instability occurring at the beginning or in the middle of

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

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

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

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

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

7 Application to Models with I(1) Variables

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

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

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

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

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 (7.2).

8 Asymptotic Results

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

8.1 Assumptions

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

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

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 non-redundant elements of Σ^{-1} .

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

Assumption 2. (a) When $v = a$ or c , $\|\widehat{\beta}_{1,n} - \beta_0\| \rightarrow_p 0$ and $\sup_{j=1, \dots, n-m+1} \|\widehat{\beta}_{(j)} - \beta_0\| \rightarrow_p 0$ with m fixed under H_0 and H_1 . When $v = b$ or d , $\|\widehat{\beta}_{1,n+m} - \beta_0\| \rightarrow_p 0$ and $\sup_{j=1, \dots, n-m+1} \|\widehat{\beta}_{2(j)} - \beta_0\| \rightarrow_p 0$ $n \rightarrow \infty$ with m fixed under H_0 and H_1 .

(b) When $v = c$ or d , $\sup_{\beta \in B(\beta_0, \varepsilon_n)} \|\widehat{\Sigma}_{1,n+m}(\beta) - \Sigma_0\| \rightarrow_p 0$ as $n \rightarrow \infty$ for some nonsingular matrix Σ_0 , for all sequences of constants $\{\varepsilon_n : n \geq 1\}$ for which $\varepsilon_n \rightarrow 0$ as $n \rightarrow \infty$.

Assumption 3. (a) When $v = a$ or b , $T_{n+1}(\beta, I_m)$ is continuously differentiable in a neighborhood of β_0 with probability one under H_0 and H_1 . When $v = c$ or d ,

$T_{n+1}(\beta, \Sigma)$ 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).

(b) When $v = a$ or b , either $E \sup_{\beta \in B(\beta_0, \varepsilon)} \|(\partial/\partial\beta)T_1(\beta, I_m)\| < \infty$ or $(n - m + 1)^{-1} \sum_{j=1}^{n-m+1} \sup_{\beta \in B(\beta_0, \varepsilon)} \|(\partial/\partial\beta)T_j(\beta, I_m)\| = O_p(1)$ for some $\varepsilon > 0$. When $v = c$ or d , the same condition holds with I_m replaced by Σ and with the suprema taken over $\Sigma \in N(\Sigma_0)$, where Σ_0 is as in Assumption 2(b) and $N(\Sigma_0)$ denotes some neighborhood of Σ_0 .

(c) When $v = a$ or b , the distribution function of $T_1(\beta_0, I_m)$ is continuous and increasing at its $1 - \alpha$ quantile. When $v = c$ or d , the same condition holds with I_m replaced by Σ_0 , where Σ_0 is as in Assumption 2(b).

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

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

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

(b) $EU_1^2 < \infty$ and $E\|X_1\|^{2+\delta} < \infty$ for some $\delta > 0$.

(c) EX_1X_1' and $\Sigma_0 = EU_{1,m}U_{1,m}'$ are positive definite.

(d) U_1 has an absolutely continuous distribution.

Assumption IV. (a) $EU_1Z_1 = 0$.

(b) $EU_1^2 < \infty$, $E\|X_1\|^2 < \infty$, and $E\|Z_1\|^{2+\delta} < \infty$ for some $\delta > 0$.

(c) EZ_1Z_1' and $\Sigma_0 = EU_{1,m}U_{1,m}'$ are positive definite and EX_1Z_1' has full row rank.

(d) U_1 has an absolutely continuous distribution.

Assumption GMM. (a) $Eg(W_1, \beta) = 0$ for $\beta \in \mathcal{B}$ if and only if $\beta = \beta_0 \in \mathcal{B}$.

(b) \mathcal{B} is compact.

(c) $g(W_1, \beta)$ is continuous on \mathcal{B} almost surely and $Eg(W_1, \beta)$ is continuous on \mathcal{B} .

(d) $E \sup_{\beta \in \mathcal{B}} \|g(W_1, \beta)\|^{1+\delta} < \infty$ for some $\delta > 0$.

(e) The one-step GMM weight matrix \mathcal{V} is non-stochastic and positive definite; the two-step GMM weight matrix functions $\mathcal{V}_{(j)}(\beta)$ and $\mathcal{V}_{r,s}(\beta)$ satisfy $\sup_{j=1, \dots, n-m+1} \sup_{\beta \in B(\beta_0, \varepsilon)} |\mathcal{V}_{(j)}(\beta) - \mathcal{V}(\beta)| \rightarrow_p 0$ and $\sup_{\beta \in B(\beta_0, \varepsilon)} |\mathcal{V}_{r,s}(\beta) - \mathcal{V}(\beta)| \rightarrow_p 0$ for some $\varepsilon > 0$ for $(r, s) = (1, n)$ and $(r, s) = (1, n + m)$, for some symmetric positive definite non-stochastic function $\mathcal{V}(\beta)$ defined on $B(\beta_0, \varepsilon)$ that is continuous at β_0 ; and the CU weight matrix functions $\mathcal{V}_{(j)}(\beta)$ and $\mathcal{V}_{r,s}(\beta)$ satisfy analogous convergence conditions but with $B(\beta_0, \varepsilon)$ replaced by \mathcal{B} and with $\mathcal{V}(\beta)$ being a symmetric non-stochastic function with eigenvalues bounded away from zero on \mathcal{B} .

(f) When $T = S$, $g(W_1, \beta)$ is continuously differentiable on a neighborhood of β_0 almost surely. When $T = S$ and $v = a$ or b , $V_j(\beta) = V(W_j, \dots, W_{j+m-1}, \beta)$ is a positive definite weight matrix that is a function of the observations W_j, \dots, W_{j+m-1} and the parameter β and is continuously differentiable in β on a neighborhood of β_0 almost surely for $j = 1, \dots, n + 1$, $E \sup_{\beta \in B(\beta_0, \varepsilon)} \|g(W_1, \beta)\|^2 < \infty$, $E \sup_{\beta \in B(\beta_0, \varepsilon)} (\|(\partial/\partial\beta')g(W_1, \beta)\| \cdot \|g(W_1, \beta)\|) < \infty$, and $\sup_{j=1, \dots, n-m+1} \sup_{\beta \in B(\beta_0, \varepsilon)} (\|V_j^{-1}(\beta)\| +$

$\|(\partial/\partial\beta_r)V_j^{-1}(\beta)\| = O_p(1)$ for $r = 1, \dots, d_\beta$ for some $\varepsilon > 0$. When $T = S$ and $v = c$ or d , the previous conditions 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 r -th element of the vector comprised of β and the non-redundant elements of Σ^{-1} . When $T = P$ and/or $v = c$ or d , $U(W_1, \beta)$ is continuously differentiable on a neighborhood of β_0 almost surely and $E \sup_{\beta \in B(\beta_0, \varepsilon)} \|U(W_1, \beta)(\partial/\partial\beta)U(W_1, \beta)\| < \infty$. When $v = c$ or d , $E\mathbf{U}_{1,m}\mathbf{U}'_{1,m}$ is positive definite and $E \sup_{\beta \in B(\beta_0, \varepsilon)} U^2(W_1, \beta) < \infty$. When $T = S$ and $v = c$ or d , $E \sup_{\beta \in B(\beta_0, \varepsilon)} \|Z(W_1, \beta)\|^2 < \infty$.

(g) $g(W_1, \beta_0)$ has an absolutely continuous distribution when $T = S$ and $v = a$ or b and $U(W_1, \beta_0)$ has an absolutely continuous distribution when $T = S$ and $v = c$ or d and when $T = P$.

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 $\widehat{\beta}_{1,n+m}$ and $\widehat{\Sigma}_{1,n+m}$ and their behavior is dominated by the pre-change observations and in Assumption 3 the post-change observations only affect $T_{n+1}(\beta_0, I_m)$ and $T_{n+1}(\beta_0, \Sigma_0)$ and whether they are continuously differentiable does not depend on the distribution of the observations.

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

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 used to verify Assumption 3(c).

Lemma 1 (a) *Assumptions 1 and LS imply that Assumptions 2 and 3 hold for $T = S$ and P and $v = a, b, c$, and d for the linear regression model estimated using the LS estimator.*

(b) *Assumptions 1 and IV imply that Assumptions 2 and 3 hold for $T = S$ and P and $v = a, b, c$, and d for the IV regression model estimated using the IV estimator.*

(c) *Assumptions 1 and GMM imply Assumptions 2 and 3 hold for $T = S$ and P and $v = a, b, c$, and d for the moment condition model estimated using a GMM estimator.*

8.2 Results

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

For $T = S$ and P and $v = a, b, c$, and d , let $\widehat{F}_{T_v, n}(x)$ denote the empirical df

based on $\{\widehat{T}_{v,j} : j = 1, \dots, n - m + 1\}$. That is,

$$\widehat{F}_{T_v,n}(x) = \frac{1}{n - m + 1} \sum_{j=1}^{n-m+1} 1(T_{v,j} \leq x). \quad (8.1)$$

For $T = S$ and P , let $F_{T_v}(x)$ denote the df of $T_1(\beta_0, I_m)$ at x when $v = a$ and b and the df of $T_1(\beta_0, \Sigma_0)$ at x when $v = c$ and d . Let $q_{T_v,1-\alpha}$ denote the $1 - \alpha$ quantile of $T_1(\beta_0, I_m)$ when $v = a$ and b and the $1 - \alpha$ quantile of $T_1(\beta_0, \Sigma_0)$ when $v = c$ and d . Let $\widehat{q}_{T_v,1-\alpha}$ denote the $1 - \alpha$ sample quantile of $\{T_{v,j} : j = 1, \dots, n - m + 1\}$ for $v = a, b, c$, and d , as defined in (2.3).

For $T = S$ and P , let $T_{v,\infty}$ be a random variable with the same distribution as $T_{n+1}(\beta_0, I_m)$ when $v = a$ and b and the same distribution as $T_{n+1}(\beta_0, \Sigma_0)$ when $v = c$ and d . Under Assumptions 1-3 and H_0 , the distributions of $T_{n+1}(\beta_0, I_m)$ and $T_{n+1}(\beta_0, \Sigma_0)$ equal those of $T_1(\beta_0, I_m)$ and $T_1(\beta_0, \Sigma_0)$, respectively. Also, the distributions of $T_{n+1}(\beta_0, I_m)$ and $T_{n+1}(\beta_0, \Sigma_0)$ 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 $\{W_{n,i} : i = n + 1, \dots, n + m\}$ to be independent of n , which is appropriate for fixed alternatives.

Theorem 1 *Suppose Assumptions 1-3 hold for $T = S$ or P and $v = a, b, c$, or d . Then, as $n \rightarrow \infty$,*

- (a) $T_v \rightarrow_d T_{v,\infty}$ under H_0 and H_1 ,
- (b) $\widehat{F}_{T_v,n}(x) \rightarrow_p F_{T_v}(x)$ for all x in a neighborhood of $q_{T_v,1-\alpha}$ under H_0 and H_1 ,
- (c) $\widehat{q}_{T_v,1-\alpha} \rightarrow_p q_{T_v,1-\alpha}$ under H_0 and H_1 , and
- (d) $P(T_v > \widehat{q}_{T_v,1-\alpha}) \rightarrow \alpha$ under H_0 .

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

2. Part (c) of the Theorem shows that the random critical value $\widehat{q}_{T_v,1-\alpha}$ 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 T_v does not diverge to infinity as $n \rightarrow \infty$ under H_1 . Hence, T_v is not a consistent test. However, if $T_{n+1}(\beta_0, I_m)$ is stochastically greater than $T_1(\beta_0, I_m)$ (or $T_{n+1}(\beta_0, \Sigma_0)$ is stochastically greater than $T_1(\beta_0, \Sigma_0)$) under H_1 , then T_v 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 P_v tests for $v = a, b, c$, and d have the correct size asymptotically. To verify Assumption 2, one could use near epoch dependence (NED) or mixing conditions in place of stationarity and ergodicity. We use the stationarity and ergodicity condition here because it allows for more general dependence, such as long-memory dependence, and is simpler and more elegant than NED or mixing conditions.

5. The idea of the proof of part (b) of the Theorem is to show that (i) the difference between $\widehat{F}_{T_v,n}(x)$ and a smoothed version of it, say, $\widehat{F}_{T_v,n}(x, h_n)$ converges in probability to zero, where h_n indexes the amount of smoothing and $h_n \rightarrow 0$ as $n \rightarrow \infty$, (ii) the difference between $\widehat{F}_{T_v,n}(x, h_n)$ and an analogous df with $\widehat{\beta}_{(j)}$ or $\widehat{\beta}_{2(j)}$ replaced by β_0 converges in probability to zero, (iii) the difference between the latter and the empirical df of $\{T_j(\beta_0, I_m) : j = 1, \dots, n-m+1\}$ or $\{T_j(\beta_0, \Sigma_0) : j = 1, \dots, n-m+1\}$ converges in probability to zero as $n \rightarrow \infty$, and (iv) the difference between the latter and its expectation, $F_{T_v}(x)$, is asymptotically negligible. The reason for considering a smoothed version of $\widehat{F}_{T_v,n}(x)$ is that it is a smooth function of $\widehat{\beta}_{(j)}$ or $\widehat{\beta}_{2(j)}$ and, hence, result (ii) can be established by taking a mean-value expansion about β_0 . Result (iv) holds by the ergodic theorem because $\{T_j(\beta_0, I_m) : j = 1, \dots, n-m+1\}$ and $\{T_j(\beta_0, \Sigma_0) : j = 1, \dots, n-m+1\}$ are finite subsets of stationary and ergodic random variables using Assumption 1.

9 Monte Carlo Experiment

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

9.1 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 $[-\sqrt{12}, \sqrt{12}]$ (which has mean zero and variance one). (Note that the test size 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 test size 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 size-corrected tests because we do not want to confound power differences with size distortions (which are quite large for the F test in some scenarios). For the F test, size correction is straightforward. By simulation we determine critical values that yield the desired test size, .05, for each distribution and each n , m , and ρ value when $\beta_0 = \beta_{1i} = 0$. These critical values are employed when computing the power of the F test.

For the other tests, size-correction is not as straightforward because their critical values are sample quantiles, not constants. For the other tests, we determine by simulation the significance levels that yield the finite sample null rejection rates to be as close to the desired test size, .05, as possible for each distribution and each n , m , and ρ value when $\beta_0 = \beta_{1i} = 0$. (The rejection rates 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 nominal .05 tests. Note that this method of size correction is equivalent to the method in which the critical value is adjusted for any test that has a non-random 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 null rejection rates of nominal .05 tests and simulated standard errors in the interval (.0020, .0025) for the simulated alternative hypothesis rejection rates when these rejection rates are in the interval (.20, .80).

9.2 Monte Carlo Results

9.2.1 Size

Table I presents the test size results for nominal .05 tests. For clarity, the results for the recommended test S_d are in bold face in Table I.

When $m = 5$, separate results are not given for P_b and P_d because $P_b = S_b$ and $P_d = S_d$ by definition. When $m = 1$, separate results are not given for S_a , S_b , P_a , P_b , P_c , or P_d because $S_a = P_a = P_c = S_c$ and $S_b = P_b = P_d = S_d$ when $m = 1$. The test P_b dominates P_a and the test P_d dominates P_c in terms of size performance across all cases in Table I. Hence, for brevity, we do not report any results for P_a or P_c in Table I.

The main results are as follows:

1. The F test has (exactly) correct size for all values of (n, m) when the distribution is normal and $\rho = 0$. The F test also has fairly good size when $m = 1$ and the distribution is normal, χ^2 , or t_3 .
2. In most other cases, the size of the F test is poor and, in some cases, it is very poor. Across all of the cases considered, the rejection rate of the F test varies between .002 and .329. The standard deviation of the rejection rate 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 rejection probability is .102; while for $\rho = 0$, $m = 5$, $n = 250$, and uniform distribution, the rejection probability 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 rejection probability is .115. For the same case except with $\rho = .8$, the rejection rate 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 size asymptotically in these cases.

3. The S_a and S_c tests perform very much better than the F test in an overall sense. For example, the standard deviation of the rejection rate of the S_c test from the desired value .05 across the 72 cases in Table I is .018, compared to .095 for the F test. But, these tests still tend to over-reject the null hypothesis. Their rejection rates vary between .050 and .090 when $\rho = 0$ or $.4$ and between .060 and .118 when $\rho = .8$.
4. The S_b and S_d tests, which incorporate the finite sample adjustments described above, perform the best in terms of size in an overall sense. Their rejection rates vary between .034 and .072 when $\rho = 0$ or $.4$ and between .049 and .082 when $\rho = .8$. The standard deviations of the rejection rates for the S_b and S_d tests from the desired value .05 are .011 and .008, respectively, compared to .018 for the S_c test. Hence, the finite sample adjustments work well. Except for the case of $m = 1$ and uniform distribution, the rejection rates of the S_d test vary between .040 and .058 when $\rho = 0$ or $.4$ and between .049 and .073 when $\rho = .8$. The S_d test has better size when $m = 10$ or 5 than when $m = 1$. When $m = 10$ or 5 , its rejection rate varies between .040 and .064.
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 often 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_d test, 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 other 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 size results of Table I show that the F test performs poorly in many of the cases considered. The T_v tests for $v = a$ and c greatly outperform the F test in an overall sense, but still over-reject the null hypothesis. The S_b and S_d tests perform best and their performance in an absolute sense is pretty good.

9.2.2 Power

Next, we consider Table II which provides the size-corrected power results. Table II does not report power results for the S_a and S_c tests because they are quite similar to those of the S_b and S_d tests, respectively. The principle findings are as follows:

1. The power of the S_b and S_d tests when $\rho = 0$ is essentially the same. Hence, the estimation of the error covariance matrix by S_d costs little when the errors are uncorrelated. On the other hand, when $\rho = .4$ or $.8$, the S_d test is more powerful than the S_b test and the estimation of the error covariance matrix pays dividends. When $\rho = .4$, the S_d test is 5.9% more powerful on average than the S_b test (over the cases in which they differ). When $\rho = .8$, it is 61.1% more powerful on average. Hence, the S_d test outperforms the S_b test in terms of power.
2. The S_d test is more powerful than the P_d 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.
3. The F test is 4.1% more powerful than the S_d test on average when $\rho = 0$. The F test is 1.8% more powerful on average when $\rho = .4$. The F and S_d tests have essentially the same power when $\rho = .8$ and $m = 1$. The S_d test is 51.3% more powerful than the F test when $\rho = .8$ and $m = 10$ or 5. Hence, the S_d 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_d test because of its size distortions.
4. 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 size and power results discussed above, we recommend using the S_d test. The S_d and S_b tests have the best size performance and the power performance of the S_d test is better than that of S_b . Not surprisingly, the size performance of the S_d test is far superior to that of the F test. In addition, the size-corrected power of

the S_d test is close to that of the F test when the errors are uncorrelated and better when the errors have noticeable correlation.

The finite sample size adjustment of the S_d test, which distinguishes it from the S_c test, works quite well. The estimation of the error covariance matrix, which distinguishes the S_d test from the S_b test, improves power when the errors are correlated without sacrificing power when the errors are uncorrelated and without sacrificing size performance. The use of a weight matrix that projects onto the column space of the transformed post-change regressors, rather than an identity weight matrix, which distinguishes the S_d test from the P_d test (when $m = 10$), is found to increase power.

9.2.3 Change in Regressor Distribution

We carry out some simulations to see whether a change in the regressor distribution alone causes the T_v tests to reject the null hypothesis more frequently than when there is no change in the regressor distribution, the parameters, or the error distribution. Six cases are considered. In each case, the pre-change regressor innovation distribution is $N(0, 1)$, recentered and rescaled χ_2^2 , or $U[-\sqrt{12}, \sqrt{12}]$ and the post-change regressor innovation distribution is one of these three distributions but a different one. We consider the same values of n , m , and ρ as above.

The rejection rates of the T_v tests in the above cases are always within .006 of their rejection rates for the corresponding cases that have the same pre-change regressor innovation distributions and no change in this distribution after $i = n$. This indicates that the T_v tests do not reject the null with probability greater than .05 when the only instability present is instability in the regressor distribution. This is a desirable feature of the tests.

10 Appendix of Proofs

Proof of Theorem 1. We start by proving parts (a)-(d) for $T = S$ and $v = a$.

We prove part (a) first. By Assumption 2(a), $\widehat{\beta}_{1,n} \rightarrow_p \beta_0$. In consequence, there exists a sequence of non-negative constants $\{\varepsilon_n : n \geq 1\}$ for which $\varepsilon_n \rightarrow 0$ and $P(K_n) \rightarrow 1$, where $K_n = \{\|\widehat{\beta}_{1,n} - \beta_0\| < \varepsilon_n\}$. Let $x \in R$ be a continuity point of the df of $S_{n+1}(\beta_0)$, defined in (3.7). Let K_n^c denote the complement of the set K_n . We have

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

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

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

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

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

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

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

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

We have

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

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

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

We have

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

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

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

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

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

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

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

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

where $k'(\cdot)$ denotes the derivative of $k(\cdot)$, $\tilde{\beta}_{(j)}$ lies between $\hat{\beta}_{(j)}$ and β_0 , $B < \infty$ denotes the bound on the derivative of $k(\cdot)$, the inequality holds with probability that goes to one because $\sup_{j=1, \dots, n-m+1} \|\hat{\beta}_{(j)} - \beta_0\| < \varepsilon$ 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 $\sup_{r=1, \dots, n-m+1} \|\hat{\beta}_{(r)} - \beta_0\|/h_n \rightarrow_p 0$. 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 $\varepsilon > 0$, $\hat{F}_{S_a, n}(q_{S_a, 1-\alpha} - \varepsilon) \rightarrow_p F_{S_a}(q_{S_a, 1-\alpha} - \varepsilon) < 1 - \alpha$ and $\hat{F}_{S_a, n}(q_{S_a, 1-\alpha} + \varepsilon) \rightarrow_p F_{S_a}(q_{S_a, 1-\alpha} + \varepsilon) > 1 - \alpha$.

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

This completes the proof for the case where $T = S$ and $v = a$.

The proof for $T = P$ and $v = a$ is essentially the same and, hence, is not given. The proof for $v = b$ is essentially the same as that for $v = a$ because Assumption 2(a) implies that the estimators $\hat{\beta}_{1, n+m}$ and $\hat{\beta}_{2(j)}$ behave like $\hat{\beta}_{1, n}$ and $\hat{\beta}_{(j)}$ asymptotically.

The proofs for the cases where $v = c$ and d are essentially the same as that for $v = a$, but with $S_j(\beta)$, β , β_0 , $\hat{\beta}_{1, n}$, $\hat{\beta}_{(j)}$, and β_j replaced by $S_j(\beta, \Sigma)$ and the vectors comprised of the non-redundant elements of (β, Σ) , (β_0, Σ_0) , $(\hat{\beta}_{1, n}, \hat{\Sigma}_{1, n+m})$, $(\hat{\beta}_{(j)}, \hat{\Sigma}_{1, n+m})$, and (β_j, Σ_j) , respectively, where Σ_j is some random or non-random $m \times m$ 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}) . Also, when $v = d$, $\hat{\beta}_{1, n}$ and $\hat{\beta}_{(j)}$ are replaced by $\hat{\beta}_{1, n+m}$ and $\hat{\beta}_{2(j)}$ respectively. \square

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

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

$$= \sup_{j=1, \dots, n-m+1} \|(n-m)^{-1} \sum_{i=j, \dots, j+m-1} \xi_i\| + o_p(1), \quad (10.8)$$

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

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

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

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

$$\begin{aligned} &\sup_{j=1, \dots, n-m+1} \|\widehat{\beta}_{(j)} - \beta_0\| \\ &= \sup_{j=1, \dots, n-m+1} \left\| (n-m)^{-1} \sum_{i=1, \dots, n; i \neq j, \dots, j+m-1} X_i X_i' \right\|^{-1} \\ &\quad \times (n-m)^{-1} \sum_{i=1, \dots, n; i \neq j, \dots, j+m-1} X_i U_i \\ &\leq \|(EX_1 X_1' + o_p(1))^{-1} (EX_1 U_1 + o_p(1))\| \\ &= O_p(1), \end{aligned} \quad (10.10)$$

where the inequality holds by applying (10.8) and (10.9) twice with $\xi_i = X_i X_i' - EX_i X_i'$ and $\xi_i = X_i U_i$ 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 (4.2). In this case, (10.8) and (10.9) are applied with $\xi_i = X_i Z_i'$, $\xi_i = Z_i Z_i'$, and $\xi_i = Z_i U_i$. (Note that $EU_1^2 < \infty$ and $E\|Z_1\|^{2+\delta} < \infty$ imply that $E\|U_1 Z_1\|^{1+\delta_1} < \infty$ for some $\delta_1 > 0$ by Hölder's inequality.)

The proof that $\sup_{j=1, \dots, n-m+1} \|\widehat{\beta}_{2(j)} - \beta_0\| \rightarrow_p 0$ for the LS and IV estimators is the same as the proof given above but with m replaced by $\lceil m/2 \rceil$ in the appropriate places.

The proof that $\|\widehat{\beta}_{r,s} - \beta_0\| \rightarrow_p 0$ under H_0 and H_1 for $(r, s) = (1, n)$ and $(r, s) = (1, n+m)$ for the LS and IV estimators is fairly standard and, hence, is not given. (Note that the proof for $\widehat{\beta}_{1, n+m}$ 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 consider $T = S$. For the LS estimator,

$$\frac{\partial}{\partial \beta} S_j(\beta, \Sigma) = -2\mathbf{X}'\Sigma^{-1}(\mathbf{Y} - \mathbf{X}\beta). \quad (10.11)$$

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,

$$\begin{aligned} & \frac{\partial}{\partial \omega_{k,\ell}} S_j(\beta, \Sigma) \\ &= 2(\mathbf{Y} - \mathbf{X}\beta)' \Sigma^{-1} \mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \frac{\partial}{\partial \omega_{k,\ell}} (\Sigma^{-1}) (\mathbf{Y} - \mathbf{X}\beta) \\ & \quad + (\mathbf{Y} - \mathbf{X}\beta)' \Sigma^{-1} \mathbf{X} \frac{\partial}{\partial \omega_{k,\ell}} [(\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1}] \mathbf{X}' \Sigma^{-1} (\mathbf{Y} - \mathbf{X}\beta). \end{aligned} \quad (10.12)$$

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,

$$\begin{aligned} \frac{\partial}{\partial \beta} S_j(\beta, \Sigma) &= -2\mathbf{X}'\Sigma^{-1}\mathbf{Z}(\mathbf{Z}'\Sigma^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\Sigma^{-1}(\mathbf{Y} - \mathbf{X}\beta) \\ &= -2\mathbf{X}'\Sigma^{-1/2}P_{\Sigma^{-1/2}\mathbf{Z}}\Sigma^{-1/2}(\mathbf{Y} - \mathbf{X}\beta), \end{aligned} \quad (10.13)$$

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

$$\begin{aligned} |\tilde{\mathbf{X}}'P_{\tilde{\mathbf{Z}}}(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\beta)| &\leq (\tilde{\mathbf{X}}'P_{\tilde{\mathbf{Z}}}\tilde{\mathbf{X}})^{1/2}((\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\beta)'P_{\tilde{\mathbf{Z}}}(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\beta))^{1/2} \\ &\leq (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{1/2}((\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\beta)'(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\beta))^{1/2}. \end{aligned} \quad (10.14)$$

The supremum of the rhs 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,

$$E \sup_{\beta \in B(\beta_0, \varepsilon), \Sigma \in N(\Sigma_0)} \left\| \frac{\partial}{\partial \beta} S_1(\beta, \Sigma) \right\| < \infty \quad (10.15)$$

for the IV estimator when $T = S$. Note that we apply the Cauchy-Schwarz inequality in (10.14) in order to eliminate the $(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}$ term that appears in the lhs, which cannot be bounded on its own.

An analogous result to (10.15) holds with $\partial/\partial \beta$ replaced by $\partial/\partial \omega_{k,\ell}$ by combining the calculations in (10.12) and (10.14). To bound the second term of (10.12), 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

$$|x'PAPx| \leq \sum_{r=1}^m \sum_{s=1}^m |a_{r,s}| \cdot \|Px\| \leq \sum_{r=1}^m \sum_{s=1}^m |a_{r,s}| \cdot \|x\|, \quad (10.16)$$

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

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

$$\begin{aligned} \frac{\partial}{\partial\beta} P_j(\beta, \Sigma) &= -2\mathbf{X}'\Sigma^{-1}(\mathbf{Y} - \mathbf{X}\beta) \text{ and} \\ \frac{\partial}{\partial\omega_{k,\ell}} P_j(\beta, \Sigma) &= (\mathbf{Y} - \mathbf{X}\beta)' \frac{\partial}{\partial\omega_{k,\ell}} (\Sigma^{-1})(\mathbf{Y} - \mathbf{X}\beta). \end{aligned} \quad (10.17)$$

The expectation of the supremum over $\beta \in B(\beta_0, \varepsilon)$ and $\Sigma \in N(\Sigma_0)$ of the rhs of second equation in (10.17) is finite because $EU_1^2 < \infty$ and $E\|X_1\|^2 < \infty$. Also, the rhs of the first equation in (10.17) is the same as in (10.11). Hence, Assumption 3(b) holds by the same argument as above.

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

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

$$\sup_{\beta \in \mathcal{B}} \sup_{j=1, \dots, n-m+1} |Q_{(j)}^{(k)}(\beta) - Q^{(k)}(\beta)| \rightarrow_p 0 \text{ and} \quad (10.18)$$

$$Q^{(k)}(\beta_0) < \inf_{\beta \notin B(\beta_0, \varepsilon) \cap \mathcal{B}} Q^{(k)}(\beta), \text{ where} \quad (10.19)$$

$$\begin{aligned} Q^{(1)}(\beta) &= Eg(W_1, \beta)' \mathcal{V}^{-1} Eg(W_1, \beta), \\ Q^{(2)}(\beta) &= Eg(W_1, \beta)' \mathcal{V}^{-1}(\beta_0) Eg(W_1, \beta), \text{ and} \\ Q^{(CU)}(\beta) &= Eg(W_1, \beta)' \mathcal{V}^{-1}(\beta) Eg(W_1, \beta). \end{aligned} \quad (10.20)$$

Condition (10.18) holds provided

$$\sup_{\beta \in \mathcal{B}} \sup_{j=1, \dots, n-m+1} |(n-m)^{-1} \sum_{i=1, \dots, n; i \neq j, \dots, j+m-1} g(W_i, \beta) - Eg(W_1, \beta)| \rightarrow_p 0, \quad (10.21)$$

because Assumption GMM(e) insures that the weight matrices are well-behaved. Equation (10.21) holds pointwise in β for all $\beta \in \mathcal{B}$ by applying (10.8) and (10.9) with $\xi_i = g(W_i, \beta) - Eg(W_1, \beta)$ using Assumptions 1 and GMM(d). Then, a generic uniform convergence result strengthens pointwise convergence to uniform convergence over $\beta \in \mathcal{B}$. In particular, Theorem 5 of Andrews (1992) using Assumption TSE-1D gives the desired result under Assumptions 1 and GMM(b)-(d).

Condition (10.19) holds by Assumption GMM(a)-(c) and (e).

Next, we use (10.18) and (10.19) to show that $\sup_{j=1,\dots,n-m+1} \|\widehat{\beta}_{(j)} - \beta_0\| \rightarrow_p 0$. By (10.19), 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

$$\begin{aligned}
& P\left(\sup_{j=1,\dots,n-m+1} \|\widehat{\beta}_{(j)} - \beta_0\| > \varepsilon\right) \\
& \leq P\left(\sup_{j=1,\dots,n-m+1} Q^{(k)}(\widehat{\beta}_{(j)}) - Q^{(k)}(\beta_0) > \delta\right) \\
& = P\left(\sup_{j=1,\dots,n-m+1} (Q^{(k)}(\widehat{\beta}_{(j)}) - Q^{(k)}(\widehat{\beta}_{(j)}) + Q^{(k)}(\widehat{\beta}_{(j)}) - Q^{(k)}(\beta_0)) > \delta\right) \\
& \leq P\left(\sup_{j=1,\dots,n-m+1} (Q^{(k)}(\widehat{\beta}_{(j)}) - Q^{(k)}(\beta_0)) > \delta\right) \\
& \leq P\left(2 \sup_{j=1,\dots,n-m+1} |Q^{(k)}(\beta) - Q^{(k)}(\beta_0)| > \delta\right) \\
& = o_p(1), \tag{10.22}
\end{aligned}$$

where the second inequality holds because $\widehat{\beta}_{(j)}$ minimizes $Q^{(k)}(\beta)$ over $\beta \in \mathcal{B}$ and the second equality holds by (10.18).

The proof that $\sup_{j=1,\dots,n-m+1} \|\widehat{\beta}_{2(j)} - \beta_0\| \rightarrow_p 0$ is essentially the same as that given above. The proofs that $\|\widehat{\beta}_{1,n} - \beta_0\| \rightarrow_p 0$ and $\|\widehat{\beta}_{1,n+m} - \beta_0\| \rightarrow_p 0$ are standard (and are special cases of the proof above) and, hence, are not given. This completes the verification of Assumption 2(a) for the GMM case.

To verify Assumption 2(b) in the GMM case, we write

$$\begin{aligned}
& \sup_{\beta \in B(\beta_0, \varepsilon_n)} \|\widehat{\Sigma}_{1,n+m}(\beta) - \Sigma_0\| \\
& \leq \sup_{\beta \in B(\beta_0, \varepsilon_n)} \|\widehat{\Sigma}_{1,n+m}(\beta) - \Sigma(\beta)\| + \sup_{\beta \in B(\beta_0, \varepsilon_n)} \|\Sigma(\beta) - \Sigma_0\|, \tag{10.23}
\end{aligned}$$

where $\Sigma(\beta) = E\mathbf{U}_{1,m}(\beta)\mathbf{U}_{1,m}(\beta)'$. The second term on the rhs 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 rhs 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 GMM estimators when $v = a$ or b , we verify that

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

For $T = S$, we have

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

$$+ \sum_{i=j}^{j+m-1} g(W_i, \beta)' \left(\frac{\partial}{\partial \beta_r} (V_j^{-1}(\beta)) \right) \sum_{i=j}^{j+m-1} g(W_i, \beta) \quad (10.24)$$

for $r = 1, \dots, d_\beta$. 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 $j = 1, \dots, n-m+1$ using Assumption GMM(f). In consequence, it suffices to show the desired result with $V_j^{-1}(\beta)$ and $(\partial/\partial\beta_r)V_j^{-1}(\beta)$ replaced by I_d . The latter holds by Markov's inequality given the moment conditions in Assumption GMM(f).

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

For the case where $v = c$ or d , 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). \square

Footnotes

¹ The author thanks Ray Fair for suggesting the testing problem considered in this paper and for helpful comments, Patrik Guggenberger for helpful comments, and 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.

² The estimator $\widehat{\beta}_{1,n}$, which appears in the statistic S_a , depends on n observations. In contrast, the estimators $\{\widehat{\beta}_{(j)} : j = 1, \dots, n - m + 1\}$, which appear in the statistics $\{S_j(\widehat{\beta}_{(j)}) : j = 1, \dots, n - m + 1\}$, only depend on $n - m$ observations. To see whether the use of the same number of observations by all estimates of β leads to better size results, we carried out some Monte Carlo simulations for the case where S_a is defined using the estimator $\widehat{\beta}_{1,n-m}$ instead of $\widehat{\beta}_{1,n}$. This had essentially no effect on the size of the test for the cases considered.

³ This holds provided the rank of $\sum_{i=n+1}^{n+m} X_i X_i'$ is greater than or equal to m , as is typically the case when $d \geq m$.

References

- Andrews, D. W. K. (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, 241-257.
- (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821-856.
- Andrews, D. W. K. and R. C. Fair (1988): "Inference in Nonlinear Econometric Models with Structural Change," *Review of Economic Studies*, 55, 615-640.
- Chow, G. C. (1960): "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 28, 591-605.
- 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. (2002): "Testing for a New Economy in the 1990s," unpublished manuscript, Cowles Foundation, Yale University.
- Ghysels, E., A. Guay, and A. Hall (1997): "Predictive Tests for Structural Change with Unknown Breakpoint," *Journal of Econometrics*, 82, 209-233.
- Ghysels, E. and A. Hall (1990): "A Test for Structural Stability of Euler Conditions Parameters Estimated Via the Generalized Method of Moments Estimator," *International Economic Review*, 31, 355-364.
- Hansen, B. E. (1992): "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business and Economic Statistics*, 10, 321-336.
- Politis, D. N., J. P. Romano, and M. Wolf (1999): *Subsampling*. Springer Series in Statistics. New York: Springer.
- Stock, J. H. (1994): "Unit Roots, Structural Breaks and Trends," Chapter 46 in *Handbook of Econometrics, Volume IV*, ed. by R. F. Engle and D. L. McFadden. New York: North-Holland.

Table I
 True Size of Tests with Nominal Significance Level .05
 Using Normal, χ_2^2 , t_3 , and Uniform Regressors and Errors

<i>m</i>	ρ	Test	Normal		χ_2^2		t_3		Uniform	
			<i>n</i>		<i>n</i>		<i>n</i>		<i>n</i>	
			100	250	100	250	100	250	100	250
10	0	S_a	.078	.063	.083	.064	.087	.064	.071	.061
		S_b	.055	.055	.068	.060	.071	.060	.046	.049
		S_c	.069	.060	.072	.061	.075	.060	.066	.060
		S_d	.046	.052	.056	.055	.058	.055	.043	.048
		P_b	.074	.063	.082	.063	.084	.068	.056	.052
		P_d	.063	.060	.072	.060	.075	.065	.048	.052
		F	.051	.050	.088	.089	.090	.085	.028	.024
10	.4	S_a	.087	.064	.086	.067	.090	.067	.081	.063
		S_b	.061	.054	.068	.061	.072	.061	.053	.052
		S_c	.072	.059	.075	.063	.077	.063	.068	.059
		S_d	.047	.050	.056	.058	.058	.057	.042	.049
		P_b	.071	.058	.079	.068	.085	.066	.060	.053
		P_d	.062	.056	.072	.065	.078	.065	.048	.050
		F	.123	.115	.131	.125	.140	.129	.105	.101
10	.8	S_a	.118	.078	.112	.073	.116	.076	.118	.078
		S_b	.075	.060	.077	.060	.082	.064	.074	.061
		S_c	.091	.066	.088	.064	.094	.067	.087	.068
		S_d	.053	.055	.061	.058	.064	.060	.049	.054
		P_b	.077	.061	.079	.061	.085	.065	.075	.062
		P_d	.062	.058	.076	.063	.080	.066	.050	.055
		F	.329	.286	.318	.270	.314	.255	.334	.288

Table I (cont).

m	ρ	Test	Normal		χ_2^2		t_3		Uniform	
			n		n		n		n	
			100	250	100	250	100	250	100	250
5	0	$S_a (= P_a)$.064	.057	.057	.060	.058	.059	.062	.055
		$S_b (= P_b)$.051	.052	.053	.057	.052	.057	.042	.047
		$S_c (= P_c)$.060	.056	.054	.058	.055	.058	.060	.053
		$S_d (= P_d)$.047	.052	.049	.056	.050	.055	.040	.045
		F	.050	.049	.103	.102	.099	.089	.007	.004
5	.4	$S_a (= P_a)$.069	.059	.060	.057	.062	.061	.067	.061
		$S_b (= P_b)$.051	.052	.052	.054	.053	.057	.044	.050
		$S_c (= P_c)$.065	.059	.055	.057	.058	.059	.061	.057
		$S_d (= P_d)$.050	.053	.050	.054	.052	.056	.041	.048
		F	.071	.068	.099	.094	.102	.093	.040	.036
5	.8	$S_a (= P_a)$.095	.070	.086	.066	.090	.071	.094	.070
		$S_b (= P_b)$.062	.058	.061	.057	.066	.061	.060	.058
		$S_c (= P_c)$.080	.063	.072	.060	.075	.062	.078	.065
		$S_d (= P_d)$.056	.055	.060	.057	.061	.058	.049	.055
		F	.146	.125	.142	.118	.147	.125	.141	.123
1	0	$S_c (= S_a = P_c = P_a)$.060	.053	.059	.056	.061	.052	.057	.050
		$S_d (= S_b = P_d = P_b)$.048	.048	.053	.053	.053	.049	.034	.039
		F	.050	.052	.055	.057	.054	.049	.008	.002
1	.4	$S_c (= S_a = P_c = P_a)$.065	.054	.060	.052	.062	.053	.063	.053
		$S_d (= S_b = P_d = P_b)$.051	.049	.053	.050	.053	.050	.046	.046
		F	.054	.052	.052	.050	.055	.052	.030	.025
1	.8	$S_c (= S_a = P_c = P_a)$.089	.063	.084	.060	.082	.062	.090	.066
		$S_d (= S_b = P_d = P_b)$.072	.058	.068	.056	.069	.058	.073	.059
		F	.074	.059	.068	.054	.075	.061	.072	.059

Table II
 Power of Significance Level .05 Size-corrected Tests
 Using Normal, χ_2^2 , t_3 , and Uniform Regressors and Errors

$\ \beta_1\ $	m	ρ	Test	Normal		χ_2^2		t_3		Uniform	
				n		n		n		n	
				100	250	100	250	100	250	100	250
1.75	10	0	S_b	.91	.95	.66	.78	.66	.78	.95	.96
			S_d	.90	.94	.67	.78	.67	.78	.93	.96
			P_b	.83	.89	.48	.52	.49	.44	.94	.96
			P_d	.80	.88	.39	.57	.39	.41	.92	.95
			F	.94	.95	.79	.82	.81	.83	.96	.97
1.75	10	.4	S_b	.80	.85	.57	.66	.56	.59	.83	.89
			S_d	.85	.91	.62	.72	.61	.70	.89	.94
			P_b	.71	.80	.42	.47	.44	.42	.82	.88
			P_d	.74	.85	.36	.45	.36	.37	.88	.93
			F	.83	.86	.67	.71	.69	.71	.87	.89
1.75	10	.8	S_b	.41	.46	.32	.39	.32	.37	.49	.43
			S_d	.76	.87	.56	.71	.53	.64	.90	.80
			P_b	.38	.45	.31	.37	.31	.32	.49	.43
			P_d	.68	.81	.35	.48	.33	.41	.90	.79
			F	.42	.47	.36	.44	.40	.40	.49	.44
1.75	5	0	$S_b (= P_b)$.68	.72	.36	.42	.35	.45	.81	.83
			$S_d (= P_d)$.66	.72	.33	.42	.34	.43	.79	.83
			F	.70	.73	.41	.44	.44	.46	.81	.83
1.75	5	.4	$S_b (= P_b)$.57	.62	.33	.39	.32	.40	.68	.71
			$S_d (= P_d)$.61	.67	.30	.39	.32	.39	.74	.79
			F	.60	.63	.36	.41	.40	.43	.68	.71
1.75	5	.8	$S_b (= P_b)$.35	.37	.30	.34	.27	.30	.37	.39
			$S_d (= P_d)$.54	.63	.31	.38	.30	.37	.65	.73
			F	.36	.38	.34	.35	.30	.32	.39	.40

Table II (cont.)

$\ \beta_1\ $	m	ρ	Test	Normal		χ_2^2		t_3		Uniform	
				n		n		n		n	
				100	250	100	250	100	250	100	250
1.75	1	0	$S_d (= S_b = P_b = P_d)$.31	.32	.25	.25	.29	.29	.41	.42
			F	.32	.32	.24	.25	.28	.29	.40	.41
1.75	1	.4	$S_d (= S_b = P_b = P_d)$.31	.32	.24	.27	.29	.29	.36	.36
			F	.32	.32	.26	.27	.29	.29	.36	.37
1.75	1	.8	$S_d (= S_b = P_b = P_d)$.30	.30	.31	.29	.27	.27	.31	.31
			F	.30	.30	.29	.29	.28	.28	.31	.31
7	1	0	$S_d (= S_b = P_b = P_d)$.77	.78	.72	.72	.76	.77	.81	.82
			F	.78	.78	.72	.72	.75	.76	.81	.82
7	1	.4	$S_d (= S_b = P_b = P_d)$.77	.78	.72	.74	.76	.76	.79	.80
			F	.77	.78	.73	.74	.75	.76	.80	.80
7	1	.8	$S_d (= S_b = P_b = P_d)$.76	.76	.76	.75	.74	.75	.77	.77
			F	.76	.76	.75	.75	.74	.75	.77	.77