Econometric Theory, **26**, 2010, 994–1031.
doi:10.1017/S0266466609990442

IMPROVED AND EXTENDED END-OF-SAMPLE INSTABILITY TESTS USING A FEASIBLE QUASI-GENERALIZED LEAST SQUARES PROCEDURE

DUKPA KIM
University of Virginia

This paper extends the Andrews (2002, *Econometrica* 71, 1661–1694) and Andrews and Kim (2006, *Journal of Business & Economic Statistics* 24, 379–394) ordinary least squares–based end-of-sample instability tests for linear regression models. The author proposes to quasi-difference the data first using a consistent estimate of the sum of the autoregressive coefficients of the error process and then test for the end-of-sample instability. For the cointegration model, the feasible quasi-generalized least squares (FQGLS) version of the Andrews and Kim (2006) P test is considered and is shown, by simulations, to be more robust to serial correlation in the error process and to have power no less than Andrews and Kim's original test. For the linear time trend model, the FQGLS version of the Andrews (2002) S test is considered with the error process allowed to be nonstationary up to one unit root, and the new test is shown to be robust to potentially nonstationary serial correlation. A simulation study also shows that the finite-sample properties of the proposed test can be further improved when the Andrews (1993, *Econometrica* 61,139–165) or Andrews and Chen (1994, *Journal of Business & Economic Statistics* 12, 187–204) median unbiased estimate of the sum of the autoregressive coefficients is used.

1. INTRODUCTION

Although the econometrics literature has paid a vast amount of attention to parameter instability,[1] relatively little work has been done for end-of-sample instability tests. These differ from other regular structural break tests in that they use an asymptotic framework that assumes that the number of postbreak observations is fixed when the sample size increases. Two important papers in this area are Andrews (2002) and Andrews and Kim (2006). The former analyzed

This paper is a refined version of the author's Ph.D. dissertation, chapter 3. I am very grateful to my adviser Pierre Perron for his guidance and encouragement. I also thank two referees and a co-editor for their useful comments and suggestions. Address correspondence to Dukpa Kim, Department of Economics, University of Virginia, Monroe Hall, McCormick Road, Charlottesville, VA 22903, USA; e-mail: dukpa@virginia.edu.

 © Cambridge University Press 2009 0266-4666/10 $15.00

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

end-of-sample parameter instability in stationary linear regressions and generalized methods of moments (GMM) models; the latter considered cointegration models.

This paper extends in several directions the Andrews (2002) and Andrews and Kim (2006) end-of-sample tests. Section 2 proposes a new end-of-sample instability test in the cointegration model that can be regarded as a modification of the Andrews and Kim (2006) ordinary least squares-based (OLS-based) P test. A crucial assumption in Andrews and Kim (2006) is the strict stationarity of the error process. However, in the cointegration model, the error process often shows a considerable amount of serial correlation, and the Andrews and Kim (2006) test can exhibit nonnegligible size distortions in finite samples. The new test, labeled $P(\hat{\alpha}, \Xi_T)$, is based on the feasible quasi-generalized least squares (FQGLS) procedure, and the motivation for this is to robustify the test procedure to persistent errors. This paper specifies the error process to have an autoregressive representation. Then, the sum of the autoregressive parameters, α_0, is a useful measure of persistence, as argued in Andrews and Chen (1994). The proposed FQGLS procedure is intended to make the time series data less persistent, for it quasi-differences the time series data by $\hat{\alpha}$, a consistent estimate of α_0.

Another contribution of this paper is to show that the end-of-sample tests can be consistent depending on the nature of the regressors with unstable coefficients. In particular, when a change occurs in the coefficient of an integrated regressor or time trend, the end-of-sample instability tests can be consistent. On the other hand, when a mean shift or a change in the coefficient of a stationary regressor occurs, the end-of-sample instability tests are not consistent.

Although most of the theoretical results are established for a general weighting matrix Ξ_T that is assumed to converge to some fixed matrix Ξ_0, the Monte Carlo simulation study focuses on $\Xi_T = I$ and $\hat{\Sigma}_{T+m}^{-1}$ where I is the identity matrix and $\hat{\Sigma}_{T+m}^{-1}$ is the inverse of the estimate for the covariance matrix of the quasi-differenced errors (a precise definition is given in Section 2), and it shows that the new test is more robust to serial correlation in the error process and has power no less than Andrews and Kim's original test.

Section 3 describes the second extension, pertaining to the linear time trend model, where the error process is allowed to be nonstationary up to one unit root. This setup is similar to that of Perron and Yabu (2009), who analyzed tests for a regular structural break. For this model, a new test, labeled $S(\hat{\alpha}, \Xi_T)$, is proposed. This test is also based on the FQGLS procedure and can be regarded as a modification of the Andrews (2002) S test. A novel feature of this test is that it has asymptotically correct size whether the error process is stationary or has a unit root. Hence, prior knowledge of the nature of the error process is not required for testing the end-of-sample instability of the linear time trend. Although this robustness is shown only pointwise in the sum of autoregressive coefficients, it is also shown that the new test has asymptotically correct size under the local-to-unity asymptotic framework, which suggests that the new test can have good finite-sample size even in the presence of strongly serially correlated errors. The

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

simulation results are in general agreement with the theoretical findings and also show that further improvements of the proposed test can be obtained when the Andrews (1993) or Andrews and Chen (1994) median unbiased estimate of the sum of the autoregressive coefficients is used.

Section 4 contains an empirical example. The $P(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1})$ test is applied to the cointegration equation where U.S. federal government spending is explained by the real gross domestic product (GDP), the consumer price index (CPI), a set of demographic variables, and a war dummy variable. The test results seem to suggest that U.S. federal spending since 2001 deviates from the past trend.

Section 5 offers some concluding remarks. The Appendix collects all technical derivations.

2. COINTEGRATION MODEL

2.1. Model and Test Statistics

This section considers the model in Andrews and Kim (2006), which is given by

$$y_t = \begin{cases} z_t'\gamma_0 + u_t = z_{1,t}'\gamma_{1,0} + z_{2,t}'\gamma_{2,0} + u_t, & \text{for } t = 1, \ldots, T \\ z_t'\gamma_t + u_t = z_{1,t}'\gamma_{1,t} + z_{2,t}'\gamma_{2,t} + u_t, & \text{for } t = T+1, \ldots, T+m, \end{cases} \tag{1}$$

where $z_{1,t}$ contains deterministic components and integrated (I(1)) random variables and $z_{2,t}$ contains stationary and ergodic random variables. The error process, $\{u_t\}$ is assumed to be stationary and ergodic also.

The null and alternative hypotheses are

$$H_0 : \begin{cases} \gamma_t = \gamma_0 \text{ for all } t = T+1, \ldots, T+m \\ \text{and } \{u_t, t = 1, \ldots, T+m\} \text{ are stationary and ergodic} \end{cases}$$

and

$$H_1 : \begin{cases} \gamma_t \neq \gamma_0 \text{ for some } t = T+1, \ldots, T+m \\ \text{and/or the distribution of } \{u_t, t = T+1, \ldots, T+m\} \text{ differs from} \\ \text{the distribution of } \{u_t, t = 1, \ldots, m\}. \end{cases}$$

The most crucial assumption is the stationarity and ergodicity of $\{u_t\}$. The Andrews and Kim (2006) simulation results show that their tests behave fairly well under various distributions of the error process. However, in practice, $\{u_t\}$ often has substantial serial correlation, and, in that case, the test may show size distortions in finite samples. The goal of this paper is to robustify the test procedure to persistent errors. What is important is not to eliminate all serial correlation but to make the series less persistent. This paper specifies $\{u_t\}$ to be a linear process, as follows.

Assumption A. Let $u_t = d(L)e_t$,with $d(L) = \sum_{i=0}^{\infty} d_i L^i$, $\sum_{i=0}^{\infty} i\,|d_i| < \infty$, $d(z) \neq 0$, $\forall |z| \leq 1$ and $\{e_t\}$ is an independently and identically distributed process with $\mathrm{E}(e_t^2) < \infty$.

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

Obviously, the error process is not allowed to be nonstationary because the model would then be a spurious regression. Under these conditions, u_t has an autoregressive representation, i.e., $A(L)u_t = e_t$ where $A(L) = 1 - \sum_{i=1}^{\infty} a_i L^i$. Let α_0 denote the sum of the autoregressive parameters, i.e.,

$$\alpha_0 = 1 - A(1) = \sum_{i=1}^{\infty} a_i. \tag{2}$$

Andrews and Chen (1994) argue that the sum of autoregressive parameters, α_0, is an appropriate scalar measure of persistence, because the cumulative impulse response is a simple function of α_0, i.e., the cumulative impact of a unit change in e_t on u_{t+s}, $s \geq 0$ is given by $\sum_{i=0}^{\infty} d_i = 1/(1-\alpha_0)$.

Note that the first-order correction by α_0 will eliminate the discrepancy between the instantaneous and cumulative impacts. To see this, let

$$v_t = (1 - \alpha_0 L)u_t = (1 - \alpha_0 L)d(L)e_t. \tag{3}$$

Then the cumulative impact of a unit change in e_t on v_{t+s}, $s \geq 0$ is again unity. This is the motivation for the FQGLS procedure described subsequently.

Let $\hat{u}_t$ be the residuals from a regression of y_t on $\{z'_{1,t}, z'_{2,t}\}$ and $\hat{\alpha}$ be the OLS estimate of α_0 in the augmented Dickey–Fuller (ADF) type regression:

$$\hat{u}_t = \hat{\alpha}\hat{u}_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta\hat{u}_{t-i} + e_{tk}, \qquad t = k+2, \ldots, T, \tag{4}$$

where $k \to \infty$ and $k^3/T \to 0$ as $T \to \infty$. Then, it is well known that $\sqrt{T}(\hat{\alpha} - \alpha_0) = O_p(1)$.

Now, define the series $\{q_t(\hat{\alpha}) : t \geq 1\}$ to be the quasi-difference of $\{y_t\}$ in (1) by $\hat{\alpha}$. More precisely, for $t = 2, \ldots, T$,

$$q_t(\hat{\alpha}) = w'_t(\hat{\alpha})\gamma_0 + v_t - (\hat{\alpha} - \alpha_0)u_{t-1}, \tag{5}$$

and, for $t = T+1, \ldots, T+m$,

$$q_t(\hat{\alpha}) = z'_t\gamma_t - \hat{\alpha}z'_{t-1}\gamma_{t-1} + v_t - (\hat{\alpha} - \alpha_0)u_{t-1},$$

where $q_t(\alpha) = (1 - \alpha L)y_t = y_t - \alpha y_{t-1}$, $w_t(\alpha) = (w'_{1,t}(\alpha), w'_{2,t}(\alpha))' = (1 - \alpha L)z_t = (1 - \alpha L)(z'_{1,t}, z'_{2,t})'$. For $t = 1$, let $q_1(\alpha) = y_1$ and $w_1(\alpha) = z_1$, and thus

$$q_1(\hat{\alpha}) = y_1 = z'_1\gamma_0 + v_1 + \alpha_0 u_0 = w'_1(\hat{\alpha})\gamma_0 + u_1. \tag{6}$$

The basic estimators are defined next. The full-sample estimator for γ_0 is defined by

$$\hat{\gamma}_{T+m} = \text{OLS estimate of } \gamma_0 \text{ in the regression of } q_j(\hat{\alpha}) \text{ on } w_j(\hat{\alpha})$$
$$\text{using observations with index } j = 1, \ldots, T+m, \tag{7}$$

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

and the partial-sample estimators are defined by

$$\hat{\gamma}_{(t)} = \text{OLS estimate of } \gamma_0 \text{ in the regression of } q_j(\hat{\alpha}) \text{ on } w_j(\hat{\alpha})$$
$$\text{using observations with index } j = 1, \ldots, T$$
$$\text{and } j \neq t, \ldots, t + [m/2] - 1$$
$$\text{for } t = 1, \ldots, T - m + 1. \tag{8}$$

This particular FQGLS estimator is often derived under the assumption that $u_0 = 0$, which is not needed in this paper and is referred to as the "conditional case." Another popular FQGLS estimator is the Prais–Winsten estimator. Canjels and Watson (1997) showed that the "conditional case" estimator works well when the variance of u_0 is small whereas the Prais–Winsten estimator works well when the variance is large. This paper will not explicitly deal with the Prais–Winsten estimator, but the asymptotic results can be easily extended to allow for its use with a simple modification of the proofs given in the Appendix. Also, simulations not reported in this paper show that using the Prais–Winsten estimator makes almost no difference in terms of the performance of the end-of-sample instability tests, at least under the data generating processes (DGPs) used in the simulations.

As a matter of notation, for any variable x_t, let the corresponding capital letter with two subscripts, $X_{r,s}$, denote a stack of x_t's indexed from r to s, i.e., $X_{r,s} = (x_r, \ldots, x_s)'$. Let the predicted errors $\hat{v}_t(\alpha, \gamma)$ be given by

$$\hat{v}_t(\alpha, \gamma) = q(\alpha)_t - w(\alpha)_t' \gamma, \quad \text{for } t = 1, \ldots, T + m \tag{9}$$

and the $P_t(\alpha, \gamma, \Xi_T)$ statistics by

$$P_t(\alpha, \gamma, \Xi_T) = \hat{V}_{t,t+m-1}'(\alpha, \gamma)\, \Xi_T\, \hat{V}_{t,t+m-1}(\alpha, \gamma), \quad \text{for } t = 2, \ldots, T + 1, \tag{10}$$

where $\hat{V}_{t,t+m-1}(\alpha, \gamma)$ is a vector of $\hat{v}_t(\alpha, \gamma)$'s. This is the square of the norm of the vector $\hat{V}_{t,t+m-1}(\alpha, \gamma)$ with respect to the weighting matrix Ξ_T, which is assumed to converge in probability to Ξ_0 under the null hypothesis. Further discussion on the choice of Ξ_T is given later. The test statistic $P(\hat{\alpha}, \Xi_T)$ is given by

$$P(\hat{\alpha}, \Xi_T) = P_{T+1}(\hat{\alpha}, \hat{\gamma}_{T+m}, \Xi_T), \tag{11}$$

and the empirical distribution function of $\{P_t(\hat{\alpha}, \hat{\gamma}_{(t)}, \Xi_T),\ t = 2, \ldots, T - m + 1\}$ is given by

$$\hat{F}_T(x) = \frac{1}{T - m} \sum_{t=2}^{T-m+1} 1(P_t(\hat{\alpha}, \hat{\gamma}_{(t)}, \Xi_T) < x).$$

Denote by $\hat{\varphi}_{1-p,T}$ the $1 - p$ quantile of $\hat{F}_T(x)$, i.e.,

$$\hat{\varphi}_{1-p,T} = \inf\{x : \hat{F}_T(x) \geq 1 - p\},$$

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

which will be used as a critical value. Note that the Andrews and Kim (2006) P test is $P(0, I)$ in terms of the notation in (11), where I is the $m \times m$ identity matrix.

Because the errors have the same variance as a result of the stationarity assumption, a natural choice for Ξ_T is the identity matrix, which imposes an equal weight on each element of $\hat{V}_{t,t+m-1}(\alpha, \gamma)$. Another choice is the inverse of the estimate for $\Sigma_0 = \mathrm{E}V_{2,m+1}V'_{2,m+1}$ to account for the serial correlation remaining in v_t after the first-order correction by $\hat{\alpha}$. In particular, we consider $\hat{\Sigma}_{T+m}$ defined by

$$\hat{\Sigma}_{T+m} = \frac{1}{T}\sum_{t=2}^{T+1} \hat{V}_{t,t+m-1}(\hat{\alpha}, \hat{\gamma}_{T+m})\hat{V}'_{t,t+m-1}(\hat{\alpha}, \hat{\gamma}_{T+m}). \tag{12}$$

Hence, we focus on the $P(\hat{\alpha}, I)$ and $P(\hat{\alpha}, \hat{\Sigma}^{-1}_{T+m})$ tests for the cointegration model, although most of our results are established for the general weighting matrix Ξ_T.

2.2. Asymptotic Results for the $P(\hat{\alpha}, \Xi_T)$ Tests

It is convenient to introduce a sequence of random vectors $\{\zeta_{0,t} : t \geq 1\}$ that is stationary and ergodic under both H_0 and H_1 and another sequence of random vectors $\{\zeta_{T,t} : T+1 \leq t \leq T+m\}$ whose joint distribution may differ from that of $\{\zeta_{0,t} : T+1 \leq t \leq T+m\}$ and is independent of T under H_1. Let $\zeta_t = \{v_t, z'_{2,t}\}'$. Assume that $\zeta_t = \zeta_{0,t}$ for $t = 1, \ldots, T+m$ under H_0, and $\zeta_t = \zeta_{0,t}$ for $t = 1, \ldots, T$ and $\zeta_t = \zeta_{T,t}$ for $t = T+1, \ldots, T+m$ under H_1.

Assumption 1.

(a) $\{\zeta_{0,t} : t \geq 1\}$ are mean zero, stationary, and ergodic random vectors under H_0 and H_1. The distribution of $\{z_{1,t} : t \leq T\}$ is the same under H_0 and H_1. Under H_1, the distribution of $\{\zeta_{T,t} : T+1 \leq t \leq T+m\}$ does not depend on T.

(b) $\mathrm{E}z_{2,t}u_t = \mathrm{E}z_{2,t-1}u_t = \mathrm{E}z_{2,t}u_{t-1} = 0$, $\mathrm{E}|u_t|^{2+2\delta} < \infty$, and $\mathrm{E}||z_{2,t}||^{2+2\delta} < \infty$ for some $\delta > 0$ and $t \leq T$.

(c) Let $\eta_T(r) = B_T^{-1} z_{1,[\mathrm{Tr}]}$ for $r \in [0, 1]$, $\{B_T : T \geq 1\}$ be nonrandom positive definite diagonal matrices, and $\eta(r)$ for $r \in [0, 1]$ be some vector stochastic process that has bounded and continuous sample path almost surely. Then, $\eta_T(r) \Rightarrow \eta(r)$ as $T \to \infty$, where $\Rightarrow$ denotes weak convergence.

(d) $T^{-1}\sum_{t=1}^{T} B_T^{-1} z_{1,t}(u_t, z'_{2,t}) = o_p(1)$, $T^{-1}\sum_{t=1}^{T} B_T^{-1} z_{1,t-1}(u_t, z'_{2,t}) = o_p(1)$, and $T^{-1}\sum_{t=1}^{T} B_T^{-1} z_{1,t}(u_{t-1}, z'_{2,t-1}) = o_p(1)$.

(e) $\max_{t=T+1,\ldots,T+m} ||B_T^{-1} z_{t,1}|| = O_p(1)$ under H_0 and H_1.

(f) Σ_0, $\int_0^1 \eta(r)\eta(r)'dr$, $\Sigma_{2,0}$, and $[(1+\alpha_0^2)\Sigma_{2,0} - \alpha_0(\Omega_{2,0} + \Omega'_{2,0})]$ are positive definite, where $\Sigma_0 = \mathrm{E}V_{2,m+1}V'_{2,m+1}$, $\Sigma_{2,0} = \mathrm{E}z_{2,t}z'_{2,t}$, and $\Omega_{2,0} = \mathrm{E}z_{2,t-1}z'_{2,t}$, for $t \leq T$.

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

(g) The distribution of $P_2(\alpha_0, \gamma_0, \Xi_0)$ is continuous and increasing at its $1-p$ quantile.

The preceding assumptions are almost the same as those in Andrews and Kim (2006), but the moment conditions in (b) are slightly stronger to accommodate the FQGLS procedure. Note that these strengthened moment conditions are not completely innocuous. For example, a lagged dependent variable is not permitted as a regressor in the stationary regression model.

Andrews and Kim (2006) stated that their test is inconsistent (see their Theorem 2(a) and the subsequent comments). However, this statement is slightly misleading, because the alternative hypothesis H_1 contains a wide range of DGPs and the asymptotic behaviors of their tests and also those proposed in this paper differ across the DGPs in H_1. In particular, these tests can be consistent depending on the nature of the regressors whose coefficients change after $t = T$. Note from (10) and (11) that the test statistic $P(\hat{\alpha}, \Xi_T)$ is constructed with the predicted residuals for $t = T+1, \ldots, T+m$. When $\hat{\gamma}_{T+m}$ and $\hat{\alpha}$ are consistent at some proper rates, we can write in large samples

$$\begin{aligned}\hat{v}_t(\hat{\alpha}, \hat{\gamma}_{T+m}) &= q(\hat{\alpha})_t - w(\hat{\alpha})_t' \hat{\gamma}_{T+m} \\ &= z_t'\gamma_t - \hat{\alpha} z_{t-1}'\gamma_{t-1} + v_t - (\hat{\alpha}-\alpha_0)u_{t-1} - (z_t - \hat{\alpha} z_{t-1})'\hat{\gamma}_{T+m} \\ &= z_t'(\gamma_t - \hat{\gamma}_{T+m}) - \hat{\alpha} z_{t-1}'(\gamma_{t-1} - \hat{\gamma}_{T+m}) + v_t - (\hat{\alpha}-\alpha_0)u_{t-1} \\ &\simeq z_t'(\gamma_t - \gamma_0) - \alpha_0 z_{t-1}'(\gamma_{t-1} - \gamma_0) + v_t \qquad \textbf{(13)}\end{aligned}$$

for $t = T+1, \ldots, T+m$. Let

$$\omega_T = \max_{T+1 \le t \le T+m} \left| z_t'(\gamma_t - \gamma_0) \right|;$$

then it is evident from (13) that the order of magnitude of the test statistic $P(\hat{\alpha}, \Xi_T)$ depends on ω_T and that the tests based on the predicted residuals can be consistent, if $\omega_T \xrightarrow{p} \infty$. The consistency of $\hat{\gamma}_{T+m}$ under the alternative hypothesis is another related issue, because it also depends on the postbreak observations.[2]

Accordingly, we consider two mutually exclusive and exhaustive subclasses of H_1. The first one is given by

$$H_2 : \begin{cases} y_t \text{ satisfies the conditions in } H_1 \text{ and} \\ \omega_T = O_p(1). \end{cases}$$

The additional condition in H_2 is not binding with the stationary regressors unless $||\gamma_{2,t}||$ is increasing in T. For $z_{1,t}$, it should be the case that $\max_{T+1 \le t \le T+m} ||\gamma_{1,t} - \gamma_{1,0}|| = o(1)$, with the rate of decay depending on the nature of the regressor. One example of a permitted coefficient change for $z_{1,t}$ is the joint broken trend alternative, given by (assuming that $z_{1,t}$ is a scalar)

$$\gamma_{1,t} = \gamma_{1,0} + (1 - z_{1,T}/z_{1,t})\gamma_{1,1}.$$

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

In this example,

$$\begin{aligned}\omega_T &= \max_{T+1\le t\le T+m} \left|z_{1,t}'(\gamma_{1,t}-\gamma_{1,0})\right| \\ &= \max_{T+1\le t\le T+m} \left|(z_{1,t}-z_{1,T})\gamma_{1,1}\right| \\ &\le \left\|\gamma_{1,1}\right\| \max_{T+1\le t\le T+m} \left\|z_{1,t}-z_{1,T}\right\| = O_p(1)\end{aligned}$$

because m is fixed. The second class of alternative hypotheses is given by

$$H_3 : \begin{cases} y_t \text{ satisfies the conditions in } H_1 \text{ and} \\ p\lim_{T\to\infty} \omega_T = \infty. \end{cases}$$

When there are only joint broken trends, the condition in H_3 is not satisfied. When $\gamma_{1,t} = \gamma_{1,1} \ne \gamma_{1,0}$ for $t \ge T+1$, the condition is satisfied.

Now we present the consistency of the basic estimators in relation to the term ω_T.

LEMMA 1. *Suppose that Assumptions A and 1 hold. Then* $||B_T(\hat{\gamma}_{1,T+m} - \gamma_{1,0})|| = O_p(\omega_T T^{-1}) + o_p(1)$, $||\hat{\gamma}_{2,T+m} - \gamma_{2,0}|| = O_p(\omega_T T^{-1}) + o_p(1)$, $\max_{2\le t\le T-m+1}||B_T(\hat{\gamma}_{1,(t)} - \gamma_{1,0})|| = o_p(1)$, *and* $\max_{2\le t\le T-m+1}||\hat{\gamma}_{2,(t)} - \gamma_{2,0}|| = o_p(1)$.

Note that the $O_p(\omega_T T^{-1})$ term in Lemma 1 shows that the consistency of $\hat{\gamma}_{T+m}$ depends on the order of magnitude of ω_T. On the other hand, the uniform consistency of $\hat{\gamma}_{(t)}$'s does not depend on ω_T because the postbreak observations are not used.

The next lemma states that the consistency of the estimate of the variance-covariance matrix defined in (12) holds if $\omega_T = o_p(T^{1/2})$.

LEMMA 2. *Suppose that Assumptions A and 1 hold. Then* $\hat{\Sigma}_{T+m} \xrightarrow{p} \Sigma_0$ *if* $\omega_T = o_p(T^{1/2})$.

Let $F(x)$ and φ_{1-p} denote the distribution function of $P_2(\alpha_0, \gamma_0, \Xi_0)$ and its $1-p$ quantile, respectively. The asymptotic results under the null hypothesis for the $P(\hat{\alpha}, \Xi_T)$ test are stated in the following theorem.

THEOREM 1. *Suppose that Assumptions A and 1 hold. Then, as* $T \to \infty$ *with* m *fixed, the following asymptotic results hold under* H_0 *if* $\Xi_T \xrightarrow{p} \Xi_0$*:*

(i) $P(\hat{\alpha}, \Xi_T) \xrightarrow{d} P_2(\alpha_0, \gamma_0, \Xi_0)$*;*

(ii) $\hat{F}_T(x) \xrightarrow{p} F(x)$ *for all* x *in a neighborhood of* φ_{1-p}*;*

(iii) $\hat{\varphi}_{1-p,T} \xrightarrow{p} \varphi_{1-p}$*;*

(iv) $\Pr(P(\hat{\alpha}, \Xi_T) > \hat{\varphi}_{1-p,T}) \to p$.

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

When the chosen Ξ_T does not depend on the postbreak observations, e.g., $\Xi_T = I$, neither do $\hat{F}_T(x)$ and $\hat{\varphi}_{1-p,T}$ and the results in (ii) and (iii) are true under H_2 and H_3 as long as $\Xi_T \xrightarrow{p} \Xi_0$. When $\Xi_T = \hat{\Sigma}_{T+m}^{-1}$, Lemma 2 implies that the results in (ii) and (iii) are true if $\omega_T = o_p(T^{1/2})$.

The asymptotic results under the alternative hypotheses are stated in the following theorem.

THEOREM 2. *Suppose that Assumptions A and 1 hold. Then, as $T \to \infty$ with m fixed, the following asymptotic results hold:*

(i) $|P(\hat{\alpha}, \Xi_T) - P_{T+1}(\alpha_0, \gamma_0, \Xi_0)| = o_p(1)$ *under* H_2, *if* $\Xi_T \xrightarrow{p} \Xi_0$;

(ii) $P(\hat{\alpha}, \Xi_T) \xrightarrow{p} \infty$ *and* $\Pr(P(\hat{\alpha}, \Xi_T) > \hat{\varphi}_{1-p,T}) \to 1$ *under* H_3, *if* $\Xi_T \xrightarrow{p} \Xi_0$;

(iii) $P(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1}) \xrightarrow{p} \infty$ *and* $\Pr(P(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1}) > \hat{\varphi}_{1-p,T}) \to 1$ *under* H_3.

Both (i) and (ii) obviously apply to the case with $\Xi_T = I$. When $\Xi_T = \hat{\Sigma}_{T+m}^{-1}$, Lemma 2 implies that they apply if $\omega_T = o_p(T^{1/2})$. The asymptotic behavior of the $P(\hat{\alpha}, \Xi_T)$ test when Ξ_T is not convergent is somewhat specific to the chosen Ξ_T, and (iii) states the result for $\Xi_T = \hat{\Sigma}_{T+m}^{-1}$. Note that the $P(\hat{\alpha}, \Xi_T)$ test is not a consistent test under H_2 but it becomes a consistent test under H_3.

2.3. Simulations

The DGP used for the simulation is as follows:

$$y_t = \beta_0 + \beta_1 t + z_{1,t}'\beta_2 + z_{2,t}'\beta_3 + u_t,$$

where $z_{1,t}$ is a (2×1) vector of integrated regressors and $z_{2,t}$ is a (2×1) vector of mean zero stationary regressors. For the first experiment, u_t, $z_{1,t} - z_{1,t-1}$, and $z_{2,t}$ are AR(1) processes defined by

$$u_t = (1 - \rho_D)^{1/2}\psi_t + \rho_D^{1/2}\eta_t,$$

$$z_{1,t} - z_{1,t-1} = (1 - \rho_D)^{1/2}\xi_t + \rho_D^{1/2}\eta_t(1, 1)',$$

$$z_{2,t} = \alpha z_{2,t-1} + z_{2,t}^*$$

with the innovations $\psi_t = \alpha\psi_{t-1} + \psi_t^*$, $\eta_t = \alpha\eta_{t-1} + \eta_t^*$, and $\xi_t = \alpha\xi_{t-1} + \xi_t^*$, where $(\psi_t^*, \eta_t^*, \xi_t^*, z_{2,t}^*)$ are i.i.d.$N(0, I)$. The parameter ρ_D dictates the extent of the correlation between the error u_t and the difference of the integrated regressors $z_{1,t}$. The values of ρ_D are set at 0, 0.4, and 0.8. For the size simulations, $(\beta_0, \beta_1, \beta_2', \beta_3') = 0$ for $1 \le t \le T + m$. For the power simulations, $(\beta_0, \beta_1, \beta_2', \beta_3') = 0$ for $1 \le t \le T$ and $(\beta_0, \beta_1, \beta_2', \beta_3')$ is proportional to a vector of ones and $||(\beta_0, \beta_1, \beta_2', \beta_3')'|| = 0.1$ for $t \ge T + 1$. Four tests are simulated: the $P(\hat{\alpha}, I)$ and $P(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1})$ tests proposed in this paper and the $P(0, I)$ and $P(0, \hat{\Sigma}_{T+m}^{-1})$ tests proposed by Andrews and Kim (2006). The sample size is 100, and the number

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

of replications is 5,000. The value of α ranges from 0 to 0.975 in increments of 0.1 when $\alpha \leq 0.9$ and in increments of 0.025 when $\alpha > 0.9$. The values of m are 2, 5, and 10.

The second experiment specifies the errors to be an ARMA(1,1) process. More precisely,

$$u_t = \alpha u_{t-1} + \psi_t^* + b\psi_{t-1}^*,$$

$$z_{1,t} - z_{1,t-1} = \xi_t,$$

$$z_{2,t} = \alpha z_{2,t-1} + z_{2,t}^*$$

with $\xi_t = \alpha\xi_{t-1} + \xi_t^*$ and $(\psi_t^*, \eta_t^*, \xi_t^*, z_{2,t}^*)$ being i.i.d.$N(0, I)$. The values of b are 0.5 and -0.5, and the other parameter values are the same as in the first experiment.

Figures 1a–c report the results from the first experiment with each figure corresponding to a different value of ρ_D, and Figures 1d and e are for the second experiment with each figure corresponding to a different value of b.

In all figures, the same pattern is observed. When $m = 2$ and 5, the $P(0, I)$ and $P(0, \hat{\Sigma}_{T+m}^{-1})$ tests are clearly inferior to the $P(\hat{\alpha}, I)$ and $P(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1})$ tests, because the $P(0, I)$ and $P(0, \hat{\Sigma}_{T+m}^{-1})$ tests show large size distortions when $\alpha > 0.7$ and their power functions are lower than those of the $P(\hat{\alpha}, I)$ and $P(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1})$ tests by a large margin when $\alpha > 0.6$. Comparing the $P(\hat{\alpha}, I)$ and $P(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1})$ tests, one can observe that the latter is more conservative and has smaller power but the difference is small. When $m = 10$, it is not obvious which test dominates. For $\alpha_0 \leq 0.6$, the $P(0, I)$ test performs the best. Its power is the highest, and its size is close to the nominal 5%. The power of the $P(\hat{\alpha}, I)$ test is comparable to that of the $P(0, I)$ test, but its size is inflated. On the other hand, for $\alpha \geq 0.7$, the $P(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1})$ test performs the best. Indeed, the $P(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1})$ test is the only one for which the size remains close enough to the nominal 5% level for $\alpha \geq 0.7$. Its power is greater than those of the $P(0, I)$ and $P(0, \hat{\Sigma}_{T+m}^{-1})$ tests in most cases. The power of the $P(\hat{\alpha}, I)$ test is greater than or similar to that of the $P\left(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1}\right)$ test but the former has size distortions. Overall, the $P(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1})$ is the preferred test, although its power is not the best uniformly in α when $m = 10$.

3. LINEAR TIME TREND MODEL AND THE *S* TESTS

3.1. Model and Test Statistics

This section considers the linear time trend model that is given by (1) with $(z_{1,t}, z_{2,t}) = (t, 1)$, $(\gamma_{1,0}, \gamma_{2,0}) = (\beta_0, \mu_0)$, and $(\gamma_{1,t}, \gamma_{2,t}) = (\beta_t, \mu_t)$, i.e.,

$$y_t = \begin{cases} \mu_0 + \beta_0 t + u_t, \text{ for } t = 1, \ldots, T \\ \mu_t + \beta_t t + u_t, \text{ for } t = T+1, \ldots, T+m. \end{cases} \tag{14}$$

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

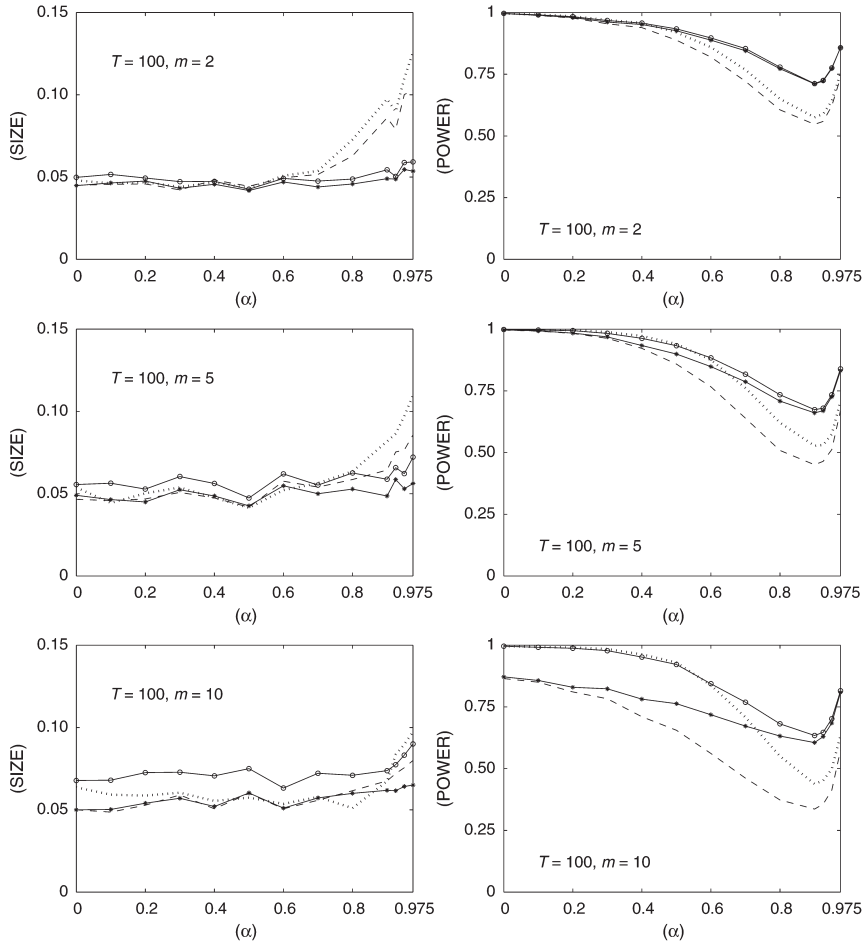


FIGURE 1a. Probabilities of rejecting the null hypothesis, $\rho_D = 0$.
$\left[-\circ\colon P(\hat{\alpha}, I);\ \cdots\colon P(0, I);\ -*\colon P(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1});\ --\colon P\left(0, \hat{\Sigma}_{T+m}^{-1}\right)\right]$

The null and alternative hypotheses are the same as in the previous section. The linear time trend assumption is not crucial. Most of the results presented subsequently can readily be extended to other deterministic functions such as higher order polynomials in time. The linear time trend is used only because it is expositionally simple and of most interest to practitioners. Now, the error process $\{u_t\}$ is allowed to be nonstationary up to one unit root.

Assumption B. $\{u_t\}$ satisfies Assumption A or $\Delta u_t = d(L)e_t$ with $d(L) = \sum_{i=0}^{\infty} d_i L^i$, $\sum_{i=0}^{\infty} i\,|d_i| < \infty$, $d(z) \neq 0$, $\forall |z| \leq 1$, and $\{e_t\}$ is an independently and identically distributed process with $\mathrm{E}(e_t^2) < \infty$.

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

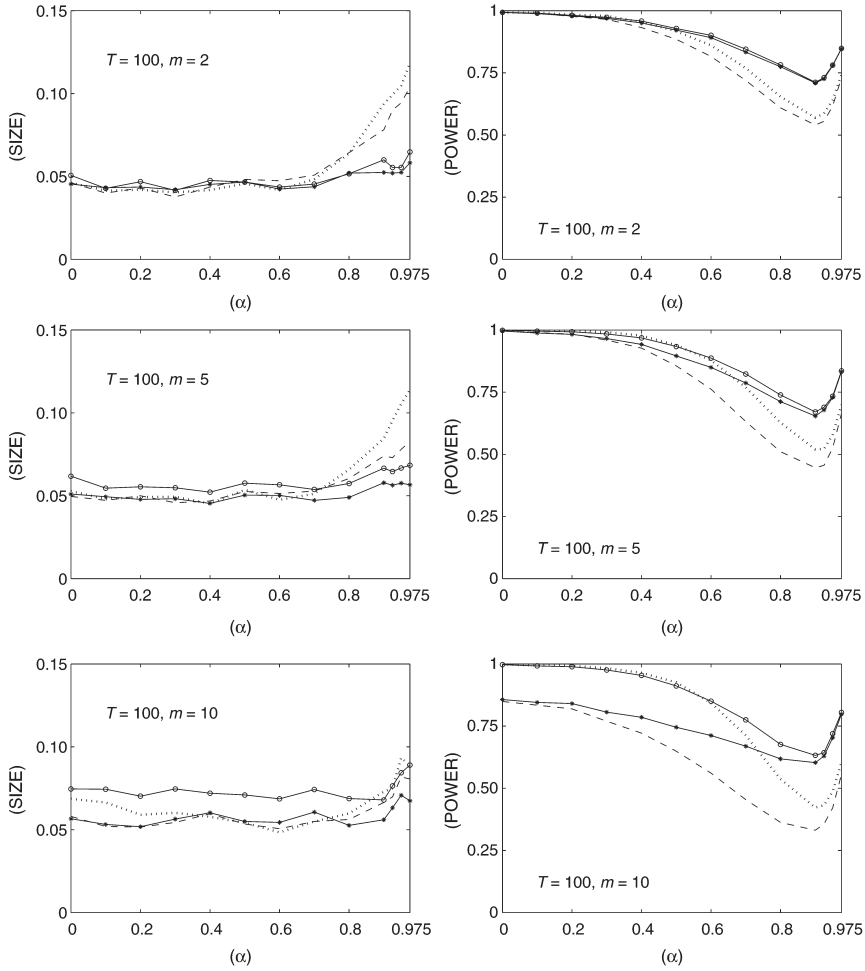


FIGURE 1b. Probabilities of rejecting the null hypothesis, $\rho_D = 0.4$.
$\left[-\circ\text{: } P(\hat{\alpha}, I);\ \cdots\text{: } P(0, I);\ -*\text{: } P\left(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1}\right);\ --\text{: } P\left(0, \hat{\Sigma}_{T+m}^{-1}\right)\right]$

Note that when the second part of Assumption B holds, $\alpha_0 = 1$ from (2) and that $\hat{\alpha}$ in (4) is T-consistent, i.e., $T(\hat{\alpha} - 1) = O_p(1)$. This kind of testing problem arises in various economic analyses. For example, the testing procedure suggested in what follows can be useful for a researcher who wants to test whether the growth rate of an economic variable such as gross domestic product (GDP) has recently changed but does not know whether it is trend stationary or difference stationary.

As in Section 2, consider the quasi-differenced series in (5) and (6) with $w_t(\hat{\alpha}) = (t - \hat{\alpha}(t-1), 1 - \hat{\alpha})'$ for $t = 2, \ldots, T + m$, and $w_1(\hat{\alpha}) = (1, 1)'$. The

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

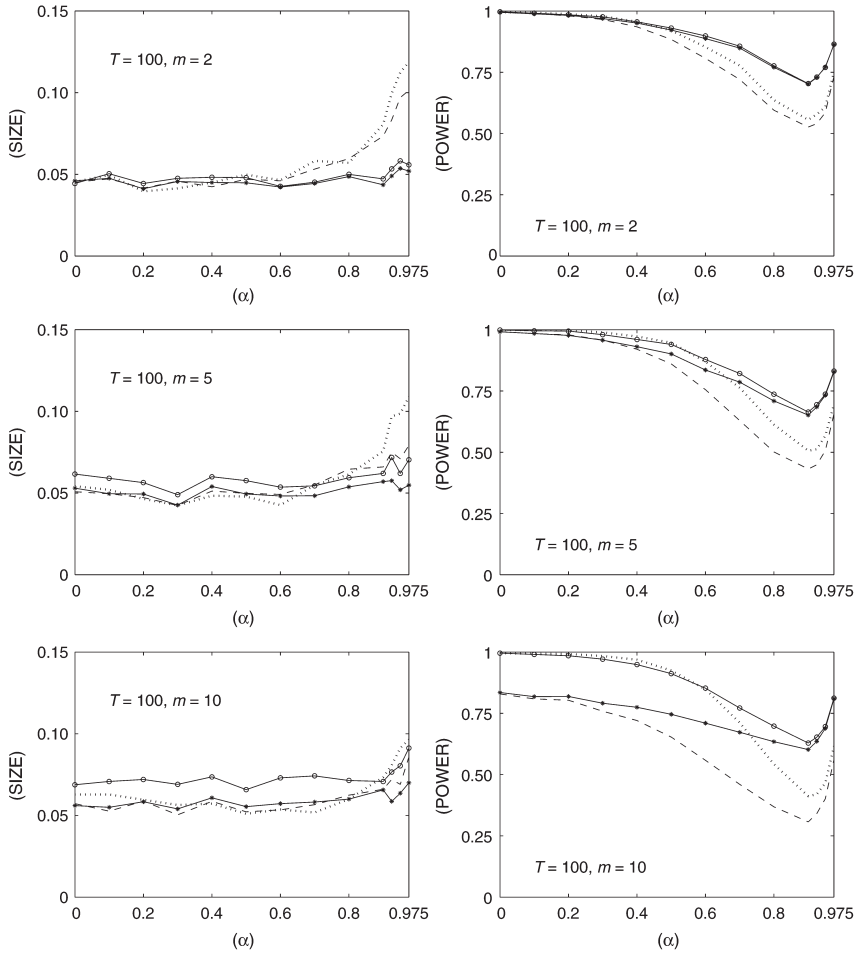


FIGURE 1c. Probabilities of rejecting the null hypothesis, $\rho_D = 0.8$.
$\left[-\circ\colon P\left(\hat{\alpha}, I\right);\ \cdots\colon P(0, I);\ -*\colon P\left(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1}\right);\ --\colon P\left(0, \hat{\Sigma}_{T+m}^{-1}\right)\right]$

regression coefficients $\hat{\gamma}_{T+m}$ and $\hat{\gamma}_{(t)}$ for $2 \leq t \leq T - m + 1$ are again defined by (7) and (8). However, instead of $P_t(\alpha, \gamma, \Xi_T)$ in (10), we consider, for $t = 2, \ldots, T + 1$,

$$S_t(\alpha, \gamma, \Xi_T) = \hat{V}'_{t,t+m-1}(\alpha, \gamma) A_t(\alpha, \Xi_T) \hat{V}_{t,t+m-1}(\alpha, \gamma),$$

with

$$A_t(\alpha, \Xi_T) = \Xi_T W_{t,t+m-1}(\alpha) \left[W'_{t,t+m-1}(\alpha) \Xi_T W_{t,t+m-1}(\alpha)\right]^{-} W_{t,t+m-1}(\alpha) \Xi_T, \tag{15}$$

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

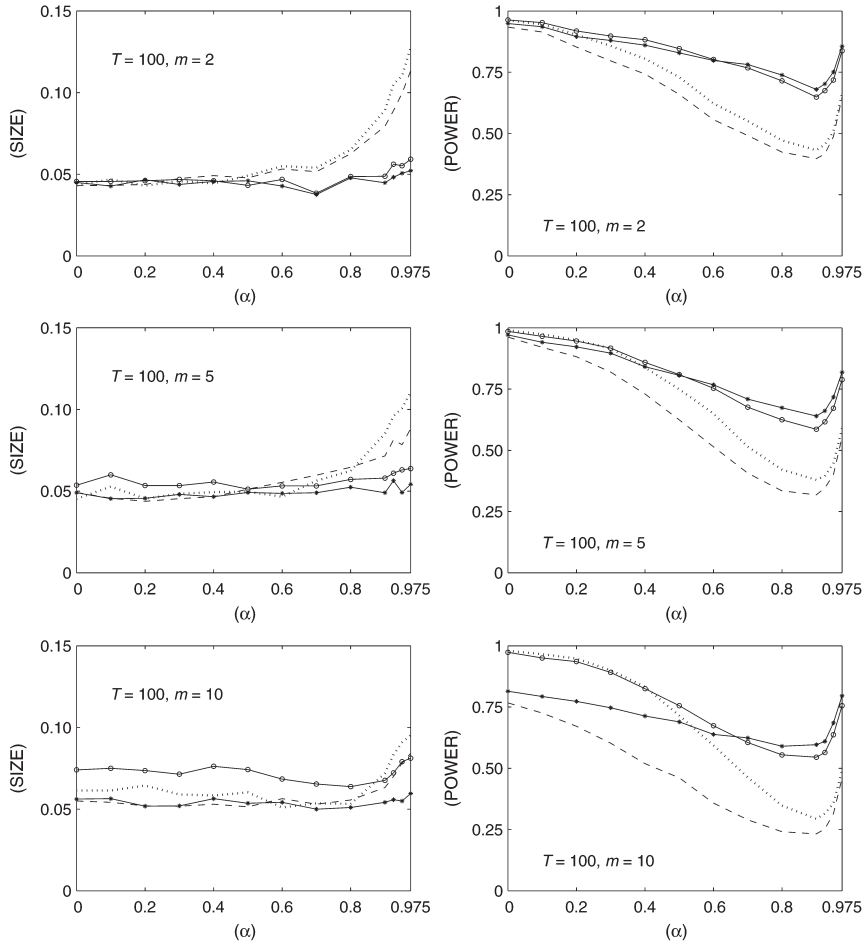


FIGURE 1d. Probabilities of rejecting the null hypothesis, $b = 0.5$.
$\left[-\circ\colon P(\hat{\alpha}, I);\ \cdots\colon P(0, I);\ -*\colon P\left(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1}\right);\ --\colon P\left(0, \hat{\Sigma}_{T+m}^{-1}\right)\right]$

where $[\cdot]^-$ denotes the Moore–Penrose generalized inverse. Recall that $W_{r,s}(\alpha)$ is a stack of $w_t(\hat{\alpha}) = (t - \hat{\alpha}(t-1), 1-\hat{\alpha})'$'s for $t = r, \ldots, s$. Finally, the test statistic $S(\hat{\alpha}, \Xi_T)$ is given by

$$S(\hat{\alpha}, \Xi_T) = S_{T+1}(\hat{\alpha}, \hat{\gamma}_{T+m}, \Xi_T).$$

Some motivation for this particular form of the statistic is given in Section 3.2. The empirical distribution function and the critical values are defined as in Section 2. Note that $A_t(\alpha, \Xi_T)$ is the projection matrix onto the space spanned by the columns of $\Xi_T^{1/2} W_{t,t+m-1}(\alpha)$, pre- and postmultiplied by $\Xi_T^{1/2}$. Given α

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

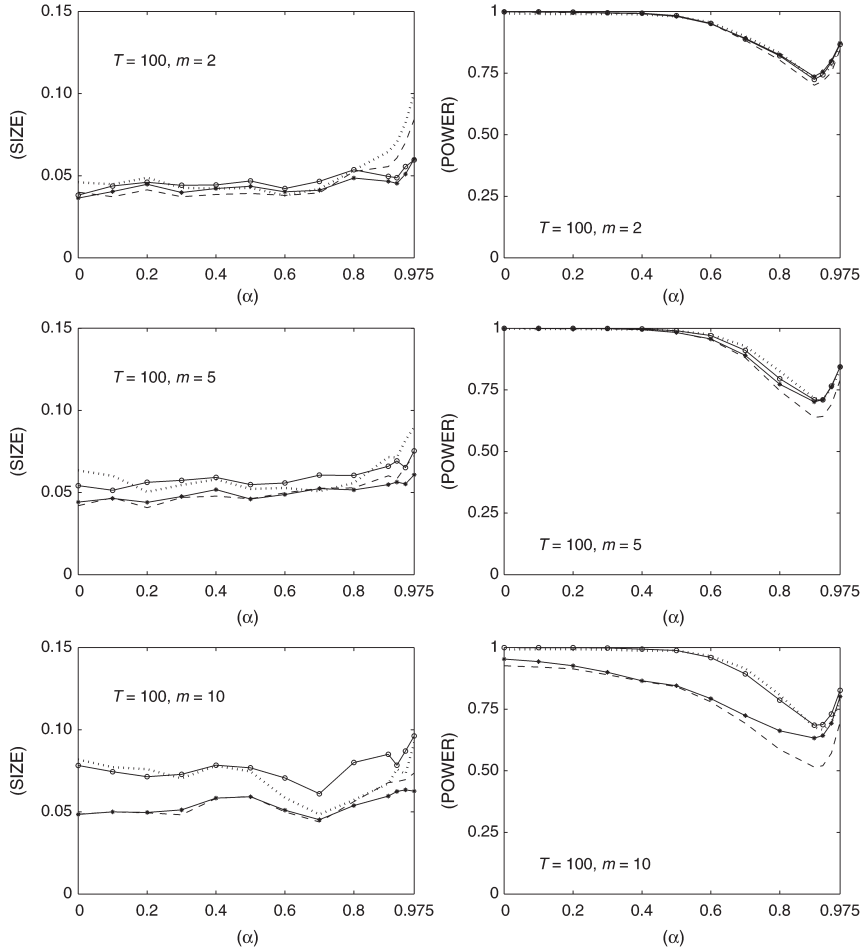


FIGURE 1e. Probabilities of rejecting the null hypothesis, $b = -0.5$.
$\left[-\circ\colon P(\hat{\alpha}, I);\ \cdots\colon P(0, I);\ -*\colon P\left(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1}\right);\ --\colon P\left(0, \hat{\Sigma}_{T+m}^{-1}\right)\right]$

and $\Xi_T^{1/2}$, the range space of $\Xi_T^{1/2} W_{t,t+m-1}(\alpha)$ does not depend on t because the regressors are $(t, 1)$, and thus $A_t(\alpha, \Xi_T)$ is invariant to the time index. Therefore, we can simply use $A(\alpha, \Xi_T)$ in the definition of $S_t(\alpha, \gamma, \Xi_T)$, i.e.,

$$S_t(\alpha, \gamma, \Xi_T) = \hat{V}'_{t,t+m-1}(\alpha, \gamma) A(\alpha, \Xi_T) \hat{V}_{t,t+m-1}(\alpha, \gamma).$$

Again, the two tests of interest are $S(\hat{\alpha}, \Xi_T)$ with $\Xi_T = I$ and $\Xi_T = \hat{\Sigma}_{T+m}^{-1}$. The proposed $S(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1})$ test is the FQGLS version of the Andrews (2002) S test.

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

3.2. Motivation for the $S(\hat{\alpha}, \Xi_T)$ Tests

One can see the motivation for using the weighting matrix $A_t(\alpha, \Xi_T)$ from the following heuristic argument. For simplicity, ignore the serial correlation in u_t for a moment and assume that γ_0 and the distribution of $\{u_t\}$ are known. Suppose that the deterministic part of the process $\{y_t\}$ for $T+1 \le t \le T+m$ is still linear in time, even though the intercept and the slope may have changed, i.e., $\gamma_t = \gamma_1$, for $T+1 \le t \le T+m$ and one wants to test if $\gamma_1 = \gamma_0$. Then, one way to test this hypothesis is to compute the vector of the predicted residuals $\hat{U}_{T+1,T+m}$ by

$$\hat{U}_{T+1,T+m} = Y_{T+1,T+m} - Z_{T+1,T+m}\gamma_0 = Z_{T+1,T+m}(\gamma_1 - \gamma_0) + U_{T+1,T+m}$$

and to measure its norm. If the hypothesis is correct, $\hat{U}_{T+1,T+m} = U_{T+1,T+m}$, the distribution of which is known. If not, the distribution of $\hat{U}_{T+1,T+m}$ will differ by the term $Z_{T+1,T+m}(\gamma_1 - \gamma_0)$. Hence, a legitimate test procedure can be established by measuring the norm of $\hat{U}_{T+1,T+m}$. It is natural to ask which norm is most useful. Let P_Z be the projection matrix associated with $Z_{T+1,T+m}$. Note that the norm $||P_Z \cdot ||$ preserves the magnitude of the change because

$$||P_Z Z_{T+1,T+m}(\gamma_1 - \gamma_0)|| = ||Z_{T+1,T+m}(\gamma_1 - \gamma_0)||,$$

whereas it reduces the magnitude of the errors because $||P_Z U_{T+1,T+m}|| \le ||U_{T+1,T+m}||$, by the properties of projection matrices. As a result, the power of the test is expected to be higher when $||P_Z \cdot ||$ is used than when the regular euclidean $||\cdot||$ norm is used. In fact, note that

$$P_Z \hat{U}_{T+1,T+m} = Z_{T+1,T+m}\hat{\gamma}_1 - Z_{T+1,T+m}\gamma_0,$$

where $\hat{\gamma}_1 = (Z'_{T+1,T+m}Z_{T+1,T+m})^{-1}Z'_{T+1,T+m}Y_{T+1,T+m}$ and

$$\begin{aligned} ||P_Z \hat{U}_{T+1,T+m}||^2 &= (\hat{\gamma}_1 - \gamma_0)' Z'_{T+1,T+m} Z_{T+1,T+m}(\hat{\gamma}_1 - \gamma_0) \\ &= (R\hat{\gamma}_1 - r)'[R(Z'_{T+1,T+m}Z_{T+1,T+m})^{-1}R']^{-1}(R\hat{\gamma}_1 - r), \end{aligned}$$

where $R = I$ and $r = \gamma_0$. The last expression is simply the numerator of the Wald statistic. When the serial correlation in $\{u_t\}$ and the quasi-difference are taken into account, the proposed weighting matrix $A(\hat{\alpha}, \Xi_T)$ naturally follows. In the deterministic trend model with higher order polynomials in t, $S_t(\alpha, \gamma, \Xi_T)$ can be defined analogously. In the cointegration model considered in the previous section, $S_t(\alpha, \gamma, \Xi_T)$ cannot be used because of the integrated regressors that make the distribution of $A_t(\alpha_0, \Xi_T)$ dependent on t.

3.3. Asymptotic Results for the $S(\hat{\alpha}, \Xi_T)$ Tests

Let $v_t = (1 - \alpha_0 L)u_t$. We make the following assumptions.

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

Assumption 2.

(a) Assumption 1(a) holds with $\zeta_t = \upsilon_t$ and $\Sigma_0 = \mathrm{E}V_{2,m+1}V_{2,m+1}' > 0$.

(b) $\mathrm{E}|\upsilon_t|^{2+2\delta} < \infty$, for some $\delta > 0$, and $\max_{T+1\le t\le T+m}\left|T^{-1/2}\sum_{j=1}^{t}\upsilon_j\right| = O_p(1)$.

(c) The distribution of $S_2(\alpha_0, \gamma_0, \Xi_0)$ is continuous and increasing at its $1-p$ quantile.

The next lemma is analogous to Lemma 1.

LEMMA 3. *Suppose that Assumptions B and 2 hold. Partition* $\hat{\gamma}_{T+m}$ *and* $\hat{\gamma}_{(t)}$ *such that* $\hat{\gamma}_{T+m} = (\hat{\beta}_{T+m}, \hat{\mu}_{T+m})$ *and* $\hat{\gamma}_{(t)} = (\hat{\beta}_{(t)}, \hat{\mu}_{1,(t)})$. *Then, when* $|\alpha_0| < 1$, $T^{-1}|\hat{\beta}_{T+m} - \beta_0| = O_p(\omega_T T^{-1}) + o_p(1)$, $|\hat{\mu}_{T+m} - \mu_0| = O_p(\omega_T T^{-1}) + o_p(1)$, $\max_{t\le T-m+1} T^{-1}|(\hat{\beta}_{(t)} - \beta_0)| = o_p(1)$, *and* $\max_{t\le T-m+1}|\hat{\mu}_{(t)} - \mu_0| = o_p(1)$. *When* $\alpha_0 = 1$, $|\hat{\beta}_{T+m} - \beta_0| = O_p(\omega_T T^{-1}) + o_p(1)$, $|\hat{\mu}_{T+m} - \mu_0| = O_p(\omega_T T^{-1}) + O_p(1)$, $\max_{t\le T-m+1}|\hat{\beta}_{(t)} - \beta_0| = o_p(1)$, *and* $\max_{t\le T-m+1}|\hat{\mu}_{(t)} - \mu_0| = O_p(1)$.

The consistency of the variance-covariance matrix estimate defined in (12) also holds when $\alpha_0 = 1$. The next lemma states the relevant result.

LEMMA 4. *Suppose that Assumptions B and 2 hold. Then,* $\hat{\Sigma}_{T+m} \xrightarrow{p} \Sigma_0$, *if* $\omega_T = o_p(T^{1/2})$.

Let $F(x)$ and φ_{1-p} denote the distribution function of $S_2(\alpha_0, \gamma_0, \Xi_0)$ and its $1-p$ quantile. The asymptotic results pertaining to the $S(\hat{\alpha}, \Xi_T)$ test are stated in the next two theorems. They are analogous to Theorems 1 and 2, respectively.

THEOREM 3. *Suppose that Assumptions B and 2 hold. Then, as* $T \to \infty$ *with* m *fixed, the following asymptotic results hold under* H_0 *if* $\Xi_T \xrightarrow{p} \Xi_0$:

(i) $S(\hat{\alpha}, \Xi_T) \xrightarrow{d} S_2(\alpha_0, \gamma_0, \Xi_0)$;

(ii) $\hat{F}_T(x) \xrightarrow{p} F(x)$ *for all* x *in a neighborhood of* φ_{1-p};

(iii) $\hat{\varphi}_{1-p,T} \xrightarrow{p} \varphi_{1-p}$;

(iv) $\Pr(S(\hat{\alpha}, \Xi_T) > \hat{\varphi}_{1-p,T}) \to p$.

Again, (ii) and (iii) hold under H_2 and H_3 as long as $\Xi_T \xrightarrow{p} \Xi_0$.

THEOREM 4. *Suppose that Assumptions B and 2 hold. Then, as* $T \to \infty$ *with* m *fixed, the following asymptotic results hold:*

(i) $|S(\hat{\alpha}, \Xi_T) - S_{T+1}(\alpha_0, \gamma_0, \Xi_0)| = o_p(1)$ *under* H_2, *if* $\Xi_T \xrightarrow{p} \Xi_0$;

(ii) $S(\hat{\alpha}, \Xi_T) \xrightarrow{p} \infty$ *and* $\Pr(S(\hat{\alpha}, \Xi_T) > \hat{\varphi}_{1-p,T}) \to 1$ *under* H_3, *if* $\Xi_T \xrightarrow{p} \Xi_0$;

(iii) $S(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1}) \xrightarrow{p} \infty$ *and* $\Pr(S(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1}) > \hat{\varphi}_{1-p,T}) \to 1$ *under* H_3.

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

As in the cointegration model, the $S(\hat{\alpha}, \Xi_T)$ test is not consistent under H_2 but is so under H_3, if $\Xi_T \xrightarrow{p} \Xi_0$. The consistency of the $S(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1})$ test under H_3 is also established. Examples for H_2 include a mean shift and a joint broken trend, i.e., $\mu_t = \mu_1 \neq \mu_0$ and $\beta_t = \beta_0 + (1 - T/t)\beta_1$ for $t > T$. When y_t in (14) is subject to these breaks, an alternative representation of y_t is

$$y_t = \mu_0 + \beta_0 t + \mu_1 DU_t + \beta_1 DT_t + u_t,$$

where $DU_t = 1$ and $DT_t = (t - T)$ for $t > T$ and 0 elsewhere. An example for H_3 is the disjoint broken trend, i.e., $\beta_t = \beta_1$, in which case an alternative representation of y_t is

$$y_t = \mu_0 + \beta_0 t + \beta_1 DT_t^* + u_t,$$

where $DT_t^* = t$ for $t > T$ and 0 elsewhere.

It is very popular in the econometrics literature to model a stochastic process with an autoregressive root local to unity to approximate heavily serially correlated processes. Because the error process in the deterministic trend model often shows substantial serial correlation, it is of great interest to show that the asymptotic results for the $S(\hat{\alpha}, \Xi_T)$ test still hold under the local-to-unity framework. We make the following assumption to adopt the standard local-to-unity asymptotics.

Assumption C. $\{u_t\}$ is such that $u_t = \alpha_0 u_{t-1} + v_t$, $\alpha_0 = 1 + c_0 T^{-1}$, $v_t = d(L)e_t$ with $d(L) = \sum_{i=0}^{\infty} d_i L^i$, $\sum_{i=0}^{\infty} i|d_i| < \infty$, $d(z) \neq 0$, $\forall |z| \leq 1$ and $\{e_t\}$ is an independently and identically distributed process with $\mathrm{E}(e_t^2) < \infty$.

Note that α_0 does not necessarily stand for the sum of the autoregressive parameters under Assumption C. However, $\hat{\alpha}$ from (4) is still convergent to unity at rate T, i.e., $T(\hat{\alpha} - 1) = O_p(1)$, though the exact limiting distribution is different from the unit root case. In fact, what matters to show the results in Theorems 3 and 4 is the T-consistency of $\hat{\alpha}$, not its limiting distribution, and this strongly suggests that the same results will hold under the local-to-unity asymptotics. The relevant results are stated in the following theorem.

THEOREM 5. *Suppose that Assumptions C and 2 hold. Then, as $T \to \infty$ with m fixed, the statements in Theorems 3 and 4 hold.*

3.4. Finite-Sample Refinement

As far as $\hat{\alpha}$ is concerned, only the consistency at the specified rate is required for the asymptotic validity of the test procedure. The finite-sample performance of the $S(\hat{\alpha}, \Xi_T)$ test can, however, be improved by using a modified estimate. It is well known that $\hat{\alpha}$ is biased. Section 3.5 will show that the use of the Andrews (1993) or Andrews and Chen (1994) median unbiased estimator improves the finite-sample properties of the $S(\hat{\alpha}, \Xi_T)$ test. The precise steps are as follows.

1. In (4), use the modified Akaike information criterion proposed by Ng and Perron (2001) to select k and get $\hat{\alpha}$ by OLS.

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

2. Compute $\hat{\alpha}_{MUE}$, the median unbiased estimate of $\hat{\alpha}$. When $k = 0$ is selected, use the method of Andrews (1993) and, when $k > 0$, use the method of Andrews and Chen (1994).

Note that the median unbiased estimator is not considered for the cointegration model because it is currently developed only for the autoregressive models with an intercept and time trend.

3.5. Simulations

The DGP for the simulations is as stated in (14). The innovation process $\{v_t\}$ is an ARMA(1,1) process with

$$u_t = \alpha_0 u_{t-1} + v_t,$$

$$v_t = a_1 v_{t-1} + e_t + b_1 e_{t-1},$$

where the value of α_0 ranges from 0 to 1 in increments of 0.1 when $\alpha_0 \leq 0.9$ and in increments of 0.025 when $\alpha_0 > 0.9$ and $(a_1, b_1) = (0,0)$, $(0.5,0)$, $(-0.5,0)$, $(0,0.5)$, or $(0,-0.5)$ and e_t is i.i.d.$N(0,1)$. For the size simulations, the linear trend, say, d_t, is set to zero without loss of generality, i.e., $d_t = 0$ for $1 \leq t \leq T+m$, whereas, for the power simulation, $d_t = 0$ for $1 \leq t \leq T$ and $0.5 + 0.5(t-T)$ for $t \geq T+1$. Six tests are compared: $S(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1})$, $S(\hat{\alpha}_{MUE}, \hat{\Sigma}_{T+m}^{-1})$, $S(\hat{\alpha}, I)$, $S(\hat{\alpha}_{MUE}, I)$, $P(0, I)$, and $P(\hat{\alpha}, I)$. Of greatest interest are the four S tests with the other two tests serving as benchmarks. The sample size is 100, and the number of replications is 5,000. The values of m are 2, 5, and 10.

Figures 2a–e report the results with each panel corresponding to a different specification of v_t. The results are basically the same across the different specifications of v_t. It is easily observed that the $P(0, I)$ and $P(\hat{\alpha}, I)$ tests are inferior to the S tests, because the $P(0, I)$ test suffers from large size distortions when $\alpha_0 > 0.7$ and the power function of the $P(\hat{\alpha}, I)$ test is lower than the others in most cases. The $S(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1})$ and $S(\hat{\alpha}, I)$ tests show similar properties. The $S(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1})$ test is more conservative and less powerful than the $S(\hat{\alpha}, I)$, test but the difference in power is small even for $m = 10$. Hence, the $S(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1})$ test is preferred to the $S(\hat{\alpha}, I)$ test. Also, the use of a median unbiased estimator improves the size property with no loss in power. Overall, the test $S(\hat{\alpha}_{MUE}, \hat{\Sigma}_{T+m}^{-1})$ performs the best with $S(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1})$ a close second.

4. EMPIRICAL APPLICATION[3]

Quarterly U.S. government spending is analyzed. Two cointegration equations are considered. The first one is

$$\log S_t = c + \beta_1 \log RGDP_t + \beta_2 \log CPI_t + \alpha' AS_t + \delta W_t + e_t, \quad \textbf{(16)}$$

where S_t is the sum of federal consumption expenditures and gross investment; $RGDP_t$ is the real GDP; CPI_t is the CPI (all urban consumers); AS_t is a (4×1)

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

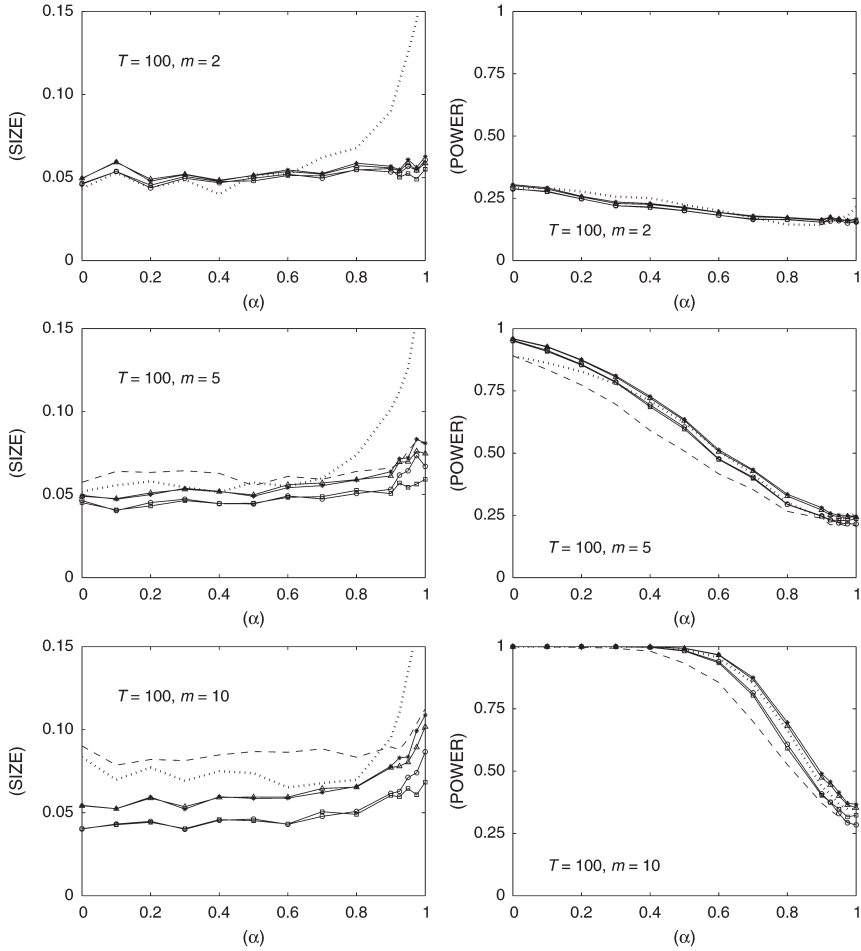


FIGURE 2a. Probabilities of rejecting the null hypothesis, $(a_1, b_1) = (0, 0)$. $\left[-\circ\colon S(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1});\ -\square\colon S(\hat{\alpha}_{MUE}, \hat{\Sigma}_{T+m}^{-1});\ --\colon P(\hat{\alpha}, I);\ \cdots\colon P(0, I);\ -*\colon S(\hat{\alpha}, I);\ -\triangle\colon S_I(\hat{\alpha}_{MUE}, I)\right]$

vector that contains the proportions of four age groups (20–24, 25–44, 45–64, and 65+); and W_t is a dummy variable for the Vietnam War (1964.I–1972.IV) and the Iraq War (2003.I–IV). For AS_t, quarterly data are not available, and thus each annual value is used for four quarters.[4] The second equation includes an additional variable P_t, which is a dummy variable for Republican presidents:

$$\log S_t = c + \beta_1 \log RGDP_t + \beta_2 \log CPI_t + \alpha' AS_t + \delta W_t + \gamma P_t + e_t. \quad \textbf{(17)}$$

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

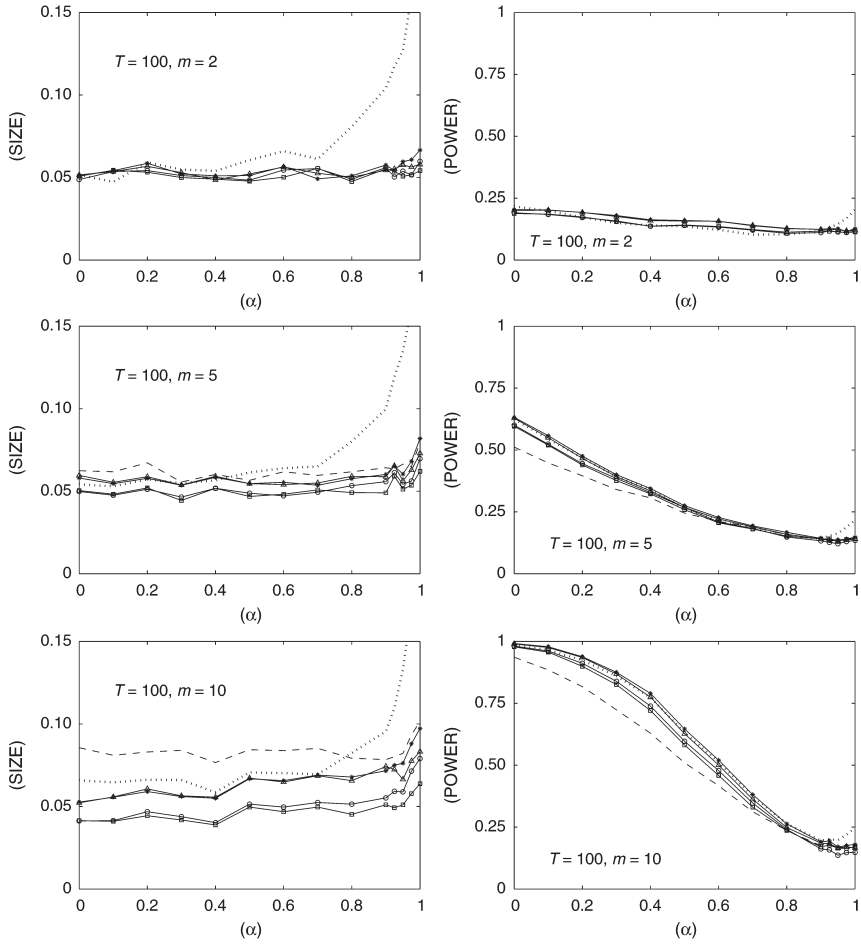


FIGURE 2b. Probabilities of rejecting the null hypothesis, $(a_1, b_1) = (0.5, 0)$. $\left[-\circ: S(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1});\ -\square: S(\hat{\alpha}_{MUE}, \hat{\Sigma}_{T+m}^{-1});\ --: P(\hat{\alpha}, I);\ \cdots: P(0, I);\ -*: S(\hat{\alpha}, I),\ -\triangle: S_I(\hat{\alpha}_{MUE}, I)\right]$

The aim is to see if U.S. government spending after 2001 (President G. W. Bush's administration) deviates from the past trend specified by the preceding cointegration regression. The prebreak sample is from 1954.I to 2000.IV ($T = 188$), and the number of postbreak observations m is set to 8–16, which is about 5–10% of the prebreak sample size.

Figure 3 shows the plots of the actual and estimated federal government spending. The estimated spending is obtained from (16) and (17) by applying the FQGLS procedure to the full sample. Because the two equations yield very similar plots, we present only the one from (17).

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

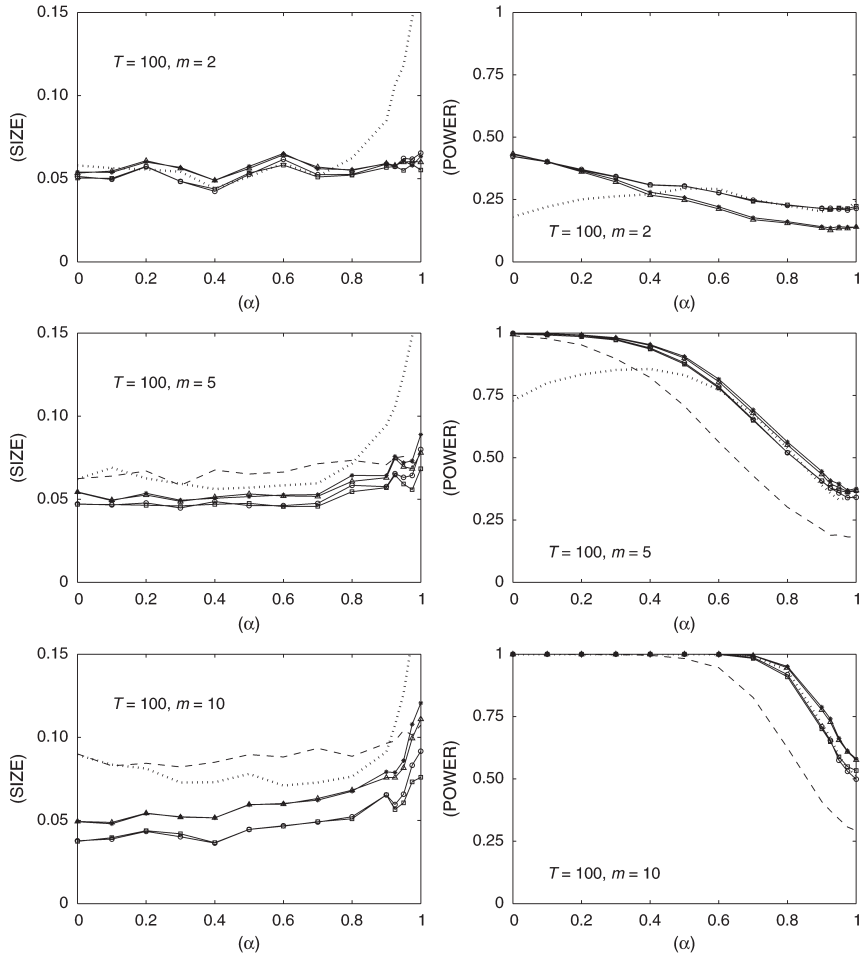


FIGURE 2c. Probabilities of rejecting the null hypothesis, $(a_1, b_1) = (-0.5, 0)$.
$\left[-\circ\colon S(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1});\ \ -\square\colon S(\hat{\alpha}_{MUE}, \hat{\Sigma}_{T+m}^{-1});\ \ --\colon P(\hat{\alpha}, I);\ \ \cdots\colon P(0, I);\ \ -*\colon S(\hat{\alpha}, I);\ \ -\triangle\colon S_I(\hat{\alpha}_{MUE}, I) \right]$

The differences between the solid and dashed lines are similar to the predicted errors used for the end-of-sample instability test but are not exactly the same because of the particular testing procedure described in the earlier section. The vertical line in the figure splits the sample into the pre- and postbreak periods. The testing problem can be restated as "Are the deviations after 2001 significantly larger than those before 2001?"

The ADF t-statistics (with an intercept and slope) for $\log S_t$, $\log RGDP_t$, and $\log CPI_t$ are -0.83, -2.76 and -2.28, respectively, which suggests that these variables are integrated. The series AS_t is a very slowly evolving process, and,

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

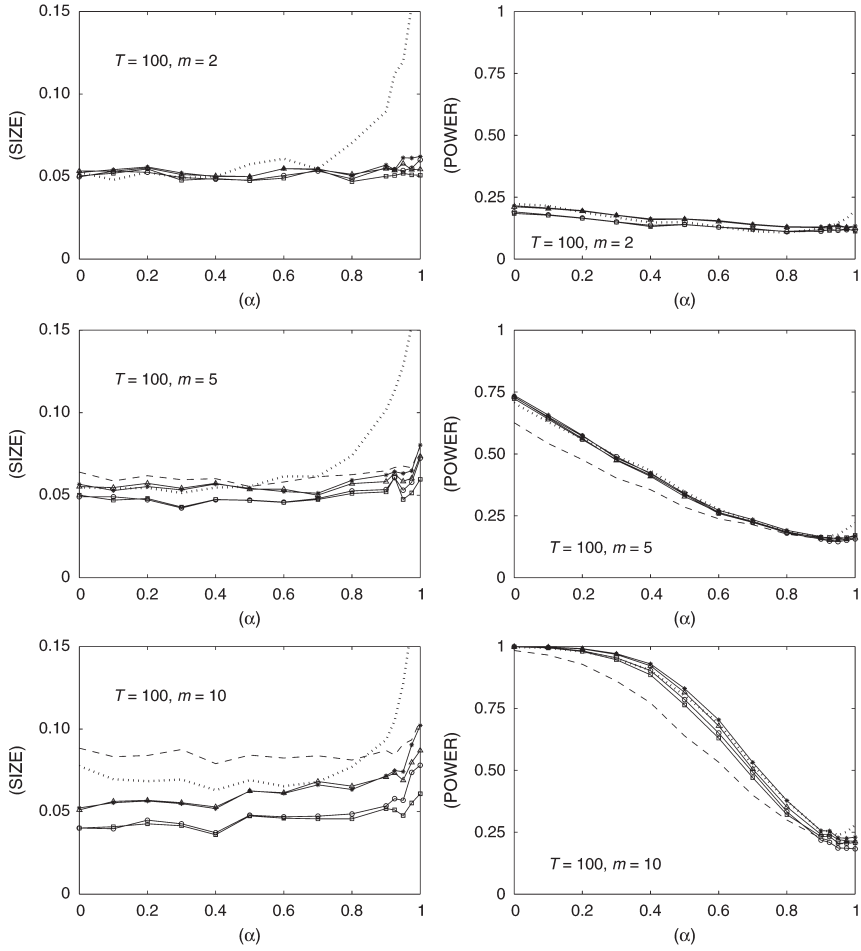


FIGURE 2d. Probabilities of rejecting the null hypothesis, $(a_1, b_1) = (0, 0.5)$. $\left[-\circ: S\left(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1}\right);\ -\square: S\left(\hat{\alpha}_{MUE}, \hat{\Sigma}_{T+m}^{-1}\right);\ --: P(\hat{\alpha}, I);\ \cdots: P(0, I);\ -*: S(\hat{\alpha}, I);\ -\triangle: S_I(\hat{\alpha}_{MUE}, I)\right]$

given the size of the sample (less than 50 years), it is difficult to decide its distributional characteristics using a statistical test. Hence, AS_t is simply assumed not to be integrated. To see if a cointegration relationship holds for the prebreak observations, the ADF t-statistic has been computed from the OLS residuals obtained from the preceding regressions using the observations up to 2001, and the values are -4.74 and -4.75, respectively, for (16) and (17), which are small enough to reject the null of no cointegration at the 1% significance level.

For the instability of the cointegration after 2001, only the results for the $P\left(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1}\right)$ test are presented, because it has the best size and power properties. The estimated p-values for each m are presented in Table 1.

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

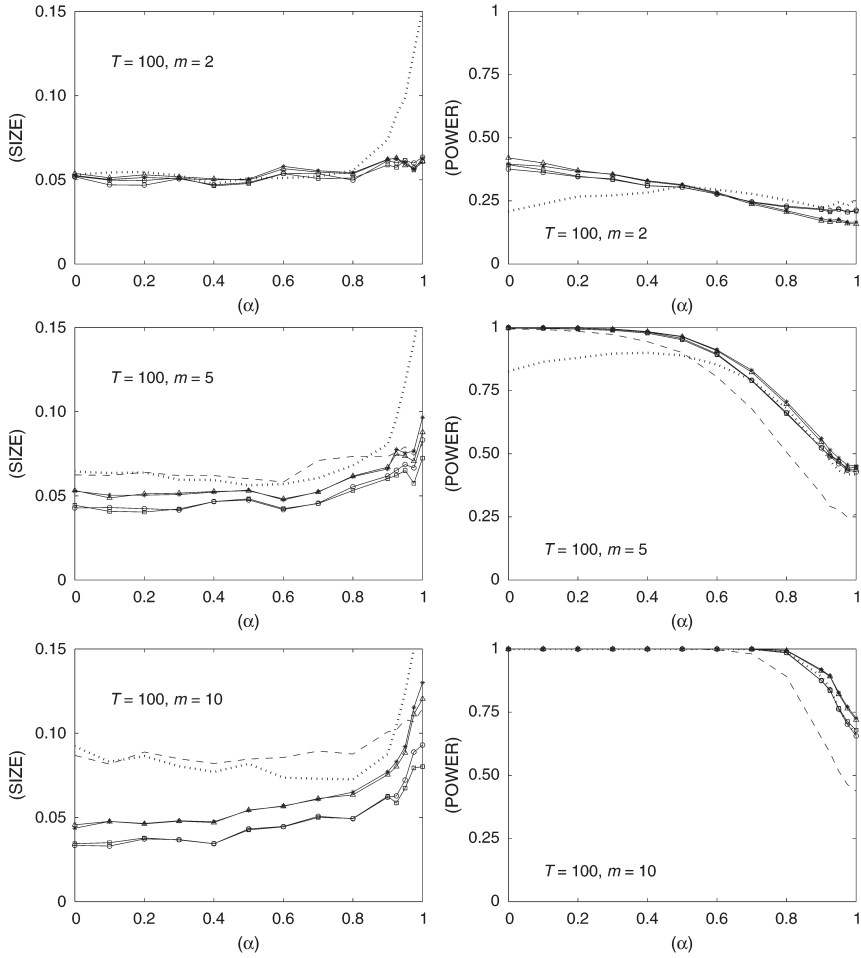


FIGURE 2e. Probabilities of rejecting the null hypothesis, $(a_1, b_1) = (0, -0.5)$.
$\left[-\circ\colon S(\hat{\alpha}, \hat{\Sigma}_{T+m}^{-1});\quad -\square\colon S(\hat{\alpha}_{MUE}, \hat{\Sigma}_{T+m}^{-1});\quad --\colon P(\hat{\alpha}, I);\quad \cdots\colon P(0, I);\quad -*\colon S(\hat{\alpha}, I);\quad -\triangle\colon S_I(\hat{\alpha}_{MUE}, I) \right]$

For equation (16), the null hypothesis that the cointegration is maintained after 2001 is rejected at the 5% significance level for $m = 13$ and 14 and at the 15% significance level for all $m \geq 10$. For equation (17), the null hypothesis is rejected at the 5% significance level for $m = 13$, 14 and 15, and at the 15% significance level for all $m \geq 10$ except for $m = 12$. These test results seem to suggest that the government spending after the Bush administration significantly deviates from the past trend.

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

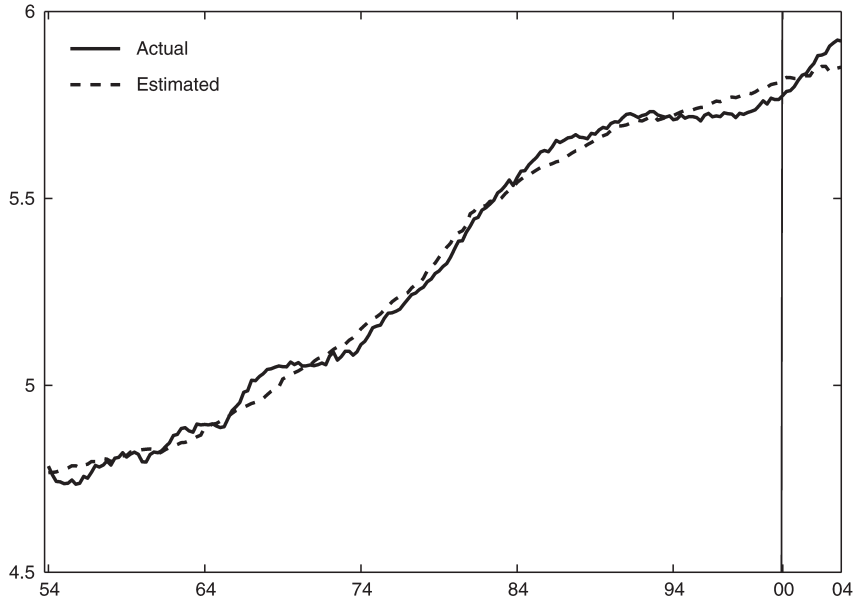


FIGURE 3. Quarterly U.S. government spending, 1954.I–2004.IV.

5. CONCLUSION

This paper proposes the $P(\hat{\alpha}, \Xi_T)$ tests and the $S(\hat{\alpha}, \Xi_T)$ tests for the cointegration model and the linear time trend model. The simulation study shows that the $P(\hat{\alpha}, \Xi_T)$ tests have better finite-sample properties than the Andrews and Kim (2006) original P test. The $S(\hat{\alpha}, \Xi_T)$ tests are shown to be robust to potentially nonstationary serial correlation. The simulation study also shows that the finite-sample properties of the $S(\hat{\alpha}, \Xi_T)$ tests can be further improved when the Andrews (1993) or Andrews and Chen (1994) median unbiased estimate of the first-order autoregressive coefficient is used. The proposed tests should be useful in practice.

TABLE 1. Estimated p-values

	m									
	8	9	10	11	12	13	14	15	16	ADF
Eqn. (16)	0.217	0.190	0.090	0.130	0.142	0.017	0.035	0.052	0.064	−4.74
Eqn. (17)	0.217	0.212	0.101	0.141	0.153	0.006	0.035	0.035	0.052	−4.75

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

NOTES

1. For a recent review, see Perron (2006).
2. Andrews and Kim (2006) also stated that the consistency of their OLS estimates of γ_0 holds under H_1 (see their Lemma A1). However, this is imprecise, because the terms in their (B.23) are not enough to capture the difference between $(\hat{\gamma}_{l,1:T} - \gamma_{l,0})$ and $(\hat{\gamma}_{l,1:T+m} - \gamma_{l,0})$ in terms of their notation.
3. I thank Professor Laurence J. Kotlikoff for his suggestions related to this empirical application.
4. The data for S_t and $RGDP_t$ are from the National Income and Product Accounts, the data for CPI_t are from the Bureau of Labor Statistics, and the data for AS_t are from the Economic Report of the President.

REFERENCES

Andrews, D.W.K. (1993) Exactly median-unbiased estimator of first order autoregressive/unit root models. *Econometrica* 61, 139–165.
Andrews, D.W.K. (2002) End-of-sample instability tests. *Econometrica* 71, 1661–1694.
Andrews, D.W.K. & H.Y. Chen (1994) Approximately median-unbiased estimation of autoregressive models. *Journal of Business & Economic Statistics* 12, 187–204.
Andrews, D.W.K. & J. Kim (2006) Tests for cointegration breakdown over a short time period. *Journal of Business & Economic Statistics* 24, 379–394.
Canjels, E. & M.W. Watson (1997) Estimating deterministic trends in the presence of serial correlated errors. *Review of Economics and Statistics* 79, 184–200.
Ng, S. & P. Perron (2001) Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69, 1519–1554.
Perron, P. (2006) Dealing with structural breaks. In K. Patterson & T.C. Mills (eds.), *Palgrave Handbook of Econometrics*, vol. 1: *Econometric Theory*, pp. 278–352. Palgrave Macmillan.
Perron, P. & T. Yabu (2009) Testing for shifts in trend with an integrated or stationary noise component. *Journal of Business & Economic Statistics* 27, 369–396.

APPENDIX

As a matter of notation, let $t \leq T+m$ and $t \leq T-m+1$ denote $1 \leq t \leq T+m$ and $2 \leq t \leq T-m+1$, respectively; $\|\cdot\|$ denotes the 2-norm.

Proof of Lemma 1. First consider $\hat{\gamma}_T$, which is defined by

$\hat{\gamma}_T =$ OLS estimate of γ_0 in the regression of $q_j(\hat{\alpha})$ on $w_j(\hat{\alpha})$

using observations with index $j = 1, \ldots, T$.

Let $A_{11} = T^{-1}\sum_{t=1}^{T} B_T^{-1} w_{1,t}(\hat{\alpha}) w_{1,t}'(\hat{\alpha}) B_T^{-1}$, $A_{22} = T^{-1}\sum_{t=1}^{T} w_{2,t}(\hat{\alpha}) w_{2,t}'(\hat{\alpha})$, $A_{21} = A_{12}'$, $A_{12} = T^{-1}\sum_{t=1}^{T} B_T^{-1} w_{1,t}(\hat{\alpha}) w_{2,t}'(\hat{\alpha})$, $B_1 = T^{-1}\sum_{t=1}^{T} B_T^{-1} w_{1,t}(\hat{\alpha})\hat{v}_t(\hat{\alpha}, \gamma_0)$, and, $B_2 = T^{-1}\sum_{t=1}^{T} w_{2,t}(\hat{\alpha})\hat{v}_t(\hat{\alpha}, \gamma_0)$. Then, $\hat{\gamma}_T = (\hat{\gamma}_{1,T}', \hat{\gamma}_{2,T}')'$ can be written as

$$B_T^{-1}(\hat{\gamma}_{1,T} - \gamma_{1,0}) = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}(B_1 - A_{12}A_{22}^{-1}B_2),$$

$$(\hat{\gamma}_{2,T} - \gamma_{2,0}) = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}(B_2 - A_{21}A_{11}^{-1}B_1). \quad \textbf{(A.1)}$$

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

Usual arguments show that $A_{11} \Rightarrow (1-\alpha_0)^2 \int_0^1 \eta(r)\eta'(r)dr > 0$, $A_{22} \xrightarrow{p} [(1+\alpha_0^2)\Sigma_{2,0} - \alpha_0(\Omega_{2,0}+\Omega'_{2,0})] > 0$, $A_{12} = o_p(1)$, $B_1 = o_p(1)$, and $B_2 = o_p(1)$. The consistency of $B_T^{-1}(\hat{\gamma}_{1,T} - \gamma_{1,0})$ and $(\hat{\gamma}_{2,T} - \gamma_{2,0})$ follows from these. Now, define for $t \le T-m+1$,

$$A_{11}^{(t)} = T^{-1} \sum_{j=t}^{t+[m/2]-1} B_T^{-1} w_{1,j}(\hat{\alpha}) w'_{1,j}(\hat{\alpha}) B_T^{-1},$$

$$A_{12}^{(t)} = T^{-1} \sum_{j=t}^{t+[m/2]-1} B_T^{-1} w_{1,j}(\hat{\alpha}) w'_{2,j}(\hat{\alpha}) \quad \text{with } A_{21}^{(t)} = A_{12}^{(t)\prime},$$

$$A_{22}^{(t)} = T^{-1} \sum_{j=t}^{t+[m/2]-1} w_{2,j}(\hat{\alpha}) w'_{2,j}(\hat{\alpha}),$$

$$B_1^{(t)} = T^{-1} \sum_{j=t}^{t+[m/2]-1} B_T^{-1} w_{1,j}(\hat{\alpha}) \hat{v}_j(\hat{\alpha}, \gamma_0) \quad \text{and}$$

$$B_2^{(t)} = T^{-1} \sum_{j=t}^{t+[m/2]-1} w_{2,j}(\hat{\alpha}) \hat{v}_j(\hat{\alpha}, \gamma_0). \tag{A.2}$$

These terms capture the difference between $\hat{\gamma}_T$ and $\hat{\gamma}_{(t)}$ for $t \le T-m+1$. First, $A_{11}^{(t)}$ is asymptotically negligible uniformly in t because, from Assumption 1(c), we have

$$\begin{aligned}
\max_{t \le T-m+1} ||A_{11}^{(t)}|| &= \max_{t \le T-m+1} T^{-1} \left\| \sum_{j=t}^{t+[m/2]-1} B_T^{-1} w_{1,j}(\hat{\alpha}) w'_{1,j}(\hat{\alpha}) B_T^{-1} \right\| \\
&\le \max_{t \le T-m+1} T^{-1} \left[\frac{m}{2}\right] \left\| B_T^{-1} w_{1,t}(\hat{\alpha}) w'_{1,t}(\hat{\alpha}) B_T^{-1} \right\| \\
&\le \max_{t \le T-m+1} T^{-1} \left[\frac{m}{2}\right] ||B_T^{-1} w_{1,j}(\hat{\alpha})||^2 = o_p(1).
\end{aligned}$$

The moment conditions in Assumption 1(b) imply that $\mathrm{E}\|z_{2,t} z'_{2,t-1}\|^{1+\delta} < \infty$, $\mathrm{E}\|z_{2,t} u_t\|^{1+\delta} < \infty$, $\mathrm{E}\|z_{2,t-1} u_t\|^{1+\delta} < \infty$, $\mathrm{E}\|z_{2,t} u_{t-1}\|^{1+\delta} < \infty$, $\mathrm{E}|v_t|^{2+\delta} < \infty$, $\mathrm{E}|u_t v_t|^{1+\delta} < \infty$, $\mathrm{E}\|z_{2,t} v_t\|^{1+\delta} < \infty$, $\mathrm{E}\|z_{2,t-1} v_t\|^{1+\delta} < \infty$, and $\mathrm{E}\|z_{2,t} v_{t-1}\|^{1+\delta} < \infty$ by Cauchy–Schwarz inequality. Also, as shown by Andrews and Kim (2006), note that, for a sequence of stationary vectors, $\{\zeta_t : t = 1, \ldots, T\}$ with $\mathrm{E}\zeta_t = 0$ and $\mathrm{E}\|\zeta_t\|^{s(1+\delta)} < \infty$, $s > 0$

$$T^{-1/s} \max_{1 \le j \le T} \left\| \sum_{j=t}^{t+[m/2]-1} \zeta_j \right\| = o_p(1). \tag{A.3}$$

For $A_{22}^{(t)}$, note that

$$\max_{t \le T-m+1} \left\| A_{22}^{(t)} \right\|$$

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

$$
\begin{aligned}
&= \max_{t\le T-m+1} T^{-1}\left\|\sum_{j=t}^{t+[m/2]-1}\left[z_{2,j}z_{2,j}' - \hat{\alpha}z_{2,j-1}z_{2,j}' - \hat{\alpha}z_{2,j}z_{2,j-1}'\right.\right.\\
&\qquad\left.\left.+\hat{\alpha}^2 z_{2,j-1}z_{2,j-1}'\right]\right\|\\
&\le \max_{t\le T-m+1} T^{-1}\left\|\sum_{j=t}^{t+[m/2]-1} z_{2,j}z_{2,j}'\right\| + \hat{\alpha}\max_{t\le T-m+1} T^{-1}\left\|\sum_{j=t}^{t+[m/2]-1} z_{2,j-1}z_{2,j}'\right\|\\
&\quad + \hat{\alpha}\max_{t\le T-m+1} T^{-1}\left\|\sum_{j=t}^{t+[m/2]-1} z_{2,j}z_{2,j-1}'\right\|\\
&\quad + \hat{\alpha}^2\max_{t\le T-m+1} T^{-1}\left\|\sum_{j=t}^{t+[m/2]-1} z_{2,j-1}z_{2,j-1}'\right\| = o_p(1),
\end{aligned}
$$

where the last equality holds from (A.3),

$$
\begin{aligned}
&\max_{t\le T-m+1} T^{-1}\left\|\sum_{j=t}^{t+[m/2]-1} z_{2,j}z_{2,j}'\right\|\\
&\quad\le \max_{t\le T-m+1} T^{-1}\left\|\sum_{j=t}^{t+[m/2]-1}\left(z_{2,j}z_{2,j}' - \mathrm{E}z_{2,j}z_{2,j}'\right)\right\| + T^{-1}\left[\frac{m}{2}\right]\|\mathrm{E}z_{2,j}z_{2,j}'\| = o_p(1),
\end{aligned}
$$

and the same argument applies to the other terms. Using Cauchy–Schwarz inequality, it follows that $\max_{t\le T-m+1}\|A_{12}^{(t)}\| = o_p(1)$. Similarly, we also have $\max_{t\le T-m+1}\|B_1^{(t)}\| = o_p(1)$ and $\max_{t\le T-m+1}\|B_2^{(t)}\| = o_p(1)$. Therefore, we obtain that $\max_{t\le T-m+1}\|B_T(\hat{\gamma}_{1,(t)} - \gamma_{1,0})\| = o_p(1)$ and $\max_{t\le T-m+1}\|\hat{\gamma}_{2,(t)} - \gamma_{2,0}\| = o_p(1)$. Note that these results hold under both H_0 and H_1 because the observations after $t = T$ are not used. Now, prove the results pertaining to $\hat{\gamma}_{T+m}$. Analogously in (A.2), define

$$A_{11}^{(T+m)} = T^{-1}\sum_{j=T+1}^{T+m} B_T^{-1}w_{1,j}(\hat{\alpha})w_{1,j}'(\hat{\alpha})B_T^{-1},$$

$$A_{12}^{(T+m)} = T^{-1}\sum_{j=T+1}^{T+m} B_T^{-1}w_{1,j}(\hat{\alpha})w_{2,j}'(\hat{\alpha}) \quad \text{with } A_{21}^{(T+m)} = A_{12}^{(T+m)\prime},$$

$$A_{22}^{(T+m)} = T^{-1}\sum_{j=T+1}^{T+m} w_{2,j}(\hat{\alpha})w_{2,j}'(\hat{\alpha}),$$

$$B_1^{(T+m)} = T^{-1}\sum_{j=T+1}^{T+m} B_T^{-1}w_{1,j}(\hat{\alpha})\hat{v}_j(\hat{\alpha},\gamma_0) \quad \text{and}$$

$$B_2^{(T+m)} = T^{-1}\sum_{j=T+1}^{T+m} w_{2,j}(\hat{\alpha})\hat{v}_j(\hat{\alpha},\gamma_0).$$

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

From Assumptions 1(a) and (e), it follows that $\|A_{11}^{(T+m)}\| = o_p(1)$, $\|A_{12}^{(T+m)}\| = o_p(1)$, and $\|A_{22}^{(T+m)}\| = o_p(1)$. Note that

$$B_1^{(T+m)} = T^{-1}\sum_{j=T+1}^{T+m} B_T^{-1} w_{1,j}(\hat{\alpha})([z_j'(\gamma_j-\gamma_0)-\hat{\alpha}z_{j-1}'(\gamma_{j-1}-\gamma_0)]+\upsilon_j-(\hat{\alpha}-\alpha_0)u_{j-1})$$

$$= T^{-1}\sum_{j=T+1}^{T+m} B_T^{-1} w_{1,j}(\hat{\alpha})[z_j'(\gamma_j-\gamma_0)-\hat{\alpha}z_{j-1}'(\gamma_{j-1}-\gamma_0)]+o_p(1)$$

$$= O_p(\omega_T T^{-1})+o_p(1)$$

and

$$B_2^{(T+m)} = T^{-1}\sum_{j=T+1}^{T+m} w_{2,j}(\hat{\alpha})\left[z_j'(\gamma_j-\gamma_0)-\hat{\alpha}z_{j-1}'(\gamma_{j-1}-\gamma_0)\right]+o_p(1)$$

$$= O_p(\omega_T T^{-1})+o_p(1).$$

From (A.1) and the preceding results, it follows that $\|B_T(\hat{\gamma}_{1,T+m}-\gamma_{1,0})\| = O_p(\omega_T T^{-1}) + o_p(1)$ and $\|\hat{\gamma}_{2,T+m}-\gamma_{2,0}\| = O_p(\omega_T T^{-1})+o_p(1)$. ■

Proof of Lemma 2. Instead of $\hat{\Sigma}_{T+m}$ in (12), let

$$\hat{\Sigma}_T = T^{-1}\sum_{t=2}^{T-m+1} \hat{V}_{t,t+m-1}(\hat{\alpha},\hat{\gamma}_{T+m})\hat{V}_{t,t+m-1}'(\hat{\alpha},\hat{\gamma}_{T+m}), \tag{A.4}$$

and then it is sufficient to show that $||\hat{\Sigma}_T-\Sigma_0|| = o_p(1)$ and $||\hat{\Sigma}_{T+m}-\hat{\Sigma}_T|| = o_p(1)$, if $\omega_T = o_p(T^{1/2})$.

First, recall that $\hat{V}_{t,t+m-1}(\hat{\alpha},\hat{\gamma}_{T+m}) = -W_{t,t+m-1}(\hat{\alpha})(\hat{\gamma}_{T+m}-\gamma_0)+V_{t,t+m-1}-(\hat{\alpha}-\alpha_0)U_{t-1,t+m-2}$ for $t \le T-m+1$ and write

$$\begin{aligned}
&\hat{V}_{t,t+m-1}(\hat{\alpha},\hat{\gamma}_{T+m})\hat{V}_{t,t+m-1}'(\hat{\alpha},\hat{\gamma}_{T+m})\\
&\quad= W_{t,t+m-1}(\hat{\alpha})(\hat{\gamma}_{T+m}-\gamma_0)(\hat{\gamma}_{T+m}-\gamma_0)'W_{t,t+m-1}'(\hat{\alpha})\\
&\qquad+ V_{t,t+m-1}V_{t,t+m-1}'+(\hat{\alpha}-\alpha_0)^2U_{t-1,t+m-2}U_{t-1,t+m-2}'\\
&\qquad- W_{t,t+m-1}(\hat{\alpha})(\hat{\gamma}_{T+m}-\gamma_0)V_{t,t+m-1}'-V_{t,t+m-1}(\hat{\gamma}_{T+m}-\gamma_0)'W_{t,t+m-1}'(\hat{\alpha})\\
&\qquad+(\hat{\alpha}-\alpha_0)W_{t,t+m-1}(\hat{\alpha})(\hat{\gamma}_{T+m}-\gamma_0)U_{t-1,t+m-2}'\\
&\qquad+(\hat{\alpha}-\alpha_0)U_{t-1,t+m-2}(\hat{\gamma}_{T+m}-\gamma_0)'W_{t,t+m-1}'(\hat{\alpha})\\
&\qquad-(\hat{\alpha}-\alpha_0)V_{t,t+m-1}U_{t-1,t+m-2}'-(\hat{\alpha}-\alpha_0)U_{t-1,t+m-2}V_{t,t+m-1}'.
\end{aligned} \tag{A.5}$$

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

We have $T^{-1}\sum_{t=2}^{T-m+1} V_{t,t+m-1}V'_{t,t+m-1} \xrightarrow{p} \Sigma_0$, by the ergodic theorem, and note that

$$\begin{aligned}
&\left\|W_{t,t+m-1}(\hat{\alpha})(\hat{\gamma}_{T+m}-\gamma_0)\right\| \\
&\quad\le \sum_{j=t}^{t+m-1}\left[\left\|B_T^{-1}w_{1,j}(\hat{\alpha})\right\|\left\|B_T(\hat{\gamma}_{1,T+m}-\gamma_0)\right\|+\left\|z_{2,j}-\hat{\alpha}z_{2,j-1}\right\|\left\|\hat{\gamma}_{2,T+m}-\gamma_0\right\|\right] \\
&\quad\le m\left(\max_{t\le T+m}\left\|B_T^{-1}z_{1,t}\right\|+|\hat{\alpha}|\max_{t\le T+m}\left\|B_T^{-1}z_{1,t-1}\right\|\right)\left\|B_T(\hat{\gamma}_{1,T+m}-\gamma_0)\right\| \\
&\qquad+\sum_{j=t}^{t+m-1}\left(\left\|z_{2,t}\right\|+|\hat{\alpha}|\left\|z_{2,t-1}\right\|\right)\left\|\hat{\gamma}_{2,T+m}-\gamma_0\right\| \\
&\quad= a_T\left\|B_T(\hat{\gamma}_{1,T+m}-\gamma_0)\right\|+b_{t,T}\left\|\hat{\gamma}_{2,T+m}-\gamma_0\right\|, \qquad \textbf{(A.6)}
\end{aligned}$$

and Assumptions 1(b), (c), and (e) imply that $a_T = O_p(1)$, $T^{-1}\sum_{t=2}^{T-m+1} b_{t,T} = O_p(1)$, and $T^{-1}\sum_{t=2}^{T-m+1} b_{t,T}^2 = O_p(1)$. Furthermore, $||B_T(\hat{\gamma}_{1,T+m}-\gamma_{1,0})|| = o_p(1)$ and $||\hat{\gamma}_{2,T+m}-\gamma_{2,0}|| = o_p(1)$, if $\omega_T = o_p(T^{1/2})$. Hence, we obtain

$$\begin{aligned}
&T^{-1}\left\|\sum_{t=2}^{T-m+1} W_{t,t+m-1}(\hat{\alpha})(\hat{\gamma}_{T+m}-\gamma_0)(\hat{\gamma}_{T+m}-\gamma_0)'W'_{t,t+m-1}(\hat{\alpha})\right\| \\
&\quad\le T^{-1}\sum_{t=2}^{T-m+1}\left\|W_{t,t+m-1}(\hat{\alpha})(\hat{\gamma}_{T+m}-\gamma_0)\right\|^2 = o_p(1).
\end{aligned}$$

Similar arguments apply to the rest of the terms in (A.5), and it follows that $||\hat{\Sigma}_T - \Sigma_0|| = o_p(1)$.

Now, let $s_t = w_t'(\hat{\alpha})(\gamma_t-\gamma_0)+\hat{\alpha}z'_{t-1}(\gamma_t-\gamma_{t-1}) = z_t'(\gamma_t-\gamma_0)-\hat{\alpha}z'_{t-1}(\gamma_{t-1}-\gamma_0)$, for $t = T+1,\ldots,T+m$ and $S_{T+1,T+m}$ be the vector of s_t's. Note that

$$\hat{\Sigma}_{T+m}-\hat{\Sigma}_T = T^{-1}\sum_{t=T-m+2}^{T+1}\hat{V}_{t,t+m-1}(\hat{\alpha},\hat{\gamma}_{T+m})\hat{V}'_{t,t+m-1}(\hat{\alpha},\hat{\gamma}_{T+m})$$

and the last term in the summation is such that

$$\begin{aligned}
&T^{-1}\left\|\hat{V}_{T+1,T+m}(\hat{\alpha},\hat{\gamma}_{T+m})\hat{V}'_{T+1,T+m}(\hat{\alpha},\hat{\gamma}_{T+m})\right\| \\
&\quad\le T^{-1}\left\|-W_{T+1,T+m}(\hat{\alpha})(\hat{\gamma}_{T+m}-\gamma_0)+S_{T+1,T+m}+V_{T+1,T+m}\right. \\
&\qquad\left.-(\hat{\alpha}-\alpha_0)U_{T,T+m-1}\right\|^2 \\
&\quad\le\left[T^{-1/2}\left\|W_{T+1,T+m}(\hat{\alpha})(\hat{\gamma}_{T+m}-\gamma_0)\right\|+T^{-1/2}\left\|S_{T+1,T+m}\right\|+o_p(1)\right]^2. \qquad \textbf{(A.7)}
\end{aligned}$$

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

A similar bound can be found for each of $t = T-m+2, \ldots, T$. Note that $\|S_{T+1,T+m}\| = O_p(\omega_T)$ and $\|W_{T+1,T+m}(\hat{\alpha})(\hat{\gamma}_{T+m} - \gamma_0)\| = O_p(\omega_T T^{-1}) + o_p(1)$, because

$$\begin{aligned}\left\|W_{T+1,T+m}(\hat{\alpha})(\hat{\gamma}_{T+m} - \gamma_0)\right\| &\le \left\|W_{1,T+1,T+m}(\hat{\alpha})B_T^{-1}\right\| \left\|B_T(\hat{\gamma}_{1,T+m} - \gamma_{1,0})\right\| \\ &\quad + \left\|W_{2,T+1,T+m}(\hat{\alpha})\right\| \left\|(\hat{\gamma}_{2,T+m} - \gamma_{2,0})\right\| \\ &= O_p(1)O_p(\omega_T T^{-1}).\end{aligned}$$

Therefore, if $\omega_T = o_p(T^{1/2})$, $||\hat{\Sigma}_{T+m} - \hat{\Sigma}_T|| = o_p(1)$, because m is fixed. ■

Proof of Theorem 1. Define the set $L_{1,T}(\varepsilon)$ by

$$L_{1,T}(\varepsilon) = \left\{ \begin{array}{c} ||B_T(\hat{\gamma}_{1,T+m} - \gamma_{1,0})|| < \varepsilon,\ ||\hat{\gamma}_{2,T+m} - \gamma_{2,0}|| < \varepsilon, \\ \max_{t \le T-m+1} ||B_T(\hat{\gamma}_{1,(t)} - \gamma_{1,0})|| < \varepsilon,\ \max_{t \le T-m+1} ||\hat{\gamma}_{2,(t)} - \gamma_{2,0}|| < \varepsilon, \\ ||\Xi_T - \Xi_0|| < \varepsilon \text{ and } |\hat{\alpha} - \alpha_0| < \varepsilon \end{array} \right\}$$

and the set $L_{2,T}(c)$ by $L_{2,T}(c) = \left\{\max_{t \le T+m} ||B_T^{-1} z_{1,t}|| \le c\right\}$. From Assumption 1 and Lemma 1, we know there exists a sequence of positive constants $\{\varepsilon_T : T \ge 1\}$ such that $\varepsilon_T \to 0$ and $\Pr(L_{1,T}(\varepsilon_T)) \to 1$, as $T \to \infty$. Let $\{c_T\}$ be a sequence of constants such that $c_T \to \infty$ and $c_T \varepsilon_T \to 0$. By Assumptions 1(c) and (e), $\Pr(L_{2,T}(c_T)) \to 1$, as $T \to \infty$. Let $L_T = L_{1,T}(\varepsilon_T) \cap L_{2,T}(c_T)$. Then, $\Pr(L_T) \to 1$ and $\Pr(\bar{L}_T) \to 0$ as $T \to \infty$, where $\bar{L}_T$ denotes the complement of L_T.

Now, from (5), (9), and (10), write for $t \le T-m+1$,

$$\begin{aligned}P_t(\hat{\alpha}, \hat{\gamma}_{(t)}(\hat{\alpha}), \Xi_T) &= (\hat{\gamma}_{(t)} - \gamma_0)' W'_{t,t+m-1}(\hat{\alpha}) \Xi_T W_{t,t+m-1}(\hat{\alpha})(\hat{\gamma}_{(t)} - \gamma_0) \\ &\quad + V'_{t,t+m-1} \Xi_T V_{t,t+m-1} + (\hat{\alpha} - \alpha_0)^2 U'_{t-1,t+m-2} \Xi_T U_{t-1,t+m-2} \\ &\quad - 2V'_{t,t+m-1} \Xi_T W_{t,t+m-1}(\hat{\alpha})(\hat{\gamma}_{(t)} - \gamma_0) \\ &\quad + 2(\hat{\alpha} - \alpha_0) U'_{t-1,t+m-2} \Xi_T W_{t,t+m-1}(\hat{\alpha})(\hat{\gamma}_{(t)} - \gamma_0) \\ &\quad - 2(\hat{\alpha} - \alpha_0) U'_{t-1,t+m-2} \Xi_T V_{t,t+m-1} \qquad \textbf{(A.8)}\end{aligned}$$

and $P_t(\alpha_0, \gamma_0, \Xi_0) = V'_{t,t+m-1} \Xi_0 V_{t,t+m-1}$. Then, it follows that

$$\begin{aligned}&|P_t(\hat{\alpha}, \hat{\gamma}_{(t)}(\hat{\alpha}), \Xi_T) - P_t(\alpha_0, \gamma_0, \Xi_0)| \\ &\quad \le \left\|W_{t,t+m-1}(\hat{\alpha})(\hat{\gamma}_{(t)} - \gamma_0)\right\|^2 \|\Xi_T\| + |\hat{\alpha} - \alpha_0|^2 \|\Xi_T\| \left\|U_{t-1,t+m-2}\right\|^2 \\ &\quad + 2\left\|V_{t,t+m-1}\right\| \|\Xi_T\| \left\|W_{t,t+m-1}(\hat{\alpha})(\hat{\gamma}_{(t)} - \gamma_0)\right\| \\ &\quad + 2|\hat{\alpha} - \alpha_0| \left\|U'_{t-1,t+m-2}\right\| \|\Xi_T\| \left\|W_{t,t+m-1}(\hat{\alpha})(\hat{\gamma}_{(t)} - \gamma_0)\right\|\end{aligned}$$

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

$$+ 2\left|\hat{\alpha}-\alpha_0\right| \left\|U_{t-1,t+m-2}\right\| \|\Xi_T\| \left\|V_{t,t+m-1}\right\|$$

$$+ \left|V_{t,t+m-1}\Xi_T V_{t,t+m-1} - V'_{t,t+m-1}\Xi_0 V_{t,t+m-1}\right|.$$

Note that, on the set L_T,

$$\max_{t\le T+m}\left\|B_T^{-1}w_{1,t}(\hat{\alpha})\right\| = \max_{t\le T+m}\left\|B_T^{-1}z_{1,t}-\hat{\alpha}B_T^{-1}z_{1,t-1}\right\|$$

$$\le \max_{t\le T+m}\left\|B_T^{-1}z_{1,t}\right\| + \left|\hat{\alpha}\right| \max_{t\le T+m}\left\|B_T^{-1}z_{1,t-1}\right\|$$

$$\le c_T + (|\alpha_0|+\varepsilon_T)c_T = (1+|\alpha_0|+\varepsilon_T)c_T$$

and, thus,

$$\left\|W_{t,t+m-1}(\hat{\alpha})(\hat{\gamma}_{(t)}-\gamma_0)\right\|$$

$$\le \sum_{j=t}^{t+m-1}\left[\left\|B_T^{-1}w_{1,j}(\hat{\alpha})\right\| \|B_T(\hat{\gamma}_{1,(t)}-\gamma_0)\| + \left\|z_{2,j}-\hat{\alpha}z_{2,j-1}\right\| \left\|\hat{\gamma}_{2,(t)}-\gamma_0\right\|\right]$$

$$\le \sum_{j=t}^{t+m-1}\left[(1+|\alpha_0|+\varepsilon_T)c_T\varepsilon_T + \left[\|z_{2,j}\| + (|\alpha_0|+\varepsilon_T)\|z_{2,j-1}\|\right]\varepsilon_T\right] \equiv a_t(\varepsilon_T, c_T). \tag{A.9}$$

Therefore, on the set L_T,

$$\left|P_t(\hat{\alpha},\hat{\gamma}_{(t)}(\hat{\alpha}),\Xi_T) - P_t(\alpha_0,\gamma_0,\Xi_0)\right|$$

$$\le a_t(\varepsilon_T,c_T)^2(\varepsilon_T+\|\Xi_0\|) + \varepsilon_T^2(\varepsilon_T+\|\Xi_0\|)\left\|U_{t-1,t+m-2}\right\|^2$$

$$+ 2\left\|V_{t,t+m-1}\right\|(\varepsilon_T+\|\Xi_0\|)a_t(\varepsilon_T,c_T) + 2\varepsilon_T\left\|U'_{t-1,t+m-2}\right\|(\varepsilon_T+\|\Xi_0\|)a_t(\varepsilon_T,c_T)$$

$$+ 2\varepsilon_T\left\|U_{t-1,t+m-2}\right\|(\varepsilon_T+\|\Xi_0\|)\left\|V_{t,t+m-1}\right\| + \varepsilon_T\left\|V_{t,t+m-1}\right\|^2$$

$$\equiv g_t(\varepsilon_T,c_T).$$

Similarly, we can write, on the set L_T, under H_0,

$$\left|P_{T+1}(\hat{\alpha},\hat{\gamma}_{T+m}(\hat{\alpha}),\Xi_T) - P_{T+1}(\alpha_0,\gamma_0,\Xi_0)\right| \le g_{T+1}(\varepsilon_T,c_T).$$

Assumptions 1(a) and (b) imply that $\mathrm{E}g_t(\varepsilon_T,c_T)$ is finite and constant for $2\le t\le T+1$. In addition, $\mathrm{E}g_t(\varepsilon_T,c_T)\to 0$, as $(\varepsilon_T,c_T\varepsilon_T)\to(0,0)$, which implies $g_t(\varepsilon_T,c_T)\xrightarrow{p}0$. The statement in (i) follows with $t=T+1$. For the statement in (ii), let

$$F_T(x) = \frac{1}{T-m}\sum_{t=2}^{T-m+1} 1(P_t(\alpha_0,\gamma_0,\Xi_0) < x).$$

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

By the ergodic theorem, $F_T(x) \xrightarrow{p} F(x)$. Now, it suffices to show $\left|\hat{F}_T(x) - F_T(x)\right| \xrightarrow{p} 0$. Let $P_t^0 = P_t(\alpha_0, \gamma_0, \Xi_0)$ and $\hat{P}_t = P_t(\hat{\alpha}, \hat{\gamma}_{(t)}, \Xi_T)$ for simplicity. Because $1(\hat{P}_t < x) = 1(P_t^0 < x - \hat{P}_t + P_t^0)$, we have

$$1(\hat{P}_t < x) \geq 1(P_t^0 < x - g_t(\varepsilon_T, c_T)) = 1(P_t^0 < x) - 1(x - g_t(\varepsilon_T, c_T) \leq P_t^0 < x),$$

$$1(\hat{P}_t < x) \leq 1(P_t^0 < x + g_t(\varepsilon_T, c_T)) = 1(P_t^0 < x) + 1(x \leq P_t^0 < x + g_t(\varepsilon_T, c_T)),$$

which implies that

$$\left|1(\hat{P}_t < x) - 1(P_t^0 < x)\right| \leq 1(x - g_t(\varepsilon_T, c_T) \leq P_t^0 < x + g_t(\varepsilon_T, c_T))$$

and thus

$$\left|\hat{F}_T(x) - F_T(x)\right| \leq \frac{1}{T-m} \sum_{t=2}^{T-m+1} 1(x - g_t(\varepsilon_T, c_T) \leq P_t^0 < x + g_t(\varepsilon_T, c_T)).$$

Note that for any $\delta > 0$,

$$P\left(\left|\hat{F}_T(x) - F_T(x)\right| > \delta\right) \leq \delta^{-1} P(x - g_t(\varepsilon_T, c_T) \leq P_t^0 < x + g_t(\varepsilon_T, c_T)) = o(1)$$

because $g_t(\varepsilon_T, c_T) \xrightarrow{p} 0$. The statements in (iii) and (iv) follow from (ii). ■

Proof of Theorem 2. Let $L_{1,T}(\varepsilon)$ be the same as in the proof of Theorem 1 and replace $L_{2,T}(c)$ by

$$L_{2,T}(c) = \left\{ \max_{t \leq T+m} ||B_T^{-1} z_{1,t}|| \leq c \quad \text{and} \quad \omega_T \leq c \right\}.$$

Again, choose a sequence of positive constants $\{\varepsilon_T : T \geq 1\}$ such that $\varepsilon_T \to 0$ and $\Pr(L_{1,T}(\varepsilon_T)) \to 1$, as $T \to \infty$ but choose a sequence of positive constants $\{c_T\}$ such that $c_T \to \infty$ and $c_T^2 \varepsilon_T \to 0$. Let $L_T = L_{1,T}(\varepsilon_T) \cap L_{2,T}(c_T)$. As in the proof of Lemma 2, let $s_t = w_t'(\hat{\alpha})(\gamma_t - \gamma_0) + \hat{\alpha} z_{t-1}'(\gamma_t - \gamma_{t-1}) = z_t'(\gamma_t - \gamma_0) - \hat{\alpha} z_{t-1}'(\gamma_{t-1} - \gamma_0)$, for $t = T+1, \ldots, T+m$ and $S_{T+1,T+m}$ be the vector of s_t's and, similarly, let $s_t^0 = w_t'(\alpha_0)(\gamma_t - \gamma_0) + \alpha_0 z_{t-1}'(\gamma_t - \gamma_{t-1}) = z_t'(\gamma_t - \gamma_0) - \alpha_0 z_{t-1}'(\gamma_{t-1} - \gamma_0)$, for $t = T+1, \ldots, T+m$ and $S_{T+1,T+m}^0$ be the vector of s_t^0's. Then

$$\begin{aligned}
P(\hat{\alpha}, \Xi_T) = {} & (\hat{\gamma}_{T+m} - \gamma_0)' W_{T+1,T+m}'(\hat{\alpha}) \Xi_T W_{T+1,T+m}(\hat{\alpha})(\hat{\gamma}_{T+m} - \gamma_0) \\
& + (S_{T+1,T+m} + V_{T+1,T+m})' \Xi_T (S_{T+1,T+m} + V_{T+1,T+m}) \\
& + (\hat{\alpha} - \alpha_0)^2 U_{T,T+m-1}' \Xi_T U_{T,T+m-1} \\
& - 2(S_{T+1,T+m} + V_{T+1,T+m})' \Xi_T W_{T+1,T+m}(\hat{\alpha})(\hat{\gamma}_{T+m} - \gamma_0) \\
& + 2(\hat{\alpha} - \alpha_0) U_{T,T+m-1}' \Xi_T W_{T,T+m}(\hat{\alpha})(\hat{\gamma}_{T+m} - \gamma_0) \\
& - 2(\hat{\alpha} - \alpha_0) U_{T,T+m-1}' \Xi_T (S_{T+1,T+m} + V_{T+1,T+m})
\end{aligned}$$

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

$$= (S_{T+1,T+m} + V_{T+1,T+m})' \Xi_T (S_{T+1,T+m} + V_{T+1,T+m})$$
$$- 2S'_{T+1,T+m} \Xi_T W_{T+1,T+m}(\hat{\alpha})(\hat{\gamma}_{T+m} - \gamma_0)$$
$$- 2(\hat{\alpha} - \alpha_0) U'_{T,T+m-1} \Xi_T S_{T+1,T+m} + o_p(1) \quad \textbf{(A.10)}$$

and

$$P_{T+1}(\alpha_0, \gamma_0, \Xi_0) = (S^0_{T+1,T+m} + V_{T+1,T+m})' \Xi_0 (S^0_{T+1,T+m} + V_{T+1,T+m}).$$

Because $\omega_T = \max_{T+1 \le t \le T+m} ||z'_t \gamma_t - z'_t \gamma_0|| = O_p(1)$ under H_2,

$$\begin{aligned} &\left\| S_{T+1,T+m} \right\| \\ &\quad \le m \max_{T+1 \le t \le T+m} |s_t| \\ &\quad \le m \max_{T+1 \le t \le T+m} \left| z'_t (\gamma_t - \gamma_0) \right| + \left| \hat{\alpha} \right| m \max_{T+1 \le t \le T+m} \left| z'_{t-1} (\gamma_{t-1} - \gamma_0) \right| \\ &\quad = O_p(1) \end{aligned}$$

and, similarly, $\left\| S^0_{T+1,T+m} \right\| = O_p(1)$. Also, $\left\| S_{T+1,T+m} - S^0_{T+1,T+m} \right\| = o_p(1)$. Therefore, $\left| P(\hat{\alpha}, \Xi_T) - P_{T+1}(\alpha_0, \gamma_0, \Xi_0) \right| = o_p(1)$, as argued in statement (i).

For statement (ii), recall that, under H_3, we have $\left\| B_T(\hat{\gamma}_{1,T+m} - \gamma_{1,0}) \right\| = O_p(\omega_T T^{-1}) + o_p(1)$ and $\left\| \hat{\gamma}_{2,T+m} - \gamma_{2,0} \right\| = O_p(\omega_T T^{-1}) + o_p(1)$ from Lemma 1, and thus $\left\| W_{T+1,T+m}(\hat{\alpha})(\hat{\gamma}_{T+m} - \gamma_0) \right\| = O_p(1) O_p(\omega_T T^{-1})$ and $\left\| S_{T+1,T+m} \right\| = O_p(\omega_T)$. This implies that $S'_{T+1,T+m} \Xi_T S_{T+1,T+m}$ is the asymptotically dominant term in (A.10), and we can focus only on this term. Obviously, it diverges under H_3, and so does $P(\hat{\alpha}, \hat{\Sigma}^{-1}_{T+m})$. This shows the first part of (ii). From Theorem 1(iii), the critical value $\hat{\varphi}_{1-p,T}$ is convergent, and the rest of statement (ii) follows immediately.

For statement (iii), note that we can write

$$P(\hat{\alpha}, \hat{\Sigma}^{-1}_{T+m}) = (S_{T+1,T+m} + o_p(\omega_T))' \hat{\Sigma}^{-1}_{T+m} (S_{T+1,T+m} + o_p(\omega_T))$$

and from (A.7)

$$\begin{aligned} \hat{\Sigma}_{T+m} &= \hat{\Sigma}_T + T^{-1} \sum_{t=T-m+2}^{T+1} \hat{V}_{t,t+m-1}(\hat{\alpha}, \hat{\gamma}_{T+m}) \hat{V}'_{t,t+m-1}(\hat{\alpha}, \hat{\gamma}_{T+m}) \\ &= O_p(1) + O_p(T^{-1} \omega_T^2). \end{aligned}$$

When $\omega_T = O_p(T^{1/2})$, we can see that the consistency of $\hat{\Sigma}_{T+m}$ does not hold but $\hat{\Sigma}_{T+m}$ is still convergent. Hence, $P(\hat{\alpha}, \hat{\Sigma}^{-1}_{T+m}) = O_p(\omega_T^2)$. When the order of magnitude of ω_T is strictly greater than $O_p(T^{1/2})$, $T\omega_T^{-2} ||\hat{\Sigma}_{T+m}|| = O_p(1)$. Assuming that $p \lim T\omega_T^{-2} \hat{\Sigma}_{T+m}$ is invertible, $||\hat{\Sigma}^{-1}_{T+m}|| = O_p(T\omega_T^{-2})$. Therefore, $P(\hat{\alpha}, \hat{\Sigma}^{-1}_{T+m}) = O_p(T)$.

When $\hat{\Sigma}_{T+m}$ is not consistent, $\hat{\varphi}_{1-p,T} \xrightarrow{p} \varphi_{1-p}$. However, this does not harm the consistency of the $P(\hat{\alpha}, \hat{\Sigma}^{-1}_{T+m})$ test. Note that $\hat{\varphi}_{1-p,T} \le \max_{T-m+1} P_t(\hat{\alpha}, \hat{\gamma}_{(t)}(\hat{\alpha}), \hat{\Sigma}^{-1}_{T+m})$ and from (A.8)

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

$$P_t(\hat{\alpha}, \hat{\gamma}_{(t)}(\hat{\alpha}), \Xi_T) \leq \left\| \hat{\Sigma}_{T+m}^{-1} \right\| \begin{pmatrix} \left\| W_{t,t+m-1}(\hat{\alpha})(\hat{\gamma}_{(t)} - \gamma_0) \right\|^2 + \left\| V_{t,t+m-1} \right\|^2 \\ + (\hat{\alpha} - \alpha_0)^2 \left\| U_{t-1,t+m-2} \right\|^2 \\ + 2 \left\| V_{t,t+m-1} \right\| \left\| W_{t,t+m-1}(\hat{\alpha})(\hat{\gamma}_{(t)} - \gamma_0) \right\| \\ + 2 \left| \hat{\alpha} - \alpha_0 \right| \left\| U_{t-1,t+m-2} \right\| \left\| W_{t,t+m-1}(\hat{\alpha})(\hat{\gamma}_{(t)} - \gamma_0) \right\| \\ + 2 \left| \hat{\alpha} - \alpha_0 \right| \left\| U_{t-1,t+m-2} \right\| \left\| V_{t,t+m-1} \right\| \end{pmatrix}.$$

We also have $T^{-1/2} \max_{T-m+1} \left\| V_{t,t+m-1} \right\| = o_p(1)$, $T^{-1/2} \max_{T-m+1} \left\| U_{t,t+m-1} \right\| = o_p(1)$, and

$$\begin{aligned} T^{-1/2} \max_{T-m+1} \left\| W_{t,t+m-1}(\hat{\alpha})(\hat{\gamma}_{(t)} - \gamma_0) \right\| &\leq T^{-1/2} \max_{T-m+1} a_T \max_{T-m+1} \left\| B_T(\hat{\gamma}_{1,(t)} - \gamma_0) \right\| \\ &\quad + T^{-1/2} \max_{T-m+1} b_{t,T} \max_{T-m+1} \left\| \hat{\gamma}_{2,(t)} - \gamma_0 \right\| \\ &= o_p(1), \end{aligned}$$

from Assumption 1(b), (A.3), and (A.6). Therefore, we obtain $\max_{T-m+1} P_t(\hat{\alpha}, \hat{\gamma}_{(t)}(\hat{\alpha}), \hat{\Sigma}_{T+m}^{-1}) = o_p(T)$, and the second part of (iii) follows. ■

Proof of Lemma 3. When $|\alpha_0| < 1$, Lemma 1 directly applies, and no separate proof is needed. When $\alpha_0 = 1$, define $\hat{\beta}_T$, $\hat{\mu}_T = \gamma_T$ as in the proof of Lemma 1. Let $(w_{1,1}(\hat{\alpha}), w_{2,1}(\hat{\alpha})) = (1, 1)$ and, for $t \geq 2$, $w_{1,t}(\hat{\alpha}) = t - \hat{\alpha}(t-1) = (1-\hat{\alpha})t + \hat{\alpha}$ and $w_{2,t}(\hat{\alpha}) = (1-\hat{\alpha})$. Then, write

$$A_{11} = \sum_{t=2}^{T} w_{1,t}^2(\hat{\alpha}) + 1, \qquad A_{12} = \sum_{t=2}^{T} w_{1,t}(\hat{\alpha}) w_{2,t}(\hat{\alpha}) + 1, \quad \text{with } A_{21} = A_{12}',$$

$$A_{22} = \sum_{t=2}^{T} w_{2,t}^2(\hat{\alpha}) + 1, \qquad B_1 = \sum_{t=2}^{T} w_{1,t}(\hat{\alpha}) \hat{v}_t(\hat{\alpha}, \gamma_0) + \hat{v}_1(\hat{\alpha}, \gamma_0),$$

and $\quad B_2 = \sum_{t=2}^{T} w_{2,t}(\hat{\alpha}) \hat{v}_t(\hat{\alpha}, \gamma_0) + \hat{v}_1(\hat{\alpha}, \gamma_0).$

It follows that, under Assumptions B and 2, $T^{-1} A_{11} \xrightarrow{d} \kappa^2/3 - \kappa + 1$, $A_{12} \xrightarrow{d} \kappa^2/2 - \kappa + 1$, $A_{22} \xrightarrow{p} 1$, $T^{-1/2} B_1 \xrightarrow{d} B(1) - \kappa \int B(r) dr - \kappa \int r dB(r) + \kappa^2 \int r B(r) dr$, and $|B_2 - \hat{v}_1(\hat{\alpha}, \gamma_0)| = o_p(1)$, where κ is a random variable to which $T(\hat{\alpha} - 1)$ converges in distribution and $B(r)$ is the standard Wiener process multiplied by a scalar. Realize

$$(\hat{\beta}_T - \beta_0) = (T^{-1} A_{11} - T^{-1} A_{12} A_{22}^{-1} A_{21})^{-1} (T^{-1} B_1 - T^{-1} A_{12} A_{22}^{-1} B_2),$$

$$(\hat{\mu}_T - \mu_0) = (A_{22} - T^{-1} A_{21} (T^{-1} A_{11})^{-1} A_{12})^{-1} (B_2 - A_{21} (T^{-1} A_{11})^{-1} T^{-1} B_1);$$

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

then, it follows that $|\hat{\beta}_T - \beta_0| = o_p(1)$ and $|\hat{\mu}_T - \mu_0| = O_p(1)$. Now, define for $t \le T - m + 1$,

$$A_{11}^{(t)} = \sum_{j=t}^{t+[m/2]-1} w_{1,j}^2(\hat{\alpha}), \qquad A_{12}^{(t)} = \sum_{j=t}^{t+[m/2]-1} w_{1,j}(\hat{\alpha}) w_{2,j}(\hat{\alpha}) \quad \text{with } A_{21}^{(t)} = A_{1,2}^{(t)\prime},$$

$$A_{22}^{(t)} = \sum_{j=t}^{t+[m/2]-1} w_{2,j}^2(\hat{\alpha}), \qquad B_1^{(t)} = \sum_{j=t}^{t+[m/2]-1} w_{1,j}(\hat{\alpha}) \hat{v}_j(\hat{\alpha}, \gamma_0), \quad \text{and}$$

$$B_2^{(t)} = \sum_{j=t}^{t+[m/2]-1} w_{2,j}(\hat{\alpha}) \hat{v}_j(\hat{\alpha}, \gamma_0). \tag{A.11}$$

These terms capture the difference between $(\hat{\mu}_T, \hat{\beta}_T)$ and $(\hat{\mu}_{(t)}, \hat{\beta}_{(t)})$ for $t = 2, \ldots, T - m + 1$, and they are asymptotically negligible uniformly in t, e.g.,

$$\begin{aligned}
&\max_{t \le T-m+1} T^{-1} |A_{11}^{(t)}| \\
&\quad = \max_{t \le T-m+1} T^{-1} \left| \sum_{j=t}^{t+[m/2]-1} [(1-\hat{\alpha})^2 j^2 + 2(1-\hat{\alpha})\hat{\alpha} j + \hat{\alpha}^2] \right| \\
&\quad \le |T(\hat{\alpha}-1)|^2 \max_{t \le T-m+1} T^{-1} \left| \sum_{j=t}^{t+[m/2]-1} \left(\frac{j}{T}\right)^2 \right| + \max_{t \le T-m+1} T^{-1} \left| \sum_{j=t}^{t+[m/2]-1} \hat{\alpha}^2 \right| \\
&\qquad + 2|\hat{\alpha}| |T(\hat{\alpha}-1)| \max_{t \le T-m+1} T^{-1} \sum_{j=t}^{t+[m/2]-1} \frac{j}{T} \\
&\quad = o_p(1).
\end{aligned}$$

Therefore, it follows that $\max_{t \le T-m+1} |\hat{\beta}_{(t)} - \beta_0| = o_p(1)$ and $\max_{t \le T-m+1} |\hat{\mu}_{(t)} - \mu_0| = O_p(1)$. For $\hat{\beta}_{T+m}$ and $\hat{\mu}_{T+m}$, define $A_{11}^{(T+m)}$, $A_{12}^{(T+m)}$, $A_{22}^{(T+m)}$, $B_1^{(T+m)}$, and $B_2^{(T+m)}$ by replacing $\sum_{j=t}^{t+[m/2]-1}$ by $\sum_{j=T+1}^{T+m}$ in (A.11), and then it follows that $T^{-1} \left\| A_{11}^{(T+m)} \right\|$, $\left\| A_{12}^{(T+m)} \right\|$, and $\left\| A_{22}^{(T+m)} \right\|$ are all $o_p(1)$. Note that

$$\begin{aligned}
B_1^{(T+m)} &= \sum_{j=T+1}^{T+m} [(1-\hat{\alpha})t + \hat{\alpha}][w_j'(\hat{\alpha})(\gamma_j - \gamma_0) + \hat{\alpha} z_{j-1}'(\gamma_j - \gamma_{j-1})] + o_p(1) \\
&= O_p(\omega_T) + o_p(1), \\
B_2^{(T+m)} &= \sum_{j=T+1}^{T+m} (1-\hat{\alpha})[w_j'(\hat{\alpha})(\gamma_j - \gamma_0) + \hat{\alpha} z_{j-1}'(\gamma_j - \gamma_{j-1})] + o_p(1) \\
&= O_p(\omega_T T^{-1}) + o_p(1).
\end{aligned}$$

By collecting terms, the assertion follows. ■

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

Proof of Lemma 4. When $|\alpha_0| < 1$, Lemma 2 directly applies. For the case $\alpha_0 = 1$, let $\hat{\Sigma}_T$ be defined in (A.4), and it is sufficient to show that $||\hat{\Sigma}_T - \Sigma_0|| = o_p(1)$ and $||\hat{\Sigma}_{T+m} - \hat{\Sigma}_T|| = o_p(1)$. Note that $\max_{2\le t\le T+m} T^{-1/2}||U_{t-1,t+m-2}|| = O_p(1)$ and $\max_{2\le t\le T+m} T^{-1/2}||V_{t,t+m-1}|| = o_p(1)$ from Assumption 2(b). From Lemma 3, if $\omega_T = o_p(T^{1/2})$,

$$\begin{aligned}
&\max_{2\le t\le T+m} ||W_{t,t+m-1}(\hat{\alpha})(\hat{\gamma}_{T+m} - \gamma_0)|| \\
&\quad \le m \max_{2\le t\le T+m} |w_t'(\hat{\alpha})(\hat{\gamma}_{T+m} - \gamma_0)| \\
&\quad \le m \max_{2\le t\le T+m} |(1-\hat{\alpha})(\hat{\mu}_{T+m} - \mu_0)| + m \max_{2\le t\le T+m} \left|((1-\hat{\alpha})t + \hat{\alpha})\left(\hat{\beta}_{T+m} - \beta_0\right)\right| \\
&\quad = o_p(1).
\end{aligned}$$

From (A.5) with $\alpha_0 = 1$, it follows that $||\hat{\Sigma}_T - \Sigma_0|| = o_p(1)$. Now, let s_t be the same as in the proof of Lemma 2, i.e., $s_t = z_t'(\gamma_t - \gamma_0) - \hat{\alpha} z_{t-1}'(\gamma_{t-1} - \gamma_0)$. We observe that, when $\alpha_0 = 1$, the orders of magnitude of $\left\|S_{T+1,T+m}\right\|$ and $\left\|W_{T+1,T+m}(\hat{\alpha})(\hat{\gamma}_{T+m} - \gamma_0)\right\|$ are the same as when $|\alpha_0| < 1$:

$$\begin{aligned}
\max_{T+1\le t\le T+m} |s_t| &\le \max_{T+1\le t\le T+m} \left|z_t'(\gamma_t - \gamma_0)\right| + \left|\hat{\alpha}\right| \max_{T+1\le t\le T+m} \left|z_{t-1}'(\gamma_{t-1} - \gamma_0)\right| \\
&= O_p(\omega_T), \\
\left\|W_{T+1,T+m}(\hat{\alpha})(\hat{\gamma}_{T+m} - \gamma_0)\right\| &\le \left\|W_{1,T+1,T+m}(\hat{\alpha})\right\| \left|\hat{\beta}_{T+m} - \beta_0\right| \\
&\quad + \left\|W_{2,T+1,T+m}(\hat{\alpha})\right\| \left|\hat{\mu}_{T+m} - \mu_0\right| \\
&= O_p(1) O_p(\omega_T T^{-1}) + O_p(T^{-1}) O_p(\omega_T T^{-1}) \\
&= O_p(\omega_T T^{-1}).
\end{aligned}$$

Therefore, it follows from (A.7) that, if $\omega_T = o_p(T^{1/2})$, $||\hat{\Sigma}_{T+m} - \hat{\Sigma}_T|| = o_p(1)$. ■

Proof of Theorem 3. The proof for the case where $|\alpha_0| < 1$ directly follows from the proof of Theorem 1. For the case where $\alpha_0 = 1$, define the set $L_{1,T}^N(\varepsilon)$ by

$$L_{1,T}^N(\varepsilon) = \left\{ \begin{array}{c} |\hat{\beta}_{T+m} - \beta_0| < \varepsilon,\ \max_{t\le T-m+1} |\hat{\beta}_{(t)} - \beta_0| < \varepsilon, \\ |||\Xi_T|| - ||\Xi_0||| \le ||\Xi_T - \Xi_0|| < \varepsilon \text{ and } \sqrt{T}|\hat{\alpha} - 1| < \varepsilon \end{array} \right\}$$

and the set $L_{2,T}^N(c)$ by

$$L_{2,T}^N(c) = \left\{ \begin{array}{c} |\hat{\mu}_{T+m} - \mu_0| < c,\ \max_{t\le T-m+1} |\hat{\mu}_{(t)} - \mu_0| < c, \\ \max_{1\le t\le T+m} |u_t/\sqrt{T}| < c \text{ and } T|\hat{\alpha} - 1| < c \end{array} \right\}.$$

By Assumptions B and 2(b), and Lemma 3, there exists a sequence of positive constants $\{\varepsilon_T : T \ge 1\}$ such that $\varepsilon_T \to 0$ and $\Pr(L_{1,T}^N(\varepsilon_T)) \to 1$, as $T \to \infty$. Let $\{c_T\}$ be a sequence

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press

of constants such that $c_T \to \infty$ and $c_T \varepsilon_T \to 0$. By Assumptions B and 2(a) and (b), $\Pr(L_{2,T}^N(c_T)) \to 1$, as $T \to \infty$. Let $L_T^N = L_{1,T}^N(\varepsilon_T) \cap L_{2,T}^N(c_T)$. Then, $\Pr(L_T^N) \to 1$ and $\Pr(\bar{L}_T^N) \to 0$ as $T \to \infty$, where $\bar{L}_T^N$ denotes the complement of L_T^N. Note that, on the set L_T^N, it holds that $|\hat{\alpha}| < 1 + \varepsilon_T/\sqrt{T}$ and $|\hat{\alpha}| < 1 + c_T/T$. Repeat the argument in the proof of Theorem 1 using the set L_T^N. ■

Proof of Theorem 4. This is similar to the proof of Theorem 2 and is omitted. ■

Proof of Theorem 5. It suffices to show that the results in Lemmas 3 and 4 hold under Assumptions C and 2, because Theorems 3 and 4 use only the T-consistency of $\hat{\alpha}$, not its limiting distribution. For the results in Lemma 3, let A_{11}, A_{12}, A_{22}, B_1, and B_2 be the same as in the proof of Lemma 3. Then under Assumptions C and 2, $T^{-1}A_{11} \xrightarrow{d} \kappa_c^2/3 - \kappa_c + 1$, $A_{12} \xrightarrow{d} \kappa_c^2/2 - \kappa_c + 1$, $A_{22} \xrightarrow{p} 1$, $T^{-1/2}B_1 \xrightarrow{d} B(1) - \kappa \int B_c(r)dr - \kappa \int r dB(r) + \kappa^2 \int r B_c(r)dr$, and $|B_2 - \hat{v}_1(\hat{\alpha}, \gamma_0)| = o_p(1)$, where κ_c is a random variable to which $T(\hat{\alpha} - 1)$ converges in distribution and $B_c(r)$ is a scalar multiple of the Ornstein–Uhlenbeck process. Repeat the rest of the argument in the proof of Lemma 3, and then the results follow. For the results in Lemma 4, from Assumptions C and 2(b), note that $\max_{2 \le t \le T+m} T^{-1/2}||U_{t-1,t+m-2}|| \le \sqrt{m} \max_{1 \le t \le T+m} T^{-1/2}|u_t| = O_p(1)$ because $\max_{1 \le t \le T} T^{-1/2}|u_t| = O_p(1)$ and $\max_{T+1 \le t \le T+m} T^{-1/2}|u_t| = \max_{1 \le j \le m} T^{-1/2}|u_{T+j}| = \max_{1 \le j \le m} T^{-1/2}|\alpha_0^j u_T + \sum_{s=1}^{j} \alpha_0^{j-s} v_{T+j}| = O_p(1)$. Therefore, we can repeat the rest of the argument with $\alpha_0 = 1 + cT^{-1}$. ■

https://doi.org/10.1017/S0266466609990442 Published online by Cambridge University Press