Econometric Theory, **20**, 2004, 56–94. Printed in the United States of America.
DOI: 10.1017/S0266466604201037

STATIONARITY TESTING WITH COVARIATES

Michael Jansson
University of California, Berkeley

Two new stationarity tests are proposed. Both tests can be viewed as generalizations of existing stationarity tests and dominate these in terms of local asymptotic power. Improvements are achieved by accommodating stationary covariates. A Monte Carlo investigation of the small sample properties of the tests is conducted, and an empirical illustration from international finance is provided.

1. INTRODUCTION

Let y_t be an observed univariate time series generated by

$$y_t = \mu_t^y + v_t^y, \qquad t = 1, \ldots, T, \tag{1}$$

where μ_t^y is deterministic component and v_t^y is an unobserved error process with initial condition $v_1^y = u_1^y$ and generating mechanism

$$\Delta v_t^y = (1 - \theta L) u_t^y, \qquad t = 2, \ldots, T, \tag{2}$$

where u_t^y is a stationary $I(0)$ process. (In this paper, a process is said to be $I(0)$ if its partial sum process converges weakly to a Brownian motion.)

The problem of testing the null hypothesis $H_0 : \theta = 1$ against $H_1 : \theta < 1$ has attracted considerable attention in the literature, as has the closely related problem of testing for parameter constancy in the "local-level" unobserved components model. Pertinent references include LaMotte and McWorther (1978), Nyblom and Mäkeläinen (1983), Nyblom (1986), Nabeya and Tanaka (1988), Tanaka (1990), Kwiatkowski, Phillips, Schmidt, and Shin (1992), Saikkonen and Luukkonen (1993a, 1993b), Choi (1994), and Leybourne and McCabe (1994). (For a review, see Stock, 1994.) Under H_0, $v_t^y = u_t^y$ and y_t is a (trend-) stationary process, whereas y_t is an integrated process with a random walk–type nonstationarity under the alternative hypothesis. For this reason, tests of H_0 are often referred to as stationarity tests. The cited papers differ somewhat

This paper has benefited from the comments of Pentti Saikkonen (the co-editor), two anonymous referees, and seminar participants at University of Aarhus, Indiana University, Purdue University, Stanford University, UC Riverside, the 2001 Nordic Econometric Meeting, and the 2001 NBER Summer Institute. A MATLAB program that implements the tests proposed in this paper is available at http://elsa.Berkeley.EDU/users/mjansson. Address correspondence to: Michael Jansson, Department of Economics, 549 Evans Hall #3880, UC Berkeley, Berkeley, CA 94720-3880, USA; e-mail: mjansson@econ.berkeley.edu.

  0266-4666/04 $12.00

with respect to the assumptions on the underlying stationary process u_t^y and the form of the deterministic component μ_t^y. On the other hand, all previous studies (of which the author is aware) have been concerned with the situation where y_t is observed in isolation. Specifically, all previously devised tests have exploited only the information contained in y_t when testing H_0.

In applications, it is extremely rare that individual time series are observed in isolation. As a consequence, it seems reasonable to ask whether more powerful stationarity tests can be obtained be utilizing the information contained in related time series. To fix ideas, suppose a k-vector time series x_t of covariates is observed, whose generating mechanism is

$$x_t = \mu_t^x + u_t^x, \qquad t = 1, \ldots, T, \tag{3}$$

where μ_t^x is deterministic component and u_t^x is an unobserved stationary $I(0)$ process. Moreover, suppose the deterministic components μ_t^y and μ_t^x are pth-order polynomial trends; that is, suppose

$$\mu_t^y = \sum_{i=0}^{p} \beta_i^y t^i, \qquad \mu_t^x = \sum_{i=0}^{p} \beta_i^x t^i, \tag{4}$$

where $\{\beta_i^y : 0 \leq i \leq p\} \subseteq \mathbb{R}$ and $\{\beta_i^x : 0 \leq i \leq p\} \subseteq \mathbb{R}^k$ are unknown parameters.

The present paper proposes two new tests that exploit the information contained in the covariates x_t when testing the null hypothesis that y_t is (trend-) stationary. Both tests are valid under mild moment and memory conditions on $u_t = (u_t^y, u_t^{x\prime})'$ and enjoy optimality properties in the special case where u_t is Gaussian white noise. The tests can be viewed as generalizations of existing univariate stationarity tests, and the new tests dominate their univariate counterparts in terms of asymptotic local power whenever the zero-frequency correlation between u_t^y and u_t^x is nonzero. (When the zero-frequency correlation equals zero, the new tests coincide with their univariate counterparts.) In fact, substantial power gains can be achieved if an appropriate set of covariates x_t can be found. The paper therefore provides an affirmative answer to the question posed in the beginning of the previous paragraph. Results complementary to those obtained here can be found in Hansen (1995) and Elliott and Jansson (2003). These papers demonstrate the usefulness of covariates in the context of testing for an autoregressive unit root.

Section 2 derives the tests and establishes their asymptotic optimality properties in the special case where the underlying innovation sequence is Gaussian white noise. In Section 3, the tests are extended to accommodate general stationary errors by means of nonparametric corrections. Section 4 shows how the tests can be applied to test the null hypothesis that a vector integrated process is cointegrated with a prespecified cointegration vector and presents an empirical illustration. Finally, Section 5 offers a few concluding remarks, and all proofs are collected in the Appendix.

2. TESTING WITH WHITE NOISE ERRORS

Let $(y_t, x_t')'$ be generated by (1)–(4) and suppose $u_t \sim i.i.d.\ \mathcal{N}(0,\Sigma)$, where

$$\Sigma = \begin{pmatrix} \sigma_{yy} & \sigma_{xy}' \\ \sigma_{xy} & \Sigma_{xx} \end{pmatrix}$$

is a known, positive definite matrix (partitioned in conformity with u_t). Consider the problem of testing

$$H_0 : \theta = 1 \quad \text{vs.} \quad H_1 : \theta < 1.$$

This problem is that of testing whether the (permanent) component $(1-\theta)\sum_{s=1}^{t-1} u_s^y$ is absent from the following permanent-transitory decomposition of y_t:

$$y_t = \mu_t^y + (1-\theta)\sum_{s=1}^{t-1} u_s^y + u_t^y.$$

To see how the use of stationary covariates x_t facilitates the testing problem, consider the series $y_t - \sigma_{xy}'\Sigma_{xx}^{-1}x_t$, whose permanent-transitory decomposition is

$$y_t - \sigma_{xy}'\Sigma_{xx}^{-1}x_t = \mu_t^{y.x} + (1-\theta)\sum_{s=1}^{t-1} u_s^y + u_t^{y.x},$$

where $\mu_t^{y.x} = \mu_t^y - \sigma_{xy}'\Sigma_{xx}^{-1}\mu_t^x$ and $u_t^{y.x} = u_t^y - \sigma_{xy}'\Sigma_{xx}^{-1}u_t^x$. Because x_t is stationary, the transformation $y_t - \sigma_{xy}'\Sigma_{xx}^{-1}x_t$ does not affect the permanent component. On the other hand, $\text{Var}(u_t^{y.x}) = (1-\rho^2)\text{Var}(u_t^y)$, so the transformation reduces the variance of the transitory component by a fraction ρ^2, where $\rho^2 = \sigma_{yy}^{-1}\sigma_{xy}'\Sigma_{xx}^{-1}\sigma_{xy}$ is the squared coefficient of multiple correlation computed from Σ. The covariates x_t can therefore be used to attenuate the transitory component of y_t without affecting the permanent component. As a consequence, the use of covariates makes it easier to detect the permanent component of y_t if it is present, thereby leading to improvements in power relative to the case where the covariates are ignored. The remainder of this section makes these heuristic ideas more precise.

2.1. Point Optimal Invariant Tests

Define $\beta = (\beta_0^y, \ldots, \beta_p^y, \beta_0^{x\prime}, \ldots, \beta_p^{x\prime})'$ and for any $t = 1, \ldots, T$, let

$$z_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix}, \qquad v_t = \begin{pmatrix} v_t^y \\ u_t^x \end{pmatrix}, \qquad d_t = \begin{pmatrix} d_t^y & 0 \\ 0 & I_k \otimes d_t^x \end{pmatrix},$$

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

where $d_t^y = d_t^x = (1,\ldots,t^p)'$. Using this notation, the model can be written as

$$z_t = d_t'\beta + v_t, \qquad t = 1,\ldots,T.$$

The problem of testing $H_0: \theta = 1$ vs. $H_1: \theta < 1$ is invariant under the group of transformations of the form $z_t \to z_t + d_t' b$, $b \in \mathbb{R}^{(k+1)(p+1)}$. A maximal invariant is $m_T = D_\perp' \operatorname{vec}(z_1,\ldots,z_T)$, where $D_\perp$ is a matrix whose columns form an orthonormal basis for the orthogonal complement of the column space of $(d_1,\ldots,d_T)'$. For any θ^*, let

$$z_t(\theta^*) = \begin{pmatrix} y_t(\theta^*) \\ x_t \end{pmatrix}, \qquad d_t(\theta^*) = \begin{pmatrix} d_t^y(\theta^*) & 0 \\ 0 & I_k \otimes d_t^x \end{pmatrix},$$

where $y_t(\theta^*)$ satisfies the difference equation $y_t(\theta^*) = \Delta y_t + \theta^* y_{t-1}(\theta^*)$ with initial condition $y_1(\theta^*) = y_1$ and $d_t^y(\theta^*)$ is defined analogously. The probability density of m_T is proportional to

$$\exp\left(-\frac{1}{2}\sum_{t=1}^{T} \tilde{v}_t(\theta;\Sigma)'\Sigma^{-1}\tilde{v}_t(\theta;\Sigma)\right),$$

where, for any θ^*,

$$\tilde{v}_t(\theta^*;\Sigma) = z_t(\theta^*) - d_t(\theta^*)'\left(\sum_{s=1}^{T} d_s(\theta^*)\Sigma^{-1}d_s(\theta^*)'\right)^{-1} \times \left(\sum_{s=1}^{T} d_s(\theta^*)\Sigma^{-1}z_s(\theta^*)\right).$$

By the Neyman–Pearson lemma, the test that rejects for large values of

$$P_T(\bar{\theta}) = P_T(\bar{\theta};\Sigma) = \sum_{t=1}^{T} \tilde{v}_t(1;\Sigma)'\Sigma^{-1}\tilde{v}_t(1;\Sigma) - \sum_{t=1}^{T} \tilde{v}_t(\bar{\theta};\Sigma)'\Sigma^{-1}\tilde{v}_t(\bar{\theta};\Sigma) \tag{5}$$

is the most powerful invariant test of $\theta = 1$ against the specific alternative $\theta = \bar{\theta}$.

Theorem 1 characterizes the limiting distribution of $P_T(\bar{\theta})$ under a local-to-unity reparameterization of θ and $\bar{\theta}$ in which $\lambda = T(1-\theta) \geq 0$ and $\bar{\lambda} = T(1-\bar{\theta}) > 0$ are held constant as T increases without bound. The limiting representation of $P_T(\bar{\theta})$ involves the random functional φ_P, the definition of which is given next.

Let $R \in [0,1)$, $\lambda \geq 0$, and $\bar{\lambda} > 0$ be given. Let $\bar{\Sigma}^{1/2}$ be the (lower triangular) Cholesky factor of the 2×2 matrix

$$\bar{\Sigma} = \bar{\Sigma}(R) = \begin{pmatrix} 1 & R \\ R & 1 \end{pmatrix}$$

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

and for $l \in \{0, \bar{\lambda}\}$, define

$$U_l^{\lambda}(r) = \bar{\Sigma}^{-1/2}\begin{pmatrix} V_l^{\lambda}(r) \\ W(r) \end{pmatrix}, \qquad D_l(r) = \begin{pmatrix} D_l^{y}(r) & 0 \\ 0 & D^{x}(r) \end{pmatrix}\bar{\Sigma}^{-1/2\prime},$$

where

$$V_l^{\lambda}(r) = V^{\lambda}(r) - l\int_0^r \exp(-l(r-s))V^{\lambda}(s)\,ds,$$

$$D_l^{y}(r) = D^{y}(r) - l\int_0^r \exp(-l(r-s))D^{y}(s)\,ds,$$

$V^{\lambda}(r) = V(r) + \lambda\int_0^r V(s)\,ds$, $D^{y}(r) = D^{x}(r) = (1,\ldots,r^{p})'$, and $(V,W)'$ is a Brownian motion with covariance matrix $\bar{\Sigma}$. (Here, and elsewhere, the dependence of U_l^{λ} and D_l on R is suppressed.) Finally, let $R_{\#} = (1-R^2)^{-1/2}$ and define

$$\begin{aligned}
&\varphi_P(\lambda;\bar{\lambda},R^2) \\
&\quad = -\bar{\lambda}^2R_{\#}^2\int_0^1 V_{\bar{\lambda}}^{\lambda}(r)^2\,dr + 2\bar{\lambda}R_{\#}^2\left(\int_0^1 V_{\bar{\lambda}}^{\lambda}(r)\,dV^{\lambda}(r) - R\int_0^1 V_{\bar{\lambda}}^{\lambda}(r)\,dW(r)\right) \\
&\qquad + \left(\int_0^1 D_{\bar{\lambda}}(r)\,dU_{\bar{\lambda}}^{\lambda}(r)\right)'\left(\int_0^1 D_{\bar{\lambda}}(r)D_{\bar{\lambda}}(r)'\,dr\right)^{-1}\left(\int_0^1 D_{\bar{\lambda}}(r)\,dU_{\bar{\lambda}}^{\lambda}(r)\right) \\
&\qquad - \left(\int_0^1 D_0(r)\,dU_0^{\lambda}(r)\right)'\left(\int_0^1 D_0(r)D_0(r)'\,dr\right)^{-1}\left(\int_0^1 D_0(r)\,dU_0^{\lambda}(r)\right).
\end{aligned}$$

THEOREM 1. *Let z_t be generated by (1)–(4). Suppose $u_t \sim$ i.i.d. $N(0,\Sigma)$ and suppose $\lambda = T(1-\theta) \geq 0$ and $\bar{\lambda} = T(1-\bar{\theta}) > 0$ are fixed as T increases without bound. Then $P_T(\bar{\theta}) \to_d \varphi_P(\lambda;\bar{\lambda},\rho^2)$.*

Corresponding to any invariant (possibly randomized) test of $H_0: \theta = 1$ there is a test function $\phi_T: \mathbb{R}^{(T-p-1)(k+1)} \to [0,1]$ such that H_0 is rejected with probability $\phi_T(m)$ whenever m_T, the maximal invariant, equals m. For any given θ and any such ϕ_T, the probability of rejecting H_0 is $\int \phi_T(m) f_T(m|\theta,\Sigma)\,dm$, where $f_T(\cdot|\theta,\Sigma)$ denotes the probability density of the maximal invariant and the domain of integration is $\mathbb{R}^{(T-p-1)(k+1)}$. A test ϕ_T is of level $\alpha \in (0,1)$ if its size, namely, $\int \phi_T(m) f_T(m|1,\Sigma)\,dm$, is less than or equal to α. Similarly, a sequence $\{\phi_T\}$ of test functions is said to be asymptotically of level α if

$$\overline{\lim_{T\to\infty}} \int \phi_T(m) f_T(m|1,\Sigma)\,dm \leq \alpha.$$

When $\overline{\lim}_{T\to\infty}$ on the left-hand side equals $\lim_{T\to\infty}$ and the inequality is an equality, $\{\phi_T\}$ is said to be asymptotically of size α.

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

The test statistic $P_T(\bar{\theta})$ is point optimal invariant (POI) in the sense that the power

$$\int \phi_T(m) f_T(m|\bar{\theta},\Sigma)\, dm$$

against the point alternative $\theta = \bar{\theta}$ is maximized over all invariant tests of level α by the test function $1(P_T(\bar{\theta}) > c_T^P(\bar{\theta},\alpha,\Sigma))$, where $1(\cdot)$ is the indicator function and $c_T^P(\bar{\theta},\alpha,\Sigma)$ is such that the test is of size α. This optimality result has an obvious asymptotic analogue. Let the function $c^P(\cdot,\cdot,\cdot)$ be implicitly defined by the relation $\Pr(\varphi_P(0;\bar{\lambda},\rho^2) > c^P(\bar{\lambda},\alpha,\rho^2)) = \alpha$. The statistic $P_T(\bar{\theta})$ is asymptotically POI under local-to-unity asymptotics in the sense that $\phi_T^P(m_T;\bar{\lambda},\alpha,\Sigma) = 1(P_T(1 - T^{-1}\bar{\lambda}) > c^P(\bar{\lambda},\alpha,\rho^2))$ maximizes

$$\lim_{T\to\infty} \int \phi_T(m) f_T(m|1 - T^{-1}\bar{\lambda},\Sigma)\, dm$$

over all invariant tests asymptotically of level α; that is,

$$\overline{\lim_{T\to\infty}} \int \phi_T(m) f_T(m|1 - T^{-1}\bar{\lambda},\Sigma)\, dm$$

$$\leq \underline{\lim}_{T\to\infty} \int \phi_T^P(m_T;\bar{\lambda},\alpha,\Sigma) f_T(m|1 - T^{-1}\bar{\lambda},\Sigma)\, dm$$

whenever $\{\phi_T\}$ is asymptotically of level α. Moreover, $\underline{\lim}_{T\to\infty}$ on the right-hand side equals $\lim_{T\to\infty}$ and is given by $\Pr(\varphi_P(\bar{\lambda};\bar{\lambda},\rho^2) > c^P(\bar{\lambda},\alpha,\rho^2))$.

Theorem 2 of Saikkonen and Luukkonen (1993a) obtains an upper bound on the asymptotic power function of any location and scale invariant stationarity test in the univariate case. Because scale invariance is not imposed, the result stated here covers a larger class of tests than Theorem 2 of Saikkonen and Luukkonen (1993a) even in the univariate case. (The present paper obviates the need to impose scale invariance by assuming that Σ is known.) Moreover, the multivariate model studied here contains the univariate model of Saikkonen and Luukkonen (1993a) as a special case.

The function $\pi^\alpha(\lambda;\rho^2) = \Pr(\varphi_P(\lambda;\lambda,\rho^2) > c^P(\lambda,\alpha,\rho^2))$ provides an upper bound on the asymptotic power function of any invariant test asymptotically of level α. The bound is sharp in the sense that it can be attained for any given λ by the test $\phi_T^P(m_T;\lambda,\alpha,\Sigma)$. Moreover, although no test statistic attains the upper bound uniformly in λ, it turns out that it is possible to construct tests whose power functions are very close to the bound. The Gaussian power envelope therefore constitutes a useful benchmark against which the power function of any invariant test (asymptotically of level α) can be compared.

The univariate counterpart of $P_T(\bar{\theta})$ is

$$P_T^y(\bar{\theta}) = P_T^y(\bar{\theta};\sigma_{yy}) = \sigma_{yy}^{-1}\left(\sum_{t=1}^{T} \hat{v}_t^y(1)^2 - \sum_{t=1}^{T} \hat{v}_t^y(\bar{\theta})^2\right),$$

where

$$\hat{v}_t^y(\theta^*) = y_t(\theta^*) - d_t^y(\theta^*)'\left(\sum_{s=1}^{T} d_s^y(\theta^*)\, d_s^y(\theta^*)'\right)^{-1}\left(\sum_{s=1}^{T} d_s^y(\theta^*) y_s(\theta^*)\right)$$

for any θ^*. When $u_t^y \sim i.i.d.\ \mathcal{N}(0,\sigma_{yy})$, the test that rejects for large values of $P_T^y(\bar{\theta})$ is more powerful against the specific alternative $\theta = \bar{\theta} < 1$ than any other invariant test of H_0 based solely on y_t, where invariance is with respect to transformations of the form $y_t \to y_t + b_y' d_t^y$, $b_y \in \mathbb{R}^{p+1}$.

When $\rho^2 = 0$, the time series y_t and x_t are independent. In that case, the covariates x_t carry no information about y_t, and the statistics $P_T(\bar{\theta})$ and $P_T^y(\bar{\theta})$ are equivalent. In contrast, the rejection regions of the tests based on the statistics $P_T(\bar{\theta})$ and $P_T^y(\bar{\theta})$ differ whenever $\rho^2 \neq 0$. These differences persist asymptotically, as $P_T^y(\bar{\theta}) \to_d \varphi_P(\lambda;\bar{\lambda},0)$ under the assumptions of Theorem 1. Comparing $\varphi_P(\lambda;\bar{\lambda},0)$ and $\varphi_P(\lambda;\bar{\lambda},\rho^2)$, the limiting distribution of $P_T(\bar{\theta})$ is seen to depend on the covariates x_t only through the parameter ρ^2. As a consequence, the "quality" of the covariates can be summarized by this scalar parameter.

Figure 1 plots $\pi^{0.05}(\lambda;\rho^2)$ for selected values of ρ^2 in the constant mean ($p = 0$) case. (The curves were generated by taking 20,000 draws from the distribution of the discrete approximation [based on 2,000 steps] to the limiting random variables.) The lowest curve corresponds to $\rho^2 = 0$ and therefore provides an upper bound on the (local asymptotic) power function of any invariant univariate stationarity test. An increase in the quality of the covariates (as measured by ρ^2) leads to an increase in the level of the power envelope. Indeed, the difference between the power envelope and its univariate counterpart is quite remarkable for most values of ρ^2. For concreteness, consider the alternative $\lambda = 5$, which corresponds to a moving average coefficient θ of 0.975 when $T = 200$. The univariate power envelope is 0.32, whereas the envelopes are 0.40 and 0.58 when ρ^2 equals 0.2 and 0.5, respectively. Because they are upper bounds, these power envelopes do not by themselves illustrate the power gains attainable by feasible tests. On the other hand, the evidence presented in Figure 1 clearly suggests that substantial power gains can be achieved by including covariates in a stationarity test provided an appropriate set of covariates can be found. The power envelopes are lower in the linear trend ($p = 1$) case, but the qualitative conclusion remains the same, as can be seen from Figure 2.

2.2. Locally Best Invariant Tests

Even asymptotically, the critical region of the test based on $P_T(1 - T^{-1}\bar{\lambda})$ depends on $\bar{\lambda}$. As a consequence, no test is asymptotically uniformly most

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

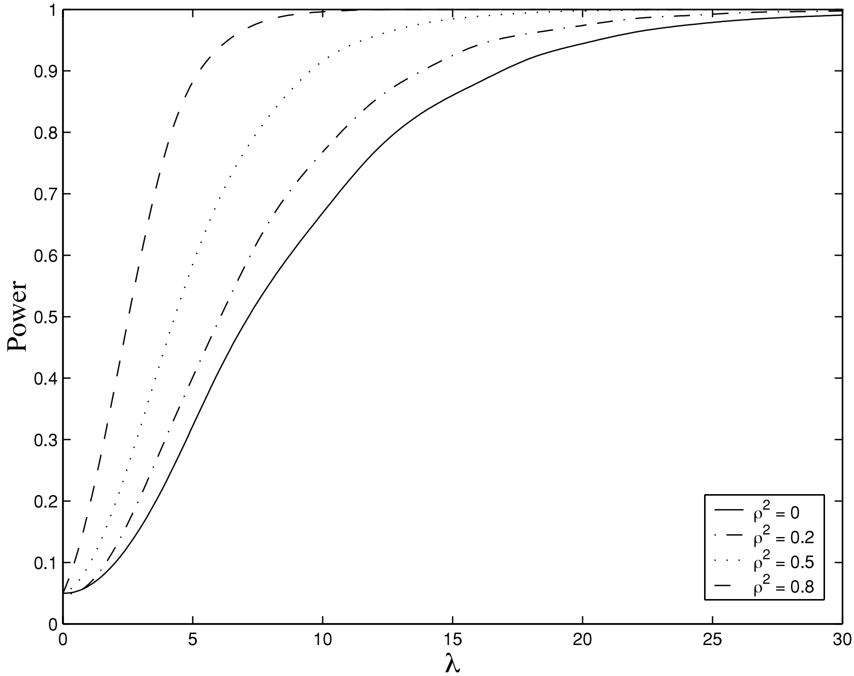

FIGURE 1. Power envelopes: 5% level tests, constant mean ($p = 0$).

powerful (with respect to the class of invariant tests) in the sense of Basawa and Scott (1983). In such cases, tests based on weaker optimality concepts seem worth considering. One such concept, the concept of point optimality, justifies the test based on $P_T(1 - T^{-1}\bar{\lambda}^{\dagger})$, where $\bar{\lambda}^{\dagger}$ is a prespecified alternative against which maximal power is desired. As an alternative to that test, the present section develops a test based on a Taylor series expansion of $P_T(1 - T^{-1}\bar{\lambda})$ around $\bar{\lambda} = 0$. The resulting test can be implemented without specifying an alternative in advance and enjoys certain local optimality properties.

Using simple algebra, it can be shown that

$$\dot{P}_T = \frac{\partial}{\partial \bar{\lambda}} \partial P_T(1 - T^{-1}\bar{\lambda})\bigg|_{\bar{\lambda}=0} = -\begin{pmatrix}1\\0\end{pmatrix}'\left(\Sigma^{-1}T^{-1}\sum_{t=1}^{T}\tilde{v}_t(1)\tilde{v}_t(1)'\right)\begin{pmatrix}1\\0\end{pmatrix}$$

and

$$\ddot{P}_T = \frac{1}{2}\frac{\partial^2}{\partial \bar{\lambda}^2}\partial P_T(1 - T^{-1}\bar{\lambda})\bigg|_{\bar{\lambda}=0} = L_T + T^{-1}\dot{P}_T,$$

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

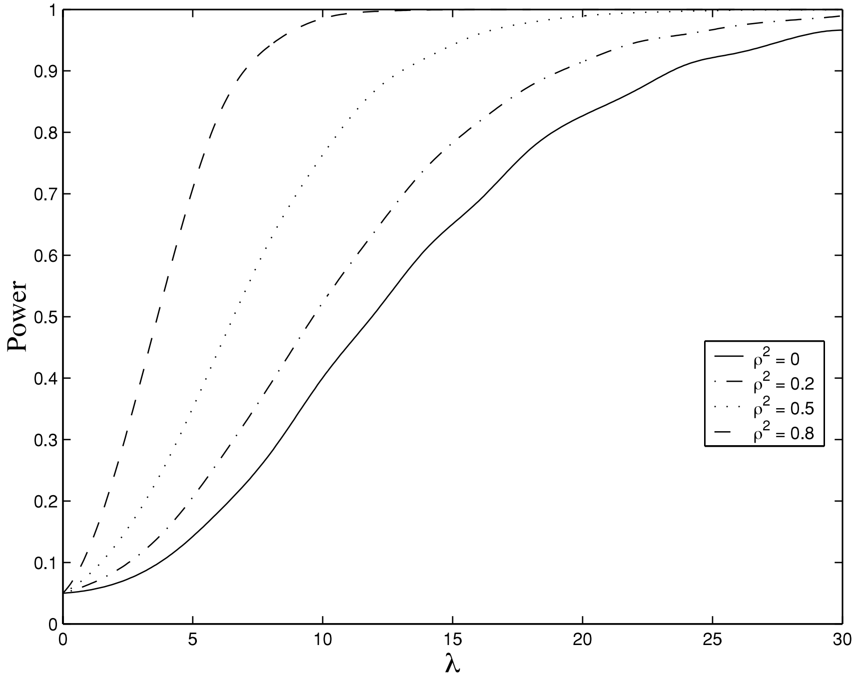

FIGURE 2. Power envelopes: 5% level tests, linear trend ($p = 1$).

where

$$L_T = L_T(\Sigma) = \sum_{t=1}^{T} \widetilde{V}_t' \Sigma^* \widetilde{V}_t + \left(\sum_{t=1}^{T} d_t \Sigma^{**} \widetilde{V}_t\right)' \left(\sum_{t=1}^{T} d_t \Sigma^{-1} d_t'\right)^{-1} \left(\sum_{t=1}^{T} d_t \Sigma^{**} \widetilde{V}_t\right),$$

$$\Sigma^* = \begin{pmatrix} \sigma_{yy.x}^{-1} & \sigma^{xy\prime} \\ \sigma^{xy} & 0 \end{pmatrix}, \quad \Sigma^{**} = \begin{pmatrix} 0 & \sigma^{xy\prime} \\ -\sigma^{xy} & 0 \end{pmatrix}, \tag{6}$$

$\widetilde{V}_t = T^{-1} \sum_{s=1}^{t-1} \tilde{v}_s(1;\Sigma)$, $\sigma_{yy.x} = \sigma_{yy} - \sigma_{xy}' \Sigma_{xx}^{-1} \sigma_{xy}$, and $\sigma^{xy} = -\sigma_{yy.x}^{-1} \Sigma_{xx}^{-1} \sigma_{xy}$. (The dependence of $\dot{P}_T$ and $\ddot{P}_T$ on Σ has been suppressed to achieve notational economy, and the notation $\widetilde{V}_t$ recognizes the fact that $\tilde{v}_t(1;\Sigma)$ does not depend on Σ.)

Under the assumptions of Theorem 1, $T^{-1} \sum_{t=1}^{T} \tilde{v}_t(1)\tilde{v}_t(1)' \to_p \Sigma$. As a consequence, the limiting distribution of $\dot{P}_T$ is degenerate: $\dot{P}_T \to_p -1$. On the other hand, Theorem 2(a), which follows, shows that under the same assumptions the limiting distribution of L_T equals that of the random variable $\varphi_L(\lambda;\rho^2)$, where, for any $0 \le R < 1$,

$$\varphi_L(\lambda;R^2) = \int_0^1 \tilde{U}^{\lambda}(r)'\bar{\Sigma}^*\tilde{U}^{\lambda}(r)\,dr + \left(\int_0^1 D(r)\bar{\Sigma}^{**}\tilde{U}^{\lambda}(r)\,dr\right)'\left(\int_0^1 D(r)D(r)'\,dr\right)^{-1} \times \left(\int_0^1 D(r)\bar{\Sigma}^{**}\tilde{U}^{\lambda}(r)\,dr\right),$$

where

$$\tilde{U}^{\lambda}(r) = U_0^{\lambda}(r) - \left(\int_0^r D(s)\,ds\right)'\left(\int_0^1 D(s)D(s)'\,ds\right)\left(\int_0^1 D(s)\,dU_0^{\lambda}(s)\right),$$

$D(r) = D_0(r)$, and

$$\bar{\Sigma}^* = \begin{pmatrix} 1-R_{\#}^2R^2 & -R_{\#}R \\ -R_{\#}R & 0 \end{pmatrix}, \qquad \bar{\Sigma}^{**} = \begin{pmatrix} 0 & -R_{\#}R \\ R_{\#}R & 0 \end{pmatrix}.$$

The test that rejects for large values of L_T is asymptotically equivalent (in an obvious sense) to the test that rejects for large values of the second-order Taylor approximation to $P_T(1 - T^{-1}\bar{\lambda})$, namely, $\dot{P}_T\bar{\lambda} + \ddot{P}_T\bar{\lambda}^2$. This observation suggests that L_T enjoys certain local optimality properties. A sequence $\{\phi_T\}$ of tests is asymptotically locally efficient (with respect to the class of invariant tests asymptotically of size α) in the sense of Basawa and Scott (1983) if it maximizes

$$\lim_{T\to\infty} \frac{\partial}{\partial\lambda}\int \phi_T(m)f_T(m|1-T^{-1}\lambda,\Sigma)\,dm\bigg|_{\lambda=0}$$

over all invariant tests asymptotically of size α. As Theorem 2(b) shows, any invariant test (asymptotically of size α) is asymptotically locally efficient according to that definition.[1] To obtain a nontrivial characterization of local optimality in the present context, the following alternative concept of asymptotic local optimality is useful. Let q^* be the smallest integer q such that

$$\underline{\lim}_{T\to\infty} \int |l_T^{(q)}(m|\Sigma)|\cdot f_T(m|1,\Sigma)\,dm > 0,$$

where $l_T^{(q)}(m|\Sigma) = \partial^q \log f_T(m|1-T^{-1}\lambda,\Sigma)/\partial\lambda^q|_{\lambda=0}$. An invariant test is said to be asymptotically locally best invariant (LBI) if it maximizes

$$\lim_{T\to\infty} \frac{\partial^{q^*}}{\partial\lambda^{q^*}}\int \phi_T(m)f_T(m|1-T^{-1}\lambda,\Sigma)\,dm\bigg|_{\lambda=0}$$

over all invariant tests asymptotically of the same size. In regular cases where partial derivatives of $\int \log f_T(m|1-T^{-1}\lambda,\Sigma)\cdot f_T(m|1,\Sigma)\,dm$ with respect to λ

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

can be obtained by differentiating under the integral sign, this concept of local asymptotic optimality agrees with that of Basawa and Scott (1983) when $q^* = 1$. The testing problem studied here has $q^* = 2$ and as Theorem 2(c) shows, L_T is asymptotically LBI in the (stronger) sense defined here.[2]

THEOREM 2. *Let z_t be generated by (1)–(4). Suppose $u_t \sim$ i.i.d. $N(0,\Sigma)$ and suppose $\lambda = T(1-\theta) \geq 0$ is fixed as T increases without bound. Then*

(a) $L_T \to_d \varphi_L(\lambda;\rho^2)$.

If $\{\phi_T\}$ is asymptotically of size $\alpha \in (0,1)$, then

(b)

$$\lim_{T\to\infty} \frac{\partial}{\partial\lambda} \int \phi_T(m) f_T(m|1 - T^{-1}\lambda, \Sigma)\, dm \bigg|_{\lambda=0} = 0,$$

(c)

$$\overline{\lim_{T\to\infty}} \frac{\partial^2}{\partial\lambda^2} \int (\phi_T(m) - \phi_T^L(m;\alpha,\Sigma)) f_T(m|1 - T^{-1}\lambda, \Sigma)\, dm \bigg|_{\lambda=0} \leq 0,$$

where $\phi_T^L(m_T;\alpha,\Sigma) = 1(L_T > c^L(\alpha,\rho^2))$ and $\Pr(\varphi_L(0;\rho^2) > c^L(\alpha,\rho^2)) = \alpha$.

The univariate counterpart of L_T is

$$L_T^y = L_T^y(\sigma_{yy}) = \sigma_{yy}^{-1} \sum_{t=1}^{T} (\hat{V}_t^y)^2,$$

where $\hat{V}_t^y = T^{-1} \sum_{s=1}^{t-1} \hat{v}_s^y(1)$. The statistics L_T and L_T^y are equivalent if and only if $\rho^2 = 0$. Moreover, $L_T^y \to_d \varphi_L(\lambda;0)$ under the assumptions of Theorem 2, so the difference between L_T and L_T^y persists asymptotically whenever $\rho^2 \neq 0$. As was the case with the power envelopes derived in the previous section, the inclusion of covariates can have a substantial effect on the power properties of the LBI test. (This will become apparent in Section 3.2.)

3. TESTING WITH WEAKLY DEPENDENT ERRORS

The analysis in the previous section proceeded under the restrictive assumption that $u_t \sim$ *i.i.d.* $\mathcal{N}(0,\Sigma)$, where Σ is known. The optimality theory seems to depend on the normality assumption. On the other hand, it is straightforward to construct feasible test statistics having limiting representations of the form φ_P and φ_L under much less stringent assumptions on u_t. For instance, the following assumption suffices.

A1. $u_t = \sum_{i=0}^{\infty} C_i \varepsilon_{t-i}$, where $\{\varepsilon_t : t \in \mathbb{Z}\}$ is *i.i.d.* $(0, I_{k+1})$, $\sum_{i=0}^{\infty} C_i$ has full rank, and $\sum_{i=1}^{\infty} i\|C_i\| < \infty$, where $\|\cdot\|$ is the Euclidean norm.

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

3.1. Feasible Tests

Define the matrices

$$\Omega = \begin{pmatrix} \omega_{yy} & \omega_{xy}' \\ \omega_{xy} & \Omega_{xx} \end{pmatrix} = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} E(u_t u_s')$$

and

$$\Gamma = \begin{pmatrix} \gamma_{yy} & \gamma_{yx} \\ \gamma_{xy} & \Gamma_{xx} \end{pmatrix} = \lim_{T\to\infty} T^{-1} \sum_{t=2}^{T} \sum_{s=1}^{t-1} E(u_t u_s'),$$

where the partitioning is in conformity with u_t. Moreover, let $\rho^2 = \omega_{yy}^{-1}\omega_{xy}'\Omega_{xx}^{-1}\omega_{xy}$ be the squared coefficient of multiple correlation computed from Ω, the long-run covariance matrix of u_t. (Because $\Omega = E(u_t u_t')$ when u_t is white noise, the present definition of ρ^2 is consistent with that of Section 2.)

Under A1 and local-to-unity asymptotics, $L_T(\Omega) \to_d \varphi_L(\lambda;\rho^2)$, so an "autocorrelation robust" version of L_T can be obtained by employing the long-run covariance matrix Ω in the definition of the test statistic. Analogously, an autocorrelation robust POI test can be based on $P_T(\bar{\theta};\Omega)$. In general, $P_T(\bar{\theta};\Omega)$ suffers from "serial correlation bias" under A1. Specifically, $P_T(\bar{\theta};\Omega) \to_d \varphi_P(\lambda;\bar{\lambda},\rho^2) + 2\bar{\lambda}\omega_{yy.x}^{-1}\gamma_{yy.x}$, where $\gamma_{yy.x} = \gamma_{yy} - \omega_{xy}'\Omega_{xx}^{-1}\gamma_{xy}$. Let

$$Q_T(\bar{\theta};\Omega,\Gamma) = P_T(\bar{\theta};\Omega) - 2T(1-\bar{\theta})\omega_{yy.x}^{-1}\gamma_{yy.x}. \tag{7}$$

The statistic $Q_T(\bar{\theta};\Omega,\Gamma)$ coincides with $P_T(\bar{\theta};\Omega)$ when u_t is white noise, because $\Gamma = 0$ in that case. More generally, $Q_T(\bar{\theta};\Omega,\Gamma)$ corrects $P_T(\bar{\theta};\Omega)$ for serial correlation bias and $Q_T(\bar{\theta};\Omega,\Gamma) \to_d \varphi_P(\lambda;\bar{\lambda},\rho^2)$ under A1 and local-to-unity asymptotics.

In most (if not all) applications, the tests based on $L_T(\Omega)$ and $Q_T(\bar{\theta};\Omega,\Gamma)$ are infeasible because Ω and Γ are unknown. It therefore seems natural to consider the test statistics $\hat{L}_T = L_T(\hat{\Omega})$ and $\hat{Q}_T(\bar{\theta}) = Q_T(\bar{\theta};\hat{\Omega},\hat{\Gamma})$, where

$$\hat{\Omega} = \begin{pmatrix} \hat{\omega}_{yy} & \hat{\omega}_{xy}' \\ \hat{\omega}_{xy} & \hat{\Omega}_{xx} \end{pmatrix}$$

and

$$\hat{\Gamma} = \begin{pmatrix} \hat{\gamma}_{yy} & \hat{\gamma}_{yx} \\ \hat{\gamma}_{xy} & \hat{\Gamma}_{xx} \end{pmatrix}$$

are estimators of Ω and Γ, respectively.

THEOREM 3. *Let z_t be generated by (1)–(4). Suppose A1 holds and suppose $\lambda = T(1-\theta) \geq 0$ and $\bar{\lambda} = T(1-\bar{\theta}) > 0$ are fixed as T increases without bound. If $(\hat{\Omega},\hat{\Gamma}) \to_p (\Omega,\Gamma)$, then $\hat{Q}_T(\bar{\theta}) \to_d \varphi_P(\lambda;\bar{\lambda},\rho^2)$ and $\hat{L}_T \to_d \varphi_L(\lambda;\rho^2)$.*

Conventional (possibly prewhitened) kernel estimators of Ω and Γ (e.g., Andrews, 1991; Andrews and Monahan, 1992) meet the consistency requirement of Theorem 3. Conditions under which VAR(1) prewhitened kernel estimators are consistent are provided in Section 3.3.

The statistics

$$\hat{Q}_T^y(\bar{\theta}) = Q_T^y(\bar{\theta};\hat{\omega}_{yy},\hat{\gamma}_{yy}) - 2T(1-\bar{\theta})\hat{\omega}_{yy}^{-1}\hat{\gamma}_{yy}$$

and $\hat{L}_T^y = L_T^y(\hat{\omega}_{yy})$ are univariate counterparts of $\hat{Q}_T(\bar{\theta})$ and $\hat{L}_T$, respectively. Under the assumptions of Theorem 3, $\hat{Q}_T^y(\bar{\theta}) \to_d \varphi_P(\lambda;\bar{\lambda},0)$ and $\hat{L}_T^y \to_d \varphi_L(\lambda;0)$. The test statistic $\hat{L}_T^y$ is well known (e.g., Kwiatkowski et al., 1992). On the other hand, the semiparametric version $\hat{Q}_T^y(\bar{\theta})$ of the univariate POI test would appear to be new.

3.2. Asymptotic Power Properties

Saikkonen and Luukkonen (1993a) considered the constant mean case and found that their test statistic $\tilde{R}(1-7/T)$, which corresponds to $\hat{Q}_T^y(1-7/T)$, has a local asymptotic power function that is almost indistinguishable from the univariate power envelope. The choice $\bar{\lambda} = 7$ produces a test that is asymptotically 0.50-optimal, level 0.05 in the sense of Davies (1969). In other words, $\lambda = 7$ is the alternative for which the univariate power envelope for 5% level tests equals 0.50. In the general case, it therefore seems natural to consider $\hat{Q}_T(1-T^{-1}\bar{\lambda}^\dagger)$, where $\bar{\lambda}^\dagger$ is such that the test statistic is asymptotically 0.50-optimal, level 0.05. Although computationally feasible, such a procedure seems cumbersome in view of the fact that the power envelope for 5% level tests depends not only on the order of the deterministic component in the model but also on the parameter ρ^2, which measures the quality of the covariates. To construct test statistics that are asymptotically 0.50-optimal, level 0.05 one would therefore have to use a new $\bar{\lambda}^\dagger$ for each ρ^2. Fortunately, a much simpler approach yields very satisfactory results. The approach taken here is to use the same $\bar{\lambda}^\dagger$ for all values of ρ^2. The value of $\bar{\lambda}^\dagger$ is chosen in such a way that the test is asymptotically 0.50-optimal, level 0.05 in the worst case scenario $\rho^2 = 0$, the case where the univariate test is optimal. This approach generates a test that has excellent power properties (relative to the power envelope) when ρ^2 is low. Moreover, $\hat{Q}_T$ dominates its univariate counterpart for all values of ρ^2. In fact, the test has a power function that is very close to the power envelope even for nonzero values of ρ^2.

Figure 3 illustrates this in the constant mean case with $\rho^2 = 0.50$. In addition to the power envelope and the local asymptotic power of $\hat{Q}_T$, Figure 3 also plots the local power function of the LBI test $\hat{L}_T$ and the univariate tests $\hat{Q}_T^y$ and $\hat{L}_T^y$. Comparing $\hat{Q}_T$ to $\hat{Q}_T^y$, it is seen that the inclusion of covariates can lead to huge gains in power in cases where an appropriate set of covariates can

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

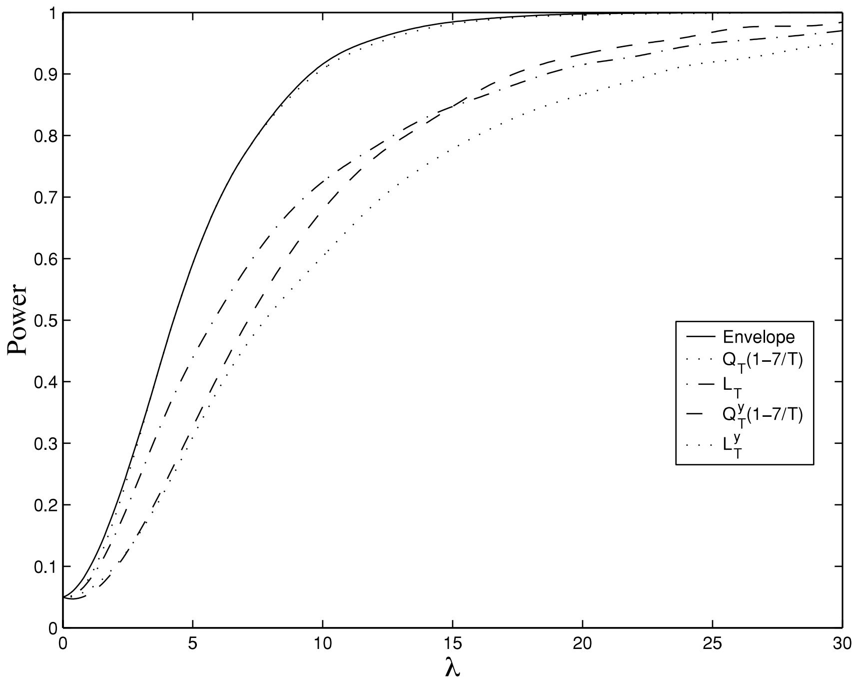

FIGURE 3. Power curves, $\rho^2 = 0.5$: 5% level tests, constant mean ($p = 0$).

be found. The Pitman asymptotic relative efficiency (ARE) of $\hat{Q}_T$ with respect to $\hat{Q}_T^y$ (evaluated at power 0.50) is 1.65, implying that in large samples the univariate test needs 65% more observations than the test using covariates to have comparable power properties when $\rho^2 = 0.50$. The case where covariates are included is qualitatively similar to the univariate case in the sense that the POI test dominates the LBI test for all but extremely small values of λ. Indeed, the inferiority (as measured by the Pitman ARE) of the LBI test is somewhat more pronounced when useful covariates are available.

Figure 4 presents results for the linear trend case. The statistics $\hat{Q}_T$ and $\hat{Q}_T^y$ use $\bar{\lambda}^\dagger = 12$, the value that yields an asymptotically 0.50-optimal, level 0.05 test in the univariate case. All power curves lie below the curves for the constant mean case, but the pattern is the same as in Figure 3. In particular, the statistic $\hat{Q}_T$ has a power function that lies close to the envelope and far above the power functions corresponding to $\hat{L}_T$ and $\hat{Q}_T^y$. For instance, the Pitman ARE of $\hat{Q}_T$ with respect to $\hat{Q}_T^y$ (evaluated at power 0.50) is 1.82, indicating that the inclusion of covariates is even more beneficial in the linear trend case than in the constant mean case.

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

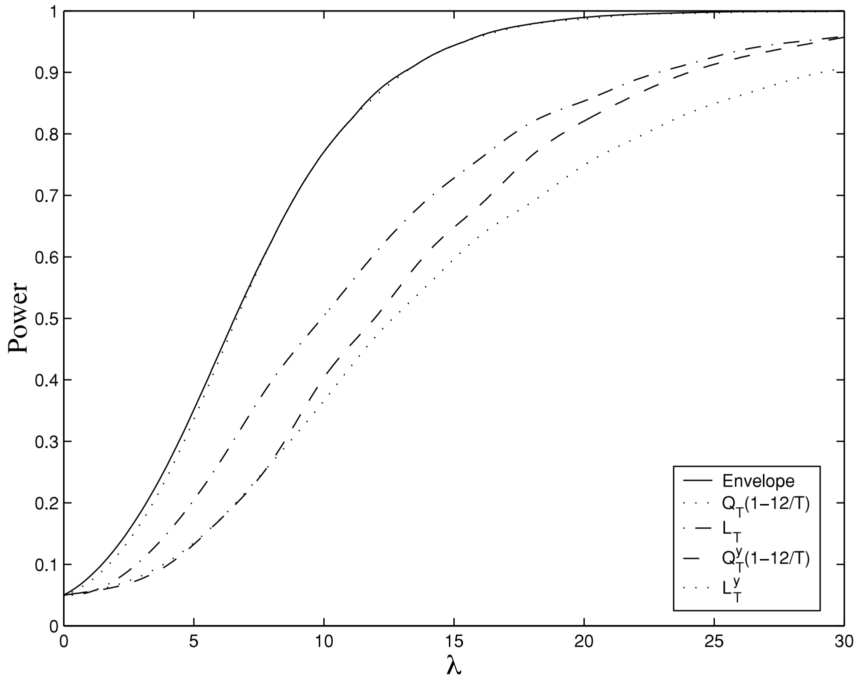

FIGURE 4. Power curves, $\rho^2 = 0.5$: 5% level tests, linear trend $(p = 1)$.

Tables 1 and 2 give various critical values for $\hat{Q}_T$ and $\hat{L}_T$ for $p \in \{0,1\}$, which seem to be the cases of empirical relevance. In the case of $\hat{Q}_T$, the critical values correspond to the recommended values of $\bar{\lambda}^\dagger$, namely, $\bar{\lambda}^\dagger = 7$ when $p = 0$ and $\bar{\lambda}^\dagger = 12$ when $p = 1$. The critical values are presented for ρ^2 in steps of 0.1. The recommendation is to use the critical value corresponding to $\hat{\rho}^2 = \hat{\omega}_{yy}^{-1} \hat{\omega}_{xy}' \hat{\Omega}_{xx}^{-1} \hat{\omega}_{xy}$ computed from $\hat{\Omega}$. Interpolation can be used to obtain critical values for values of $\hat{\rho}^2$ between those given in the tables.

In general, point optimal and locally optimal tests may fail to be consistent in curved statistical models (van Garderen, 2000). In view of the following fixed parameter result, the tests based on $\hat{Q}_T$ and $\hat{L}_T$ are consistent if $\hat{\Omega}$ and $\hat{\gamma}_{yy.x}$ are well behaved under fixed alternatives.

THEOREM 4. *Let z_t be generated by (1)–(4). Suppose A1 holds and suppose $\theta < 1$ and $\bar{\lambda} = T(1 - \bar{\theta}) > 0$ are fixed as T increases without bound. If $\hat{\gamma}_{yy.x} = o_p(T^2)$, $\hat{\omega}_{yy} = o_p(T^2)$, $\hat{\omega}_{xy} = o_p(T)$, and $\hat{\Omega}_{xx}^{-1} = O_p(1)$, then*

$$\lim_{T\to\infty} \Pr(\hat{Q}_T(\bar{\theta}) > c) = \lim_{T\to\infty} \Pr(\hat{L}_T > c) = 1$$

for any $c \in R$.

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

TABLE 1. Percentiles of $\hat{L}_T$ and $\hat{Q}_T(1-7/T)$, constant mean case $(p=0)$

	$\hat{L}_T$				$\hat{Q}_T(1-7/T)$			
ρ^2	90%	95%	97.5%	99%	90%	95%	97.5%	99%
0	0.348	0.458	0.589	0.748	−1.969	−0.973	0.055	1.451
0.1	0.362	0.484	0.622	0.804	−1.938	−0.854	0.244	1.588
0.2	0.382	0.516	0.652	0.867	−1.880	−0.787	0.361	1.663
0.3	0.404	0.571	0.725	0.940	−1.887	−0.694	0.345	1.968
0.4	0.444	0.621	0.797	1.059	−1.989	−0.761	0.460	2.049
0.5	0.493	0.701	0.924	1.216	−2.146	−0.740	0.575	2.110
0.6	0.572	0.838	1.124	1.541	−2.518	−0.964	0.448	2.249
0.7	0.665	0.999	1.337	1.812	−3.079	−1.458	0.028	2.058
0.8	0.942	1.430	1.930	2.583	−4.821	−2.813	−0.841	1.216
0.9	1.750	2.736	3.743	5.126	−9.932	−7.054	−4.650	−1.805

Note: The percentiles were computed by generating 20,000 draws from the discrete time approximation (based on 2,000 steps) to the limiting random variables.

3.3. Covariance Matrix Estimation

Under fairly general conditions, the requirements of Theorems 3 and 4 are met by VAR(1) prewhitened kernel estimators with plug-in bandwidths. These estimators are defined as follows.

TABLE 2. Percentiles of $\hat{L}_T$ and $\hat{Q}_T(1-12/T)$, linear trend case $(p=1)$

	$\hat{L}_T$				$\hat{Q}_T(1-12/T)$			
ρ^2	90%	95%	97.5%	99%	90%	95%	97.5%	99%
0	0.118	0.147	0.176	0.214	−5.019	−3.927	−2.959	−1.634
0.1	0.120	0.151	0.185	0.228	−4.944	−3.807	−2.660	−1.208
0.2	0.117	0.148	0.180	0.226	−5.162	−3.970	−2.736	−1.405
0.3	0.115	0.149	0.185	0.236	−5.317	−4.035	−2.686	−1.115
0.4	0.115	0.153	0.197	0.251	−5.600	−4.224	−2.975	−1.377
0.5	0.112	0.157	0.207	0.273	−6.106	−4.431	−3.121	−1.141
0.6	0.114	0.170	0.222	0.297	−6.993	−5.130	−3.410	−1.209
0.7	0.115	0.183	0.258	0.358	−8.546	−6.303	−4.442	−2.060
0.8	0.128	0.222	0.339	0.485	−11.941	−9.278	−6.858	−4.062
0.9	0.143	0.336	0.545	0.839	−23.141	−18.951	−15.393	−11.340

Note: The percentiles were computed by generating 20,000 draws from the discrete time approximation (based on 2,000 steps) to the limiting random variables.

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

For $t = 2,\ldots,T$, let $\hat{v}_t^{PW} = \hat{v}_t - \hat{A}\hat{v}_{t-1}$, where $\hat{A}$ is a $(k+1)\times(k+1)$ matrix and $\hat{v}_t = z_t - d_t'(\sum_{s=1}^T d_s d_s')^{-1}(\sum_{s=1}^T d_s z_s)$. Define

$$\hat{\Sigma} = T^{-1}\sum_{t=1}^T \hat{v}_t\hat{v}_t',$$

$$\hat{\Lambda} = (T-1)^{-1}\sum_{t=2}^T \hat{v}_t^{PW}\hat{v}_{t-1}',$$

$$\hat{\Omega}^{PW} = (T-1)^{-1}\sum_{t=2}^T\sum_{s=2}^T k\left(\frac{|t-s|}{\hat{b}_T}\right)\hat{v}_t^{PW}\hat{v}_s^{PW\prime},$$

and

$$\hat{\Gamma}^{PW} = (T-1)^{-1}\sum_{t=3}^T\sum_{s=2}^{t-1} k\left(\frac{|t-s|}{\hat{b}_T}\right)\hat{v}_t^{PW}\hat{v}_s^{PW\prime},$$

where $k(\cdot)$ is a kernel and $\{\hat{b}_T\}$ is a sequence of (possibly sample-dependent) bandwidth parameters. The proposed estimators of Ω and Γ are

$$\hat{\Omega} = (I-\hat{A})^{-1}\hat{\Omega}^{PW}(I-\hat{A}')^{-1}$$

and

$$\hat{\Gamma} = (I-\hat{A})^{-1}\hat{\Gamma}^{PW}(I-\hat{A}')^{-1} + (I-\hat{A})^{-1}\hat{A}\hat{\Sigma} - (I-\hat{A})^{-1}\hat{\Lambda}\hat{A}'(I-\hat{A}')^{-1},$$

respectively. Consider the following assumption.

A2.

(i) $k(0) = 1, k(\cdot)$ is continuous at zero, $\sup_{s\geq 0}|k(s)| < \infty$, and $\int_0^\infty \bar{k}(r)\,dr < \infty$, where $\bar{k}(r) = \sup_{s\geq r}|k(s)|$ (for every $r \geq 0$).
(ii) $\hat{b}_T = \hat{a}_T b_T$, where $\hat{a}_T$ and b_T are positive with $\hat{a}_T + \hat{a}_T^{-1} = O_p(1)$ and $b_T^{-1} + T^{-1/2}b_T = o(1)$.
(iii) $T^{1/2}(\hat{A} - A) = O_p(1)$ for some A such that $(I - A)$ is nonsingular.
(iv) The matrix A in (iii) is block upper triangular.

Assumption A2(i) is discussed in Jansson (2002), whereas Assumptions A2(ii) and (iii) are adapted from Andrews and Monahan (1992). Assumption A2(iv) is helpful when studying the behavior of $\hat{\Omega}$ and $\hat{\Gamma}$ under fixed alternatives. When $\hat{A} = 0$, $\hat{\Omega}$ and $\hat{\Gamma}$ are standard kernel estimators and A2(iii) and (iv) are trivially satisfied. A nondegenerate prewhitening matrix satisfying A2(iii) is discussed subsequently.

LEMMA 5. *Let z_t be generated by (1)–(4). Suppose A1 and A2(i)–(iii) hold and suppose $\lambda = T(1-\theta) \geq 0$ and $\bar{\lambda} = T(1-\bar{\theta}) > 0$ are fixed as T increases without bound. Then* $(\hat{\Omega},\hat{\Gamma}) \to_p (\Omega,\Gamma)$.

LEMMA 6. *Let z_t be generated by (1)–(4). Suppose A1 and A2 hold and suppose $\theta < 1$ and $\bar{\lambda} = T(1-\bar{\theta}) > 0$ are fixed as T increases without bound. Then $\hat{\gamma}_{yy.x} = o_p(T^2)$, $\hat{\omega}_{yy} = o_p(T^2)$, $\hat{\omega}_{xy} = o_p(T)$, and $\hat{\Omega}_{xx}^{-1} = O_p(1)$.*

Under local alternatives (i.e., under the assumptions of Theorem 3 and Lemma 5), A2(iii) is satisfied by the least squares estimator

$$\hat{A}_{LS} = \left(\sum_{t=2}^{T} \hat{v}_t \hat{v}_{t-1}'\right)\left(\sum_{t=2}^{T} \hat{v}_{t-1} \hat{v}_{t-1}'\right)^{-1}.$$

On the other hand, standard cointegration arguments can be used to show that the first column of $\hat{A}_{LS}$ converges at rate T to first unit vector in $\mathbb{R}^{k+1}$ under fixed alternatives (i.e., under the assumptions of Theorem 4 and Lemma 6). As a consequence, $\hat{A}_{LS}$ violates A2(iii) under fixed alternatives.

An estimator $\hat{A}$ satisfying A2(iii) under both local and fixed alternatives can be obtained by modifying $\hat{A}_{LS}$ as follows. Let $\hat{M}_{LS}\hat{J}_{LS}\hat{M}_{LS}^{-1}$ be the Jordan decomposition of $\hat{A}_{LS}$. Define $\hat{A} = \hat{M}_{LS}\hat{J}\hat{M}_{LS}^{-1}$, where $\hat{J}$ is a Jordan matrix obtained from $\hat{J}_{LS}$ by dividing the diagonal elements of each Jordan block by $\max(1,|\mu|/0.97)$, where μ is the eigenvalue (real or complex) associated with the Jordan block and $|\cdot|$ denotes absolute value. This adjustment preserves the eigenvectors of $\hat{A}_{LS}$ and bounds the eigenvalues of $\hat{A}$ away from unity. By construction, $\hat{A} = \hat{A}_{LS}$ whenever the eigenvalues of $\hat{A}_{LS}$ do not exceed 0.97. More generally, the properties of $\hat{A}$ are easily deduced once the properties of $\hat{A}_{LS}$ have been established. In particular, $\hat{A}$ satisfies A2(iii) whenever $T^{1/2}(\hat{A}_{LS} - A_{LS}) = O_p(1)$ for some A_{LS} (as is true under both local and fixed alternatives), whereas A2(iv) holds if the matrix A_{LS} is block upper triangular (as is the case under fixed alternatives). Lemmas 5 and 6 therefore demonstrate the plausibility of the high-level assumptions on $\hat{\Omega}$ and $\hat{\Gamma}$ made in Theorems 3 and 4, respectively.

3.4. Finite Sample Properties

To investigate the finite sample properties of the test statistics introduced in Section 3.1, a small Monte Carlo experiment is conducted. Samples of size $T = 200$ are generated according to (1)–(4). The errors u_t are generated by the bivariate model

$$\begin{pmatrix} u_t^y \\ u_t^x \end{pmatrix} = \begin{pmatrix} c_{yy}(L) & 0 \\ \rho & (1-\rho^2)^{1/2} \end{pmatrix}\begin{pmatrix} \varepsilon_t^y \\ \varepsilon_t^x \end{pmatrix}, \tag{8}$$

where $(\varepsilon_t^y, \varepsilon_t^x)' \sim i.i.d.\ \mathcal{N}(0, I_2)$ and $c_{yy}(1) = 1$. Two specifications of $c_{yy}(L)$ are considered:

$$c_{yy}^{AR}(L) = (1-a)\sum_{i=0}^{\infty} a^i L^i, \qquad a \in \{-0.8, -0.5, -0.2, 0.2, 0.5, 0.8\}$$

and

$$c_{yy}^{MA}(L) = \frac{1}{1+b}(1+bL), \qquad b \in \{-0.8,-0.5,-0.2,0,0.2,0.5,0.8\},$$

corresponding to an AR(1) and an MA(1) model for u_t^y, respectively. In both cases,

$$\Omega = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T}\sum_{s=1}^{T} E(u_t u_s') = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

In particular, the parameter ρ in (8) is the correlation coefficient computed from Ω.

The parameters Ω and Γ are estimated using VAR(1) prewhitened kernel estimators. Specifically, $\hat{\Omega}^{PW}$ and $\hat{\Gamma}^{PW}$ are constructed using the quadratic spectral kernel (which clearly satisfies Assumption A2(i)) along with a plug-in bandwidth. The value of the plug-in bandwidth is obtained by setting $b_T = 1.3221 \cdot T^{1/5}$ (following Andrews, 1991) and $\hat{a}_T = \max(\min(\hat{\alpha}_{AR(1)}(2)^{1/5}, 5), 0.05)$, where $\hat{\alpha}_{AR(1)}(2)$ is computed from Andrews's (1991) equation (6.4) (with $w_a = 1$ for all a). Because $0.05 \le \hat{a}_T \le 5$ is imposed, A2(ii) is automatically satisfied. In particular, the condition $\hat{a}_T \le 5$ controls the behavior of the estimated bandwidth under fixed alternatives, thereby circumventing the problems discussed by Choi (1994). Finally, the matrix $\hat{A}$ used in the prewhitening procedure was computed by modifying the ordinary least squares (OLS) estimator in the manner described in Section 3.3.

Tables 3 and 4 and 5 and 6 summarize the results for the constant mean and linear trend cases, respectively. The tables report the observed rejection rates of 5% level tests implemented using critical values based on the estimate $\hat{\rho}^2$ computed from $\hat{\Omega}$. As was the case with the asymptotic analysis of Section 3.2, the simulation evidence is favorable to the tests developed in this paper. The rejection rates of the new tests are quite similar to those of their univariate counterparts under the null hypothesis. No noticeable loss in power is observed in the case where the covariates are uninformative (when $\rho^2 = 0$), whereas substantial power gains are achieved in the cases where the covariates do carry information about y_t.

In addition to documenting the superiority of the new tests, the simulation evidence also points out some problems with the small sample properties of the new tests and their univariate counterparts. Rejection rates under the null tend to fall far short of the nominal level in the MA(1) model with $|b| \ge 0.5$, which leads to an unnecessary reduction in power when asymptotic critical values are used. Likewise, power is very low in the AR(1) model with $a = 0.8$, especially so for the point optimal tests. Moreover, the pattern exhibited by the rejection rates in the AR(1) model with $a = 0.8$ is rather peculiar. In part, the latter phenomenon appears to be due to imprecision of the estimates of Ω and Γ, because simulation results (not reported here) show that the power of the

TABLE 3. Monte Carlo rejection rates (AR(1) model, 5% level tests, constant mean, $T = 200$)

				$\hat{L}_T, \rho^2 =$				$\hat{Q}_T, \rho^2 =$			
a	θ	$\hat{L}_T^y$	$\hat{Q}_T^y$	0	0.2	0.5	0.8	0	0.2	0.5	0.8
−0.8	1	6.0	2.1	5.9	5.9	6.3	8.1	2.0	1.8	1.6	1.0
	0.975	31.1	22.6	30.8	35.9	43.5	59.7	22.4	28.3	39.9	63.6
	0.950	58.7	55.8	58.5	64.9	70.3	79.8	55.0	66.8	81.5	95.8
	0.925	75.2	77.4	74.8	78.1	82.0	87.2	76.4	85.5	94.3	99.5
	0.900	83.4	88.6	83.1	85.4	87.3	91.0	88.2	92.9	97.8	99.9
−0.5	1	5.1	3.9	5.1	5.4	5.0	5.1	4.2	1.8	3.3	2.4
	0.975	31.0	28.6	30.9	34.7	42.6	59.7	27.8	28.3	51.7	81.7
	0.950	60.2	64.1	60.0	63.2	69.3	79.7	63.7	66.8	87.5	99.1
	0.925	75.9	82.3	75.7	77.2	82.2	88.2	81.7	85.5	96.5	100.0
	0.900	84.2	91.1	84.0	85.7	88.3	92.8	91.0	92.9	99.1	100.0
−0.2	1	5.0	4.5	5.0	5.4	4.7	4.6	4.4	4.6	4.8	4.3
	0.975	31.9	31.7	31.2	34.0	42.0	58.1	31.2	37.6	53.9	84.3
	0.950	59.7	65.1	59.2	62.0	68.8	78.3	64.5	73.1	88.8	99.2
	0.925	75.4	82.7	75.0	76.9	80.9	86.9	82.3	88.7	97.4	99.9
	0.900	83.6	91.0	83.5	84.9	87.6	91.0	90.5	95.0	99.3	100.0
0.2	1	5.1	5.0	5.3	4.5	4.3	4.0	5.1	4.2	4.2	4.3
	0.975	30.7	30.1	30.2	32.7	41.1	55.4	29.2	36.5	52.1	83.0
	0.950	59.1	64.1	58.4	59.5	65.7	72.5	63.0	71.4	86.2	98.5
	0.925	73.6	80.9	72.8	73.9	76.9	77.7	80.3	87.2	95.4	99.6
	0.900	81.1	89.3	80.7	81.0	81.5	78.7	88.6	93.4	98.2	98.9
0.5	1	4.8	3.9	4.7	4.7	5.2	4.9	4.0	3.5	3.8	4.4
	0.975	28.3	26.2	27.8	31.8	39.1	55.4	25.1	32.5	48.3	77.6
	0.950	53.2	56.4	52.3	56.1	62.2	71.1	54.9	65.0	81.4	96.5
	0.925	64.9	72.8	63.8	67.2	70.4	73.9	70.1	79.5	90.3	96.0
	0.900	69.5	79.2	68.2	69.9	71.3	71.1	75.4	81.6	88.2	88.1
0.8	1	3.6	1.3	3.8	3.7	4.1	6.1	1.5	1.6	1.5	2.4
	0.975	18.0	6.6	17.0	20.3	28.7	49.9	6.3	9.6	19.0	45.8
	0.950	24.6	6.3	22.2	27.0	36.2	56.8	5.7	8.8	18.6	49.9
	0.925	16.3	2.7	15.1	18.0	26.3	50.3	2.5	3.8	7.3	27.8
	0.900	11.2	2.3	10.2	12.3	19.3	40.3	2.0	2.5	3.8	13.0

Note: Based on 5,000 Monte Carlo replications.

infeasible tests using the true values of Ω and Γ is monotonic in θ. It follows from Theorem 4 that the low power in the AR(1) model with $a = 0.8$ is a finite sample phenomenon. In an attempt to quantify the effect of a change in the sample size for moderate values of T, Tables 7 and 8 investigate the power

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

TABLE 4. Monte Carlo rejection rates (MA(1) model, 5% level tests, constant mean, $T = 200$)

				$\hat{L}_T$, $\rho^2 =$				$\hat{Q}_T$, $\rho^2 =$			
b	θ	$\hat{L}_T^y$	$\hat{Q}_T^y$	0	0.2	0.5	0.8	0	0.2	0.5	0.8
−0.8	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.975	0.1	0.0	0.1	0.4	1.3	5.4	0.0	0.0	0.0	0.3
	0.950	6.5	1.5	6.4	8.8	13.8	28.8	1.4	2.6	4.3	13.3
	0.925	19.2	10.1	19.2	23.8	32.2	47.3	9.6	13.2	20.0	40.3
	0.900	32.5	22.6	32.2	36.9	45.6	58.7	22.0	27.6	39.7	63.7
−0.5	1	0.7	0.1	0.7	0.3	0.6	0.7	0.2	0.0	0.0	0.0
	0.975	13.3	8.0	13.4	16.5	22.8	39.8	7.7	10.3	14.7	31.1
	0.950	40.9	36.5	40.7	45.0	53.2	66.7	35.9	43.5	59.1	84.8
	0.925	60.3	60.8	59.7	62.8	69.7	78.4	60.1	67.7	82.5	97.1
	0.900	71.9	75.9	71.6	73.7	79.0	85.2	75.3	81.9	92.7	99.4
−0.2	1	3.7	3.1	3.8	3.7	3.8	3.2	3.3	3.0	2.6	2.2
	0.975	28.8	26.9	28.5	31.4	40.0	55.4	26.2	32.0	48.2	78.2
	0.950	56.9	61.0	56.6	60.4	68.2	76.2	60.7	69.9	86.0	98.8
	0.925	73.4	80.6	73.0	75.9	81.2	85.8	79.9	87.2	96.3	99.9
	0.900	82.4	89.4	82.0	83.7	87.7	90.1	89.0	94.2	99.0	100.0
0	1	5.2	5.0	5.3	5.0	4.5	4.2	5.3	4.3	4.1	4.3
	0.975	31.4	31.0	30.9	34.3	40.3	56.4	30.2	38.2	54.1	84.0
	0.950	60.3	65.9	59.7	62.4	67.5	75.9	64.7	73.8	88.6	99.1
	0.925	75.1	82.5	74.8	76.4	79.4	83.2	82.0	88.6	96.7	99.9
	0.900	83.1	90.9	82.6	84.4	85.4	86.7	90.3	94.6	98.8	100.0
0.2	1	4.3	3.7	4.2	3.7	4.2	4.5	3.7	3.1	3.3	3.6
	0.975	28.1	26.7	28.1	30.3	38.8	56.2	26.5	32.1	48.5	79.9
	0.950	56.1	60.6	55.6	58.3	65.4	74.8	59.8	69.1	85.7	98.5
	0.925	70.4	78.7	69.9	72.5	76.6	80.6	77.7	85.8	95.0	99.7
	0.900	78.9	87.6	78.0	79.8	81.7	82.2	87.0	92.5	98.2	99.4
0.5	1	1.7	0.9	1.7	2.0	2.8	5.8	0.8	0.8	1.0	2.1
	0.975	19.1	14.6	19.0	24.1	36.1	58.4	14.4	19.6	35.5	76.0
	0.950	45.4	44.5	44.8	51.7	62.3	76.8	43.3	55.6	77.1	98.1
	0.925	60.9	65.8	60.6	65.3	72.7	82.0	64.2	75.6	91.8	99.8
	0.900	68.4	77.4	67.7	72.1	76.9	83.8	75.8	84.9	95.9	99.4
0.8	1	1.0	0.3	1.1	1.2	2.9	6.9	0.4	0.3	0.6	2.3
	0.975	14.7	8.8	14.4	20.7	34.7	57.5	8.4	14.5	31.5	75.6
	0.950	41.4	36.6	40.2	46.8	59.5	75.9	35.4	46.8	72.9	98.1
	0.925	56.2	56.6	55.3	60.1	70.7	81.5	54.3	67.3	88.2	99.7
	0.900	63.3	67.0	62.5	66.9	76.0	84.3	63.8	77.3	93.5	99.7

Note: Based on 5,000 Monte Carlo replications.

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

Table 5. Monte Carlo rejection rates (AR(1) model, 5% level tests, linear trend, $T = 200$)

				$\hat{L}_T$, $\rho^2 =$				$\hat{Q}_T$, $\rho^2 =$			
a	θ	$\hat{L}_T^y$	$\hat{Q}_T^y$	0	0.2	0.5	0.8	0	0.2	0.5	0.8
−0.8	1	7.4	0.5	7.4	7.7	9.5	11.9	0.4	0.3	0.1	0.0
	0.975	16.1	1.9	16.0	18.9	24.8	37.5	1.8	2.9	4.7	6.7
	0.950	39.9	13.2	39.3	43.8	52.8	66.3	13.0	19.9	33.5	57.7
	0.925	59.9	36.0	59.1	65.1	73.6	83.1	35.0	46.1	66.8	90.0
	0.900	73.9	57.1	73.5	78.7	85.0	91.0	55.7	68.8	85.3	97.7
−0.5	1	5.6	2.4	5.7	5.6	5.5	6.1	2.5	2.1	1.7	0.6
	0.975	14.4	8.0	14.2	15.7	20.7	33.9	7.5	10.7	18.6	36.6
	0.950	35.3	28.4	34.8	41.0	51.1	64.4	27.5	40.1	61.3	91.5
	0.925	57.3	54.1	56.8	63.7	71.4	80.7	53.2	68.8	86.2	99.2
	0.900	72.6	73.2	72.3	77.5	83.0	89.8	72.0	84.8	95.6	99.9
−0.2	1	5.5	4.1	5.5	5.2	4.9	4.9	4.0	3.4	3.1	3.0
	0.975	14.0	10.7	13.7	14.7	19.9	32.0	10.5	14.9	26.0	53.8
	0.950	36.3	33.9	36.0	40.0	48.8	62.5	33.0	45.1	69.6	96.1
	0.925	58.4	60.1	57.9	61.9	70.3	79.8	59.0	72.4	90.5	99.7
	0.900	73.0	77.4	72.5	76.5	82.1	89.2	76.6	86.9	97.2	100.0
0.2	1	4.7	4.4	4.5	4.6	5.3	5.0	4.1	3.9	4.6	3.8
	0.975	12.8	10.6	12.7	13.4	20.3	31.5	10.1	16.1	29.3	52.9
	0.950	32.5	33.3	32.1	36.4	47.1	58.0	32.3	44.5	69.0	94.9
	0.925	53.0	56.3	52.0	55.9	65.6	71.5	54.4	68.4	89.2	99.4
	0.900	65.4	72.0	64.7	69.2	76.0	78.6	70.5	82.1	95.9	99.9
0.5	1	4.4	3.2	4.4	4.2	5.4	5.2	3.6	3.2	3.3	3.2
	0.975	11.1	8.6	11.2	11.8	17.2	30.3	8.7	12.6	21.7	42.6
	0.950	27.7	25.5	26.9	30.8	38.9	53.6	25.1	35.1	56.7	86.5
	0.925	43.6	44.2	42.8	47.8	54.6	65.8	41.6	54.5	77.0	96.1
	0.900	55.3	58.0	53.5	56.9	61.6	71.1	54.0	66.0	84.0	96.0
0.8	1	2.9	0.7	2.8	3.7	4.2	7.5	0.7	0.7	0.9	0.9
	0.975	6.3	1.2	6.2	8.4	13.5	29.4	1.2	2.1	4.1	11.0
	0.950	11.8	1.6	11.3	15.4	23.5	43.3	1.8	2.5	5.8	22.0
	0.925	11.2	0.8	11.0	13.9	21.6	44.2	1.3	1.2	2.5	12.9
	0.900	7.0	0.3	7.5	9.1	15.0	35.8	0.6	0.4	0.7	3.4

Note: Based on 5,000 Monte Carlo replications.

against the (fixed) alternative $\theta = 0.9$ for $T \in \{200,300,400,500\}$ in the AR(1) model with $a = 0.8$. As the sample size increases, power increases in all cases but remains disappointingly low in the case of the point optimal test. Indeed, even in samples of size $T = 500$ the point optimal test fails to dominate the

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

TABLE 6. Monte Carlo rejection rates (MA(1) model, 5% level tests, linear trend, $T = 200$)

				$\hat{L}_T$, $\rho^2 =$				$\hat{Q}_T$, $\rho^2 =$			
b	θ	$\hat{L}_T^y$	$\hat{Q}_T^y$	0	0.2	0.5	0.8	0	0.2	0.5	0.8
−0.8	1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.0
	0.975	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	1.8	0.0
	0.950	0.1	0.0	0.1	0.2	1.2	5.1	0.0	0.0	12.8	0.0
	0.925	1.1	0.1	1.1	2.2	5.7	16.5	0.0	0.0	32.0	1.1
	0.900	5.2	0.3	5.0	7.6	47.4	29.8	0.3	0.7	47.8	8.3
−0.5	1	0.4	0.0	0.4	5.2	0.6	0.7	0.0	0.0	0.0	0.0
	0.975	2.2	0.2	2.2	3.0	5.3	12.0	0.3	0.5	0.7	1.4
	0.950	13.1	5.1	12.9	16.1	25.4	41.2	4.8	7.2	15.3	36.7
	0.925	31.7	20.5	31.5	35.7	47.3	62.3	19.7	26.8	45.5	79.0
	0.900	49.1	39.9	48.7	52.3	63.0	77.3	39.2	50.2	72.0	95.5
−0.2	1	3.7	2.5	3.7	3.4	3.9	3.3	2.3	2.1	1.8	1.1
	0.975	10.8	7.4	10.7	12.5	16.2	28.0	7.0	11.0	18.6	38.3
	0.950	30.6	27.4	30.1	35.8	44.5	58.9	26.7	38.0	61.1	92.3
	0.925	52.3	52.5	51.4	57.8	66.4	77.5	51.0	65.9	87.2	99.4
	0.900	68.4	73.0	67.7	72.6	80.1	87.0	71.9	82.2	96.0	100.0
0	1	4.7	4.0	4.9	5.2	4.9	4.5	4.1	4.3	3.8	3.8
	0.975	12.7	10.7	12.5	14.9	20.5	31.0	10.3	16.1	27.6	56.1
	0.950	34.1	34.2	33.3	38.5	47.9	59.9	33.2	46.6	69.5	96.5
	0.925	56.0	60.0	55.1	59.6	67.5	75.6	58.5	71.5	89.8	99.7
	0.900	70.8	77.4	70.1	73.5	78.6	84.3	75.8	86.1	96.7	100.0
0.2	1	3.4	2.9	3.5	4.0	4.2	4.7	2.8	2.6	2.3	2.5
	0.975	9.7	7.9	9.6	13.7	17.7	30.6	7.7	13.9	23.0	46.0
	0.950	29.9	28.8	29.6	36.3	44.2	57.3	27.5	41.3	62.5	92.9
	0.925	50.6	52.8	50.1	55.4	63.2	71.7	51.8	66.9	85.6	99.2
	0.900	65.6	71.0	65.0	69.2	74.6	79.8	69.9	82.0	94.1	99.9
0.5	1	0.9	0.3	1.0	1.7	3.3	6.3	0.3	0.7	0.6	1.6
	0.975	4.0	1.7	3.6	6.9	14.1	32.5	1.7	4.6	9.9	38.6
	0.950	17.2	11.9	16.7	23.7	38.4	59.2	11.5	20.9	44.2	90.2
	0.925	34.2	29.9	33.4	40.9	56.2	73.3	28.8	43.0	71.9	98.8
	0.900	48.0	46.3	46.5	54.8	67.7	81.0	44.5	61.4	87.2	99.8
0.8	1	0.4	0.0	0.5	1.0	2.8	6.5	0.1	0.1	0.4	1.3
	0.975	1.9	0.4	2.0	4.6	13.3	33.0	0.5	1.8	6.9	36.7
	0.950	11.3	4.9	11.0	19.3	35.7	58.9	4.8	12.8	37.1	90.5
	0.925	25.1	16.1	24.2	35.3	53.9	74.3	15.3	32.0	65.9	98.9
	0.900	37.4	29.0	36.3	47.4	65.3	82.4	27.6	47.8	81.6	99.8

Note: Based on 5,000 Monte Carlo replications.

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

TABLE 7. Monte Carlo rejection rates (AR(1) model, $a = 0.8$, $\theta = 0.9$, 5% level tests, constant mean)

			$\hat{L}_T$, $\rho^2 =$				$\hat{Q}_T$, $\rho^2 =$			
T	$\hat{L}_T^y$	$\hat{Q}_T^y$	0	0.2	0.5	0.8	0	0.2	0.5	0.8
200	10.1	2.1	9.0	12.3	19.9	39.0	1.9	2.4	4.5	12.5
300	26.3	9.1	22.8	27.4	34.7	53.2	7.9	10.4	15.0	30.8
400	42.3	22.6	38.8	41.7	47.4	62.0	21.1	24.2	30.0	51.6
500	53.3	37.8	49.9	52.1	57.6	69.6	34.4	38.8	48.0	70.7

Note: Based on 5,000 Monte Carlo replications.

locally optimal test. As a consequence, the locally optimal test is likely to be superior to the point optimal test in cases where the time series is believed to be highly persistent under the null hypothesis.

4. COINTEGRATION TESTING WITH A PRESPECIFIED COINTEGRATION VECTOR

An example of the applicability of the tests proposed in this paper can be obtained from the theory of cointegrated time series. Suppose $(Y_t, X_t')'$ is a $(k+1)$-vector integrated process generated by the cointegrated system

$$Y_t = \mu_t^Y + \psi' X_t + u_t^Y,$$

$$\Delta X_t = \Delta\mu_t^X + u_t^X,$$

where Y_t is a scalar, X_t is a k-vector, μ_t^Y and μ_t^X are deterministic components, and $(u_t^Y, u_t^{X\prime})'$ satisfies A1. Setting $y_t = Y_t - \psi' X_t$, $\mu_t^y = \mu_t^Y - \psi' \mu_t^X$, $x_t = \Delta X_t$, and $\mu_t^x = \mu_t^X$, the cointegration model reduces to (1)–(4) with $(u_t^y, u_t^{x\prime})' =$

TABLE 8. Monte Carlo rejection rates (AR(1) model, $a = 0.8$, $\theta = 0.9$, 5% level tests, linear trend)

			$\hat{L}_T$, $\rho^2 =$				$\hat{Q}_T$, $\rho^2 =$			
T	$\hat{L}_T^y$	$\hat{Q}_T^y$	0	0.2	0.5	0.8	0	0.2	0.5	0.8
200	6.6	0.3	6.9	8.6	14.3	35.0	0.3	0.4	0.7	2.9
300	14.0	0.7	13.0	15.4	23.0	46.0	1.0	1.2	2.5	12.2
400	21.1	2.5	19.6	24.1	33.3	54.8	2.2	3.8	6.9	23.6
500	31.9	6.0	29.1	31.6	39.3	61.8	5.9	8.1	12.5	36.7

Note: Based on 5,000 Monte Carlo replications.

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

$(u_t^Y, u_t^{X\prime})'$ and $\theta = 1$. In this context, the null hypothesis $\theta = 1$ is the hypothesis that $(Y_t, X_t')'$ is cointegrated with cointegrating vector $(1, -\psi')'$, whereas the alternative $\theta < 1$ is the hypothesis that $(Y_t, X_t')'$ is not cointegrated.

In many applications, the (potentially) cointegrating vector $(1, -\psi')'$ is known a priori from economic theory (e.g., Horvath and Watson, 1995; Zivot, 2000).[3] In such cases, the null hypothesis that $(Y_t, X_t')'$ is cointegrated with cointegrating vector $(1, -\psi')'$ is invariably tested by applying a univariate stationarity test to the series $Y_t - \psi' X_t$, thereby discarding the potentially useful information contained in the series ΔX_t. As indicated by the results of the previous sections, this empirical practice may lead to a dramatic and unnecessary reduction in power in situations where the zero-frequency correlation between ΔX_t and $Y_t - \psi' X_t$ is nonzero. In economic applications, such nonzero correlations are the rule rather than the exception.[4] When interpreted as tests of the null hypothesis of cointegration with a prespecified cointegrating vector, the stationarity tests proposed in the present paper therefore seem much more attractive than their univariate counterparts currently used in empirical work.

As an illustration, the tests are used to examine the relevance of long-run purchasing power parity (PPP). Specifically, the bilateral intercountry relationship between the United States, the domestic country, and the United Kingdom, the foreign country, is considered. The aim is to test the following version of the PPP hypothesis (e.g., Froot and Rogoff, 1995):

$$s_t = \beta_0 + \beta_1 t + \psi^D p_t^D + \psi^F p_t^F + u_t, \tag{9}$$

where s_t is the logarithm of domestic currency price of a unit of foreign exchange, p_t^D and p_t^F are the logarithms of the price indices in the domestic and foreign countries, and u_t is a stationary error term capturing deviations from PPP. In this setup, a rejection of the null hypothesis of cointegration is interpreted as evidence against long-run PPP. Upon imposing the symmetry and proportionality restriction $\psi^D = -\psi^F = 1$, the problem reduces to that of testing whether the real exchange rate $s_t - p_t^D + p_t^F$ is (trend-)stationary. The data consist of $s_t - p_t^D + p_t^F$ and $(\Delta p_t^D, \Delta p_t^F)$, where the inflation rates Δp_t^D and Δp_t^F serve as covariates.

The tests are implemented using quarterly data from the Global Financial Database (GFD). The exchange rate data is from GFD series __GBP_D, and the price series are consumer price indices. Prices for the United States and the United Kingdom are from GFD series CPUSAM and CPGBRM, respectively. When implementing the tests, the nuisance parameters are estimated in the same way as in the Monte Carlo experiment of Section 3.4. The linear trend version of the test statistics is used. In other words, $p = 1$ is imposed.[5] Two sample periods are considered. One sample period, covering the period from January 1900 through January 2001, spans the twentieth century, whereas the other sample period, covering January 1974 through January 2001, corresponds to the period of the recent float. Table 9 summarizes the results.

Table 9. Tests of long-run PPP

Sample	Univariate Tests		Using Covariates		
	$\hat{L}_T^y$	$\hat{Q}_T^y$	$\hat{L}_T$	$\hat{Q}_T$	$\hat{\rho}^2$
1,900.1–2,001.1	0.081	−8.788	**0.170**	−6.795	0.514
	(0.147)	(−3.927)	**(0.159)**	(−4.529)	
1,974.1–2,001.1	0.019	−9.728	**0.283**	−20.012	0.830
	(0.147)	(−3.927)	**(0.255)**	(−12.150)	

Note: Numbers in parentheses are 5% critical values.

In agreement with other studies (e.g., Culver and Papell, 1999; Kuo and Mikkola, 1999), the tests fail to reject the null hypothesis of stationarity when the covariates are ignored. The tests using covariates, in contrast, provide mixed evidence regarding the validity of long-run PPP. The locally optimal test based on $\hat{L}_T$ rejects the null at the 5% level in both cases, whereas the point optimal test based on $\hat{Q}_T$ fails to reject in both cases. To the extent that the stationary component of $s_t - p_t^D + p_t^F$ might be well approximated by a highly persistent autoregressive process (e.g., Engel, 2000; Kuo and Mikkola, 1999), the fact that $\hat{Q}_T$ fails to reject is to be expected in view of the simulation results reported in Section 3.4. The estimates $\hat{\rho}^2$ are large, suggesting that substantial power gains are achieved by using covariates, which in turn might explain why the $\hat{L}_T$ test reaches different conclusions than the univariate tests.

5. CONCLUSION

The tests proposed here enable researchers to utilize the information contained in related (stationary) time series when testing the null hypothesis of stationarity. Substantial power gains can be achieved by doing so. The new tests are easy to implement and are applicable whenever a set of stationary covariates is available. In particular, they are useful when testing the null hypothesis that a vector integrated process is cointegrated with a prespecified cointegrating vector, because an obvious set of covariates is available in that case.

NOTES

1. In fact, the conclusion of Theorem 2(b) holds whenever $\{\phi_T\}$ is asymptotically of level α.
2. An alternative sufficient condition for the conclusion of Theorem 2(c) is that $\{\phi_T\}$ is asymptotically of level α and $\alpha \leq \Pr(\varphi_L(0;\rho^2) > E(\varphi_L(0;\rho^2)))$.
3. The stationarity tests considered here cannot be used to test the null hypothesis of cointegration if the (potentially) cointegrating vector is unknown. For that testing problem, Shin (1994), Choi and Ahn (1995), and Nyblom and Harvey (2000) propose consistent tests, whereas Jansson (2003) derives a Gaussian power envelope and develops (nearly) efficient tests.

4. In part, this is the raison d'être of the huge literature on efficient inference in cointegrated systems (e.g., Phillips and Hansen, 1990; Phillips, 1991; Saikkonen, 1991, 1992; Park, 1992; Stock and Watson, 1993).

5. Empirical tests of long-run PPP are typically conducted using the constant mean versions of the univariate stationarity tests. The reasons for not imposing $\beta_1 = 0$ in (9) are twofold. First, as pointed out to the author by Maurice Obstfeld, the presence of a deterministic trend component in (9) cannot be ruled out on theoretical grounds. Indeed, a simple Harrod–Balassa–Samuelson model (e.g., Obstfeld and Rogoff, 1996, Chap. 4) in which the differential between productivity growth in tradables and nontradables differs between the home and foreign countries might produce a nonzero β_1 in (9). Second, the real exchange rate appears to have a nonconstant mean, suggesting that β_1 should be unrestricted in (9).

REFERENCES

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Andrews, D.W.K. & J.C. Monahan (1992) An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953–966.

Basawa, I.V. & D.J. Scott (1983) *Asymptotic Optimal Inference for Non-Ergodic Models*. Springer-Verlag.

Billingsley, P. (1999) *Convergence of Probability Measures*, 2nd ed. Wiley.

Choi, I. (1994) Residual-based tests for the null of stationarity with application to U.S. macroeconomic time series. *Econometric Theory* 10, 720–746.

Choi, I. & B.C. Ahn (1995) Testing for cointegration in a system of equations. *Econometric Theory* 11, 952–983.

Culver, S.E. & D.H. Papell (1999) Long-run purchasing power parity with short-run data: Evidence with *Journal of International Money and Finance* 18, 751–768.

Davies, R.B. (1969) Beta-optimal tests and an application to the summary evaluation of experiments. *Journal of the Royal Statistical Society, Series B* 31, 524–538.

Elliott, G. & M. Jansson (2003) Testing for unit roots with stationary covariates. *Journal of Econometrics* 115, 75–89.

Engel, C. (2000) Long-run PPP may not hold after all. *Journal of International Economics* 57, 243–273.

Froot, K.A. & K. Rogoff (1995) Perspectives on PPP and long-run real exchange rates. In G. Grossman & K. Rogoff (eds.), *Handbook of International Economics*, vol. 3, pp. 1647–1688. North-Holland.

Hansen, B.E. (1992) Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 489–500.

Hansen, B.E. (1995) Rethinking the univariate approach to unit root testing: Using covariates to increase power. *Econometric Theory* 11, 1148–1171.

Horvath, M.T.K. & M.W. Watson (1995) Testing for cointegration when some of the cointegrating vectors are prespecified. *Econometric Theory* 11, 984–1014.

Jansson, M. (2002) Consistent covariance matrix estimation for linear processes. *Econometric Theory* 18, 1449–1459.

Jansson, M. (2003) Point Optimal Tests of the Null Hypothesis of Cointegration. Working paper, UC Berkeley.

Jansson, M. & N. Haldrup (2002) Regression theory for nearly cointegrated time series. *Econometric Theory* 18, 1309–1335.

Kuo, B.-S. & A. Mikkola (1999) Re-examining long-run purchasing power parity. *Journal of International Money and Finance* 18, 251–266.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, & Y. Shin (1992) Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics* 54, 159–178.

LaMotte, L.R. & A. McWorther (1978) An exact test for the presence of random walk coefficients in a linear regression model. *Journal of the American Statistical Association* 73, 816–820.
Leybourne, S.J. & B.P.M. McCabe (1994) A consistent test for a unit root. *Journal of Business and Economic Statistics* 12, 157–166.
Nabeya, S. & K. Tanaka (1988) Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative. *Annals of Statistics* 16, 218–235.
Nyblom, J. (1986) Testing for deterministic linear trend in time series. *Journal of the American Statistical Association* 81, 545–549.
Nyblom, J. & A. Harvey (2000) Tests of common stochastic trends. *Econometric Theory* 16, 176–199.
Nyblom, J. & T. Mäkeläinen (1983) Comparisons of tests for the presence of random walk coefficients in a simple linear model. *Journal of the American Statistical Association* 78, 856–864.
Obstfeld, M. & K. Rogoff (1996) *Foundations of International Macroeconomics.* MIT Press.
Park, J.Y. (1992) Canonical cointegrating regressions. *Econometrica* 60, 119–143.
Phillips, P.C.B. (1988) Weak convergence of sample covariance matrices to stochastic integrals via martingale approximations. *Econometric Theory* 4, 528–533.
Phillips, P.C.B. (1991) Optimal inference in cointegrated systems. *Econometrica* 59, 283–306.
Phillips, P.C.B. & B.E. Hansen (1990) Statistical inference in instrumental variables regression with I(1) variables. *Review of Economic Studies* 57, 99–125.
Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.
Saikkonen, P. (1991) Asymptotically efficient estimation of cointegration regressions. *Econometric Theory* 7, 1–21.
Saikkonen, P. (1992) Estimation and testing of cointegrated systems by an autoregressive approximation. *Econometric Theory* 8, 1–27.
Saikkonen, P. & R. Luukkonen (1993a) Point optimal tests for testing the order of differencing in ARIMA models. *Econometric Theory* 9, 343–362.
Saikkonen, P. & R. Luukkonen (1993b) Testing for a moving average unit root in autoregressive integrated moving average models. *Journal of the American Statistical Association* 88, 596–601.
Shin, Y. (1994) A residual-based test of the null of cointegration against the alternative of no cointegration. *Econometric Theory* 10, 91–115.
Stock, J.H. (1994) Unit roots, structural breaks and trends. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2739–2841. North-Holland.
Stock, J.H. & M.W. Watson (1993) A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61, 783–820.
Tanaka, K. (1990) Testing for a moving average unit root. *Econometric Theory* 6, 433–444.
van Garderen, K.J. (2000) An alternative comparison of classical tests: Assessing the effects of curvature. In P. Marriott & M. Salmon (eds.), *Applications of Differential Geometry to Econometrics*, pp. 230–280. Cambridge University Press.
Zivot, E. (2000) The power of single equation tests for cointegration when the cointegrating vector is prespecified. *Econometric Theory* 16, 407–439.

APPENDIX

The proofs of Theorems 1–4 make use of Lemma 7, which shows how functional laws for sample moments of the transformed data $z_t(\bar{\theta})$ and $d_t(\bar{\theta})$ can be deduced from functional laws for z_t and d_t. Because these preliminary results might be of independent interest, they are presented in greater generality than needed for the proofs of Theorems 1–4.

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

In Lemma 7 and elsewhere in the Appendix, $\lfloor \cdot \rfloor$ denotes the integer part of the argument, and all functions are understood to be CADLAG functions defined on the unit interval (equipped with the Skorohod topology).

LEMMA 7. *Let* $\{F_{Tt}: 0 \leq t \leq T, T \geq 1\}$ *and* $\{(g_{Tt}', h_{Tt}')': 1 \leq t \leq T, T \geq 1\}$ *be triangular arrays of (vector) random variables with* $F_{T0} = 0$ *for all T. Let* $l > 0$ *be given and define* $F_{Tt}(l) = \Delta F_{Tt} + (1 - T^{-1}l)F_{T,t-1}(l)$, $g_{Tt}(l) = \Delta g_{Tt} + (1 - T^{-1}l)g_{T,t-1}(l)$, *and* $h_{Tt}(l) = \Delta h_{Tt} + (1 - T^{-1}l)h_{T,t-1}(l)$ *with initial conditions* $F_{T0}(l) = F_{T0}$, $g_{T1}(l) = g_{T1}$, *and* $h_{T1}(l) = h_{T1}$.

(a) *Suppose*

$$\begin{pmatrix} F_{T,\lfloor T\cdot \rfloor} \\ T^{-1}\sum_{t=1}^{\lfloor T\cdot \rfloor} g_{Tt} \end{pmatrix} \to_d \begin{pmatrix} F(\cdot) \\ G(\cdot) \end{pmatrix}, \tag{A.1}$$

where F and G are continuous. Then

$$\begin{pmatrix} F_{T,\lfloor T\cdot \rfloor}(l) \\ g_{T,\lfloor T\cdot \rfloor} - g_{T,\lfloor T\cdot \rfloor}(l) \\ T^{-1}\sum_{t=1}^{\lfloor T\cdot \rfloor} g_{Tt}(l) \end{pmatrix} \to_d \begin{pmatrix} F_l(\cdot) \\ lG_l(\cdot) \\ G_l(\cdot) \end{pmatrix} \tag{A.2}$$

jointly with (A.1), where $F_l(r) = F(r) - l\int_0^r \exp(-l(r-s))F(s)\,ds$ *and* $G_l(r) = G(r) - l\int_0^r \exp(-l(r-s))G(s)\,ds$.

(b) *Suppose*

$$\begin{pmatrix} T^{-1}\sum_{t=1}^{\lfloor T\cdot \rfloor} h_{Tt}' \\ T^{-1}\sum_{t=1}^{\lfloor T\cdot \rfloor} F_{Tt}h_{Tt}' \\ T^{-2}\sum_{t=2}^{\lfloor T\cdot \rfloor}\left(\sum_{i=1}^{t-1} g_{Ti}\right)h_{Tt}' \end{pmatrix} \to_d \begin{pmatrix} H(\cdot)' \\ \int_0^{\cdot} F(s)\,dH(s)' + \Gamma_{FH}(\cdot) \\ \int_0^{\cdot} G(s)\,dH(s)' + \Gamma_{GH}(\cdot) \end{pmatrix} \tag{A.3}$$

jointly with (A.1), where H, Γ_{FH}, *and* Γ_{GH} *are continuous and H is a semimartingale. Then*

$$\begin{pmatrix} T^{-1}\sum_{t=1}^{\lfloor T\cdot \rfloor} F_{Tt}h_{Tt}(l)' \\ T^{-1}\sum_{t=1}^{\lfloor T\cdot \rfloor} F_{Tt}(l)h_{Tt}' \\ T^{-1}\sum_{t=1}^{\lfloor T\cdot \rfloor} F_{Tt}(l)h_{Tt}(l)' \\ T^{-1}\sum_{t=1}^{\lfloor T\cdot \rfloor} (g_{Tt} - g_{Tt}(l))h_{Tt}' \\ T^{-1}\sum_{t=1}^{\lfloor T\cdot \rfloor} (g_{Tt} - g_{Tt}(l))h_{Tt}(l)' \end{pmatrix} \to_d \begin{pmatrix} \int_0^{\cdot} F(s)\,dH_l(s)' + \Gamma_{FH}(\cdot) \\ \int_0^{\cdot} F_l(s)\,dH(s)' + \Gamma_{FH}(\cdot) \\ \int_0^{\cdot} F_l(s)\,dH_l(s)' + \Gamma_{FH}(\cdot) \\ l\left(\int_0^{\cdot} G_l(s)\,dH(s)' + \Gamma_{GH}(\cdot)\right) \\ l\left(\int_0^{\cdot} G_l(s)\,dH_l(s)' + \Gamma_{GH}(\cdot)\right) \end{pmatrix} \tag{A.4}$$

jointly with (A.1)–(A.3), where $H_l(r) = H(r) - l\int_0^r \exp(-l(r-s))H(s)\,ds$.

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

Proof of Lemma 7. For $t = 0,\ldots,T$, $F_{Tt}(l)$ can expressed as

$$F_{Tt}(l) = F_{Tt} - lT^{-1}\sum_{i=1}^{t-1}(1 - T^{-1}l)^{t-1-i}F_{Ti}.$$

This relation can be restated as follows:

$$F_{T,\lfloor Tr\rfloor}(l) = F_{T,\lfloor Tr\rfloor} - l(1 - T^{-1}l)^{\lfloor Tr\rfloor - 1}\int_0^{\lfloor Tr\rfloor/T}(1 - T^{-1}l)^{-\lfloor Ts\rfloor}F_{T,\lfloor Ts\rfloor}\,ds, \qquad r \in [0,1].$$

Now, $\lim_{T\to\infty}\sup_{0\le r\le 1}|(1 - T^{-1}l)^{\lfloor Tr\rfloor} - \exp(-lr)| = 0$ and $F_{T,\lfloor T\cdot\rfloor} \to_d F(\cdot)$, where F is continuous, so

$$F_{T,\lfloor T\cdot\rfloor}(l) \to_d F(\cdot) - l\exp(-l\cdot)\int_0^{\cdot}\exp(ls)F(s)\,ds = F_l(\cdot)$$

by the continuous mapping theorem.

Next, using summation by parts,

$$g_{Tt} - g_{Tt}(l) = lG_{T,t-1}(l), \tag{A.5}$$

for $t = 1,\ldots,T$, where $G_{Tt} = T^{-1}\sum_{i=1}^{t} g_{Ti}$ and $G_{Tt}(l) = \Delta G_{Tt} + (1 - T^{-1}l)G_{T,t-1}(l)$ with initial conditions $G_{T0}(l) = G_{T0} = 0$. A second application of the proof of $F_{T,\lfloor T\cdot\rfloor}(l) \to_d F_l(\cdot)$ yields $G_{T,\lfloor T\cdot\rfloor}(l) \to_d G_l(\cdot)$. Moreover, using Billingsley (1999, Theorem 13.4), $\max_{1\le t\le T}\|G_{Tt}(l) - G_{T,t-1}(l)\| \to_d 0$, so

$$g_{T,\lfloor T\cdot\rfloor} - g_{T,\lfloor T\cdot\rfloor}(l) = lG_{T,\lfloor T\cdot\rfloor}(l) - l(G_{T,\lfloor T\cdot\rfloor}(l) - G_{T,\lfloor T\cdot\rfloor-1}(l)) \to_d lG_l(\cdot),$$

as claimed.

Finally, using $(G_{T,\lfloor T\cdot\rfloor}, g_{T,\lfloor T\cdot\rfloor} - g_{T,\lfloor T\cdot\rfloor}(l)) \to_d (G(\cdot), lG_l(\cdot))$, the continuous mapping theorem (CMT), and the relation $\int_0^r G_l(s)\,ds = G(r) - l\int_0^r \exp(-l(r-s))G(s)\,ds$,

$$T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} g_{Tt}(l) = G_{T,\lfloor T\cdot\rfloor} - T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor}(g_{Tt} - g_{Tt}(l)) \to_d G(\cdot) - l\int_0^{\cdot} G_l(s)\,ds = G_l(\cdot).$$

The proof of part (a) is completed by noting that the convergence results in the preceding displays hold jointly with (A.1).

Using the assumption on $T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} F_{Tt}h'_{Tt}$, part (a), and CMT,

$$\begin{aligned} T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} F_{Tt}h_{Tt}(l)' &= T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} F_{Tt}h'_{Tt} - T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} F_{Tt}(h_{Tt} - h_{Tt}(l))' \\ &\to_d \int_0^{\cdot} F(s)\,dH(s)' + \Gamma_{FH}(\cdot) - l\int_0^{\cdot} F(s)H_l(s)'\,ds \\ &= \int_0^{\cdot} F(s)\,dH_l(s)' + \Gamma_{FH}(\cdot). \end{aligned}$$

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

Next,

$$
\begin{aligned}
&T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} F_{Tt}(l)h_{Tt}' \\
&\quad = T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} F_{Tt}h_{Tt}' - l\left(\left(T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} F_{Tt}(l)\right)H_{T,\lfloor T\cdot\rfloor}' - T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} F_{Tt}(l)H_{Tt}'\right) \\
&\quad \to_d \int_0^{\cdot} F(s)\,dH(s)' + \Gamma_{FH}(\cdot) - l\left(\left(\int_0^{\cdot} F_l(s)\,ds\right)H(\cdot)' - \int_0^{\cdot} F_l(s)H(s)'\,ds\right) \\
&\quad = \int_0^{\cdot} F_l(s)\,dH(s)' + \Gamma_{FH}(\cdot),
\end{aligned}
$$

where the equalities follow from summation by parts and integration by parts, respectively.

This result, part (a), and CMT can be used to show that

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} F_{Tt}(l)h_{Tt}(l)' &= T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} F_{Tt}(l)h_{Tt}' - T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} F_{Tt}(l)(h_{Tt} - h_{Tt}(l))' \\
&\to_d \int_0^{\cdot} F_l(s)\,dH(s)' + \Gamma_{FH}(\cdot) - l\int_0^{\cdot} F_l(s)H_l(s)'\,ds \\
&= \int_0^{\cdot} F_l(s)\,dH_l(s)' + \Gamma_{FH}(\cdot).
\end{aligned}
$$

Similar reasoning yields

$$
\begin{aligned}
\begin{pmatrix} T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor}(g_{Tt} - g_{Tt}(l))h_{Tt}' \\ T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor}(g_{Tt} - g_{Tt}(l))h_{Tt}(l)' \end{pmatrix} &= l\begin{pmatrix} T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} G_{T,t-1}(l)h_{Tt}' \\ T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} G_{T,t-1}(l)h_{Tt}(l)' \end{pmatrix} \\
&\to_d l\begin{pmatrix} \int_0^{\cdot} G_l(s)\,dH(s)' + \Gamma_{GH}(\cdot) \\ \int_0^{\cdot} G_l(s)\,dH_l(s)' + \Gamma_{GH}(\cdot) \end{pmatrix}.
\end{aligned}
$$

The convergence results in the preceding displays hold jointly with (A.1)–(A.3). ∎

Proof of Theorems 1 and 2. The proof proceeds under the assumptions of Theorem 3, strengthening A1 only when necessary. Define Ω and Γ as in Section 3. Let

$$
\Psi_T = \begin{pmatrix} \omega_{yy}^{1/2} & 0' \\ 0 & \Omega_{xx}^{1/2} \end{pmatrix} \otimes \operatorname{diag}(T^{-1/2},\ldots,T^{-(p+1/2)}).
$$

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

Because $\lim_{T\to\infty}\max_{0\le i\le p}\sup_{0\le r\le 1}|T^{-i}\lfloor Tr\rfloor^i - r^i| = 0$ and $\Omega^{1/2} = \Omega_0^{1/2}\breve{\Omega}^{1/2}$, where

$$\Omega_0 = \begin{pmatrix}\omega_{yy} & 0' \\ 0 & \Omega_{xx}\end{pmatrix}, \qquad \breve{\Omega} = \begin{pmatrix}1 & \delta' \\ \delta & I_k\end{pmatrix}, \qquad \delta = \Omega_{xx}^{-1/2}\omega_{xy}\omega_{yy}^{-1/2},$$

it follows from Lemma 7 that

$$\lim_{T\to\infty}\sup_{0\le r\le 1}\|T^{1/2}\Psi_T d^{\dagger}_{\lfloor Tr\rfloor}(l) - \breve{D}_l(r)\| = 0, \tag{A.6}$$

where $d_t^{\dagger}(l) = d_t(1 - T^{-1}l)\cdot\Omega^{-1/2\prime}$,

$$\breve{D}_l(r) = \begin{pmatrix} D_l^y(r) & 0' \\ 0 & I_k \otimes D^x(r)\end{pmatrix}\breve{\Omega}^{-1/2\prime},$$

and $D_l^y(r)$ and $D^x(r)$ are defined as in the text.

Standard weak convergence results (e.g., Phillips and Solo, 1992; Phillips, 1988; Hansen, 1992) for linear processes can be used to show that the following hold jointly:

$$T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor} u_t \to_d \Omega_0^{1/2}\begin{pmatrix}V(\cdot) \\ \breve{W}(\cdot)\end{pmatrix}, \tag{A.7}$$

$$T^{-1}\sum_{t=2}^{\lfloor T\cdot\rfloor}\left(\sum_{s=1}^{t-1} u_s\right)u_t' \to_d \Omega_0^{1/2}\int_0^{\cdot}\begin{pmatrix}V(r) \\ \breve{W}(r)\end{pmatrix} d\begin{pmatrix}V(r) \\ \breve{W}(r)\end{pmatrix}'\Omega_0^{1/2\prime} + \Gamma'\int_0^{\cdot} dr, \tag{A.8}$$

where $(V,\breve{W}')'$ is a Brownian motion with covariance matrix $\breve{\Omega}$. By (A.7), Lemma 7, and the relation $v_t^y = T^{-1}\lambda\sum_{s=1}^{t-1}u_s^y + u_t^y$, simple algebra yields

$$T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor} v_t^{\dagger}(l) \to_d \breve{U}_l^{\lambda}(\cdot) = \breve{\Omega}^{-1/2}\begin{pmatrix}V_l^{\lambda}(\cdot) \\ \breve{W}(\cdot)\end{pmatrix}, \tag{A.9}$$

where $v_t^{\dagger}(l) = \Omega^{-1/2}(z_t(1 - T^{-1}l) - d_t(1 - T^{-1}l)'\beta)$ and V_l^{λ} is defined in terms of V as in the text. Similarly, using (A.7), (A.8), and Lemma 7, the following results can be verified:

$$\sum_{t=1}^{T}(v_t^{\dagger}(0) - v_t^{\dagger}(\bar{\lambda}))'(v_t^{\dagger}(0) - v_t^{\dagger}(\bar{\lambda})) \to_d \bar{\lambda}^2\rho_{\#}^2\int_0^1 V_{\bar{\lambda}}^{\lambda}(r)^2\,dr, \tag{A.10}$$

$$\sum_{t=1}^{T}(v_t^{\dagger}(0) - v_t^{\dagger}(\bar{\lambda}))'v_t^{\dagger}(0)$$

$$\to_d \bar{\lambda}\rho_{\#}^2\left(\int_0^1 V_{\bar{\lambda}}^{\lambda}(r)\,dV^{\lambda}(r) - \int_0^1 V_{\bar{\lambda}}^{\lambda}(r)\,d\breve{W}(r)'\delta + \omega_{yy}^{-1}\gamma_{yy.x}\right), \tag{A.11}$$

where $\rho_{\#} = (1 - \rho^2)^{-1/2}$, $\rho = (\omega_{yy}^{-1}\omega_{xy}'\Omega_{xx}^{-1}\omega_{xy})^{1/2}$, and $\gamma_{yy.x} = \gamma_{yy} - \omega_{xy}'\Omega_{xx}^{-1}\gamma_{xy}$.

The limiting distributions of $P_T(\bar{\theta};\Omega)$ and $L_T(\Omega)$ do not depend on k, the dimension of x_t. The remainder of the proof proceeds under the assumption that $k=1$ and $\delta = \|\delta\| = \rho$, because these assumptions simplify the algebra without leading to a loss of generality. When $k=1$ and $\delta=\rho$, the processes $\breve{D}_l$, $\breve{U}_l$ and $\breve{W}$ coincide with the processes D_l, U_l and W defined in the text (with $R=\rho$). Now,

$$P_T(\bar{\theta};\Omega) = \sum_{t=1}^{T} \tilde{v}_t^{\dagger}(0)'\tilde{v}_t^{\dagger}(0) - \sum_{t=1}^{T} \tilde{v}_t^{\dagger}(\bar{\lambda})'\tilde{v}_t^{\dagger}(\bar{\lambda}),$$

where $\tilde{v}_t^{\dagger}(l) = v_t^{\dagger}(l) - d_t^{\dagger}(l)'(\sum_{s=1}^{T} d_s^{\dagger}(l)\, d_s^{\dagger}(l)')^{-1}(\sum_{s=1}^{T} d_s^{\dagger}(l)v_s^{\dagger}(l))$. By the algebra of OLS, (A.6), and (A.9),

$$\begin{aligned}
&\sum_{t=1}^{T} \tilde{v}_t^{\dagger}(l)'\tilde{v}_t^{\dagger}(l) - \sum_{t=1}^{T} v_t^{\dagger}(l)'v_t^{\dagger}(l)\\
&\quad = -\left(\Psi_T \sum_{t=1}^{T} d_t^{\dagger}(l)v_t^{\dagger}(l)\right)'\left(\Psi_T \sum_{t=1}^{T} d_t^{\dagger}(l)\, d_t^{\dagger}(l)'\,\Psi_T'\right)^{-1}\left(\Psi_T \sum_{t=1}^{T} d_t^{\dagger}(l)v_t^{\dagger}(l)\right)\\
&\quad \to_d -\left(\int_0^1 D_l(r)\, dU_l^{\lambda}(r)\right)'\left(\int_0^1 D_l(r)D_l(r)'\, dr\right)^{-1}\left(\int_0^1 D_l(r)\, dU_l^{\lambda}(r)\right)
\end{aligned}$$

for $l \in \{0,\bar{\lambda}\}$. Using this along with (A.10) and (A.11) and the relation

$$\begin{aligned}
&\sum_{t=1}^{T} v_t^{\dagger}(0)'v_t^{\dagger}(0) - \sum_{t=1}^{T} v_t^{\dagger}(\bar{\lambda})'v_t^{\dagger}(\bar{\lambda})\\
&\quad = -\sum_{t=1}^{T} (v_t^{\dagger}(0) - v_t^{\dagger}(\bar{\lambda}))'(v_t^{\dagger}(0) - v_t^{\dagger}(\bar{\lambda})) + 2\sum_{t=1}^{T} (v_t^{\dagger}(0) - v_t^{\dagger}(\bar{\lambda}))'v_t^{\dagger}(0),
\end{aligned}$$

it follows that

$$P_T(\bar{\theta};\Omega) \to_d \varphi_P(\lambda;\bar{\lambda},\rho^2) + 2\bar{\lambda}\omega_{yy.x}^{-1}\gamma_{yy.x}.$$

Because $\gamma_{yy.x}=0$ and $\Sigma=\Omega$ under the assumptions of Theorem 1, the proof of that theorem is now complete.

Next, $L_T(\Omega)$ can be written as $L_T^{*}(\Omega) + L_T^{**}(\Omega)$, where

$$L_T^{*}(\Omega) = \sum_{t=1}^{T} \widetilde{V}_t^{\dagger\prime}\breve{\Omega}^{*}\widetilde{V}_t^{\dagger},$$

$$L_T^{**}(\Omega) = \left(\sum_{t=1}^{T} d_t^{\dagger}\breve{\Omega}^{**}\widetilde{V}_t^{\dagger}\right)'\left(\sum_{t=1}^{T} d_t^{\dagger}d_t^{\dagger\prime}\right)^{-1}\left(\sum_{t=1}^{T} d_t^{\dagger}\breve{\Omega}^{**}\widetilde{V}_t^{\dagger}\right),$$

$\widetilde{V}_t^{\dagger} = T^{-1}\sum_{s=1}^{t-1}\tilde{v}_s^{\dagger}(0)$, $d_t^{\dagger} = d_t^{\dagger}(0)$, $\breve{\Omega}^{*} = \Omega^{1/2\prime}\Omega^{*}\Omega^{1/2}$, and $\breve{\Omega}^{**} = \Omega^{1/2\prime}\Omega^{**}\Omega^{1/2}$. When $k=1$, $\breve{\Omega}^{*}$ and $\breve{\Omega}^{**}$ coincide with $\bar{\Sigma}^{*}$ and $\bar{\Sigma}^{**}$ defined in the text.

The result $L_T(\Omega) \to_d \varphi_L(\lambda;\rho^2)$ now follows from simple algebra and the fact that $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor};\widetilde{V}_t^{\dagger} \to_d \widetilde{U}^{\lambda}(\cdot)$ under the assumptions of Theorem 3, where $\widetilde{U}^{\lambda}$ is defined as in the text (with $R=\rho$). In particular, Theorem 2(a) follows because $\Sigma=\Omega$ under the assumptions of Theorem 2.

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

Under the assumptions of Theorem 2, integrals such as

$$\int \phi_T(m) f_T(m|1 - T^{-1}\lambda, \Sigma)\, dm$$

can be differentiated with respect to λ by differentiating under the integral sign. As a consequence,

$$\begin{aligned}\left| \frac{\partial}{\partial\lambda} \int \phi_T(m) f_T(m|1 - T^{-1}\lambda, \Sigma)\, dm \Big|_{\lambda=0} \right| &= \left| \int \phi_T(m) l^{(1)}(m|\Sigma) f_T(m|1, \Sigma)\, dm \right| \\ &\leq \int |l^{(1)}(m|\Sigma)|\, f_T(m|1, \Sigma)\, dm \\ &\leq \left(\int l^{(1)}(m|\Sigma)^2 f_T(m|1, \Sigma)\, dm \right)^{1/2} \\ &= (\mathrm{Var}_0(\dot{P}_T))^{1/2},\end{aligned}$$

where $\mathrm{Var}_0(\cdot)$ denotes the variance under H_0. The first inequality uses $|\phi_T| \leq 1$ and the modulus inequality for integrals, the second inequality uses the Cauchy–Schwarz inequality, and the last equality uses $\int l^{(1)}(m|\Sigma) f_T(m|1, \Sigma)\, dm = 0$ and the fact that $l^{(1)}(m_T|\Sigma)$ differs from $\dot{P}_T$ by an additive constant. Using the fact that u_t is Gaussian white noise, it is easy to show that $\lim_{T\to\infty} \mathrm{Var}_0(\dot{P}_T) = 0$. Therefore, the $\underline{\lim}_{T\to\infty}$ of the left-hand side of the preceding display is zero, as claimed in Theorem 2(b).

For any T, let $\tilde{\phi}_T^L(m_T; \alpha, \Sigma) = 1(L_T > \tilde{c}_T^L(\alpha, \Sigma))$, where $\tilde{c}_T^L(\alpha, \Sigma)$ is such that

$$\int \tilde{\phi}_T^L(m_T; \alpha, \Sigma) f_T(m|1, \Sigma)\, dm = \int \phi_T(m) f_T(m|1, \Sigma)\, dm.$$

By the Neyman–Pearson lemma and the fact that $l^{(2)}(m_T|\Sigma) - 2T^{-1} l^{(1)}(m_T|\Sigma)$ differs from $2L_T$ by an additive constant,

$$\int (\phi_T(m) - \tilde{\phi}_T^L(m; \alpha, \Sigma))(l^{(2)}(m|\Sigma) - 2T^{-1} l^{(1)}(m|\Sigma)) f_T(m|1, \Sigma)\, dm \leq 0.$$

Moreover, for any sequence $\{\eta_T\}$ of bounded functions,

$$\begin{aligned}&\frac{\partial^2}{\partial\lambda^2} \int \eta_T(m) f_T(m|1 - T^{-1}\lambda, \Sigma)\, dm \Big|_{\lambda=0} \\ &\quad = \int \eta_T(m)(l^{(2)}(m|\Sigma) + l^{(1)}(m|\Sigma)^2) f_T(m|1, \Sigma)\, dm \\ &\quad = \int \eta_T(m)(l^{(2)}(m|\Sigma) - 2T^{-1} l^{(1)}(m|\Sigma)) f_T(m|1, \Sigma)\, dm + o(1),\end{aligned}$$

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

where the second equality uses $\int l^{(1)}(m|\Sigma)^2 f_T(m|1,\Sigma)\,dm = o(1)$. Combining the preceding displays, it follows that

$$\overline{\lim_{T\to\infty}} \frac{\partial^2}{\partial\lambda^2} \int (\phi_T(m) - \tilde{\phi}_T^L(m;\alpha,\Sigma)) f_T(m|1-T^{-1}\lambda,\Sigma)\,dm \Big|_{\lambda=0} \le 0.$$

The proof of 2(c) can be completed by showing that

$$\overline{\lim_{T\to\infty}} \frac{\partial^2}{\partial\lambda^2} \int (\tilde{\phi}_T^L(m;\alpha,\Sigma) - \phi_T^L(m;\alpha,\Sigma)) f_T(m|1-T^{-1}\lambda,\Sigma)\,dm \Big|_{\lambda=0} \le 0,$$

which, because $\{\tilde{\phi}_T^L(\cdot;\alpha,\Sigma) - \phi_T^L(\cdot;\alpha,\Sigma)\}$ is bounded, holds if

$$\overline{\lim_{T\to\infty}} E_0((\tilde{\phi}_T^L(m_T;\alpha,\Sigma) - \phi_T^L(m_T;\alpha,\Sigma))(l^{(2)}(m_T|\Sigma) - 2T^{-1}l^{(1)}(m_T|\Sigma))) \le 0,$$

where $E_0(\cdot)$ denotes expectation under H_0. Now, using $E_0(l^{(1)}(m_T|\Sigma)) = 0$ and $E_0(l^{(2)}(m_T|\Sigma)) = -\text{Var}_0(\dot{P}_T)$ and the fact that $l^{(2)}(m_T|\Sigma) - 2T^{-1}l^{(1)}(m_T|\Sigma)$ differs from $2L_T$ by an additive constant,

$$l^{(2)}(m_T|\Sigma) - 2T^{-1}l^{(1)}(m_T|\Sigma) = 2L_T^{\mu} - \text{Var}_0(\dot{P}_T),$$

where $L_T^{\mu} = L_T - E_0(L_T)$. Using this relation and $\lim_{T\to\infty}\text{Var}_0(\dot{P}_T) = 0$,

$$\overline{\lim_{T\to\infty}} E_0((\tilde{\phi}_T^L(m_T;\alpha,\Sigma) - \phi_T^L(m_T;\alpha,\Sigma))(l^{(2)}(m_T|\Sigma) - 2T^{-1}l^{(1)}(m_T|\Sigma)))$$
$$= 2\,\overline{\lim_{T\to\infty}} E_0((\tilde{\phi}_T^L(m_T;\alpha,\Sigma) - \phi_T^L(m_T;\alpha,\Sigma))L_T^{\mu}).$$

Because $\{\phi_T\}$ is asymptotically of level α, it can be shown (using Theorem 2(a)) that $\lim_{T\to\infty}\tilde{c}_T^L(\alpha,\Sigma) = c^L(\alpha,\rho^2)$. Therefore, $\tilde{\phi}_T^L(m_T;\alpha,\Sigma) - \phi_T^L(m_T;\alpha,\Sigma) \to_p 0$. Moreover, $\{L_T^{\mu}\}$ is uniformly integrable under H_0, so

$$\lim_{T\to\infty} E_0((\tilde{\phi}_T^L(m_T;\alpha,\Sigma) - \phi_T^L(m_T;\alpha,\Sigma))L_T^{\mu}) = 0,$$

as was to be shown. ■

Proof of Theorem 3. The proof of Theorems 1 and 2(a) carries over to the case where Ω and Γ are replaced with consistent estimators if the following analogues of equations (A.6) and (A.9)–(A.11) can be established:

$$\sup_{0\le r\le 1} \|T^{1/2}\Psi_T \hat{d}^{\dagger}_{\lfloor Tr\rfloor}(l) - \breve{D}_l(r)\| \to_p 0, \quad \textbf{(A.12)}$$

$$T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor} \hat{v}_t^{\dagger}(l) \to_d \breve{U}_l^{\lambda}(\cdot) = \breve{\Omega}^{-1/2}\begin{pmatrix} V_l^{\lambda}(\cdot) \\ \breve{W}(\cdot)\end{pmatrix}, \quad \textbf{(A.13)}$$

$$\sum_{t=1}^{T} (\hat{v}_t^{\dagger}(0) - \hat{v}_t^{\dagger}(\bar{\lambda}))'(\hat{v}_t^{\dagger}(0) - \hat{v}_t^{\dagger}(\bar{\lambda})) \to_d \bar{\lambda}^2\rho_{\#}^2 \int_0^1 V_{\bar{\lambda}}^{\lambda}(r)^2\,dr, \quad \textbf{(A.14)}$$

$$\sum_{t=1}^{T} (\hat{v}_t^{\dagger}(0) - \hat{v}_t^{\dagger}(\bar{\lambda}))'\hat{v}_t^{\dagger}(0)$$
$$\to_d \bar{\lambda}\rho_{\#}^2\left(\int_0^1 V_{\bar{\lambda}}^{\lambda}(r)\,dV^{\lambda}(r) - \int_0^1 V_{\bar{\lambda}}^{\lambda}(r)\,d\breve{W}(r)'\delta + \omega_{yy}^{-1}\gamma_{yy.x}\right), \quad \textbf{(A.15)}$$

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

where $\hat{d}_t^{\dagger}(l) = d_t(1 - T^{-1}l) \cdot \hat{\Omega}^{-1/2\prime}$ and $\hat{v}_t^{\dagger}(l) = \hat{\Omega}^{-1/2}(z_t(1 - T^{-1}l) - d_t(1 - T^{-1}l)'\beta)$.

Now,

$$
\begin{aligned}
&\sup_{0\le r\le 1} \|T^{1/2}\Psi_T \hat{d}^{\dagger}_{\lfloor Tr\rfloor}(l) - \check{D}_l(r)\| \\
&\quad \le \sup_{0\le r\le 1} (\|T^{1/2}\Psi_T(\hat{d}^{\dagger}_{\lfloor Tr\rfloor}(l) - d^{\dagger}_{\lfloor Tr\rfloor}(l))\| + \|T^{1/2}\Psi_T d^{\dagger}_{\lfloor Tr\rfloor}(l) - \check{D}_l(r)\|) \\
&\quad = \sup_{0\le r\le 1} \|T^{1/2}\Psi_T d^{\dagger}_{\lfloor Tr\rfloor}(l)(\Omega^{1/2\prime}\hat{\Omega}^{-1/2\prime} - I_{k+1})\| + o(1) \\
&\quad \le \sup_{0\le r\le 1} \|T^{1/2}\Psi_T d^{\dagger}_{\lfloor Tr\rfloor}(l)\| \cdot \|\Omega^{1/2\prime}\hat{\Omega}^{-1/2\prime} - I_{k+1}\| + o(1) \\
&\quad = o_p(1),
\end{aligned}
$$

where the first inequality uses the triangle inequality, the first equality uses the relation $\hat{d}_t^{\dagger}(l) = d_t^{\dagger}(l)\Omega^{1/2\prime}\hat{\Omega}^{-1/2\prime}$ and (A.6), the second inequality uses the properties of $\|\cdot\|$, and the last equality uses (A.6) and the assumption $\hat{\Omega} \rightarrow_p \Omega$.

Similar reasoning establishes (A.13)–(A.15). ■

Proof of Theorem 4. By the properties of seemingly unrelated regressions, $\tilde{v}_t(1;\hat{\Omega})$ does not depend on $\hat{\Omega}$:

$$
\begin{pmatrix} \tilde{v}_t^y(1;\hat{\Omega}) \\ \tilde{v}_t^x(1;\hat{\Omega}) \end{pmatrix} = \begin{pmatrix} \tilde{v}_t^y(1) \\ \tilde{v}_t^x(1) \end{pmatrix} = \begin{pmatrix} v_t^y \\ u_t^x \end{pmatrix} - \begin{pmatrix} \left(\sum_{s=1}^T d_s^y v_s^y\right)' \left(\sum_{s=1}^T d_s^y d_s^{y\prime}\right)^{-1} d_t^y \\ \left(\sum_{s=1}^T d_s^x u_s^{x\prime}\right)' \left(\sum_{s=1}^T d_s^x d_s^{x\prime}\right)^{-1} d_t^x \end{pmatrix}
$$

because $d_t^y(1) = d_t^y = d_t^x$. Partition $\widetilde{V}_t = T^{-1}\sum_{s=1}^{t-1} \tilde{v}_t(1;\hat{\Omega})$ after the first row as $(\widetilde{V}_t^y, \widetilde{V}_t^{x\prime})'$.

Under the assumptions of Theorem 4, it follows from standard results for linear processes that

$$
T^{-1}\sum_{t=1}^T \widetilde{V}_t^x \widetilde{V}_t^y = O_p(1) \tag{A.16}
$$

and

$$
T^{-1/2}\tilde{v}^y_{\lfloor T\cdot\rfloor} \rightarrow_d (1-\theta)\omega_{yy}^{1/2} W^d(\cdot), \tag{A.17}
$$

where

$$
W^d(r) = W(r) - \left(\int_0^1 D^y(s)W(s)\,ds\right)' \left(\int_0^1 D^y(s)D^y(s)'\,ds\right)^{-1} D^y(r),
$$

W is a Wiener process, and D^y is defined as in the text. By (A.17) and CMT,

$$
T^{-2}\sum_{t=1}^T (\widetilde{V}_t^y)^2 \rightarrow_d (1-\theta)^2\omega_{yy}\int_0^1 \overline{W}^d(r)^2\,dr, \tag{A.18}
$$

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

where $\overline{W}^d(r) = \int_0^r W^d(s)\,ds$.

For any $c \in \mathbb{R}$,

$$\begin{aligned}
&\underline{\lim}_{T\to\infty} \Pr(\hat{L}_T > c) \\
&\quad \geq \underline{\lim}_{T\to\infty} \Pr\left(\sum_{t=1}^T \widetilde{V}_t' \hat{\Omega}^* \widetilde{V}_t > c\right) \\
&\quad = \underline{\lim}_{T\to\infty} \Pr\left(T^{-2}\sum_{t=1}^T (\widetilde{V}_t^y)^2 - 2(T^{-1}\hat{\omega}_{xy}')\hat{\Omega}_{xx}^{-1}\left(T^{-1}\sum_{t=1}^T \widetilde{V}_t^x \widetilde{V}_t^y\right) - cT^{-2}\hat{\omega}_{yy.x} > 0\right) \\
&\quad \geq \Pr\left((1-\theta)^2\omega_{yy}\int_0^1 \overline{W}^d(r)^2\,dr > 0\right) \\
&\quad = 1,
\end{aligned}$$

where

$$\hat{\Omega}^* = \begin{pmatrix} \hat{\omega}_{yy.x}^{-1} & -\hat{\omega}_{yy.x}^{-1}\hat{\omega}_{xy}'\hat{\Omega}_{xx}^{-1} \\ -\hat{\omega}_{yy.x}^{-1}\hat{\Omega}_{xx}^{-1}\hat{\omega}_{xy} & 0 \end{pmatrix}$$

and the first inequality uses the fact that $(\sum_{t=1}^T d_t \hat{\Omega}^{-1} d_t')$ is positive definite, whereas the second inequality uses $\hat{\omega}_{yy.x} \leq \hat{\omega}_{yy} = o_p(T^2)$, $\hat{\Omega}_{xx}^{-1} = O_p(1)$, $\hat{\omega}_{xy} = o_p(T)$, (A.16), (A.18), and the portmanteau theorem (e.g., Billingsley, 1999).

Next, consider

$$\hat{Q}_T(\bar{\theta}) = \sum_{t=1}^T \tilde{v}_t(1;\hat{\Omega})'\hat{\Omega}^{-1}\tilde{v}_t(1;\hat{\Omega}) - \sum_{t=1}^T \tilde{v}_t(\bar{\theta};\hat{\Omega})'\hat{\Omega}^{-1}\tilde{v}_t(\bar{\theta};\hat{\Omega}).$$

Now,

$$\begin{aligned}
\sum_{t=1}^T \tilde{v}_t(\bar{\theta};\hat{\Omega})'\hat{\Omega}^{-1}\tilde{v}_t(\bar{\theta};\hat{\Omega}) &= \min_b\left(\sum_{t=1}^T (z_t(\bar{\theta}) - d_t(\bar{\theta})'b)'\hat{\Omega}^{-1}(z_t(\bar{\theta}) - d_t(\bar{\theta})'b)\right) \\
&\leq \sum_{t=1}^T \check{v}_t(\bar{\theta})'\hat{\Omega}^{-1}\check{v}_t(\bar{\theta}),
\end{aligned}$$

where $\check{v}_t(\bar{\theta}) = z_t(\bar{\theta}) - d_t(\bar{\theta})'(\sum_{s=1}^T d_s d_s')^{-1}(\sum_{s=1}^T d_s z_s)$. Partition $\check{v}_t(\bar{\theta})$ after the first row as $(\check{v}_t^y(\bar{\theta}), \tilde{v}_t^x(1)')'$. The series $\check{v}_t^y(\bar{\theta})$ satisfies the difference equation $\check{v}_t^y(\bar{\theta}) = \Delta\tilde{v}_t^y(1) + \bar{\theta}\check{v}_{t-1}^y(\bar{\theta})$ with initial condition $\check{v}_1^y(\bar{\theta}) = \tilde{v}_1^y(1)$. As a consequence,

$$\begin{aligned}
T^{-2}\hat{\omega}_{yy.x}\hat{Q}_T(\bar{\theta}) &\geq T^{-2}\hat{\omega}_{yy.x}\left(\sum_{t=1}^T \tilde{v}_t(1)'\hat{\Omega}^{-1}\tilde{v}_t(1) - \sum_{t=1}^T \check{v}_t(\bar{\theta})'\hat{\Omega}^{-1}\check{v}_t(\bar{\theta})\right) \\
&= T^{-2}\left(\sum_{t=1}^T (\tilde{v}_t^y(1) - \hat{\omega}_{xy}'\hat{\Omega}_{xx}^{-1}\tilde{v}_t^x(1))^2 - \sum_{t=1}^T (\check{v}_t^y(\bar{\theta}) - \hat{\omega}_{xy}'\hat{\Omega}_{xx}^{-1}\tilde{v}_t^x(1))^2\right) \\
&= T^{-2}\left(\sum_{t=1}^T \tilde{v}_t^y(1)^2 - \sum_{t=1}^T \check{v}_t^y(\bar{\theta})^2 - 2\hat{\omega}_{xy}'\hat{\Omega}_{xx}^{-1}\sum_{t=1}^T \tilde{v}_t^x(1)(\tilde{v}_t^y(1) - \check{v}_t^y(\bar{\theta}))\right) \\
&= T^{-2}\sum_{t=1}^T \tilde{v}_t^y(1)^2 - T^{-2}\sum_{t=1}^T \check{v}_t^y(\bar{\theta})^2 + o_p(1) \\
&\to_d (1-\theta)^2\omega_{yy}\left(\int_0^1 W^d(r)^2\,dr - \int_0^1 W_\lambda^d(r)^2\,dr\right),
\end{aligned}$$

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

where

$$W_{\bar{\lambda}}^{d}(r) = W^{d}(r) - \bar{\lambda}\int_0^r \exp(-\bar{\lambda}(r-s))W^{d}(s)\, ds$$

and the last equality uses $\hat{\omega}_{xy} = o_p(T)$ and $\sum_{t=1}^{T} \tilde{v}_t^x(1)(\tilde{v}_t^y(1) - \check{v}_t^y(\bar{\theta})) = O_p(T)$, whereas the convergence result follows from (A.17), Lemma 7, and CMT.

Now,

$$\int_0^1 W^{d}(r)^2\, dr - \int_0^1 W_{\bar{\lambda}}^{d}(r)^2\, dr = \int_0^1\int_0^1 K_{\bar{\lambda}}(r,s)W^{d}(r)W^{d}(s)\, drds,$$

where

$$K_{\bar{\lambda}}(r,s) = \frac{\bar{\lambda}}{2}(\exp(-\bar{\lambda}(2-r-s)) + \exp(-\bar{\lambda}|r-s|)).$$

By the portmanteau theorem and the fact that the function $K_{\bar{\lambda}}(\cdot,\cdot)$ is positive definite in the sense that $\int_0^1\int_0^1 K_{\bar{\lambda}}(r,s)f(r)f(s)\, drds > 0$ for any nonzero, continuous function $f(\cdot)$,

$$\begin{aligned}\underline{\lim}_{T\to\infty} \Pr(\hat{Q}_T(\bar{\theta}) > c) &= \underline{\lim}_{T\to\infty} \Pr(T^{-2}\hat{\omega}_{yy.x}\hat{Q}_T(\bar{\theta}) - cT^{-2}\hat{\omega}_{yy.x} > 0)\\ &\geq \Pr\left((1-\theta)^2\omega_{yy}\left(\int_0^1 W^{d}(r)^2\, dr - \int_0^1 W_{\bar{\lambda}}^{d}(r)^2\, dsr\right) > 0\right)\\ &= 1\end{aligned}$$

for any $c \in \mathbb{R}$. ■

Proof of Lemma 5. Let $u_t^{PW} = u_t - Au_{t-1}$, where A is the matrix appearing in A2(iii). The equations defining $\hat{\Gamma}$ and $\hat{\Omega}$ are sample counterparts of the relations

$$\Gamma = (I-A)^{-1}\Gamma^{PW}(I-A')^{-1} + (I-A)^{-1}A\Sigma - (I-A)^{-1}\Lambda A'(I-A')^{-1}$$

and

$$\Omega = (I-A)^{-1}\Omega^{PW}(I-A')^{-1},$$

where

$$\Gamma^{PW} = \lim_{T\to\infty}(T-1)^{-1}\sum_{t=3}^{T}\sum_{s=2}^{t-1} E(u_t^{PW}u_s^{PW\prime}), \qquad \Sigma = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E(u_t u_t'),$$

$$\Lambda = \lim_{T\to\infty} T^{-1}\sum_{t=2}^{T} E(u_t^{PW}u_{t-1}'), \qquad \Omega^{PW} = \lim_{T\to\infty} T^{-1}\sum_{t=2}^{T}\sum_{s=2}^{T} E(u_t^{PW}u_s^{PW\prime}).$$

Because $(I-\hat{A})^{-1} \to_p (I-A)^{-1}$ and $\hat{A} \to_p A$ under A2(iii), it therefore suffices to show that $\hat{\Gamma}^{PW} \to_p \Gamma^{PW}$, $\hat{\Sigma} \to_p \Sigma$, $\hat{\Lambda} \to_p \Lambda$, and $\hat{\Omega}^{PW} \to_p \Omega^{PW}$.

Let $\hat{v}_t^{\dagger} = u_t - d_t'(\hat{\beta} - \beta)$, where $\hat{\beta} = (\sum_{t=1}^{T} d_t d_t')^{-1}(\sum_{t=1}^{T} d_t z_t)$. Let $\hat{v}_t^{\dagger\dagger} = \hat{v}_t - \hat{v}_t^{\dagger}$, $\hat{v}_t^{PW,\dagger} = \hat{v}_t^{\dagger} - \hat{A}\hat{v}_{t-1}^{\dagger}$, and $\hat{v}_t^{PW,\dagger\dagger} = \hat{v}_t^{PW} - \hat{v}_t^{PW,\dagger} = \hat{v}_t^{\dagger\dagger} - \hat{A}\hat{v}_{t-1}^{\dagger\dagger}$. Using notation typified by

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press

$$\hat{\Gamma}^{PW,\dagger,\dagger\dagger} = (T-1)^{-1} \sum_{t=3}^{T} \sum_{s=2}^{t-1} k\left(\frac{|t-s|}{\hat{b}_T}\right) \hat{v}_t^{PW,\dagger} \hat{v}_s^{PW,\dagger\dagger\prime},$$

$\hat{\Gamma}^{PW}$ can be written as $\hat{\Gamma}^{PW,\dagger,\dagger} + \hat{\Gamma}^{PW,\dagger\dagger,\dagger} + \hat{\Gamma}^{PW,\dagger,\dagger\dagger} + \hat{\Gamma}^{PW,\dagger\dagger,\dagger\dagger}$. Now, $\hat{\Gamma}^{PW,\dagger,\dagger} \to_p \Gamma^{PW}$ by Corollary 4 of Jansson (2002). The proof of $\hat{\Gamma}^{PW} \to_p \Gamma^{PW}$ is completed by using the relation

$$\hat{v}_t^{\dagger\dagger} = v_t - u_t = T^{-1} \begin{pmatrix} \lambda \sum_{s=1}^{t-1} u_s^y \\ 0 \end{pmatrix}$$

and straightforward, but tedious, bounding arguments to show that $\hat{\Gamma}^{PW,\dagger\dagger,\dagger}$, $\hat{\Gamma}^{PW,\dagger,\dagger\dagger}$, and $\hat{\Gamma}^{PW,\dagger\dagger,\dagger\dagger}$ are $o_p(1)$. Indeed, the proof of Lemma 5 of Jansson and Haldrup (2002) carries over to the present case. The details are omitted for brevity.

Proceeding in analogous fashion, it can be shown that $\hat{\Sigma} \to_p \Sigma$, $\hat{\Lambda} \to_p \Lambda$, and $\hat{\Omega}^{PW} \to_p \Omega^{PW}$. ■

Proof of Lemma 6. In view of A2(iii) and (iv), it suffices to show that $\hat{\omega}_{yy}^{PW} = o_p(T^2)$, $\hat{\gamma}_{yy.x}^{PW} = o_p(T^2)$, $\hat{\omega}_{xy}^{PW} = o_p(T)$, and $(\hat{\Omega}_{xx}^{PW})^{-1} = O_p(1)$, where $\hat{\omega}_{yy}^{PW}$, $\hat{\gamma}_{yy.x}^{PW}$, $\hat{\omega}_{xy}^{PW}$, and $\hat{\Omega}_{xx}^{PW}$ are defined in the obvious way. Now, $(\hat{\Omega}_{xx}^{PW})^{-1} = O_p(1)$ because $\hat{\Omega}_{xx}^{PW} \to_p \Omega_{xx}^{PW}$. Moreover,

$$\begin{aligned} |\hat{\omega}_{yy}^{PW}| &= \left| \sum_{i=-(T-2)}^{T-2} k\left(\frac{|i|}{\hat{b}_T}\right) \left((T-1)^{-1} \sum_{t=2}^{T-|i|} \hat{v}_{t+|i|}^{y,PW} \hat{v}_t^{y,PW} \right) \right| \\ &\le \frac{1}{T-1} \sum_{i=-(T-2)}^{T-2} \left| k\left(\frac{|i|}{\hat{b}_T}\right) \right| \left| \sum_{t=2}^{T-|i|} \hat{v}_{t+|i|}^{y,PW} \hat{v}_t^{y,PW} \right| \\ &\le \frac{1}{T-1} \sum_{i=-(T-2)}^{T-2} \left| k\left(\frac{|i|}{\hat{b}_T}\right) \right| \left(\sum_{t=2}^{T-|i|} (\hat{v}_{t+|i|}^{y,PW})^2 \right)^{1/2} \left(\sum_{t=2}^{T-|i|} (\hat{v}_t^{y,PW})^2 \right)^{1/2} \\ &\le \frac{1}{T-1} \left(\sum_{t=2}^{T} (\hat{v}_t^{y,PW})^2 \right) \left(\sum_{i=-(T-2)}^{T-2} \left| k\left(\frac{|i|}{\hat{b}_T}\right) \right| \right) \\ &= o_p(T^{3/2}), \end{aligned}$$

where the second inequality uses the Cauchy–Schwarz inequality and the last equality uses $\sum_{i=0}^{T-2} |k(i/\hat{b}_T)| = o_p(T^{1/2})$ (Jansson, 2002) and $\sum_{t=2}^{T} (\hat{v}_t^{y,PW})^2 = O_p(T^2)$.

Similar reasoning can be used to show that $\hat{\gamma}_{yy.x}^{PW} = o_p(T^2)$ and $\hat{\omega}_{xy}^{PW} = o_p(T)$. ■

https://doi.org/10.1017/S0266466604201037 Published online by Cambridge University Press