Stationarity Testing with Covariates

MICHAEL JANSSON*
DEPARTMENT OF ECONOMICS, UC BERKELEY

June 14, 2001
Preliminary draft

ABSTRACT. Two new stationarity tests are proposed. Both tests enjoy optimality properties. The tests can be viewed as generalizations of existing stationary 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.

KEYWORDS: Cointegration, local power function, power envelope, purchasing power parity, stationarity tests.

1. INTRODUCTION

Let $\{y_{Tt} : 1 \leq t \leq T, T \geq 1\}$ be an observed univariate triangular array generated by

$$y_{Tt} = \mu_t^y + v_{Tt}^y, \tag{1}$$

where $\{\mu_t^y\}$ is deterministic component and $\{v_{Tt}^y\}$ is an unobserved error process with initial condition $v_{T1}^y = u_1^y$ and generating mechanism

$$\Delta v_{Tt}^y = (1 - \vartheta_T L)\, u_t^y, \qquad 2 \leq t \leq T. \tag{2}$$

In (2), $\{\vartheta_T\}$ is a sequence of fixed (scalar) parameters and $\{u_t^y\}$ is a stationary I(0) process. The specification involves triangular arrays $\{y_{Tt}\}$ rather than sequences $\{y_t\}$ in anticipation of the need to employ a parameterization in which $\{\vartheta_T\}$ is modeled as a sequence of parameters.

The problem of testing the null hypothesis $H_0 : \vartheta_T = 1$ against the alternative $|\vartheta_T| < 1$ 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

*Email: mjansson@econ.berkeley.edu.
A MATLAB program that implements the tests proposed in this paper is available from the author upon request.

and Mäkeläinen (1983), Nyblom (1986), Nabeya and Tanaka (1988), Kwiatkowski, Phillips, Schmidt, and Shin (1992), Saikkonen and Luukkonen (1993a, 1993b) and Stock and Watson (1998) (for a review, see Stock (1994)). Under H_0, $v^y_{Tt} = u^y_t$ and $\{y_{Tt}\}$ is a (trend-)stationary process, while $\{y_{Tt}\}$ 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 with respect to the assumptions on the underlying stationary process $\{u^y_t\}$ and the form of the deterministic component $\{\mu^y_t\}$. On the other hand, all previous studies (of which the author is aware) have been concerned with the situation where $\{y_{Tt}\}$ is observed in isolation. Specifically, all previously devised tests have exploited only the information contained in $\{y_{Tt}\}$ 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. For concreteness, suppose a k-vector time series $\{x_t\}$ of covariates is observed, whose generating mechanism is

$$x_t = \mu^x_t + u^x_t, \qquad (3)$$

where $\{\mu^x_t\}$ is deterministic component and $\{u^x_t\}$ is an unobserved stationary I(0) process. Moreover, suppose the deterministic components $\{\mu^y_t\}$ and $\{\mu^x_t\}$ are polynomial trends of orders p_y and p_x, respectively. That is, suppose

$$\mu^y_t = \sum_{i=0}^{p_y} \beta^y_i t^i, \quad \mu^x_t = \sum_{i=0}^{p_x} \beta^x_i t^i, \qquad (4)$$

where $\{\beta^y_i : 0 \leq i \leq p_y\} \subseteq \mathbb{R}$ and $\{\beta^x_i : 0 \leq i \leq p_x\} \subseteq \mathbb{R}^k$ are fixed parameters.

The present article proposes two new tests that exploit the information contained in the covariates $\{x_t\}$ when testing the null hypothesis that $\{y_{Tt}\}$ is (trend-)stationary. Both tests enjoy optimality properties in the special case where $\{u_t = (u^y_t, u^{x\prime}_t)'\}$ is Gaussian white noise but are valid under mild moment and memory conditions on $\{u_t\}$. The tests can be viewed as generalizations of existing univariate stationarity tests and the new tests dominate their univariate counterparts in terms asymptotic local power whenever the zero-frequency correlation between u^y_t and u^x_t is non-zero (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 article 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 the studies of Hansen

(1995) and Elliott and Jansson (2000). 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 of the usefulness of the new tests for that purpose. Finally, Section 5 offers a few concluding remarks, while mathematical derivations appear in an Appendix.

2. TESTING WITH WHITE NOISE ERRORS

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

$$\Omega = \begin{pmatrix} \omega_{yy} & \omega_{xy}' \\ \omega_{xy} & \Omega_{xx} \end{pmatrix}$$

is a known, positive definite matrix (partitioned in conformity with u_t). Consider the problem of testing the null hypothesis $H_0 : \theta_T = 1$ against the alternative $|\theta_T| < 1$. This problem is that of testing whether the permanent component $(1 - \theta_T) \sum_{s=1}^{t-1} u_s^y$ is absent from the permanent-transitory decomposition

$$y_{Tt} = \mu_t^y + (1 - \theta_T) \sum_{s=1}^{t-1} u_s^y + u_t^y$$

of y_{Tt}. If $\{u_t^y\}$ was observed, the transformation $y_{Tt} - u_t^y$ would completely remove the transitory component u_t^y from y_{Tt} without affecting the permanent component and the testing problem would be trivial. Of course, u_t^y is unobserved and the aforementioned transformation is infeasible in practice. Nonetheless, the stationary covariates x_t can be used to attenuate the transitory component of y_{Tt} without affecting the permanent component. The use of covariates therefore makes it easier to detect the permanent component of y_{Tt} if it is present, hereby 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.

Any reasonable sequence of tests of H_0 will be consistent in the sense that the power against a fixed alternative $\theta_T = \bar{\theta} \in (-1, 1)$ tends to unity as T, the sample size, increases without bound. To obtain nondegenerate asymptotic results, a reparameterization of the model is required. The approach taken here is to employ local-to-unity asymptotics and model $\{\theta_T\}$ as a sequence of parameters lying in a shrinking neighborhood of unity as T, the sample size, tends to infinity. The appropriate local alternatives are of the form $1-\theta_T = O\left(T^{-1}\right)$, which motivates the following assumption.

A1. $\theta_T = 1 - T^{-1}\lambda$ for some $\lambda \geq 0$.

Under A1, the null and alternative hypotheses are $\lambda = 0$ and $\lambda > 0$, respectively.

2.1. The Point Optimal Invariant Test.

Define $\beta = \left(\beta_0^y, \ldots, \beta_{p_y}^y, \beta_0^{x\prime}, \ldots, \beta_{p_x}^{x\prime}\right)'$ and let

$$z_{Tt} = \begin{pmatrix} y_{Tt} \\ x_t \end{pmatrix}, \quad v_{Tt} = \begin{pmatrix} v_{Tt}^y \\ u_t^x \end{pmatrix}, \quad d_t = \begin{pmatrix} d_t^y & 0 \\ 0 & I_k \otimes d_t^x \end{pmatrix},$$

where $d_t^y = \left(1, \ldots, t^{p_y}\right)'$ and $d_t^x = \left(1, \ldots, t^{p_x}\right)'$. Using this notation, the model under consideration can be written as

$$z_{Tt} = d_t' \beta + v_{Tt}.$$

The problem of testing $H_0 : \lambda = 0$ vs. $H_A : \lambda > 0$ is invariant under the group of transformations of the form $z_{Tt} \to z_{Tt} + d_t' b$, $b \in \mathbb{R}^{p_y+1+k(p_x+1)}$. A maximal invariant is $m_T = D_T^{\perp\prime} vec\left(z_{T1}, \ldots, z_{TT}\right)$, where $D_T^{\perp}$ is a matrix whose columns form an orthonormal basis for the orthogonal complement of the column space of $\left(d_1, \ldots, d_T\right)'$. For any $l \geq 0$ and any $1 \leq t \leq T$, let

$$z_{Tt}\left(l\right) = \begin{pmatrix} y_{Tt}\left(l\right) \\ x_t \end{pmatrix}, \quad d_{Tt}\left(l\right) = \begin{pmatrix} d_{Tt}^y\left(l\right) & 0 \\ 0 & I_k \otimes d_t^x \end{pmatrix},$$

where $y_{Tt}\left(l\right) = \Delta y_{Tt} + \left(1 - T^{-1} l\right) y_{T,t-1}\left(l\right)$ with initial condition $y_{T1}\left(l\right) = y_{T1}$ and $\left\{d_{Tt}^y\left(l\right)\right\}$ is defined analogously. The probability density of m_T is proportional to

$$\exp\left(-\frac{1}{2}\sum_{t=1}^{T} \tilde{v}_{Tt}\left(\lambda\right)' \Omega^{-1} \tilde{v}_{Tt}\left(\lambda\right)\right),$$

where $\tilde{v}_{Tt}\left(l\right) = z_{Tt}\left(l\right) - d_{Tt}\left(l\right)' \tilde{\beta}_T\left(l; \Omega\right)$,

$$\tilde{\beta}_T\left(l; \Omega\right) = \left(\sum_{t=1}^{T} d_{Tt}\left(l\right) \Omega^{-1} d_{Tt}\left(l\right)'\right)^{-1} \left(\sum_{t=1}^{T} d_{Tt}\left(l\right) \Omega^{-1} z_{Tt}\left(l\right)\right)$$

is the maximum likelihood estimator of β when $\lambda = l$ and the dependence of $\tilde{v}_{Tt}\left(l\right)$

on Ω has been suppressed to achieve notational economy. By the Neyman-Pearson Lemma, the test which rejects for large values of

$$P_T\left(\bar{\lambda};\Omega\right)=\sum_{t=1}^{T}\tilde{v}_{Tt}\left(0\right)'\Omega^{-1}\tilde{v}_{Tt}\left(0\right)-\sum_{t=1}^{T}\tilde{v}_{Tt}\left(\bar{\lambda}\right)'\Omega^{-1}\tilde{v}_{Tt}\left(\bar{\lambda}\right) \tag{5}$$

is the most powerful invariant test of H_0 against the specific alternative $\lambda=\bar{\lambda}>0$.

Let $0\leq R<1$ and $l\geq 0$ be given. Let $\bar{\Omega}^{1/2}$ be the (lower triangular) Cholesky factor of the 2×2 matrix

$$\bar{\Omega}=\begin{pmatrix}1 & R\\ R & 1\end{pmatrix}.$$

Define

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

where $dV_l^{\lambda}\left(r\right)=dV^{\lambda}\left(r\right)-lV_l^{\lambda}\left(r\right)dr$, $dD_l^{y}\left(r\right)=dD^{y}\left(r\right)-lD_l^{y}\left(r\right)dr$ with initial conditions $V_l^{\lambda}\left(0\right)=0$, $D_l^{y}\left(0\right)=0$, while $D^{y}\left(r\right)=\left(1,\ldots,r^{p_y}\right)'$, $D^{x}\left(r\right)=\left(1,\ldots,r^{p_x}\right)'$, $V^{\lambda}\left(r\right)=V\left(r\right)+\lambda\int_0^r V\left(s\right)ds$ and $\left(V,W\right)'$ is a Brownian motion with covariance matrix $\bar{\Omega}$. Here, and elsewhere, the dependence of U_l^{λ} and D_l on R is suppressed. Finally, let $R_{\#}=\left(1-R^2\right)^{-1/2}$ and define

$$\varphi_P\left(\lambda;\bar{\lambda},R^2\right)=$$

$$-\bar{\lambda}^2R_{\#}^2\int_0^1 V_{\bar{\lambda}}^{\lambda}\left(r\right)^2dr+2\bar{\lambda}R_{\#}^2\left(\int_0^1 V_{\bar{\lambda}}^{\lambda}\left(r\right)dV^{\lambda}\left(r\right)-R\int_0^1 V_{\bar{\lambda}}^{\lambda}\left(r\right)dW\left(r\right)\right)$$
$$+\left(\int_0^1 D_{\bar{\lambda}}\left(r\right)dU_{\bar{\lambda}}^{\lambda}\left(r\right)\right)'\left(\int_0^1 D_{\bar{\lambda}}\left(r\right)D_{\bar{\lambda}}\left(r\right)'dr\right)^{-1}\left(\int_0^1 D_{\bar{\lambda}}\left(r\right)dU_{\bar{\lambda}}^{\lambda}\left(r\right)\right)$$
$$-\left(\int_0^1 D_0\left(r\right)dU_0^{\lambda}\left(r\right)\right)'\left(\int_0^1 D_0\left(r\right)D_0\left(r\right)'dr\right)^{-1}\left(\int_0^1 D_0\left(r\right)dU_0^{\lambda}\left(r\right)\right).$$

Theorem 1. Let $\{z_{Tt}\}$ be generated by (1) – (4), suppose $u_t\sim i.i.d.\ N\left(0,\Omega\right)$ and suppose A1 holds. Then $P_T\left(\bar{\lambda};\Omega\right)\rightarrow_d\varphi_P\left(\lambda;\bar{\lambda},\rho^2\right)$, where "$\rightarrow_d$" signifies convergence in distribution as $T\rightarrow\infty$ and $\rho=\left(\omega_{yy}^{-1}\omega_{xy}'\Omega_{xx}^{-1}\omega_{xy}\right)^{1/2}$ is the correlation coefficient computed from Ω.

For any $T > \max(p_y, p_x) + 1$ and corresponding to any invariant test of $H_0 : \lambda = 0$ there is a test function $\phi_T : \mathbb{R}^{(T-p_x-1)k+T-p_y-1} \to [0,1]$ such that H_0 is rejected with probability $\phi_T(m)$ whenever $m_T = m$. For any given λ and Ω and any such ϕ_T, the probability of rejecting H_0 is $\int \phi_T(m) f_T(m \mid \lambda, \Omega)\, dm$, where $f_T(\cdot \mid \lambda, \Omega)$ denotes the probability density of the maximal invariant. A test ϕ_T is of level $\alpha \in (0,1)$ if its size, viz. $\int \phi_T(m) f_T(m \mid 0, \Omega)\, dm$, is less than or equal to α. Similarly, a sequence $\{\phi_T\}$ of test functions is said to be asymptotically of level α if

$$\overline{\lim}_{T\to\infty} \int \phi_T(m) f_T(m \mid 0, \Omega)\, dm \leq \alpha.$$

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.

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

$$\int \phi_T(m) f_T(m \mid \bar{\lambda}, \Omega)\, dm$$

against the point alternative $\lambda = \bar{\lambda}$ is maximized over all invariant tests of level α by the test function $1\left(P_T(\bar{\lambda}; \Omega) > c_T^P(\bar{\lambda}, \alpha, \Omega)\right)$, where $1(\cdot)$ is the indicator function and $c_T^P(\bar{\lambda}, \alpha, \Omega)$ 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 $\Pr\left[\varphi_P(0; \bar{\lambda}, \rho^2) > c^P(\bar{\lambda}, \alpha, \rho^2)\right] = \alpha$. The statistic $P_T(\bar{\lambda}; \Omega)$ is asymptotically POI in the sense that $\phi_T^P(m_T; \bar{\lambda}, \alpha, \Omega) = 1\left(P_T(\bar{\lambda}; \Omega) > c^P(\bar{\lambda}, \alpha, \rho^2)\right)$ maximizes

$$\lim_{T\to\infty} \int \phi_T(m) f_T(m \mid \bar{\lambda}, \Omega)\, dm$$

over all invariant tests asymptotically of level α. That is,

$$\overline{\lim}_{T\to\infty} \int \phi_T(m) f_T(m \mid \lambda, \Omega)\, dm \leq \underline{\lim}_{T\to\infty} \int \phi_T^P(m_T; \lambda, \alpha, \Omega) f_T(m \mid \lambda, \Omega)\, dm$$

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 $\Pr\left(\varphi_P(\lambda; \lambda, \rho^2) > c^P(\lambda, \alpha, \rho^2)\right)$.

Theorem 2 of Saikkonen and Luukkonen (1993a) obtained an upper bound on the asymptotic power function of any location and scale invariant stationarity test in the univariate case. The present optimality result generalizes that result in two respects. Since scale invariance is not imposed, the result stated here covers a larger class of tests than Saikkonen and Luukkonen's (1993a) Theorem 2 even in the univariate case. Additional gains in terms of generality are achieved by studying a multivariate model which contains Saikkonen and Luukkonen's (1993a) univariate model as a special case. While the desirability of achieving the latter increase in generality should be obvious, a couple of remarks on the role of scale invariance (or lack thereof) seem appropriate. The assumption of scale invariance has been omitted as it is unnecessary in the sense that power bound obtained here can attained even when the relevant scale parameter is unknown. Moreover, imposing scale invariance does not obviate the need to estimate a scale parameter in the more general model considered in the next section. That is, the scale invariant (in the present model) test statistic is not asymptotically pivotal when less restrictive assumptions on $\{u_t\}$ are introduced.

The function $\Pr\left(\varphi_P\left(\lambda;\lambda,\rho^2\right) > c^P\left(\lambda,\alpha,\rho^2\right)\right)$ 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 $\phi_T^P\left(m_T;\lambda,\alpha,\Omega\right)$. 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. This Gaussian power envelope therefore provides us with a useful benchmark against which the power function of any invariant test (asymptotically of level α) can be compared.

The univariate counterpart of $P_T\left(\bar{\lambda};\Omega\right)$ is

$$P_T^y\left(\bar{\lambda};\omega_{yy}\right) = \omega_{yy}^{-1}\left(\sum_{t=1}^{T}\hat{\vartheta}_{Tt}^y\left(0\right)^2 - \sum_{t=1}^{T}\hat{\vartheta}_{Tt}^y\left(\bar{\lambda}\right)^2\right),$$

where $\hat{\vartheta}_{Tt}^y\left(l\right) = y_{Tt}\left(l\right) - d_{Tt}^y\left(l\right)'\hat{\beta}_T^y\left(l\right)$ and

$$\hat{\beta}_T^y\left(l\right) = \left(\sum_{t=1}^{T} d_{Tt}^y\left(l\right)d_{Tt}^y\left(l\right)'\right)^{-1}\left(\sum_{t=1}^{T} d_{Tt}^y\left(l\right)y_{Tt}\left(l\right)\right).$$

When $u_t^y \sim i.i.d.\ \mathcal{N}\left(0,\omega_{yy}\right)$, the test which rejects for large values of $P_T^y\left(\bar{\lambda};\omega_{yy}\right)$ is more powerful against the specific alternative $\lambda = \bar{\lambda} > 0$ than any other invariant test of H_0 based solely on $\{y_{Tt}\}$, where invariance is with respect to transformations of the form $y_{Tt} \to y_{Tt} + b_y' d_t^y$, $b_y \in \mathbb{R}^{p_y+1}$.

When $\rho^2 = 0$, $\{y_{Tt}\}$ and $\{x_t\}$ are independent. In that case, the covariates $\{x_t\}$ carry no information about $\{y_{Tt}\}$ and the statistics $P_T\left(\bar{\lambda};\Omega\right)$ and $P_T^y\left(\bar{\lambda};\omega_{yy}\right)$ are equivalent. In fact, $P_T\left(\bar{\lambda};\Omega\right) = P_T^y\left(\bar{\lambda};\omega_{yy}\right)$ when $\rho^2 = 0$. In contrast, the rejection regions of the tests based on the statistics $P_T\left(\bar{\lambda};\Omega\right)$ and $P_T^y\left(\bar{\lambda};\omega_{yy}\right)$ differ whenever $\rho^2 \neq 0$. These differences persist asymptotically as $P_T^y\left(\bar{\lambda};\omega_{yy}\right) \to_d \varphi_P\left(\lambda;\bar{\lambda},0\right)$ under the assumptions of Theorem 1. Comparing $\varphi_P\left(\lambda;\bar{\lambda},0\right)$ and $\varphi_P\left(\lambda;\bar{\lambda},\rho^2\right)$, the limiting distribution of $P_T(\bar{\lambda};\Omega)$ 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 $\Pr\left(\varphi_P(\lambda;\lambda,\rho^2) > c^P(\lambda, 0.05, \rho^2)\right)$ for selected values of ρ^2 in the constant mean case $(p_y = p_x = 0)$. 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. As the quality of the covariates (as measured by ρ^2) increases, so does 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_T of 0.975 when $T = 200$. The univariate power envelope is 0.32, while the envelopes are 0.40 and 0.58 when ρ^2 equals 0.2 and 0.5, respectively. Being 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 case $(p_y = p_x = 1)$, but the qualitative conclusion remains, as can be seen from Figure 2.

FIGURE 1 ABOUT HERE
FIGURE 2 ABOUT HERE

2.2. The Locally Best Invariant Test.

Even asymptotically, the critical region for the test based on $P_T(\bar{\lambda};\Omega)$ depends on $\bar{\lambda}$. As a consequence, no test is asymptotically uniformly most powerful (with respect to the class of invariant tests) in the sense of Basawa and Scott (1983, p. 69). 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 $P_T(\bar{\lambda};\Omega)$, where $\bar{\lambda}$ is a prespecified alternative against which maximal power is desired. As an alternative to that test, the present subsection develops a test based on a Taylor series expansion of $P_T(\bar{\lambda};\Omega)$ around $\bar{\lambda} = 0$. The resulting test can be implemented without specifying an alternative in advance and enjoys certain local optimality properties.

Let $\dot{P}_T(0;\Omega) = \partial P(\bar{\lambda};\Omega)/\partial\bar{\lambda}\,|_{\bar{\lambda}=0}$ and $\ddot{P}_T(0;\Omega) = \frac{1}{2}\partial^2 P(\bar{\lambda};\Omega)/\partial\bar{\lambda}^2\,|_{\bar{\lambda}=0}$. Using simple algebra, it can be shown that

$$
\begin{aligned}
\dot{P}_T(0;\Omega) &= -\begin{pmatrix} 1 \\ 0 \end{pmatrix}' \left(\Omega^{-1} T^{-1} \sum_{t=1}^{T} \tilde{v}_{Tt}(0)\, \tilde{v}_{Tt}(0)' \right) \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \\
\ddot{P}_T(0;\Omega) &= L_T(\Omega) + T^{-1} \dot{P}_T(0;\Omega),
\end{aligned}
$$

where

$$
L_T(\Omega) = \sum_{t=1}^{T} \tilde{V}_{Tt}' \Omega^* \tilde{V}_{Tt} + \left(\sum_{t=1}^{T} d_t \Omega^{**} \tilde{V}_{Tt} \right)' \left(\sum_{t=1}^{T} d_t \Omega^{-1} d_t' \right)^{-1} \left(\sum_{t=1}^{T} d_t \Omega^{**} \tilde{V}_{Tt} \right), \quad (6)
$$

$$
\Omega^* = \begin{pmatrix} \omega_{yy.x}^{-1} & \omega^{xy\prime} \\ \omega^{xy} & 0 \end{pmatrix}, \quad \Omega^{**} = \begin{pmatrix} 0 & \omega^{xy\prime} \\ -\omega^{xy} & 0 \end{pmatrix},
$$

$\tilde{V}_{Tt} = T^{-1} \sum_{s=1}^{t-1} \tilde{v}_{Ts}(0)$, $\omega_{yy.x} = \omega_{yy} - \omega_{xy}' \Omega_{xx}^{-1} \omega_{xy}$ and $\omega^{xy} = -\omega_{yy.x}^{-1} \Omega_{xx}^{-1} \omega_{xy}$.

Since $T^{-1} \sum_{t=1}^{T} \tilde{v}_{Tt}(0)\, \tilde{v}_{Tt}(0)' \to_p \Omega$, the limiting distribution of $\dot{P}_T(0;\Omega)$ is degenerate. Indeed, $\dot{P}_T(0;\Omega) \to_p -1$, where " $\to_p$ " signifies convergence in probability as $T \to \infty$. On the other hand, Theorem 2 (a) below shows that the limiting distribution of $L_T(\Omega)$ equals that of the random variable $\varphi_L(\lambda; R^2)$, where

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

$\tilde{U}^\lambda(r) = U_0^\lambda(r) - \left(\int_0^r D(s)\, ds \right)' \left(\int_0^1 D(s) D(s)'\, ds \right)^{-1} \left(\int_0^1 D(s)\, dU_0^\lambda(s) \right)$, $D(r) = D_0(r)$ and

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

The test which rejects for large values of $L_T(\Omega)$ is therefore asymptotically equivalent (in an obvious sense) to the test that rejects for large values of the second-order Taylor approximation to $P_T(\bar{\lambda};\Omega)$, viz. $\dot{P}_T(0;\Omega)\bar{\lambda}+\ddot{P}_T(0;\Omega)\bar{\lambda}^2$. As a consequence, $L_T(\Omega)$ might be expected to enjoy certain local optimality properties.

A test $\{\phi_T\}$ is asymptotically locally efficient (with respect to the class of invariant tests asymptotically of size α) in the sense of Basawa and Scott (1983, p. 70) if it maximizes

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

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. 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

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

where $l_T^{(q)}(m\mid\Omega)=\partial^q\log f_T(m\mid\lambda,\Omega)/\partial\lambda^q\mid_{\lambda=0}$. An invariant test is said to be asymptotically locally best invariant (LBI) if it maximizes

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

over all invariant tests asymptotically of the same size. In regular cases where partial derivatives of $\int\log f_T(m\mid\lambda,\Omega)\cdot f_T(m\mid 0,\Omega)\,dm$ with respect to λ 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, $L_T(\Omega)$ is asymptotically LBI in the (stronger) sense defined here.

Theorem 2. *Let $\{z_{Tt}\}$ be generated by (1) – (4), suppose $u_t\sim i.i.d.\ \mathcal{N}(0,\Omega)$ and suppose A1 holds.*
(a) $L_T(\Omega)\to_d\varphi_L(\lambda;\rho^2)$.
(b) Let $\alpha\in(0,1)$ be given and suppose $\{\phi_T\}$ is asymptotically of level α. Then

$$\lim_{T\to\infty}\frac{\partial}{\partial\lambda}\int\phi_T(m)\,f_T(m\mid\lambda,\Omega)\,dm\bigg|_{\lambda=0}=0.$$

(c) If $\{\phi_T\}$ is asymptotically of size α or $\alpha\leq\Pr\left[\varphi_L(0;\rho^2)>E\varphi_L(0;\rho^2)\right]$, then

$$\overline{\lim}_{T\to\infty} \frac{\partial^2}{\partial\lambda^2} \int \left(\phi_T(m) - \phi_T^L(m;\alpha,\Omega)\right) f_T(m \mid \lambda,\Omega)\, dm\Big|_{\lambda=0} \leq 0,$$

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

The univariate counterpart of $L_T(\Omega)$ is

$$L_T^y(\omega_{yy}) = \omega_{yy}^{-1} \sum_{t=1}^{T} \left(\hat{V}_{Tt}^y\right)^2, \tag{7}$$

where $\hat{V}_{Tt}^y = T^{-1}\sum_{s=1}^{t-1} \hat{v}_{Ts}^y(0)$. The statistics $L_T(\bar{\lambda};\Omega)$ and $L_T^y(\omega_{yy})$ are equivalent if and only if $\rho^2 = 0$. Moreover, $L_T^y(\omega_{yy}) \to_d \varphi_L(\lambda;0)$ under the assumptions of Theorem 2, so the difference between $L_T(\bar{\lambda};\Omega)$ and $L_T^y(\omega_{yy})$ persist asymptotically whenever $\rho^2 \neq 0$. As was the case with the power envelopes derived in the previous subsection, 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 $u_t \sim i.i.d.\ \mathcal{N}(0,\Omega)$, where Ω is known. The optimality theory seems to depend the normality assumption. On the other hand, it is straightforward to construct feasible test statistics having limiting distributions of the form $\varphi_L(\lambda;\bar{\lambda},\rho^2)$ and $\varphi_L(\lambda;\rho^2)$ under much less stringent assumptions. For instance, it suffices to assume that $\{u_t\}$ is generated by the linear process

$$u_t = \sum_{i=0}^{\infty} C_i \varepsilon_{t-i},$$

where

A2. (i) $\{\varepsilon_t : t \in \mathbb{Z}\}$ is *i.i.d.* with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t\varepsilon_t') = I_{k+1}$.
(ii) $\sum_{i=0}^{\infty} C_i$ has full rank and $\sum_{i=1}^{\infty} i\,\|C_i\| < \infty$, where $\|\cdot\|$ is the Euclidean norm.

Under A2, the limiting distributions of interest depend on the parameters $\{C_i\}$ only through

$$\Omega = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T}\sum_{s=1}^{T} E(u_t u_s')$$

and

$$\Gamma = \lim_{T\to\infty} T^{-1} \sum_{t=2}^{T} \sum_{s=1}^{t-1} E\left(u_t u_s'\right).$$

Notice that Ω, the long-run covariance matrix of u_t, equals $E\left(u_t u_t'\right)$ when $\{u_t\}$ is white noise, so the present notation is consistent with that of Section 2.

3.1. Feasible Tests.

Under A1-A2, $L_T(\Omega) \to_d \varphi_L(\lambda; \rho^2)$, where ρ is the correlation coefficient computed from Ω. Therefore, no modifications are needed in order to obtain a version of $L_T(\Omega)$ which has a limiting distribution of the desired form. In contrast, $P_T(\bar{\lambda}; \Omega)$ suffers from "serial correlation bias" in the more general case considered here. Specifically, $P_T(\bar{\lambda}; \Omega) \to_d \varphi_P(\lambda; \bar{\lambda}, \rho^2) + 2\bar{\lambda}\omega_{yy.x}^{-1}\gamma_{yy.x}$, where $\gamma_{yy.x} = \gamma_{yy} - \omega_{xy}'\Omega_{xx}^{-1}\gamma_{xy}$ and Ω and Γ have been partitioned in the obvious way. Let

$$Q_T\left(\bar{\lambda}; \Omega, \Gamma\right) = P_T\left(\bar{\lambda}; \Omega\right) - 2\bar{\lambda}\omega_{yy.x}^{-1}\gamma_{yy.x}. \tag{8}$$

Of course, $Q_T\left(\bar{\lambda}; \Omega, \Gamma\right) = P_T\left(\bar{\lambda}; \Omega\right)$ when $\{u_t\}$ is white noise, since $\Gamma = 0$ in that case. More generally, $Q_T\left(\bar{\lambda}; \Omega, \Gamma\right)$ corrects $P_T\left(\bar{\lambda}; \Omega\right)$ for serial correlation bias and $Q_T\left(\bar{\lambda}; \Omega, \Gamma\right) \to_d \varphi_P\left(\lambda; \bar{\lambda}, \rho^2\right)$ under A1-A2.

In most (if not all) applications, the tests based on $L_T(\Omega)$ and $Q_T\left(\bar{\lambda}; \Omega, \Gamma\right)$ are infeasible because Ω and Γ are unknown. It therefore seems natural to consider test statistics of the form $L_T\left(\hat{\Omega}_T\right)$ and $Q_T\left(\bar{\lambda}; \hat{\Omega}_T, \hat{\Gamma}_T\right)$, where $\hat{\Omega}_T$ and $\hat{\Gamma}_T$ are consistent estimators of Ω and Γ, respectively. Numerous estimators have been proposed in the literature. For concreteness, the following discussion focuses on VAR(1) prewhitened kernel estimators with plug-in bandwidths. These estimators are defined as follows.

For $2 \le t \le T$, let $\hat{v}_{Tt}^{PW} = \hat{v}_{Tt} - \hat{A}_T\hat{v}_{T,t-1}$, where $\hat{A}_T$ is a $(k+1)\times(k+1)$ matrix and $\hat{v}_{Tt} = z_{Tt} - d_t'\hat{\beta}_T$, where $\hat{\beta}_T = \left(\sum_{t=1}^T d_t d_t'\right)^{-1}\left(\sum_{t=1}^T d_t z_{Tt}\right)$ is the OLS estimator of β. Let $\hat{\Sigma}_T = T^{-1}\sum_{t=1}^T \hat{v}_t\hat{v}_t'$ and $\hat{\Sigma}_T^* = (T-1)^{-1}\sum_{t=2}^T \hat{v}_t^{PW}\hat{v}_{t-1}'$ and define

$$\hat{\Omega}_T^{PW} = (T-1)^{-1}\sum_{t=2}^{T}\sum_{s=2}^{T} k\left(\frac{|t-s|}{\hat{b}_T}\right)\hat{v}_{Tt}^{PW}\hat{v}_{Ts}^{PW\prime}, \tag{9}$$

$$\hat{\Gamma}_T^{PW} = (T-1)^{-1}\sum_{t=3}^{T}\sum_{s=2}^{t-1} k\left(\frac{|t-s|}{\hat{b}_T}\right)\hat{v}_{Tt}^{PW}\hat{v}_{Ts}^{PW\prime}, \tag{10}$$

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

$$\begin{aligned}\hat{\Omega}_T &= \left(I-\hat{A}_T\right)^{-1}\hat{\Omega}_T^{PW}\left(I-\hat{A}_T'\right)^{-1},\\ \hat{\Gamma}_T &= \left(I-\hat{A}_T\right)^{-1}\hat{\Gamma}_T^{PW}\left(I-\hat{A}_T'\right)^{-1}+\left(I-\hat{A}_T\right)^{-1}\hat{A}_T\hat{\Sigma}_T\\ &\quad-\left(I-\hat{A}_T\right)^{-1}\hat{\Sigma}_T^{*}\hat{A}_T'\left(I-\hat{A}_T'\right)^{-1}.\end{aligned}$$

Guidance on the choice of $\{\hat{A}_T\}$, $k(\cdot)$ and $\{\hat{b}_T\}$ will be provided in Section 3.3. For now, only the following high-level assumption is required.

A3. (i) $T^{1/2}\left(\hat{A}_T-A\right)=O_p(1)$ for some A such that $(I-A)$ is nonsingular.
(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) $k(0)=1$, $k(\cdot)$ is continuous at zero and $\sup_{s\geq 0}|k(s)|<\infty$.
(iv) $\int_0^\infty \bar{k}(r)\,dr<\infty$, where $\bar{k}(r)=\sup_{s\geq r}|k(s)|$ (for all $r\geq 0$).

Parts (i) and (ii) of A3 are adapted from Andrews and Monahan (1992), while A3(iii)-(iv) have been used by Jansson (2001). Under A1-A3, $\hat{\Omega}_T$ and $\hat{\Gamma}_T$ are consistent estimators of Ω and Γ, respectively. This property turns out to be sufficient for $L_T\left(\hat{\Omega}_T\right)$ and $Q_T\left(\bar{\lambda};\hat{\Omega}_T,\hat{\Gamma}_T\right)$ to have the same limiting distributions as $L_T(\Omega)$ and $Q_T\left(\bar{\lambda};\Omega,\Gamma\right)$, respectively.

Theorem 3. Let $\{z_{Tt}\}$ be generated by (1) – (4) and suppose A1-A3 hold. Then $Q_T\left(\bar{\lambda};\hat{\Omega}_T,\hat{\Gamma}_T\right)\to_d \varphi_P\left(\lambda;\bar{\lambda},\rho^2\right)$ and $L_T\left(\hat{\Omega}_T\right)\to_d \varphi_L\left(\lambda;\rho^2\right)$, where ρ is the correlation coefficient computed from Ω.

The univariate counterparts of $Q_T\left(\bar{\lambda};\hat{\Omega}_T,\hat{\Gamma}_T\right)$ and $L_T\left(\hat{\Omega}_T\right)$ are $Q_T^y\left(\bar{\lambda};\hat{\omega}_{yy,T},\hat{\gamma}_{yy,T}\right)$ and $L_T^y\left(\hat{\omega}_{yy,T}\right)$, respectively, where

$$Q_T^y\left(\bar{\lambda};\omega_{yy},\gamma_{yy}\right)=P_T^y\left(\bar{\lambda};\omega_{yy}\right)-2\bar{\lambda}\omega_{yy}^{-1}\gamma_{yy}.$$

Under A1-A3, $Q_T^y\left(\bar{\lambda};\hat{\omega}_{yy,T},\hat{\gamma}_{yy,T}\right)\to_d \varphi_P\left(\lambda;\bar{\lambda},0\right)$ and $L_T^y\left(\hat{\omega}_{yy,T}\right)\to_d \varphi_L\left(\lambda;0\right)$. The test statistic $L_T^y\left(\hat{\omega}_{yy,T}\right)$ is well known (e.g. Kwiatkowski, Phillips, Schmidt, and Shin (1992)). On the other hand, the semiparametric version $Q_T^y\left(\bar{\lambda};\hat{\omega}_{yy,T},\hat{\gamma}_{yy,T}\right)$ of the univariate POI test would appear to be new.

3.2. Asymptotic Local Power.

Saikkonen and Luukkonen (1993a) considered the constant mean case and found that their test statistic $\tilde{R}(1-7/T)$, which corresponds to $Q_T^y(7;\hat{\omega}_{yy,T},\hat{\gamma}_{yy,T})$, has a local asymptotic power function which is almost indistinguishable from the univariate power envelope. The choice $\bar{\lambda} = 7$ produces a test which 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. It therefore seems natural to consider $Q_T\left(\bar{\lambda};\hat{\Omega}_T,\hat{\Gamma}_T\right)$, where $\bar{\lambda}$ 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 orders of the deterministic components 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 $\bar{\lambda}$ for each ρ^2. Fortunately, a much simpler approach yields very satisfactory results. The approach taken here is to use the same $\bar{\lambda}$ for all values of ρ^2. The value of $\bar{\lambda}$ 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 which has excellent power properties (relative to the power envelope) when ρ^2 is low. Moreover, Q_T dominates its univariate counterpart for all values of ρ^2. In fact, the test has a power function which is very close to the power envelope even for non-zero values of ρ^2.

Figure 3 illustrates this is in the constant mean case with $\rho^2 = 0.50$. In addition to the power envelope and the local asymptotic power of Q_T, Figure 3 also plots the local power function of the LBI test L_T and the univariate tests Q_T^y and L_T^y. Comparing Q_T to 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 be found. The Pitman asymptotic relative efficiency (ARE) of Q_T with respect to Q_T^y (evaluated at power 0.50) is 1.65, implying that in large samples the univariate test needs 65% more observations as the test using covariates in order 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 even more pronounced when useful covariates are available.

FIGURE 3 ABOUT HERE

Figure 4 presents results for the linear trend case. The statistics Q_T and Q_T^y use $\bar{\lambda} = 12$, the value which yields an asymptotically 0.50-optimal, level 0.05 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 Q_T has a power function which lies close to the envelope and way above the power functions corresponding to L_T and Q_T^y. For instance, the Pitman ARE of Q_T with respect to 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.

FIGURE 4 ABOUT HERE

Tables 1a-1d gives various critical values for Q_T and L_T for $0 \leq p_y, p_x \leq 1$, which seem to be the cases of empirical relevance. The critical values are presented for ρ^2 in steps of 0.1. The recommendation is to use the critical value corresponding to $\hat{\rho}_T^2 = \hat{\omega}_{yy,T}^{-1}\hat{\omega}_{xy,T}'\hat{\Omega}_{xx,T}^{-1}\hat{\omega}_{xy,T}$ computed from $\hat{\Omega}_T$. Interpolation can be used to obtain critical values for values of $\hat{\rho}_T^2$ between those given in the tables.

TABLES 1a-1d ABOUT HERE

3.3. 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

$$\begin{pmatrix} u_t^y \\ u_t^x \end{pmatrix} = \begin{pmatrix} c_{yy}(L) & 0 \\ \rho & (1-\rho^2)^{1/2} \end{pmatrix} \begin{pmatrix} \varepsilon_t^y \\ \varepsilon_t^x \end{pmatrix}, \tag{11}$$

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:

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

and

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

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 (11) is the correlation coefficient computed from Ω.

The parameters Ω and Γ are estimated using VAR(1) prewhitened kernel estimators with plug-in bandwidths. The estimator $\hat{A}_T$ used in the prewhitening procedure is obtained by adjusting the least squares estimator

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

in such a way that the eigenvalues of $\hat{A}_T$ do not exceed 0.97 in absolute value. Let $\hat{M}_T^{LS} \hat{J}_T^{LS} \left(\hat{M}_T^{LS}\right)^{-1}$ be the Jordan decomposition of $\hat{A}_T^{LS}$ and set $\hat{A}_T = \hat{M}_T^{LS} \hat{J}_T \left(\hat{M}_T^{LS}\right)^{-1}$, where $\hat{J}_T$ is a Jordan matrix obtained from $\hat{J}_T^{LS}$ 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 $\hat{A}_T^{LS}$ and yields $\hat{A}_T = \hat{A}_T^{LS}$ whenever the eigenvalues of $\hat{A}_T^{LS}$ do not exceed 0.97. The estimators $\hat{\Omega}_T^{PW}$ and $\hat{\Gamma}_T^{PW}$ in (9) – (10) are constructed using the Quadratic Spectral kernel along with a plug-in bandwidth. The value of the plug-in bandwidth is obtained by setting $\hat{b}_T = 1.3221 \cdot T^{1/5}$ (following Andrews (1991)) and $\hat{a}_T = \max\left(\min\left(\hat{\alpha}_{AR(1)}(2)^{1/5}, 5\right), 0.05\right)$, where $\hat{\alpha}_{AR(1)}(2)$ is computed from Andrews's (1991) equation (6.4) (with $w_a = 1$ for all a). Since $0.05 \le \hat{a}_T \le 5$, is imposed, A3 (ii) is automatically satisfied. In particular, the condition $\hat{a}_T \le 5$ controls the behavior of the estimated bandwidth under fixed alternatives, hereby circumventing the problems discussed by Choi (1994, pp. 734-735).

Tables 2a-2b and 2c-2d summarize the results for the constant mean and linear trend cases, respectively. The tables report the observed rejection rates (based on 5,000 Monte Carlo replications) of 5% level tests implemented using critical values based on the estimate $\hat{\rho}_T^2$ computed from $\hat{\Omega}_T$. 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 $\{y_{Tt}\}$. 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 way short of the nominal level in the MA(1) model with $|\delta| \ge 0.5$, which leads to an unnecessary reduction in power when asymptotic critical values are used. Moreover, the pattern exhibited by the rejection rates in the AR(1) model with $a = 0.8$ is rather peculiar. The non-increasing (in λ) power would appear to be a finite sample phenomenon caused by the estimation of Ω and Γ. Specifically, the power of the infeasible tests using the true values of Ω and Γ increases with λ, as does the power of the feasible tests when the sample size is 500. A thorough investigation of these problems is beyond the scope of the present paper and is left for future research.

TABLES 2a-2d ABOUT HERE

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 (Engle and Granger (1987)). 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, $\{\mu_t^Y\}$ and $\{\mu_t^X\}$ are deterministic components and $(u_t^Y, u_t^{X\prime})'$ satisfies A2. Setting $y_{Tt} = Y_t - \psi' X_t$, $\mu_t^y = \mu_t^Y - \psi' \mu_t^X$, $x_t = \Delta X_t$ and $\mu_t^x = \mu_t^X$, the cointegration model reduces to (1) – (4) with $(u_t^y, u_t^{x\prime})' = (u_t^Y, u_t^{X\prime})'$ and $\theta_T = 1$. In this context, the null hypothesis $\theta_T = 1$ is the hypothesis that $\{(Y_t, X_t')'\}$ is cointegrated with cointegrating vector $\begin{pmatrix} 1 & -\psi' \end{pmatrix}'$ (see also Saikkonen and Luukkonen (1993b, p. 597)).

In some economic applications, the (potentially) cointegrating vector $\begin{pmatrix} 1 & -\psi' \end{pmatrix}'$ is known a priori from economic theory. In such cases, the null hypothesis that $\{(Y_t, X_t')'\}$ is cointegrated with cointegrating vector $\begin{pmatrix} 1 & -\psi' \end{pmatrix}'$ is invariably tested by applying a univariate stationarity test to the series $\{Y_t - \psi' X_t\}$, hereby discarding the potentially useful information contained in the series $\{\Delta X_t\}$. 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 non-zero. In economic applications, such non-zero correlations are the rule rather than the exception. Indeed, this is the raison d'être of the huge literature on efficient inference in cointegrated systems (e.g. Park (1992), Phillips (1991), Phillips and Hansen (1990), Saikkonen (1991, 1992) and Stock and Watson (1993)). 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) over the period of the recent float. 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)):

$$s_t = \beta_0 + \beta_1 t + \psi^D p_t^D + \psi^F p_t^F + u_t, \tag{12}$$

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 consists 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 monthly data covering the period from January 1974 through January 2001. All data series are from the Global Financial Database (GFD). The exchange rate data is from GFD series __GBP_D. Results of previous studies have been found to be sensitive to the choice of price levels (Froot and Rogoff (1995)), suggesting that both a wholesale price index (WPI) and a consumer price index (CPI) should be considered. In the case of the United States, the WPI data is from GFD series WPUSAICM, while the CPI data is from GFD series CPUSAM. Data for the United Kingdom is from GFD series WPGBRM (WPI) and CPGBRM (CPI), respectively. When implementing the tests, the nuisance parameters are estimated in the same way as in the Monte Carlo experiment of Section 3.3. The linear trend version of the test statistics is used. In other words, $p_y = p_x = 1$ is imposed. Table 3 summarizes the results.

TABLE 3 ABOUT HERE

In agreement with other studies (e.g. Culver and Papell (1999)), the tests fail to reject the null hypothesis of stationarity when the covariates are ignored. The tests using covariates, in contrast, do provide evidence against long-run PPP. With the exception of the Q_T test using CPI data, the null hypothesis of stationarity is rejected at the 5% level. The estimates $\hat{\rho}_T^2$ are large, suggesting that substantial power gains are achieved by using covariates, which in turn explains why the L_T and Q_T tests reach different conclusions than their univariate counterparts.

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 a 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, in which case an obvious set of covariates is available.

6. APPENDIX

The proofs of Theorems 1-3 make use of Lemma 4, which shows how functional laws for sample moments of the transformed data $\{z_{Tt}(\bar{\lambda})\}$ and $\{d_{Tt}(\bar{\lambda})\}$ can be deduced from functional laws for $\{z_{Tt}\}$ and $\{d_t\}$. Since these preliminary results might be of independent interest, they are presented in greater generality than needed for the proofs of Theorems 1-3.

In Lemma 4 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 4. *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

$$\begin{pmatrix} F_{T,\lfloor T\cdot\rfloor} \\ T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} g_{Tt} \end{pmatrix} \to_d \begin{pmatrix} F(\cdot) \\ G(\cdot) \end{pmatrix}, \tag{13}$$

where F and G are continuous. Then

$$\begin{pmatrix} F_{T,\lfloor T\cdot\rfloor}(l) \\ g_{T,\lfloor T\cdot\rfloor} - g_{T,\lfloor T\cdot\rfloor}(l) \\ T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} g_{Tt}(l) \end{pmatrix} \to_d \begin{pmatrix} F_l(\cdot) \\ lG_l(\cdot) \\ G_l(\cdot) \end{pmatrix} \tag{14}$$

jointly with (13), where $F_l(r) = F(r) - l\int_0^r \exp(-l(r-s)) F(s)\, ds$ and $G_l(r)$ is defined analogously.

(b) Suppose

$$\begin{pmatrix} T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} h_{Tt}' \\ T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} F_{Tt}h_{Tt}' \\ T^{-2}\sum_{t=2}^{\lfloor T\cdot\rfloor}\left(\sum_{i=1}^{t-1} g_{Ti}\right) h_{Tt}' \end{pmatrix} \to_d \begin{pmatrix} H(\cdot)' \\ \int_0^{\cdot} F(s)\, dH(s)' + \Gamma_{FH}(\cdot) \\ \int_0^{\cdot} G(s)\, dH(s)' + \Gamma_{GH}(\cdot) \end{pmatrix}, \qquad (15)$$

jointly with (13) *where* H, Γ_{FH} *and* Γ_{GH} *are continuous and* H *is a semimartingale. Then*

$$\begin{pmatrix} T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} F_{Tt}(l)\, h_{Tt}' \\ T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} F_{Tt}h_{Tt}(l)' \\ T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} F_{Tt}(l)\, h_{Tt}(l)' \\ T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} \left(g_{Tt} - g_{Tt}(l)\right) h_{Tt}' \\ T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} \left(g_{Tt} - g_{Tt}(l)\right) h_{Tt}(l)' \end{pmatrix} \to_d \begin{pmatrix} \int_0^{\cdot} F_l(s)\, dH(s)' + \Gamma_{FH}(\cdot) \\ \int_0^{\cdot} F(s)\, dH_l(s)' + \Gamma_{FH}(\cdot) \\ \int_0^{\cdot} F_l(s)\, dH_l(s)' + \Gamma_{FH}(\cdot) \\ l\left(\int_0^{\cdot} G_l(s)\, dH(s)' + \Gamma_{GH}(\cdot)\right) \\ l\left(\int_0^{\cdot} G_l(s)\, dH_l(s)' + \Gamma_{GH}(\cdot)\right) \end{pmatrix} \qquad (16)$$

jointly with (13) – (15), *where* $H_l(r) = H(r) - l\int_0^r \exp(-l(r-s))\, H(s)\, ds$.

Proof of Lemma 4. The relation

$$F_{Tt}(l) = F_{Tt} - lT^{-1}\sum_{i=1}^{t-1}\left(1 - T^{-1}l\right)^{t-1-i} F_{Ti}, \quad 0 \le t \le T,$$

can be restated as follows:

$$F_{T,\lfloor Tr\rfloor}(l) = F_{T,\lfloor Tr\rfloor} - l\left(1 - T^{-1}l\right)^{\lfloor Tr\rfloor - 1}\int_0^{\lfloor Tr\rfloor/T}\left(1 - T^{-1}l\right)^{-\lfloor Ts\rfloor} F_{T,\lfloor Ts\rfloor}\, ds, \quad 0 \le r \le 1.$$

Now, $\lim_{T\to\infty}\sup_{0\le r\le 1}\left|\left(1 - T^{-1}l\right)^{\lfloor Tr\rfloor} - \exp(-lr)\right| = 0$ 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,

$$g_{Tt} - g_{Tt}(l) = lG_{T,t-1}(l), \qquad (17)$$

where $G_{Tt} = T^{-1}\sum_{i=1}^{t} g_{Ti}$ and $G_{Tt}(l) = \Delta G_{Tt} + \left(1 - T^{-1}l\right) G_{T,t-1}(l)$ 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, $\max_{1\le t\le T} \|G_{Tt}(l) - G_{T,t-1}(l)\| \to_d 0$ (Billingsley (1999, Theorem 13.4)), so

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

as claimed. Finally, using $G_{T,\lfloor T\cdot\rfloor} \to_d G(\cdot)$, $g_{T,\lfloor T\cdot\rfloor} - g_{T,\lfloor T\cdot\rfloor}(l) \to_d lG_l(\cdot)$ and CMT,

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

The proof of part (a) is completed by noting that the convergence results in the preceding displays hold jointly with (13).

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

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

Next,

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

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

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

Similar reasoning yields

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

The convergence results in the preceding displays hold jointly with (13) – (15). ∎

Proof of Theorems 1-2. The proof proceeds under the assumptions of Theorem 3, strengthening A2 only when necessary. Define Ω and Γ as in Section 3. Let

$$
\Psi_T = \begin{pmatrix} \omega_{yy}^{1/2}\Psi_T^y & 0' \\ 0 & \Omega_{xx}^{1/2}\otimes\Psi_T^x \end{pmatrix},
$$

where $\Psi_T^y = \mathrm{diag}\left(T^{-1/2},\ldots,T^{-(p_y+1/2)}\right)$ and $\Psi_T^x = \mathrm{diag}\left(T^{-1/2},\ldots,T^{-(p_x+1/2)}\right)$. Since $\lim_{T\to\infty}\max_{0\le i\le\max(p_y,p_x)}\sup_{0\le r\le 1}\left|T^{-i}\lfloor Tr\rfloor^i - r^i\right| = 0$ and $\Omega^{1/2} = \Omega_0^{1/2}\breve{\Omega}^{1/2}$, where

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

it follows from Lemma 4 that

$$
\lim_{T\to\infty}\sup_{0\le r\le 1}\left\|T^{1/2}\Psi_T d_{T,\lfloor T\cdot\rfloor}^{\dagger}(l) - \breve{D}_l(r)\right\| = 0, \tag{18}
$$

where $d^{\dagger}_{Tt}(l) = d_{Tt}(l)\,\Omega^{-1/2\prime}$ and

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

and $D^y_l(r)$ 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:

$$T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor} u_t \rightarrow_d \Omega_0^{1/2}\begin{pmatrix} V(\cdot) \\ \breve{W}(\cdot) \end{pmatrix}, \tag{19}$$

$$T^{-1}\sum_{t=2}^{\lfloor T\cdot\rfloor}\left(\sum_{s=1}^{t-1} u_s\right)u_t' \rightarrow_d \Omega_0^{1/2}\int_0^{\cdot}\begin{pmatrix} V(r) \\ \breve{W}(r) \end{pmatrix} d\begin{pmatrix} V(r) \\ \breve{W}(r) \end{pmatrix}'\Omega_0^{1/2\prime} + \Gamma'\int_0^{\cdot} dr, \tag{20}$$

where $\left(V, \breve{W}'\right)'$ is a Wiener processes with covariance matrix $\breve{\Omega}$. By (19), Lemma 4 and the relation $v^y_{Tt} = T^{-1}\lambda\sum_{s=1}^{t-1} u^y_s + u^y_t$, simple algebra yields

$$T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor} v^{\dagger}_{Tt}(l) \rightarrow_d \breve{U}^{\lambda}_l(\cdot) = \breve{\Omega}^{-1/2}\begin{pmatrix} V^{\lambda}_l(\cdot) \\ \breve{W}(\cdot) \end{pmatrix}, \tag{21}$$

where $v^{\dagger}_{Tt}(l) = \Omega^{-1/2}v_{Tt}(l)$ and V^{λ}_l is defined in terms of V as in the text. Similarly, using (19) – (20) and Lemma 4, the following results can be verified:

$$\sum_{t=1}^{T}\left(v^{\dagger}_{Tt}(0) - v^{\dagger}_{Tt}(\bar{\lambda})\right)'\left(v^{\dagger}_{Tt}(0) - v^{\dagger}_{Tt}(\bar{\lambda})\right) \rightarrow_d \bar{\lambda}^2\rho^2_{\#}\int_0^1 V^{\lambda}_{\bar{\lambda}}(r)^2\,dr. \tag{22}$$

$$\sum_{t=1}^{T}\left(v^{\dagger}_{Tt}(0) - v^{\dagger}_{Tt}(\bar{\lambda})\right)' v^{\dagger}_{Tt}(0) \rightarrow_d$$

$$\bar{\lambda}\rho^2_{\#}\left(\int_0^1 V^{\lambda}_{\bar{\lambda}}(r)\,dV^{\lambda}(r) - \int_0^1 V^{\lambda}_{\bar{\lambda}}(r)\,d\breve{W}(r)'\delta + \omega^{-1}_{yy}\gamma_{yy.x}\right), \tag{23}$$

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

The limiting distributions of $P_T\left(\bar{\lambda};\Omega\right)$ 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 $\delta = \|\delta\| = \rho$, as 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,

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

where $\tilde{v}_{Tt}^{\dagger}(l) = v_{Tt}^{\dagger}(l) - d_{Tt}^{\dagger}(l)' \left(\sum_{t=1}^{T} d_{Tt}^{\dagger}(l)\, d_{Tt}^{\dagger}(l)'\right)^{-1} \left(\sum_{t=1}^{T} d_{Tt}^{\dagger}(l)\, v_{Tt}^{\dagger}(l)\right)$. By the algebra of OLS, (18) and (21),

$$\sum_{t=1}^{T} \tilde{v}_{Tt}^{\dagger}(l)' \tilde{v}_{Tt}^{\dagger}(l) - \sum_{t=1}^{T} v_{Tt}^{\dagger}(l)' v_{Tt}^{\dagger}(l)$$

$$= -\left(\Psi_T \sum_{t=1}^{T} d_{Tt}^{\dagger}(l)\, v_{Tt}^{\dagger}(l)\right)' \left(\Psi_T \sum_{t=1}^{T} d_{Tt}^{\dagger}(l)\, d_{Tt}^{\dagger}(l)' \Psi_T'\right)^{-1} \left(\Psi_T \sum_{t=1}^{T} d_{Tt}^{\dagger}(l)\, v_{Tt}^{\dagger}(l)\right)$$

$$\to_d \left(\int_0^1 D_l(r)\, dU_l^{\lambda}(r)\right)' \left(\int_0^1 D_l(r)\, D_l(r)'\, dr\right)^{-1} \left(\int_0^1 D_l(r)\, dU_l^{\lambda}(r)\right)$$

for $l \in \left\{0, \bar{\lambda}\right\}$. Using this along with $(22) - (23)$ and the relation

$$\sum_{t=1}^{T} v_{Tt}^{\dagger}(0)' v_{Tt}^{\dagger}(0) - \sum_{t=1}^{T} v_{Tt}^{\dagger}\left(\bar{\lambda}\right)' v_{Tt}^{\dagger}\left(\bar{\lambda}\right) =$$

$$-\sum_{t=1}^{T} \left(v_{Tt}^{\dagger}(0) - v_{Tt}^{\dagger}\left(\bar{\lambda}\right)\right)' \left(v_{Tt}^{\dagger}(0) - v_{Tt}^{\dagger}\left(\bar{\lambda}\right)\right) + 2\sum_{t=1}^{T} \left(v_{Tt}^{\dagger}(0) - v_{Tt}^{\dagger}\left(\bar{\lambda}\right)\right)' v_{Tt}^{\dagger}(0),$$

it follows that

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

The proof of Theorem 1 is completed by noting that $\gamma_{yy.x} = 0$ when $\left(u_t^y, u_t^{x\prime}\right)' \sim i.i.d.$ $\mathcal{N}(0,\Omega)$.

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

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

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

and $\tilde{V}_{Tt}^{\dagger} = T^{-1} \sum_{s=1}^{t-1} \tilde{v}_{Ts}^{\dagger}(0)$, $d_t^{\dagger} = d_{Tt}^{\dagger}(0)$, $\breve{\Omega}^* = \Omega^{1/2\prime} \Omega^* \Omega^{1/2}$ and $\breve{\Omega}^{**} = \Omega^{1/2\prime} \Omega^{**} \Omega^{1/2}$. When $k = 1$, $\breve{\Omega}^*$ and $\breve{\Omega}^{**}$ coincide with $\bar{\Omega}^*$ and $\bar{\Omega}^{**}$ defined in the text. Theorem 2 (a) now follows from simple algebra and the fact that $T^{-1/2} \sum_{t=1}^{\lfloor T \cdot \rfloor} \tilde{V}_{Tt}^{\dagger} \to_d \tilde{U}^{\lambda}(\cdot)$ under the assumptions of Theorem 2, where $\tilde{U}^{\lambda}$ is defined as in the text (with $R = \rho$).

Under the assumptions of Theorem 2, integrals such as $\int \phi_T(m) f_T(m \mid \lambda, \Omega)\, dm$ can be differentiated with respect to λ by differentiating under the integral sign. As a consequence,

$$\begin{aligned}
\left| \frac{\partial}{\partial \lambda} \int \phi_T(m) f_T(m \mid \lambda, \Omega)\, dm \right|_{\lambda=0} \Bigg| &= \left| \int \phi_T(m) l^{(1)}(m \mid \Omega) f_T(m \mid 0, \Omega)\, dm \right| \\
&\leq \int \left| l^{(1)}(m \mid \Omega) \right| f_T(m \mid 0, \Omega)\, dm \\
&\leq \left(\int l^{(1)}(m \mid \Omega)^2 f_T(m \mid 0, \Omega)\, dm \right)^{1/2},
\end{aligned}$$

where the first inequality uses $|\phi_T| \leq 1$ and the modulus inequality for integrals, while the second inequality uses the Cauchy-Schwarz inequality. Now,

$$\int l^{(1)}(m \mid \Omega) f_T(m \mid 0, \Omega)\, dm = \frac{\partial}{\partial \lambda} \int f_T(m \mid \lambda, \Omega)\, dm \Big|_{\lambda=0} = 0$$

and $l^{(1)}(m_T \mid \Omega)$ differs from $\dot{P}(0; \Omega)$ by an additive constant. As a consequence, $\int l^{(1)}(m \mid \Omega)^2 f_T(m \mid 0, \Omega)\, dm = V_0\left[\dot{P}_T(0; \Omega)\right]$, where $V_0[\cdot]$ denotes the variance under $H_0 : \lambda = 0$. Using the fact that $\{u_t\}$ is Gaussian white noise, it is easy to show that $\lim_{T\to\infty} V_0\left[\dot{P}_T(0; \Omega)\right] = 0$. Therefore, the $\overline{\lim}_{T\to\infty}$ of the left hand side of the preceding display is zero, establishing Theorem 2 (b).

For any T, let $\tilde{\phi}_T^L(m_T; \alpha, \Omega) = 1\left(L_T(\Omega) > \tilde{c}_T^L(\alpha, \Omega)\right)$, where $\tilde{c}_T^L(\alpha, \Omega)$ is such that ϕ_T and $\tilde{\phi}_T^L$ are of the same size. By the Neyman-Pearson lemma and the fact that $l^{(2)}(m_T \mid \Omega) - 2T^{-1} l^{(1)}(m_T \mid \Omega)$ differs from $2L_T(\Omega)$ by an additive constant,

$$\int \left(\phi_T(m) - \tilde{\phi}_T^L(m) \right) \left(l^{(2)}(m \mid \Omega) - 2T^{-1} l^{(1)}(m \mid \Omega) \right) f_T(m \mid 0, \Omega)\, dm \leq 0.$$

Moreover,

$$\begin{aligned} & \frac{\partial^2}{\partial \lambda^2} \int \phi_T(m) f_T(m \mid \lambda, \Omega)\, dm \bigg|_{\lambda=0} \\ = & \int \phi_T(m) \left(l^{(2)}(m \mid \Omega) + l^{(1)}(m \mid \Omega)^2 \right) f_T(m \mid 0, \Omega)\, dm \\ = & \int \phi_T(m) \left(l^{(2)}(m \mid \Omega) - 2T^{-1} l^{(1)}(m \mid \Omega) \right) f_T(m \mid 0, \Omega)\, dm + o(1), \end{aligned}$$

where the second equality uses $\int l^{(1)}(m \mid \Omega)^2 f_T(m \mid 0, \Omega)\, dm = o(1)$. Combining these results, it follows that

$$\overline{\lim}_{T \to \infty} \frac{\partial^2}{\partial \lambda^2} \int \left(\phi_T(m) - \tilde{\phi}_T^L(m; \alpha, \Omega) \right) f_T(m \mid \lambda, \Omega)\, dm \bigg|_{\lambda=0} \leq 0.$$

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

$$\overline{\lim}_{T \to \infty} E_0 \left[\left(\tilde{\phi}_T^L(m_T; \alpha, \Omega) - \phi_T^L(m_T; \alpha, \Omega) \right) \left(l^{(2)}(m_T \mid \Omega) - 2T^{-1} l^{(1)}(m_T \mid \Omega) \right) \right] \leq 0,$$

where $E_0[\cdot]$ denotes expectation under H_0. Now,

$$l^{(2)}(m_T \mid \Omega) - 2T^{-1} l^{(1)}(m_T \mid \Omega) = 2 \left(L_T(\Omega) - E_0[L_T(\Omega)] \right) - V_0 \left[\dot{P}_T(0; \Omega) \right].$$

To see this, notice that the additive constant by which $l^{(2)}(m_T \mid \Omega) - 2T^{-1} l^{(1)}(m_T \mid \Omega)$ differs from $2L_T(\Omega)$ can be computed by subtracting expected values from both expressions, rearranging and using $E_0 \left[l^{(1)}(m_T \mid \Omega) \right] = 0$ and $E_0 \left[l^{(2)}(m_T \mid \Omega) \right] = -V_0 \left[\dot{P}_T(0; \Omega) \right]$. Since $\{\phi_T\}$ is asymptotically of level α, it can be shown (using Theorem 2 (a)) that $\underline{\lim}_{T \to \infty} \tilde{c}_T^L(\alpha, \Omega) \geq c^L(\alpha, \rho^2)$. Moreover, $\lim_{T \to \infty} V_0 \left(\dot{P}(0; \Omega) \right) = 0$. Therefore,

$$
\begin{aligned}
& \overline{\lim}_{T\to\infty} E_0 \left[\left(\tilde{\phi}_T^L (m_T; \alpha, \Omega) - \phi_T^L (m_T; \alpha, \Omega) \right) \left(l^{(2)} (m_T \mid \Omega) - 2T^{-1} l^{(1)} (m_T \mid \Omega) \right) \right] \\
= \; & -2 \underline{\lim}_{T\to\infty} E_0 \left[1 \left(c^L \left(\alpha, \rho^2 \right) < L_T (\Omega) \leq \tilde{c}_T^L (\alpha, \Omega) \right) \left(L_T (\Omega) - E_0 \left[L_T (\Omega) \right] \right) \right]
\end{aligned}
$$

and it suffices to show that

$$
\underline{\lim}_{T\to\infty} E_0 \left[1 \left(c^L \left(\alpha, \rho^2 \right) < L_T (\Omega) \leq \tilde{c}_T^L (\alpha, \Omega) \right) \left(L_T (\Omega) - E_0 \left[L_T (\Omega) \right] \right) \right] \geq 0.
$$

If $\{\phi_T\}$ is asymptotically of size α, $\lim_{T\to\infty} \tilde{c}_T^L (\alpha, \Omega) = c^L \left(\alpha, \rho^2 \right)$, so the left hand side equals zero since $1 \left(c^L \left(\alpha, \rho^2 \right) < L_T (\Omega) \leq \tilde{c}_T^L (\alpha, \Omega) \right) \to_p 0$ and $\{ L_T (\Omega) - E_0 [L_T (\Omega)] \}$ is uniformly integrable under H_0. Otherwise, if $\alpha \leq \Pr \left[\varphi_L \left(0; \rho^2 \right) > E \varphi_L \left(0; \rho^2 \right) \right]$, then $\lim_{T\to\infty} E_0 [L_T (\Omega)] = E \varphi_L \left(0; \rho^2 \right) \leq c^L \left(\alpha, \rho^2 \right)$, so

$$
\begin{aligned}
& \underline{\lim}_{T\to\infty} E_0 \left[1 \left(c^L \left(\alpha, \rho^2 \right) < L_T (\Omega) \leq \tilde{c}_T^L (\alpha, \Omega) \right) \left(L_T (\Omega) - E_0 \left[L_T (\Omega) \right] \right) \right] \\
= \; & \underline{\lim}_{T\to\infty} E_0 \left[1 \left(c^L \left(\alpha, \rho^2 \right) < L_T (\Omega) \leq \tilde{c}_T^L (\alpha, \Omega) \right) \left(L_T (\Omega) - E \varphi_L \left(0; \rho^2 \right) \right) \right] \\
& + \underline{\lim}_{T\to\infty} E_0 \left[1 \left(c^L \left(\alpha, \rho^2 \right) < L_T (\Omega) \leq \tilde{c}_T^L (\alpha, \Omega) \right) \left(E \varphi_L \left(0; \rho^2 \right) - E_0 \left[L_T (\Omega) \right] \right) \right] \\
\geq \; & \underline{\lim}_{T\to\infty} E_0 \left[1 \left(c^L \left(\alpha, \rho^2 \right) < L_T (\Omega) \leq \tilde{c}_T^L (\alpha, \Omega) \right) \left(E \varphi_L \left(0; \rho^2 \right) - E_0 \left[L_T (\Omega) \right] \right) \right] \\
= \; & 0,
\end{aligned}
$$

as was to be shown. ∎

Proof of Theorem 3. All derivations in the proof of Theorems 1 and 2 (a) remain valid if Ω and Γ are replaced with consistent estimators. The proof of Theorem 3 can therefore be completed by showing that $\hat{\Omega}_T \to_p \Omega$ and $\hat{\Gamma}_T \to_p \Gamma$ under A1-A3.

Let $u_t^{PW} = u_t - A u_{t-1}$, where A is the matrix appearing in the statement of A3(i). The equations defining $\hat{\Gamma}_T$ and $\hat{\Omega}_T$ are sample counterparts of the relations

$$
\begin{aligned}
\Gamma &= (I - A)^{-1} \Gamma^{PW} (I - A')^{-1} + (I - A)^{-1} A \Sigma - (I - A)^{-1} \Sigma^* A' (I - A')^{-1}, \\
\Omega &= (I - A)^{-1} \Omega^{PW} (I - A')^{-1},
\end{aligned}
$$

where

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

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

Since $\left(I - \hat{A}_T\right)^{-1} \to_p (I-A)^{-1}$ and $\hat{A}_T \to_p A$ under A3(i), it therefore suffices to show that $\hat{\Gamma}_T^{PW} \to_p \Gamma^{PW}$, $\hat{\Sigma}_T \to_p \Sigma$, $\hat{\Sigma}_T^* \to_p \Sigma^*$ and $\hat{\Omega}_T^{PW} \to_p \Omega^{PW}$.

Let $\hat{v}_{Tt}^{\dagger} = u_t - d_t'\left(\hat{\beta}_T - \beta\right)$, $\hat{v}_{Tt}^{\dagger\dagger} = \hat{v}_{Tt} - \hat{v}_{Tt}^{\dagger}$, $\hat{v}_{Tt}^{PW,\dagger} = \hat{v}_{Tt}^{\dagger} - \hat{A}_T \hat{v}_{T,t-1}^{\dagger}$ and $\hat{v}_{Tt}^{PW,\dagger\dagger} = \hat{v}_{Tt}^{PW} - \hat{v}_{Tt}^{PW,\dagger} = \hat{v}_{Tt}^{\dagger\dagger} - \hat{A}_T \hat{v}_{T,t-1}^{\dagger\dagger}$. Using notation typified by

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

$\hat{\Gamma}_T^{PW}$ can be written as $\hat{\Gamma}_T^{PW,\dagger,\dagger} + \hat{\Gamma}_T^{PW,\dagger\dagger,\dagger} + \hat{\Gamma}_T^{PW,\dagger,\dagger\dagger} + \hat{\Gamma}_T^{PW,\dagger\dagger,\dagger\dagger}$. Now, $\hat{\Gamma}_T^{PW,\dagger,\dagger} \to_p \Gamma^{PW}$ by Corollary 3 of Jansson (2001). The proof of $\hat{\Gamma}_T^{PW} \to_p \Gamma^{PW}$ is completed by using the relation

$$\hat{v}_{Tt}^{\dagger\dagger} = v_{Tt} - u_t = T^{-1}\left(\frac{\lambda \sum_{s=1}^{t-1} u_s^y}{0}\right)$$

and straightforward, but tedious, bounding arguments to show that $\hat{\Gamma}_T^{PW,\dagger\dagger,\dagger}$, $\hat{\Gamma}_T^{PW,\dagger,\dagger\dagger}$ and $\hat{\Gamma}_T^{PW,\dagger\dagger,\dagger\dagger}$ are $o_p(1)$. Indeed, the proof of Lemma 6 of Jansson and Haldrup (2001) carries over to the present case. The details are omitted for brevity.

Proceeding in analogous fashion, it can be shown that $\hat{\Sigma}_T \to_p \Sigma$, $\hat{\Sigma}_T^* \to_p \Sigma^*$ and $\hat{\Omega}_T^{PW} \to_p \Omega^{PW}$. Once more, the details are omitted to conserve space. ∎

REFERENCES

ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

ANDREWS, D. W. K., AND J. C. MONAHAN (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953–966.

BASAWA, I. V., AND D. J. SCOTT (1983): *Asymptotic Optimal Inference for Non-Ergodic Models*. Springer-Verlag, New York.

BILLINGSLEY, P. (1999): *Convergence Of Probability Measures*. 2nd edn., New York: 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.

CULVER, S. E., AND D. H. PAPELL (1999): "Long-Run Purchasing Power Parity with Short-Run Data: Evidence with a Null Hypothesis of Stationarity," *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., AND M. JANSSON (2000): "Testing for Unit Roots with Stationary Covariates," Department of Economics Working Paper 2000-08, University of California, San Diego.

ENGLE, R. F., AND C. W. J. GRANGER (1987): "Cointegration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251–276.

FROOT, K. A., AND K. ROGOFF (1995): "Perspectives on PPP and Long-Run Real Exchange Rates," in *Handbook of International Economics, Volume III*, ed. by G. Grossman, and K. Rogoff. New York: North Holland, 1647-1688.

HANSEN, B. E. (1992): "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes," *Econometric Theory*, 8, 489–500.

——— (1995): "Rethinking the Univariate Approach to Unit Root Testing: Using Covariates to Increase Power," *Econometric Theory*, 11, 1148–1171.

JANSSON, M. (2001): "Consistent Covariance Matrix Estimation for Linear Processes," Manuscript, UC Berkeley.

JANSSON, M., AND N. HALDRUP (2001): "Regression Theory for Nearly Cointegrated Time Series," Manuscript, UC Berkeley.

KWIATKOWSKI, D., P. C. B. PHILLIPS, P. SCHMIDT, AND 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., AND A. MCWORTHER (1978): "An Exact Test for the Presence of Random Walk Coefficiants in a Linear Regession Model," *Journal of the American Statistical Association*, 73, 816–820.

NABEYA, S., AND 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., AND 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.

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.

——— (1991): "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283–306.

PHILLIPS, P. C. B., AND 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., AND 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.

——— (1992): "Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation," *Econometric Theory*, 8, 1–27.

SAIKKONEN, P., AND R. LUUKKONEN (1993a): "Point Optimal Tests for Testing the Order of Differencing in ARIMA Models," *Econometric Theory*, 9, 343–362.

——— (1993b): "Testing for a Moving Average Unit Root in Autoregressive Integrated Moving Average Models," *Journal of the American Statistical Association*, 88, 596–601.

STOCK, J. H. (1994): "Unit Roots, Structural Breaks and Trends," in *Handbook of Econometrics, Volume IV*, ed. by R. F. Engle, and D. L. McFadden. New York: North Holland, 2739-2841.

STOCK, J. H., AND M. W. WATSON (1993): "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61, 783–820.

——— (1998): "Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model," *Journal of the American Statistical Association*, 93, 349–358.

7. TABLES

TABLE 1a

PERCENTILES OF L_T AND $Q_T(\bar{\lambda})$
CONSTANT MEAN IN y ($p_y = 0$), CONSTANT MEAN IN x ($p_x = 0$)

	L_T				$Q_T(7)$			
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	1.633
0.3	0.404	0.571	0.725	0.940	-1.887	-0.694	0.346	1.962
0.4	0.444	0.621	0.797	1.059	-1.989	-0.731	0.460	2.049
0.5	0.493	0.701	0.924	1.216	-2.143	-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	-9.332	-7.054	-4.650	-1.805

TABLE 1b

PERCENTILES OF L_T AND $Q_T(\bar{\lambda})$
CONSTANT MEAN IN y ($p_y = 0$), LINEAR TREND IN x ($p_x = 1$)

	L_T				$Q_T(7)$			
ρ^2	90%	95%	97.5%	99%	90%	95%	97.5%	99%
0	0.349	0.464	0.588	0.786	-1.944	-0.912	0.075	1.349
0.1	0.332	0.444	0.560	0.713	-1.914	-0.890	0.050	1.438
0.2	0.327	0.440	0.555	0.705	-1.820	-0.776	0.360	1.784
0.3	0.309	0.420	0.526	0.680	-1.818	-0.687	0.447	1.920
0.4	0.292	0.398	0.510	0.658	-1.835	-0.584	0.511	1.826
0.5	0.270	0.362	0.467	0.604	-1.900	-0.582	0.592	2.244
0.6	0.231	0.322	0.412	0.539	-2.119	-0.659	0.641	2.331
0.7	0.178	0.258	0.336	0.443	-2.581	-0.925	0.502	2.347
0.8	0.084	0.151	0.216	0.305	-3.754	-1.762	0.025	2.137
0.9	-0.104	-0.042	0.024	0.107	-7.592	-4.818	-2.582	0.264

TABLE 1c

PERCENTILES OF L_T AND $Q_T(\bar{\lambda})$
LINEAR TREND IN y $(p_y = 1)$, CONSTANT MEAN IN x $(p_x = 0)$

	L_T				$Q_T(12)$			
ρ^2	90%	95%	97.5%	99%	90%	95%	97.5%	99%
0	0.120	0.148	0.178	0.215	-4.999	-3.981	-2.937	-1.565
0.1	0.114	0.142	0.171	0.209	-5.099	-3.922	-2.875	-1.550
0.2	0.104	0.132	0.160	0.200	-5.261	-4.109	-2.925	-1.554
0.3	0.095	0.124	0.153	0.194	-5.458	-4.072	-2.899	-1.346
0.4	0.082	0.110	0.139	0.180	-5.980	-4.522	-3.111	-1.328
0.5	0.066	0.094	0.121	0.165	-6.675	-4.981	-3.543	-1.749
0.6	0.047	0.078	0.106	0.143	-7.317	-5.778	-4.201	-2.181
0.7	0.015	0.043	0.073	0.113	-9.331	-7.382	-5.527	-3.168
0.8	-0.038	-0.010	0.018	0.054	-13.820	-10.943	-8.628	-5.362
0.9	-0.182	-0.138	-0.103	-0.061	-26.956	-22.307	-18.686	-14.090

TABLE 1d

PERCENTILES OF L_T AND $Q_T(\bar{\lambda})$
LINEAR TREND IN y $(p_y = 1)$, LINEAR TREND IN x $(p_x = 1)$

	L_T				$Q_T(12)$			
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	0.116	0.149	0.185	0.236	-5.517	-4.085	-2.686	-1.115
0.4	0.116	0.153	0.197	0.251	-5.800	-4.224	-2.975	-1.377
0.5	0.112	0.157	0.207	0.273	-6.106	-4.491	-3.121	-1.141
0.6	0.114	0.170	0.222	0.297	-6.996	-5.180	-3.410	-1.209
0.7	0.115	0.183	0.258	0.358	-8.546	-6.306	-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	0.869	-23.141	-18.961	-15.393	-11.340

TABLE 2a

MONTE CARLO REJECTION RATES: AR(1) MODEL
5% LEVEL TESTS, CONSTANT MEAN ($p_y = p_x = 0$), $T = 200$

a	λ	L_T^y	Q_T^y	$L_T, \rho^2 =$ 0	0.2	0.5	0.8	$Q_T, \rho^2 =$ 0	0.2	0.5	0.8
	0	6.0	2.1	5.9	5.9	6.6	8.1	2.0	1.8	1.6	1.0
	5	31.1	22.6	30.8	35.9	43.5	59.7	22.4	28.3	39.9	63.6
-0.8	10	58.7	55.8	58.5	64.9	70.2	79.8	55.0	66.8	81.5	95.8
	15	75.2	77.4	74.8	78.1	81.0	87.2	76.4	85.5	94.6	99.5
	20	83.4	88.6	83.1	85.4	87.3	91.0	88.2	92.9	97.8	99.9
	0	5.1	3.9	5.1	5.4	6.0	5.1	4.2	1.8	3.5	2.4
	5	31.0	28.6	30.9	34.7	42.6	59.7	27.8	28.3	51.7	81.7
-0.5	10	30.2	64.1	60.0	63.2	69.3	79.7	63.7	33.8	87.6	99.1
	15	75.9	82.3	75.7	77.2	82.2	88.2	81.7	86.5	96.5	100.0
	20	84.2	91.1	84.0	85.7	88.5	92.8	91.0	92.9	99.1	100.0
	0	5.0	4.5	5.0	5.4	4.7	4.3	4.4	4.6	4.8	4.3
	5	31.9	31.7	31.2	34.0	42.0	58.1	31.2	37.6	53.9	84.3
-0.2	10	59.7	65.1	59.2	62.0	68.8	78.3	64.5	73.1	88.8	99.2
	15	75.4	82.7	75.0	76.9	80.9	86.9	82.3	88.7	97.4	99.9
	20	83.6	91.0	83.5	84.9	87.6	91.0	90.5	96.0	99.3	100.0
	0	5.1	5.0	5.3	4.5	4.3	4.0	5.1	4.2	4.2	4.3
	5	30.7	30.1	30.2	32.7	41.1	55.4	29.2	36.5	52.1	83.0
0.2	10	59.1	64.1	53.4	59.5	65.7	72.5	63.0	71.4	86.2	98.5
	15	73.6	80.9	72.8	73.9	76.9	77.7	80.3	87.2	95.4	99.6
	20	81.1	89.3	80.7	81.0	81.5	78.7	88.6	95.4	98.2	98.9
	0	4.8	3.9	4.7	4.7	5.2	4.9	4.0	3.5	3.8	4.4
	5	28.3	26.2	27.8	31.8	39.1	55.4	25.1	32.5	48.3	77.6
0.5	10	53.2	56.4	52.3	56.1	62.2	71.1	54.9	65.0	81.4	96.5
	15	64.9	72.8	63.8	67.2	70.4	73.9	70.1	79.5	90.3	96.0
	20	69.5	79.2	68.2	69.9	71.3	71.1	75.4	81.6	88.2	88.1
	0	3.6	1.3	3.8	3.7	4.1	6.1	1.5	1.6	1.5	2.4
	5	18.0	6.6	17.0	20.3	28.7	49.9	6.3	9.6	19.0	45.8
0.8	10	24.6	6.3	22.2	27.0	36.2	56.8	5.7	8.8	18.6	49.9
	15	16.3	2.7	15.1	18.0	26.3	50.3	2.5	5.8	7.3	27.8
	20	11.2	2.5	10.2	12.3	19.3	40.3	2.0	2.5	3.8	13.0

TABLE 2b

MONTE CARLO REJECTION RATES: MA (1) MODEL
5% LEVEL TESTS, CONSTANT MEAN $(p_y = p_x = 0)$, $T = 200$

		L_T^y	Q_T^y	L_T, $\rho^2 =$				Q_T, $\rho^2 =$			
δ	λ			0	0.2	0.5	0.8	0	0.2	0.5	0.8
-0.8	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	5	0.1	0.0	0.1	0.4	1.3	5.4	0.0	0.0	0.0	0.3
	10	6.5	1.5	6.4	8.3	13.8	28.8	1.4	2.6	4.8	13.3
	15	19.2	10.1	19.2	26.3	32.2	47.3	9.3	13.2	20.0	40.3
	20	32.5	22.6	32.2	36.9	45.6	58.7	22.0	27.6	39.7	63.7
-0.5	0	0.7	0.1	0.7	0.6	0.6	0.7	0.2	0.0	0.0	0.0
	5	13.3	8.0	13.4	16.5	22.3	39.8	7.7	10.3	14.7	31.1
	10	40.9	36.5	40.7	45.0	53.2	66.7	35.9	43.5	59.1	84.8
	15	60.3	60.8	59.7	62.8	69.7	78.4	60.1	67.7	82.5	97.1
	20	71.9	76.9	71.6	73.7	79.0	85.2	76.3	81.9	92.7	99.4
-0.2	0	3.7	3.1	3.8	3.7	3.8	3.2	3.3	3.0	2.6	2.2
	5	28.8	26.9	28.5	31.4	40.0	55.4	26.2	32.0	48.2	78.2
	10	56.9	61.0	56.6	60.4	68.2	76.2	60.7	69.9	86.0	98.8
	15	73.4	80.6	73.0	75.9	81.2	85.8	79.9	87.2	96.6	99.9
	20	82.4	89.4	82.0	83.7	87.7	90.1	89.0	94.2	99.0	100.0
0	0	5.2	5.0	5.3	5.0	4.5	4.2	5.5	4.5	4.1	4.3
	5	31.4	31.0	30.9	34.3	40.3	56.4	30.2	38.2	64.1	84.0
	10	60.3	65.9	59.7	62.4	67.5	75.9	64.7	73.8	88.6	99.1
	15	75.1	82.5	74.8	76.4	79.4	83.2	82.0	88.6	96.7	99.9
	20	83.1	90.9	82.3	84.4	85.4	86.7	90.3	94.6	98.3	100.0
0.2	0	4.3	3.7	4.2	3.7	4.5	4.5	3.7	3.1	3.3	3.6
	5	28.1	26.7	28.1	30.6	38.3	53.2	26.5	32.1	48.5	79.9
	10	53.1	60.6	55.6	58.5	65.4	74.3	59.8	69.1	85.7	98.5
	15	70.4	78.7	69.9	72.5	76.6	80.6	77.7	85.8	95.0	99.7
	20	78.9	87.6	78.0	79.3	81.7	82.2	87.0	92.5	98.2	99.4
0.5	0	1.7	0.9	1.7	2.0	2.8	5.8	0.8	0.8	1.0	2.1
	5	19.1	14.6	19.0	24.1	36.1	58.4	14.4	19.6	36.5	76.0
	10	46.4	44.5	44.8	51.7	62.5	76.8	43.5	55.6	77.1	98.1
	15	60.9	65.8	60.6	65.3	73.7	82.0	64.2	75.6	91.8	99.8
	20	68.4	77.4	67.7	72.1	76.9	83.8	75.8	84.9	96.9	99.4
0.8	0	1.0	0.3	1.1	1.2	2.9	6.9	0.4	0.3	0.6	2.0
	5	14.7	8.8	14.4	20.7	34.7	57.5	8.4	14.5	31.5	75.6
	10	41.4	36.6	40.2	46.8	59.5	75.9	35.4	46.8	72.9	98.1
	15	56.2	56.6	55.3	60.1	70.7	81.5	54.3	67.5	88.2	99.7
	20	63.5	67.0	62.5	66.9	76.0	84.3	63.8	77.3	93.5	99.7

TABLE 2c

MONTE CARLO REJECTION RATES: AR(1) MODEL
5% LEVEL TESTS, LINEAR TREND ($p_y = p_x = 1$), $T = 200$

		L_T^y	Q_T^y	L_T, $\rho^2 =$				Q_T, $\rho^2 =$			
α	λ			0	0.2	0.5	0.8	0	0.2	0.5	0.8
	0	7.4	0.5	7.4	7.7	9.5	11.9	0.4	0.3	0.1	0.0
	5	16.1	1.9	16.0	18.9	24.8	37.5	1.8	2.9	4.7	6.7
-0.8	10	39.9	13.2	39.3	43.3	52.8	66.3	13.0	19.9	33.5	57.7
	15	59.9	36.0	59.1	65.1	73.3	83.1	35.0	46.1	66.8	90.0
	20	73.9	57.1	73.5	78.7	85.0	91.0	55.7	68.8	85.3	97.7
	0	5.6	2.4	5.7	5.6	5.5	6.1	2.5	2.1	1.7	0.6
	5	14.4	8.0	14.2	15.7	20.7	33.9	7.5	10.7	18.6	36.6
-0.5	10	35.3	28.4	34.8	41.0	51.1	64.4	27.5	40.1	61.3	91.5
	15	57.3	54.1	56.8	63.7	71.4	80.7	53.2	68.8	86.2	99.2
	20	72.6	73.2	72.3	77.5	83.0	89.8	72.0	84.8	95.6	99.9
	0	5.5	4.1	5.5	5.2	4.9	4.9	4.0	3.4	3.1	3.0
	5	14.0	10.7	13.7	14.7	19.9	32.0	10.5	14.9	26.0	53.8
-0.2	10	36.3	33.9	36.0	40.0	48.8	62.5	33.0	45.1	69.6	96.1
	15	58.4	60.1	57.9	61.9	70.3	79.8	59.0	72.4	90.5	99.7
	20	73.0	77.4	72.5	76.5	82.1	89.2	76.3	86.9	97.2	100.0
	0	4.7	4.4	4.5	4.6	5.3	5.0	4.1	3.9	4.6	3.8
	5	12.8	10.6	12.7	13.4	20.3	31.5	10.1	16.1	29.3	52.9
0.2	10	32.5	33.3	32.1	36.4	47.1	58.0	32.3	44.5	69.0	94.9
	15	53.0	56.3	52.0	56.9	65.6	71.5	54.4	68.4	89.2	99.4
	20	65.4	72.0	64.7	69.2	76.0	78.3	70.5	82.1	95.9	99.9
	0	4.4	3.2	4.4	4.2	5.4	5.2	3.6	3.2	3.3	3.2
	5	11.1	8.6	11.2	11.8	17.2	30.3	8.7	12.6	21.7	42.6
0.5	10	27.7	25.5	26.9	30.8	38.9	53.6	25.1	35.1	56.7	86.5
	15	45.3	44.2	42.8	47.8	54.3	65.3	41.6	54.5	77.0	96.1
	20	55.3	58.0	53.5	56.9	61.3	71.1	54.0	66.0	84.0	96.0
	0	2.9	0.7	2.8	3.7	4.2	7.5	0.7	0.7	0.9	0.9
	5	6.3	1.2	6.2	8.4	16.5	29.4	1.2	2.1	4.1	11.0
0.8	10	11.8	1.6	11.3	16.4	28.5	45.3	1.8	2.5	5.8	22.0
	15	11.2	0.8	11.0	13.9	21.6	44.2	1.3	1.2	2.5	12.9
	20	7.0	0.3	7.5	9.1	15.0	35.8	0.6	0.4	0.7	3.4

TABLE 2d

MONTE CARLO REJECTION RATES: MA(1) MODEL
5% LEVEL TESTS, LINEAR TREND ($\rho_y = \rho_x = 1$), $T = 200$

		L_T^y	Q_T^y	L_T, $\rho^2 =$				Q_T, $\rho^2 =$			
δ	λ			0	0.2	0.5	0.8	0	0.2	0.5	0.8
	0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.0
	5	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	1.8	0.0
-0.8	10	0.1	0.0	0.1	0.2	1.2	5.1	0.0	0.0	12.8	0.0
	15	1.1	0.1	1.1	2.2	5.7	16.5	0.0	0.0	32.0	1.1
	20	5.2	0.3	5.0	7.6	47.4	29.8	0.3	0.7	47.8	8.3
	0	0.4	0.0	0.4	5.2	0.6	0.7	0.0	0.0	0.0	0.0
	5	2.2	0.2	2.2	3.0	5.5	12.0	0.3	0.5	0.7	1.4
-0.5	10	13.1	5.1	12.9	13.1	25.4	41.2	4.8	7.2	15.3	36.7
	15	31.7	20.5	31.5	35.7	47.3	62.3	19.7	26.2	45.5	79.0
	20	49.1	39.9	48.7	52.3	65.0	77.3	39.2	50.2	72.0	95.5
	0	3.7	2.5	3.7	3.4	3.9	3.3	2.3	2.1	1.8	1.1
	5	10.8	7.4	10.7	12.5	16.2	28.0	7.0	11.0	18.6	38.3
-0.2	10	30.6	27.4	30.1	35.3	44.5	58.9	26.7	38.0	61.1	92.3
	15	52.3	52.5	51.4	57.8	66.4	77.5	51.0	65.9	87.2	99.4
	20	68.4	73.0	67.7	72.6	80.1	87.0	71.9	82.2	96.0	100.0
	0	4.7	4.0	4.9	5.2	4.9	4.5	4.1	4.3	3.8	3.8
	5	12.7	10.7	12.5	14.9	20.5	31.0	10.3	16.1	27.6	56.1
0	10	34.1	34.2	33.3	38.5	47.9	59.9	33.2	46.6	69.5	96.5
	15	56.0	60.0	55.1	59.6	67.5	75.6	58.5	71.5	89.8	99.7
	20	70.8	77.4	70.1	73.6	78.3	84.3	75.8	86.1	96.7	100.0
	0	3.4	2.9	3.5	4.0	4.2	4.7	2.8	2.6	2.3	2.5
	5	9.7	7.9	9.6	13.7	17.7	30.6	7.7	13.9	23.0	46.0
0.2	10	29.9	28.8	29.6	33.3	44.2	57.5	27.5	41.3	62.5	92.9
	15	50.6	52.8	50.1	55.4	63.2	71.7	51.8	66.9	85.6	99.2
	20	65.6	71.0	65.0	69.1	74.6	79.8	69.9	82.0	94.1	99.9
	0	0.9	0.6	1.0	1.7	3.5	6.8	0.8	0.7	0.6	1.6
	5	4.0	1.7	3.6	6.9	14.1	32.5	1.7	4.6	9.9	38.6
0.5	10	17.2	11.9	16.7	23.7	38.4	59.2	11.5	20.9	44.2	90.2
	15	34.2	29.9	33.4	40.9	56.2	73.8	28.8	43.0	71.9	98.8
	20	48.0	46.5	45.5	54.8	67.7	81.0	44.5	61.4	87.2	99.8
	0	0.4	0.0	0.5	1.0	2.8	6.5	0.1	0.1	0.4	1.6
	5	1.9	0.4	2.0	4.3	13.3	33.0	0.5	1.8	6.9	36.7
0.8	10	11.3	4.9	11.0	19.3	35.7	58.9	4.8	12.8	37.1	90.5
	15	25.1	16.1	24.2	36.3	53.9	74.3	15.3	32.0	65.9	98.9
	20	37.4	29.0	36.3	47.4	65.3	82.4	27.5	47.8	81.6	99.8

TABLE 3

TESTS OF LONG-RUN PPP

	Univariate Tests		Using Covariates		
Price Level	L_T^y	$Q_T^y(12)$	L_T	$Q_T(12)$	$\hat{\rho}_T^2$
CPI	0.017	-10.486	0.246*	-15.219	0.834
WPI	0.017	-9.758	0.362**	-4.795*	0.601

Notes: Rejection at the 5% (1%) level is indicated by * (**).

8. FIGURES

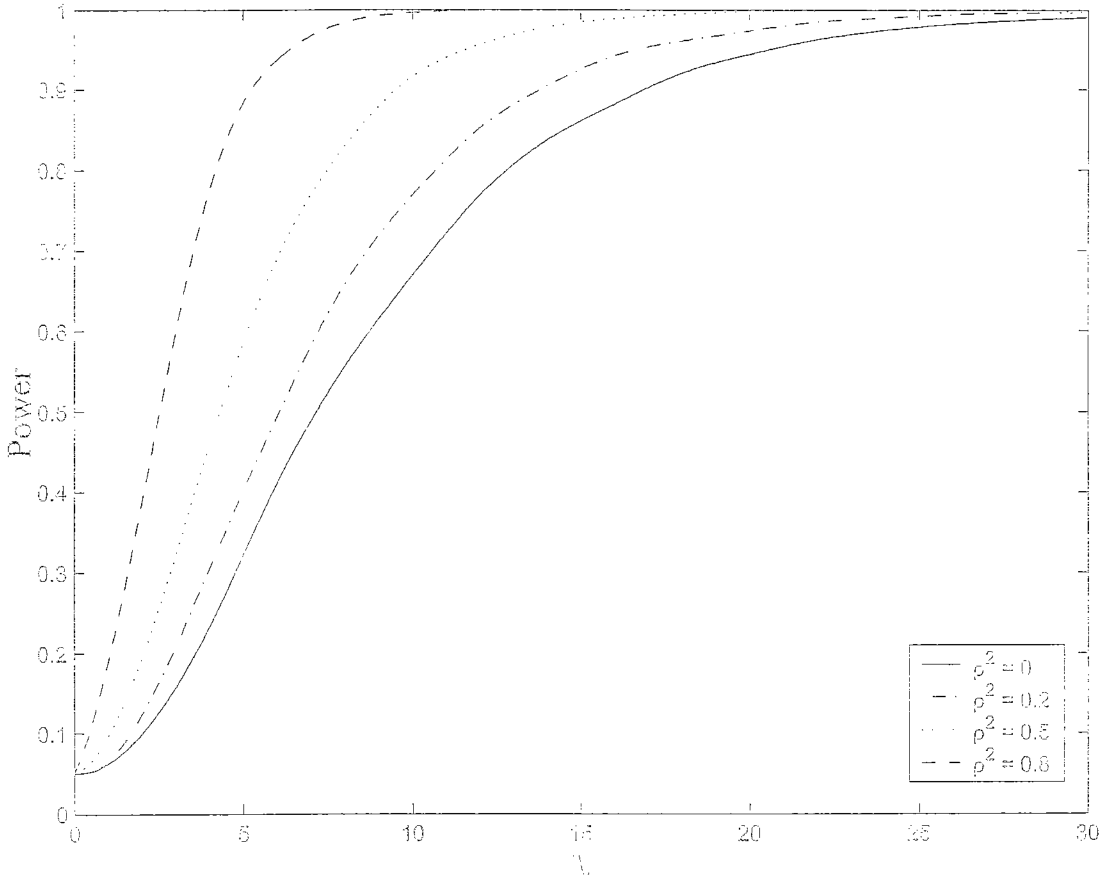

FIGURE 1: POWER ENVELOPES

5% LEVEL TESTS, CONSTANT MEAN ($p_y = p_x = 0$).

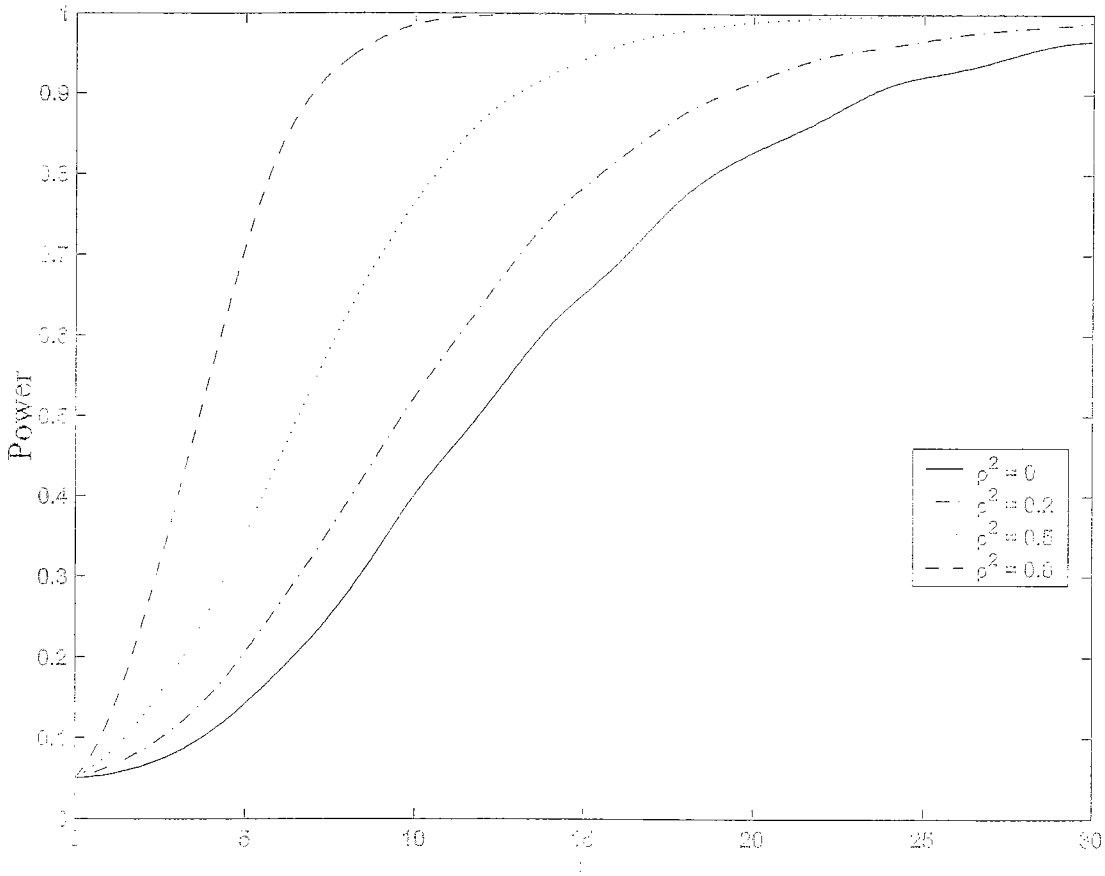

FIGURE 2: POWER ENVELOPES

5% LEVEL TESTS, LINEAR TREND $(p_y = p_x = 1)$.

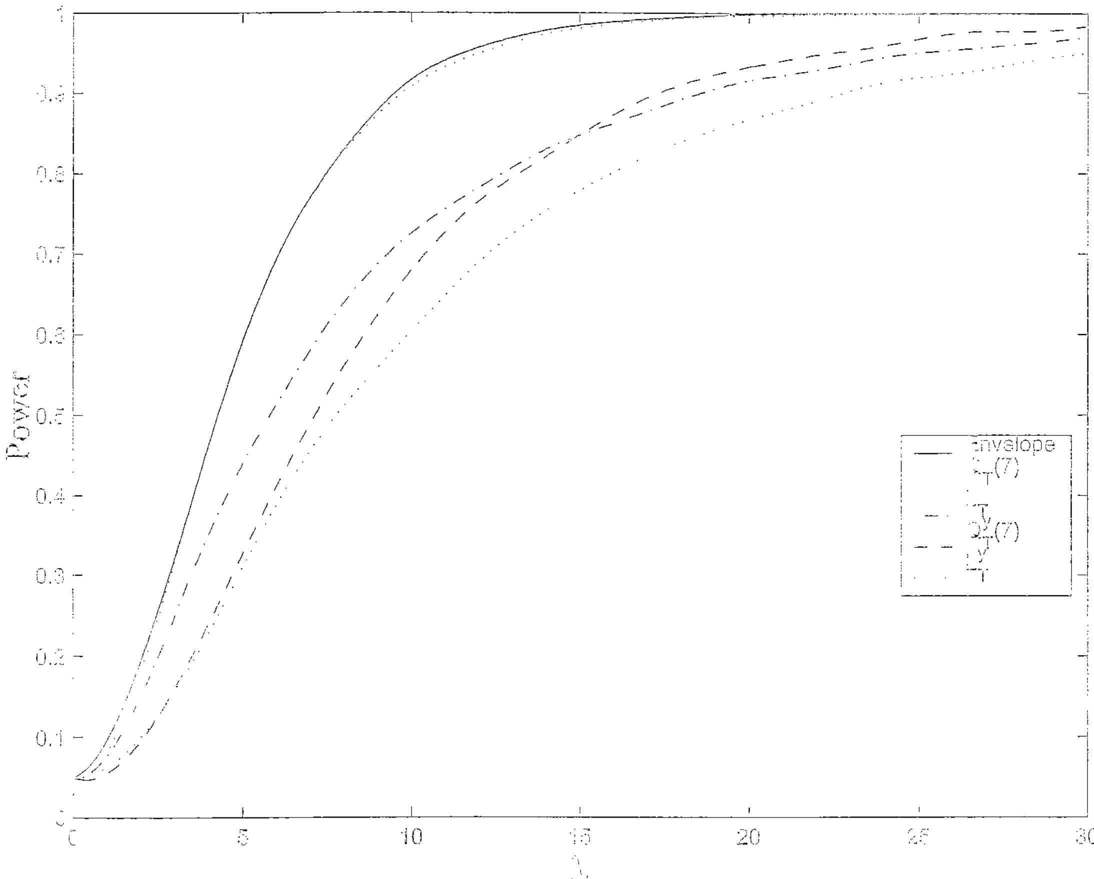

FIGURE 8: POWER CURVES, $\rho^2 = 0.5$

5% LEVEL TESTS, CONSTANT MEAN $(p_y = p_x = 0)$.

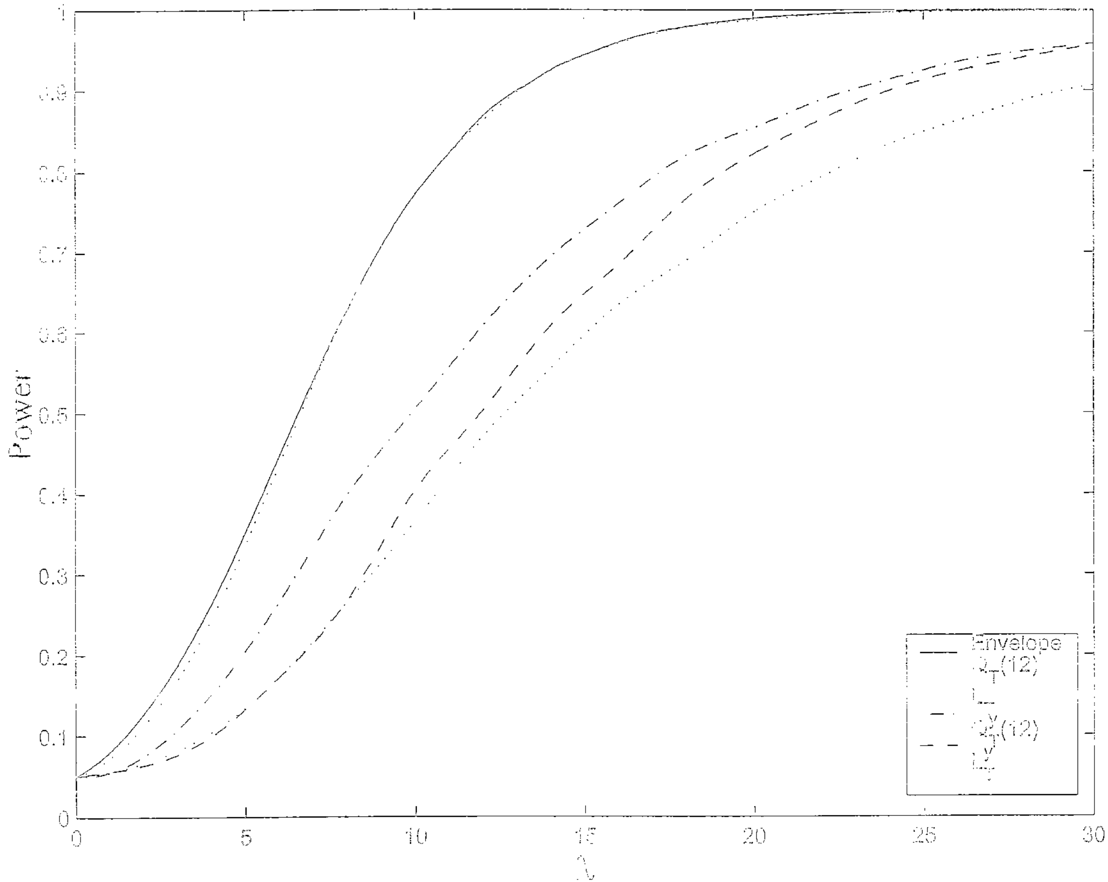

FIGURE 4: POWER CURVES, $\rho^2 = 0.5$

5% LEVEL TESTS, LINEAR TREND $(p_y = p_x = 1)$.