Econometric Theory, **11**, 1995, 1131–1147. Printed in the United States of America.

INFERENCE IN MODELS WITH NEARLY INTEGRATED REGRESSORS

CHRISTOPHER L. CAVANAGH
Columbia University

GRAHAM ELLIOTT
University of California, San Diego

JAMES H. STOCK
Kennedy School of Government
Harvard University
and
National Bureau of Economic Research

This paper examines regression tests of whether x forecasts y when the largest autoregressive root of the regressor is unknown. It is shown that previously proposed two-step procedures, with first stages that consistently classify x as I(1) or I(0), exhibit large size distortions when regressors have local-to-unit roots, because of asymptotic dependence on a nuisance parameter that cannot be estimated consistently. Several alternative procedures, based on Bonferroni and Scheffe methods, are therefore proposed and investigated. For many parameter values, the power loss from using these conservative tests is small.

1. INTRODUCTION

In a bivariate model, the asymptotic null distribution of the F-statistic testing whether x is a useful predictor of y depends on whether the largest autoregressive root α of the regressor is 1 or less than 1. The application that motivates this paper is a special case of the general Granger causality testing problem, tests of the linear rational expectations hypothesis in finance. Examples include tests of the predictability of stock returns using lagged information—for example, the lagged dividend yield or, alternatively, the lagged slope of the term structure. A large body of research (see Campbell and Shiller, 1988; for a review, see Fama, 1991) finds significant predictive content in such relations using conventional critical values. However, with regressors such as the dividend yield, there is reason to suspect a large, pos-

The authors thank Jean-Marie Dufour, Mark Watson, and three anonymous referees for helpful comments. The research was supported in part by National Science Foundation grant SES-91-22463. Address correspondence to: Professor James Stock, Kennedy School, Harvard University, Cambridge, MA 02138, USA.

 0266-4666/95 $9.00 + .10

sibly unit, autoregressive root. The presence of this arguably large autoregressive root calls into question the applicability of conventional critical values.

There are several approaches to this problem that, at least in large samples, satisfactorily handle the cases $\alpha = 1$ or, alternatively, $|\alpha| < 1$, where α is fixed; an example is using a consistent sequence of pretests for a unit root in x (cf. Elliott and Stock, 1994; Kitamura and Phillips, 1992; Phillips, 1995). However, these results are pointwise in α rather than uniform over $|\alpha| \leq 1$. This distinction matters because controlling size in the sense of Lehmann (1959, Ch. 3) and constructing an asymptotically similar test require controlling the size not just for α fixed but also for sequences of α. The sequence that we focus on in this paper is the local-to unity model $\alpha = 1 + c/T$, where c is a fixed constant. It has been established elsewhere that the resulting local-to-unity asymptotic distributions provide good approximations to finite-sample distributions when the root is close to 1 (cf. Chan, 1988; Nabeya and Sørensen, 1994). It is shown in Section 2 that a typical procedure that asymptotically controls size pointwise fails to control size in Lehmann's uniform sense because the asymptotic critical values depend on the nuisance parameter c. The consequence is substantial overrejection of the null hypothesis, both in finite samples and asymptotically.

The specific model for which formal results are developed is the recursive system:

$$x_t = \mu_x + v_t, \qquad (1 - \alpha L)b(L)v_t = \epsilon_{1t}, \tag{1.1}$$

$$y_t = \mu_y + \gamma x_{t-1} + \epsilon_{2t}, \tag{1.2}$$

where $b(L) = \sum_{i=0}^{k} b_i L^i$, $b_0 = 1$, and $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t})'$ is a martingale difference sequence with $E(\epsilon_t \epsilon_t' \mid \epsilon_{t-1}, \epsilon_{t-2}, \ldots) = \Sigma$ (with typical element σ_{ij}) and with $\sup_t E\epsilon_{it}^4 < \infty$, $i = 1,2$. Let $\delta = \text{corr}(\epsilon_{1t}, \epsilon_{2t})$. Assume that $Ev_0^2 < \infty$. The roots of $b(L)$ are assumed to be fixed and less than 1 in absolute value.

If $|\alpha| < 1$ and α is fixed, then x_t is integrated of order 0 (is I(0)), whereas if $\alpha = 1$, then x_t is integrated of order 1 (is I(1)). Thus, α can be taken to be the largest autoregressive root of the univariate representation of x_t. Accordingly, it is useful to write (1.1) in standard augmented Dickey–Fuller (Dickey and Fuller, 1979) (ADF) form:

$$\Delta x_t = \tilde{\mu}_x + \beta x_{t-1} + a(L)\Delta x_{t-1} + \epsilon_{1t}, \tag{1.3}$$

where $\tilde{\mu}_x = (1 - \alpha)b(1)\mu_x$, $\beta = (\alpha - 1)b(1)$, and $a_j = -\sum_{i=j+1}^{k} \tilde{a}_i$, where $\tilde{a}(L) = L^{-1}[1 - (1 - \alpha L)b(L)]$.

We consider the problem of testing the null hypothesis that $\gamma = \gamma_0$ or, equivalently, constructing confidence intervals for γ. For this problem, the root α is a nuisance parameter. In the motivating application to tests of the linear rational expectations hypothesis, $\gamma_0 = 0$, although the theoretical results here hold for general γ_0.

Limiting representations are presented for the case that a constant is included when (1.2) and (1.3) are estimated (the "demeaned" case). These results can be extended to regressions that include polynomials in time of general order, using the techniques in, for example, Park and Phillips (1988) and Sims, Stock, and Watson (1990). In practice, much empirical work includes a linear time trend in the specification. For this reason, although formulas are only given for the demeaned case, some numerical results are also presented for the "detrended" case, in which a constant and linear time trend are included in the regressions of (x_t, y_t) on x_{t-1}.

The paper is organized as follows. The asymptotic size of the conventional t-test of $\gamma = \gamma_0$ based on a consistent pretest of $\alpha = 1$ is derived and computed in Section 2. Section 3 describes several procedures for the construction of tests and confidence intervals that are asymptotically valid, in the sense that size is controlled for local-to-unity sequences of α as well as for α fixed. The asymptotic power of these tests against local alternatives of the form $\gamma = \gamma_0 + g/T$ is also derived in this section. These tests are based on bounds that generally result in asymptotically conservative tests. Bounds tests are a classical device that has been used in related time series problems (e.g., Dufour, 1990), and these tests are applied here to handle the nuisance parameter c. Numerical results on the asymptotic size and power are presented in Section 4. Section 5 concludes the paper.

2. ASYMPTOTIC REPRESENTATIONS AND SIZES OF PROCEDURES WITH A CONSISTENT PRETEST

2.1. Asymptotic Representations of Test Statistics

Let t_γ denote the t-statistic testing $\gamma = \gamma_0$ in (1.2), and let t_β denote the ADF t-statistic testing $\beta = 0$ in (1.3). The joint limiting distribution of (t_γ, t_β) is obtained by applying the theory of local-to-unity asymptotics developed by Bobkoski (1983), Cavanagh (1985), Chan (1988), Chan and Wei (1987), and Phillips (1987). Let $B = (B_1, B_2)'$ be a two-dimensional Brownian motion with covariance matrix $\bar{\Sigma}$, where $\bar{\Sigma}_{11} = \bar{\Sigma}_{22} = 1$ and $\bar{\Sigma}_{12} = \bar{\Sigma}_{21} = \delta$; let J_c be the diffusion process defined by $dJ_c(s) = cJ_c(s)\,ds + dB_1(s)$, where $J_c(0) = 0$; and let $J_c^\mu(s) = J_c(s) - \int_0^1 J_c(r)\,dr$. Also, let $\equiv$ denote equality in distribution, let $\Rightarrow$ denote weak convergence on $D[0,1]$, and let $[\bullet]$ denote the greatest lesser integer function. Under the local-to-unity model $\alpha = 1 + c/T$,

$$\left\{\sigma_{11}^{-1/2} T^{-1/2} \sum_{t=1}^{[T\bullet]} \epsilon_{1t}, \sigma_{22}^{-1/2} T^{-1/2} \sum_{t=1}^{[T\bullet]} \epsilon_{2t}, \omega^{-1} T^{-1/2} x_{[T\bullet]}^\mu \right\}$$
$$\Rightarrow \{B_1(\bullet), B_2(\bullet), J_c^\mu(\bullet)\}$$

jointly, where $\omega^2 = \sigma_{11}/b(1)^2$, $x_t^\mu = x_t - (T-1)^{-1} \sum_{t=2}^T x_{t-1}$ (cf. Chan and Wei, 1987; Phillips, 1987). It follows that t_β and t_γ have the joint limiting representation

$$(t_\beta, t_\gamma) \Rightarrow \{\tau_{1c} + c\Theta_c, \tau_{2c}\} \equiv \{\tau_{1c} + c\theta_c,\ \delta\tau_{1c} + (1-\delta^2)^{1/2} z\}, \qquad \textbf{(2.1)}$$

where $\tau_{1c} = (\int J_c^{\mu 2})^{-1/2} \int J_c^{\mu} dB_1$, $\tau_{2c} = (\int J_c^{\mu 2})^{-1/2} \int J_c^{\mu} dB_2$, $\Theta_c = (\int J_c^{\mu 2})^{1/2}$, and z is a standard normal random variable distributed independently of (B_1, J_c) (cf. Stock, 1991, Appendix A). The final expression in (2.1) is obtained by writing $B_2 = \delta B_1 + (1-\delta^2)^{1/2} \tilde{B}_2$, where $\tilde{B}_2$ is a standard Brownian motion distributed independently of B_1.

The limiting distribution of t_γ depends on both c and δ; however, δ is consistently estimated by the sample correlation between $\hat{\epsilon}_{1t}$ and $\hat{\epsilon}_{2t}$, so we can treat δ as known for the purposes of the asymptotic theory.

A joint test of c and γ can be performed using an appropriate Wald statistic for the system (1.2) and (1.3). Let $\phi_T(\gamma_0, c_0) = [T\hat{\beta} - c_0 \hat{b}(1),\ T(\hat{\gamma} - \gamma_0)]'$, where $\hat{b}(1) = 1 - \sum_{j=1}^{k} \hat{a}_{j-1}$, where $\{\hat{a}_j\}$ are the estimators of $\{a_j\}$ from the OLS estimation of (1.3). Also, let $\hat{\Sigma}$ be the 2×2 matrix with typical element $\hat{\sigma}_{ij} = (T-1)^{-1} \sum_{t=1}^{T} e_{it} e_{jt}$, where e_{1t} and e_{2t} are the residuals from (1.3) and (1.2), respectively. Consider the test statistic

$$W(\gamma_0, c_0) = \frac{1}{2} \phi_T(\gamma_0, c_0)' \left(\hat{\Sigma}^{-1} T^{-2} \sum_{t=2}^{T} x_{t-1}^{\mu 2} \right) \phi_T(\gamma_0, c_0). \tag{2.2}$$

Extensions of the calculations in Stock (1991) show that, under the null hypothesis $(\gamma, c) = (\gamma_0, c_0)$,

$$W(\gamma_0, c_0) \Rightarrow \tfrac{1}{2}(\tau_{1c_0}^2 + z^2). \tag{2.3}$$

The key difficulty for tests of the hypothesis $\gamma = \gamma_0$ using either t_γ or $W(\gamma_0, c_0)$ is that the limiting distributions of these statistics depend on the local-to-unity parameter c. (The exception is if $\delta = 0$, in which event t_γ has a standard normal distribution for all values of c, as well as for α fixed, $|\alpha| < 1$.) Although α is consistently estimable, c is not, so that asymptotic inference cannot in general rely on simply substituting a suitable estimator $\hat{c}$ for c when selecting critical values for tests of γ.

2.2. Asymptotic Size Distortions of Pretest-Based Procedures

This section illustrates the size distortions of two-step tests of $\gamma = \gamma_0$ when the critical values are selected using a consistent first-stage pretest. To make the discussion concrete, consider pretesting using the ADF t-statistic. Let $d_{t_\gamma, c, \eta}$ denote the $100\eta\%$ quantile of the distribution of $\delta\tau_{1c} + (1-\delta^2)^{1/2} z$ for a given value of δ. Consider the following sequential testing procedure based on a consistent ADF pretest, with an equal-tailed second-stage test with nominal level 5%:

$$\begin{aligned} &\text{if } t_\beta < b_1 - b_2 \ln T, \text{ reject } \gamma = \gamma_0 \text{ if } |t_\gamma| > 1.96,\\ &\text{if } t_\beta \geq b_1 - b_2 \ln T, \text{ reject } \gamma = \gamma_0 \text{ if } t_\gamma \notin (d_{t_\gamma,0,.025}, d_{t_\gamma,0,.975}), \end{aligned} \tag{2.4}$$

where b_1 and b_2 are constants with $b_2 > 0$. The asymptotic size of this test of $\gamma = \gamma_0$ is $\lim_{T\to\infty} \sup_{|\alpha| \leq 1} \Pr[(2.4) \text{ rejects } \gamma = \gamma_0 \,|\, \gamma_0 \text{ is true}]$.

To compute a lower bound on this size, consider three possibilities: $\alpha = 1$, α fixed and $|\alpha| < 1$, and $\alpha = 1 + c/T$. Evidently, if $\alpha = 1$ the first-stage test asymptotically rejects with probability 0 and the second-stage test asymptotically rejects with probability 5%. If α is fixed and $|\alpha| < 1$, then $|t_\beta| = O_p(T^{1/2})$; it follows that the first-stage test asymptotically rejects with probability 1, so again the correct critical values are used and the second-stage asymptotic rejection rate is 5%. If, however, $\alpha = 1 + c/T$, the probability of rejecting $\alpha = 1$ goes to 0 because, from (2.1), t_β is $O_p(1)$ for c finite. Thus, asymptotically the $\alpha = 1$ second-stage critical values are used. In this event, the rejection probability is $\Pr[|\delta\tau_{1c} + (1-\delta^2)^{1/2}z| \notin (d_{t_\gamma,0,.025}, d_{t_\gamma,0,.975})]$. Numerical evaluation reveals that, given δ, this is monotone increasing in $-c$ for $c < 0$. In the limit $c \to -\infty$, $\delta\tau_{1c} + (1-\delta^2)^{1/2}z$ is distributed as a standard normal random variable (this follows from the normality of $\int J_c^\mu \, dB_1 / \int (J_c^\mu)^2$ for $c \ll 0$; see Nabeya and Sørensen, 1994). It follows that the size of procedure (2.4) is at least

$$\text{size}(2.4) \geq \Phi(d_{t_\gamma,0,.025}) + \Phi(-d_{t_\gamma,0,.975}). \tag{2.5}$$

Asymptotic rejection rates for (2.4) and their limit, the size (2.5), are given in Table 1 for various values of δ for tests of purported level 5%. When $\delta = 0$, t_γ has a standard normal asymptotic distribution for all c so the size is the asymptotic level. For $\delta \leq 0.3$, the asymptotic size distortions of the two-step procedure are small and are arguably negligibly important for empirical work. However, for larger values of δ, the size distortions can be substantial. For example, for $\delta = 0.9$ in the detrended case, the rejection rate for test (2.4) with "level" 5% is 37% when $c = -20$, and the maximal rejection rate over all c (the size) is 64%.

To provide additional evidence of the dependence of the distribution of t_γ as a function of c, the median and upper and lower 5% quantiles of this distribution are plotted in Figure 1 for $\delta = 0.7$. Evidently, the critical distribution of t_γ is shifted most negatively in the region of $c = 0$, although a negative shift is evident even for $c = -20$. Evidently, if $c < -10$, the $c = 0$ percentiles will provide poor critical values for testing $\gamma = \gamma_0$, which is the source of the size distortions in Table 1.

Whether or not the issues addressed in this paper are important in a particular application evidently depends on the correlation δ. In many applications to financial markets, this correlation can reasonably be expected to be large. For example, the innovations in stock returns plausibly will be (negatively) correlated with the innovation in the log dividend yield, because the log stock price enters each of these variables.

These distortions are not an artifact of using an ADF-based procedure but, rather, arise because local-to-unity processes are classified as I(1) with asymptotic probability 1. Thus, these size distortions are present in other sequential procedures that share this feature. It is worth noting that one such class of procedures are the Bayesian selection rules proposed by Phillips and

Table 1. Asymptotic rejection rates and size of two-step procedure with consistent Dickey–Fuller pretest

c	$\delta = 0.0$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
						A. Demeaned case					
0.0	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
−2.5	0.050	0.051	0.051	0.052	0.053	0.056	0.058	0.062	0.065	0.066	0.069
−5.0	0.050	0.051	0.052	0.056	0.060	0.066	0.073	0.084	0.094	0.103	0.117
−10.0	0.050	0.052	0.054	0.060	0.068	0.079	0.092	0.110	0.134	0.154	0.178
−15.0	0.050	0.052	0.054	0.062	0.073	0.085	0.102	0.128	0.156	0.181	0.215
−20.0	0.050	0.053	0.055	0.065	0.076	0.091	0.109	0.137	0.169	0.199	0.235
−25.0	0.050	0.053	0.056	0.066	0.077	0.095	0.114	0.144	0.177	0.210	0.251
−30.0	0.050	0.053	0.056	0.067	0.078	0.097	0.119	0.151	0.185	0.221	0.263
Limit	0.050	0.053	0.060	0.075	0.095	0.126	0.162	0.211	0.269	0.329	0.400
Size	0.050	0.053	0.060	0.075	0.095	0.126	0.162	0.211	0.269	0.329	0.400
						B. Detrended case					
0.0	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
−2.5	0.050	0.051	0.052	0.055	0.058	0.063	0.068	0.077	0.088	0.102	0.115
−5.0	0.050	0.051	0.054	0.059	0.067	0.077	0.090	0.110	0.137	0.169	0.205
−10.0	0.050	0.052	0.058	0.069	0.084	0.103	0.129	0.165	0.211	0.268	0.335
−15.0	0.050	0.054	0.061	0.075	0.096	0.120	0.154	0.201	0.259	0.333	0.407
−20.0	0.050	0.054	0.063	0.079	0.102	0.130	0.172	0.225	0.291	0.373	0.457
−25.0	0.050	0.055	0.064	0.081	0.107	0.139	0.184	0.241	0.311	0.399	0.491
−30.0	0.050	0.055	0.065	0.084	0.111	0.147	0.193	0.254	0.331	0.420	0.513
Limit	0.050	0.055	0.074	0.107	0.158	0.226	0.310	0.414	0.527	0.644	0.748
Size	0.050	0.055	0.074	0.107	0.158	0.226	0.310	0.414	0.527	0.644	0.748

Notes: Rejection rates of the two-step procedure in (2.4) are based on asymptotic representation (2.1). "Limit" is computed for $c \ll 0$, so $t_\gamma \sim N(0,1)$. Size is the maximum of the preceding rows and this limit. In the demeaned case, regresssions (1.2) and (1.3) contain a constant. In the detrended case, they contain a constant and a linear time trend. Based on 20,000 Monte Carlo replications of the limiting representations, simulated with $T = 500$.

Ploberger (1991) and Stock (1994), which both classify local-to-unity processes as I(1) with probability 1 asymptotically (see Elliott and Stock, 1994).

3. ASYMPTOTICALLY VALID CONFIDENCE INTERVALS AND TESTS

This section describes several alternative procedures for the construction of asymptotically valid tests of $\gamma = \gamma_0$ and confidence intervals for γ. Because the size distortions found in Section 2 occur when these first-stage procedures

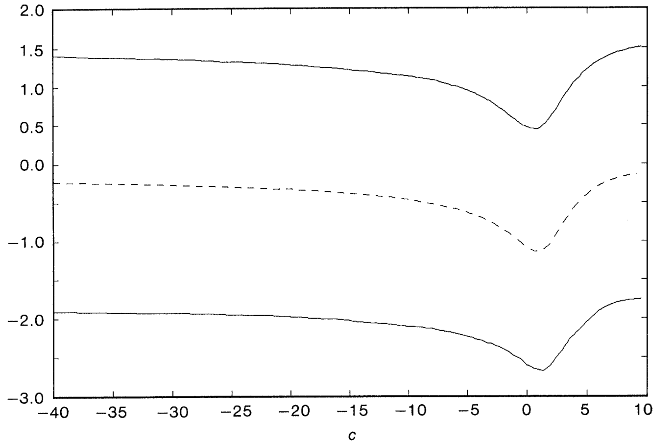

FIGURE 1. The 5, 50, and 95% percentiles of t_γ, demeaned case, $\delta = 0.7$.

classify the process as I(1), but α is in fact large but less than 1, this section focuses on asymptotically valid inference on γ in the local-to-unity case. This provides an alternative to the second line of (2.4) while leaving the first line unchanged. While the procedures apply to general c, the analysis focuses on the mean-reverting case $c \le 0$ for two reasons. First, the economic debate in the unit roots area has, in general, focused on the stationary vs. unit root model. Second, unit root tests typically have high power against close explosive alternatives, so with high probability the I(1) specification would be rejected in these cases against an explosive model, which would take us outside the range of applicability of the dichotomous treatment in (2.4).

Three types of procedures are considered: sup-bound intervals, Bonferroni intervals, and Scheffe-type intervals. Without subsequent adjustment, each can be shown to produce asymptotically conservative tests of $\gamma = \gamma_0$. However, the critical values for each procedure can be adjusted so that its nominal size equals its level asymptotically.

3.1. Sup-Bound Intervals

A simple asymptotically valid test or confidence interval can be constructed by using the extrema of the asymptotic local-to-unity critical values of t_γ. Let

$$(\underline{d}_\eta, \bar{d}_\eta) = \left(\inf_c d_{t_\gamma,c,\eta}, \sup_c d_{t_\gamma,c,\eta}\right). \tag{3.1}$$

A conservative test of $\gamma = \gamma_0$ with asymptotic level at most η can be performed by rejecting if $t_\gamma \notin (\underline{d}_{1/2\eta}, \bar{d}_{1-1/2\eta})$. An asymptotically conservative confidence interval with confidence level at least $100(1-\eta)\%$ can be constructed by inverting the acceptance region of this test, that is, as

$$\{\gamma:\hat{\gamma}-\bar{d}_{1-1/2\eta}SE(\hat{\gamma})\leq\gamma\leq\hat{\gamma}-\underline{d}_{1/2\eta}SE(\hat{\gamma})\}, \tag{3.2}$$

where $SE(\hat{\gamma})=[\hat{\sigma}_{22}/(\sum_{t=2}^{T}x_{t-1}^{\mu 2})]^{1/2}$.

In contrast, if $t_\gamma\in(\bar{d}_{1/2\eta},\underline{d}_{1-1/2\eta})$, a test of $\gamma=\gamma_0$ with asymptotic level η will accept for any value of c. Thus, a confidence interval with confidence level of at most $100(1-\eta)\%$ can be constructed by inverting this test statistic. Values of t_γ within the conservative acceptance region, $(\underline{d}_{1/2\eta},\bar{d}_{1-1/2\eta})$, but outside the acceptance region, $(\bar{d}_{1/2\eta},\underline{d}_{1-1/2\eta})$, constitute an indeterminate region in which ambiguity remains about whether a test of exactly size η would accept or reject.

The actual size of the test of $\gamma=\gamma_0$ using the upper and lower bounds is

$$\begin{aligned}\Pr[t_\gamma\notin(\underline{d}_{1/2\eta},\bar{d}_{1-1/2\eta})]&\to\Pr[\delta\tau_{1c}+(1-\delta^2)^{1/2}z\notin(\underline{d}_{1/2\eta},\bar{d}_{1-1/2\eta})]\\&=S_s(c,\eta),\end{aligned} \tag{3.3}$$

where $S_s(c,\eta)\leq\eta$. Because the size depends on only one asymptotically unknown nuisance parameter (c), it is possible to construct alternative sup-bound confidence intervals with the correct size asymptotically. Specifically, a test of $\gamma=\gamma_0$ with an asymptotic rejection rate of, say, $\bar{\eta}$ can be constructed by choosing η to satisfy $\sup_c S_s(c,\eta)=\bar{\eta}$. Evidently, the resulting value of η, say η', will be at least $\bar{\eta}$. The sup-bound confidence interval (test) with this additional size adjustment will be referred to as the size-adjusted sup-bound confidence interval (test). Note that the critical values used to construct the size-adjusted sup-bound confidence intervals depend on δ.

For the numerical work, the size-adjusted upper and lower bounds were computed by Monte Carlo simulation with $T=1{,}000$ and 20,000 replications over a grid of c, $-40\leq c\leq 10$; $\sup_c S_s(c,\eta')=\bar{\eta}$ was solved numerically for η', and the resulting bounds, as a function of δ, were stored in a lookup table.

3.2. Bonferroni Intervals

The sup-bound confidence regions do not use sample information on α. An alternative, potentially more powerful approach is to construct intervals by inverting Bonferroni tests, where the bounds are determined by taking the extrema of the critical values of t_γ, evaluated over a first-stage confidence interval for α. Let $C_c(\eta_1)$ denote a $100(1-\eta_1)\%$ confidence region for c, and let $C_{\gamma|c}(\eta_2)$ denote a $100(1-\eta_2)\%$ confidence region for γ, which depends on c. Then, a confidence region for γ that does not depend on c can be constructed as

$$C_\gamma^B(\eta)=\bigcup_{c\in C_c(\eta_1)}C_{\gamma|c}(\eta_2). \tag{3.4}$$

By Bonferroni's inequality, the region $C_\gamma^B(\eta)$ has confidence level of at least $100(1-\eta)\%$, where $\eta = \eta_1 + \eta_2$.

Asymptotically valid confidence intervals for c can be constructed by inverting the Dickey–Fuller t-statistic as developed in Stock (1991), which produces an equal-tailed confidence interval of the form, $c_l(\eta_1) \leq c \leq c_u(\eta_1)$. Given the upper and lower limits of this confidence interval, the confidence region of (3.4) can be computed by inverting t_γ. Let

$$(d_l^B(\eta_1,\eta_2), d_u^B(\eta_1,\eta_2)) = \Big(\min_{c_l \leq c \leq c_u} d_{t_\gamma,c,1/2\eta_2}, \max_{c_l \leq c \leq c_u} d_{t_\gamma,c,1-1/2\eta_2}\Big). \tag{3.5}$$

The Bonferroni confidence interval is given by

$$\hat{\gamma} - d_u^B(\eta_1,\eta_2) SE(\hat{\gamma}) \leq \gamma \leq \hat{\gamma} - d_l^B(\eta_1,\eta_2) SE(\hat{\gamma}). \tag{3.6}$$

In principle, this confidence interval can be constructed using graphical methods. First, the interval (c_l, c_u) is obtained by the method of confidence belts using t_β as in Stock (1991). Next, given this confidence interval for c, $d_l^B(\eta_1,\eta_2)$ and $d_u^B(\eta_1,\eta_2)$ are read off a plot of the asymptotic critical values of t_γ, such as Figure 1. In practice, this is more efficiently implemented using computerized lookup tables.

The asymptotic size of Bonferroni test (3.6) is

$$\begin{aligned}&\Pr[t_\gamma \notin (d_l^B(\eta_1,\eta_2), d_u^B(\eta_1,\eta_2))] \\ &\quad \to \Pr[\delta\tau_{1c} + (1-\delta^2)^{1/2} z \notin (d_l^B(\eta_1,\eta_2), d_u^B(\eta_1,\eta_2))] = S_B(c,\eta_1,\eta_2),\end{aligned} \tag{3.7}$$

where, by Bonferroni's inequality, $S_B(c,\eta_1,\eta_2) \leq \eta_1 + \eta_2$. Due to the correlation between the tests, these intervals can be quite conservative. As is the case with the sup-bound intervals, asymptotically valid size-adjusted confidence intervals can be constructed by choosing η_1 and η_2 (where $\eta_2 \leq \bar{\eta}$) so that they achieve some desired level, say $\bar{\eta}$. In practice, this size-adjustment computation is lengthy because of the need to compute first-stage confidence intervals for each realization of a Bonferroni test statistic. After some experimentation, it was found that letting $\eta_2 = \bar{\eta} = 10\%$ and choosing η_1 to solve $S_B(c,\eta_1,\bar{\eta}) = \bar{\eta}$, so that η_1 depends on δ, yielded a test with size 10% for $\delta = 0$, 0.5 ($\eta_1 = 30\%$), 0.7 ($\eta_1 = 24\%$), and 0.9 ($\eta_1 = 13\%$).

3.3. Scheffe-Type Intervals

A Scheffe-type confidence interval, say $C_\gamma^S(\eta)$, can be constructed by projecting an asymptotically valid $100(1-\eta)\%$ joint confidence set for (γ, c), say $C_{\gamma,c}(\eta)$, onto the γ axis; that is,

$$C_\gamma^S(\eta) = \{\gamma : \exists c \text{ such that } (\gamma, c) \in C_{\gamma,c}(\eta)\}. \tag{3.8}$$

This set will have asymptotic confidence level at least $100(1-\eta)\%$. The joint confidence set $C_{\gamma,c}(\eta)$ can be constructed by inverting a level-η test of the joint hypothesis, $(\gamma,c)=(\gamma_0,c_0)$.

The Wald statistic $W(\gamma,c)$ in (2.2) is a natural statistic to use to perform this test. For c finite, the limit distribution of W based on (2.3) is nonstandard, although for $c \ll 0$, τ_{1c} approaches a normal distribution so the distribution of $W(\gamma_0,c_0)$ becomes well approximated by a $\chi_2^2/2$. Let $w_{c_0,1-\eta}$ denote the $100(1-\eta)\%$ quantile of the distribution of the limiting random variable in (2.3). Selected critical values are presented in Table 2.

The set of (γ_0,c_0) for which $W(\gamma_0,c_0) < w_{c_0,1-\eta}$ constitutes a confidence set with asymptotic confidence level $100(1-\eta)\%$. To construct a Scheffe-type confidence interval for γ, it is not necessary to construct this region for (γ,c) but, rather, simply to find the set of γ_0 for which there exists some c_0 such that (γ_0,c_0) is not rejected. Thus, the Scheffe interval is constructed as

$$C_\gamma^S(\eta)=\left\{\gamma_0:\min_{c_0}\left[W(\gamma_0,c_0)-w_{c_0,1-\eta}\right]<0\right\}. \tag{3.9}$$

If the critical value did not depend on c_0, solving (3.9) would reduce to manipulating a quadratic in c_0. However, because of the nonlinear dependence of $w_{c_0,1-\eta}$ on c_0, in our numerical work (3.9) is solved on a grid. Note that, like the Bonferroni interval, interval (3.9) does not restrict the range of c and so implicitly admits positive values of c.

An advantage of this Scheffe-type interval over the Bonferroni intervals is that, because δ does not appear in the limit in (3.9), the critical values for the Scheffe interval do not depend on δ. Like the other intervals, the Scheffe tests are conservative, and size adjustment can be expected to increase power. Let $S_S(c,\eta)$ be the asymptotic size of the Scheffe test of $\gamma=\gamma_0$; then, a size-adjusted Scheffe test can be constructed by using critical values $\{w_{c_0,1-\eta}\}$, where η is chosen such that $\sup_c S_S(c,\eta)=\bar{\eta}$. Note that although $\{w_{c_0,1-\eta}\}$ does not depend on δ, the size $S_S(c,\eta)$ does, so the critical values of the size-adjusted Scheffe test will depend on δ. Because a practical advantage of the unadjusted Scheffe test is that the critical values do not depend on δ, the numerical work focuses on the unadjusted (conservative) test.

3.4. Local Asymptotic Power

A natural way to compare these tests is to compare their power against local alternatives. Because $\sum_{t=2}^T x_{t-1}^2$ is $O_p(T^2)$ in the local-to-unity setting, we consider the local alternative

$$\gamma=\gamma_0+g/T, \tag{3.10}$$

where g is a constant. The limiting representation of t_γ under (3.10) is

$$t_\gamma \Rightarrow \tilde{g}\theta_c+\delta\tau_{1c}+(1-\delta^2)^{1/2}z, \tag{3.11}$$

TABLE 2. Asymptotic critical values $w_{c_0,1-\eta}$ of $W(\gamma_0, c_0)$

c_0	$w_{c_0,.90}$	$w_{c_0,.95}$	$w_{c_0,.975}$	$w_{c_0,.99}$
		A. Demeaned case		
0.0	4.07	4.98	5.79	6.81
−1.0	3.69	4.57	5.43	6.50
−2.5	3.33	4.18	5.05	6.13
−5.0	3.03	3.84	4.63	5.75
−7.5	2.82	3.66	4.44	5.54
−10.0	2.69	3.50	4.31	5.41
−12.5	2.64	3.41	4.25	5.33
−15.0	2.59	3.35	4.19	5.25
−17.5	2.55	3.33	4.13	5.19
−20.0	2.52	3.29	4.11	5.11
−22.5	2.50	3.26	4.06	5.08
−25.0	2.48	3.26	4.02	5.08
−27.5	2.47	3.24	4.00	5.08
−30.0	2.45	3.24	3.98	5.06
Limit	2.31	3.00	3.65	4.62
		B. Detrended case		
0.0	5.58	6.55	7.53	8.68
−1.0	5.11	6.07	7.04	8.18
−2.5	4.59	5.55	6.53	7.62
−5.0	4.04	4.96	5.89	7.00
−7.5	3.65	4.57	5.46	6.67
−10.0	3.39	4.30	5.23	6.37
−12.5	3.25	4.09	4.99	6.17
−15.0	3.13	3.94	4.86	6.01
−17.5	3.02	3.83	4.73	5.87
−20.0	2.93	3.75	4.63	5.73
−22.5	2.88	3.67	4.54	5.65
−25.0	2.83	3.60	4.45	5.55
−27.5	2.79	3.57	4.37	5.49
−30.0	2.75	3.53	4.30	5.45
Limit	2.31	3.00	3.65	4.62

Notes: Entries $w_{c_0,1-\eta}$ are the 100 $(1-\eta)$% quantiles of the limiting null distribution of $W(\gamma_0, c_0)$. "Limit" refers to the case $c_0 \ll 0$ and is computed using the $\chi_2^2/2$ distribution. Entries for $c_0 \geq -30$ are computed by asymptotic simulation with $T = 1{,}000$ and 20,000 Monte Carlo replications.

where $\tilde{g} = (\omega^2/\sigma_{22})^{1/2} g$. Because there is no feedback from y_t to x_t by assumption, the distribution of t_β is the same under the local alternative as under the null.

Let d_l and d_u generically denote the lower and upper critical values used to construct the sup-bound and Bonferroni tests. For $\delta < 1$, the local asymptotic power of the test based on these critical values is

$$P[\text{Reject } H_0 : \gamma = \gamma_0 \,|\, \gamma = \gamma_0 + g/T]$$
$$= E\{\Phi[(d_l - \tilde{g}\theta_c - \delta\tau_{1c})/(1-\delta^2)^{1/2}]$$
$$+ \Phi[(-d_u + \tilde{g}\Theta_c + \delta\tau_{1c})/(1-\delta^2)^{1/2}]\}. \tag{3.12}$$

The derivation of the asymptotic power function of the Scheffe test proceeds similarly. Suppose that the true value of α is $1 + c'/T$. Under local alternative (3.10), $W(\gamma_0, c_0)$ has the limiting representation

$$W(\gamma_0, c_0) \Rightarrow \tfrac{1}{2}(1-\delta^2)^{-1}\{[\tau_{1c'} + (c'-c_0)\Theta_{c'}]^2 + [\tau_{2c'} + \tilde{g}\Theta_{c'}]^2$$
$$- 2\delta[\tau_{1c'} + (c'-c_0)\Theta_{c'}][\tau_{2c'} + \tilde{g}\Theta_{c'}]\} \equiv W^*_{c',\tilde{g}}(\gamma_0, c_0), \tag{3.13}$$

where $\tau_{1c'}$, $\tau_{2c'}$, and $\Theta_{c'}$ are as defined following (2.1), evaluated for $c = c'$. Thus, the asymptotic power of the Scheffe test with level η against the alternative $(\gamma, c) = (\gamma_0 + g/T, c')$ is

$$P[\text{Reject } H_0 : \gamma = \gamma_0 \,|\, \gamma = \gamma_0 + g/T,\ \alpha = 1 + c'/T]$$
$$= P[\min_{c_0}(W^*_{c',\tilde{g}}(\gamma_0, c_0) - w_{c_0, 1-\eta}) \geq 0]. \tag{3.14}$$

4. NUMERICAL RESULTS: SIZE AND POWER

This section evaluates the performance of the procedures in Section 3. Results are reported in terms of asymptotic size and power of tests of $\gamma = \gamma_0$; coverage rates for the corresponding confidence intervals for γ are 1 minus the size. For the cases in which the distributions are nonstandard, asymptotic size and power results were computed by numerical evaluation of (3.12) and (3.14) using the asymptotic representations, which in turn were computed by Monte Carlo simulation of the various functionals of Brownian motion with $T = 500$. All asymptotic results are based on 20,000 Monte Carlo replications for each set of parameters.

Asymptotic rejection rates of the various procedures as a function of the true values of c and of δ are summarized in Table 3 for tests with asymptotic level 10%. For a given value of δ, the size is the maximum (over c) rejection rate. Because the distribution of t_γ tends to a $N(0,1)$ for $c \ll 0$, rejection rates for this limiting case can be computed using the standard normal c.d.f., and these results are reported in the row labeled "limit." (The Scheffe limit is computed using the $\chi^2_2/2$ limit for $W(\gamma_0, c_0)$.) For each δ, the size is computed as the maximum rejection rate in the preceding rows for that procedure.

A striking feature of Table 3 is that the rejection rates of the size-adjusted procedures (the sup-bound and Bonferroni tests) do not drop below 5% and often are close to 10%. For $c \leq -5$, the size-adjusted Bonferroni test rejection rates exceed 9%. Because the Scheffe test is not size-adjusted, its rejection rates are substantially below 10%. Interestingly, the rejection rates

TABLE 3. Asymptotic null rejection rates of tests of $\gamma = \gamma_0$ with asymptotic size $\leq 10\%$: Demeaned case

		δ			
Procedure	c	0	0.5	0.7	0.9
Sup-bound	0	0.100	0.090	0.081	0.080
	−5	0.100	0.075	0.055	0.042
	−10	0.100	0.078	0.059	0.046
	−20	0.100	0.082	0.066	0.055
	−30	0.100	0.085	0.071	0.061
	Limit	0.100	0.100	0.100	0.100
	Size	0.100	0.100	0.100	0.100
Bonferroni	0	0.100	0.084	0.081	0.067
	−5	0.100	0.091	0.096	0.093
	−10	0.100	0.096	0.100	0.099
	−20	0.100	0.100	0.100	0.099
	−30	0.100	0.100	0.099	0.098
	Limit	0.100	0.100	0.100	0.100
	Size	0.100	0.100	0.100	0.100
Scheffe	0	0.027	0.028	0.039	0.052
	−5	0.036	0.024	0.031	0.041
	−10	0.043	0.026	0.031	0.038
	−20	0.051	0.029	0.031	0.037
	−30	0.066	0.036	0.036	0.035
	Limit	0.032	0.032	0.032	0.032
	Size	0.066	0.036	0.036	0.052

Notes: See the text for definitions of the various procedures. The column $\delta = 0$ was computed using the $N(0,1)$ distribution of t_γ. Entires for $\delta > 0$, $c \geq -30$ were computed by numerical evaluation of the limiting representations with 20,000 Monte Carlo replications and $T = 1{,}000$. "Limit" refers to $c \ll 0$, and entries in those rows are computed using the $N(0,1)$ approximation for t_γ and the $\chi_2^2/2$ approximation for $W(\gamma_0, c_0)$. "Size" is the maximum rejection rate in the column for the indicated test.

of the sup-bound test decline for moderate c, as the distribution moves from a Dickey–Fuller type distribution with rejections in the lower tail to a normal with rejections in the upper tail.

Local asymptotic power functions of the various procedures against alternative (3.10) with $g < 0$ are plotted in Figure 2 for various values of δ. As points of reference, Figure 2 includes two additional power functions of infeasible tests in which c is taken as known. The first is the "simultaneous equations" test, which is asymptotically most powerful unbiased. This test is implemented using the cross-equation restrictions available if c is known, which, when $b(L) = 1$, amounts to including $x_t - \alpha x_{t-1}$ as a regressor in (1.2). The second, termed "t_γ with c known," is the equal-tailed test of $\gamma = \gamma_0$ based on t_γ, which uses critical values for the true c, that is, the test with

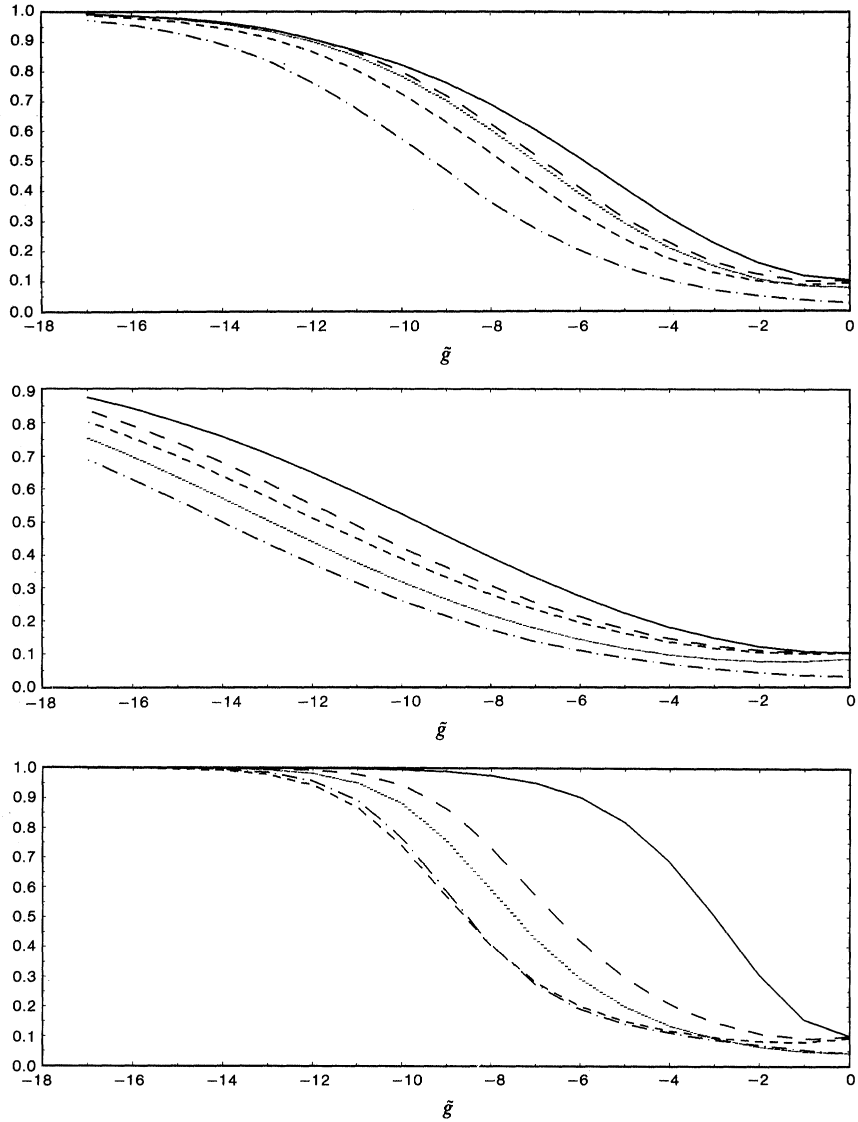

FIGURE 2. Local asymptotic power of 10% level tests of $\gamma = \gamma_0$ against $\gamma = \gamma_0 + g/T$, demeaned case, $\delta = 0$. Top: $c = -5$, $\delta = 0.5$. Middle: $c = -20$, $\delta = 0.5$. Bottom: $c = -5$, $\delta = 0.9$. Key: Simultaneous equations test (solid line); t_γ with c known (long dashes); sup-bound (dots); Bonferroni (short dashes); Scheffe (dashes and dots). $\tilde{g} = (\omega^2/\sigma_{22})^{1/2} g$.

acceptance region $d_{t_\gamma,c,.05} \leq t_\gamma \leq d_{t_\gamma,c,.95}$. For $\delta = 0$, these two tests are asymptotically equivalent, but for $\delta \neq 0$ the power function of the simultaneous equations test (the Gaussian power envelope) lies above the power function of the c-known test. Because these tests are infeasible when c is unknown, the relative power loss of the other procedures indicates the cost of lack of knowledge of c. As can be seen in Figure 2, as δ increases the relative performance of the infeasible simultaneous equations test improves. For $c=-5$, the sup-bound test has higher power than the Bonferroni or Scheffe test, although for $c=-20$ the Bonferroni has the highest power of these three. The relatively better performance for $c \ll 0$ of the Bonferroni test is to be expected, because in this case the quantiles of t_γ depend only weakly on c (cf. Figure 1). In no case does the Scheffe test have power as high as the sup-bound test, which is not surprising considering the sup-bound test is size-adjusted, whereas the Scheffe test is not. Although the power functions are not, in general, symmetric in g, the qualitative results for $g>0$ are similar.

In practice, δ is typically unknown. Table 4 therefore reports test rejection rates found in a Monte Carlo experiment with $T=100$ and 2,000 replications. The data are generated according to (1.2) and (1.3) with $\Sigma_{11}=\Sigma_{22}=1$, $\Sigma_{12}=\delta$, $\gamma=0$, and $b(L)=(1+\theta L)^{-1}$, so $(1-\alpha L)v_t$ is an MA(1). The estimated system was (1.2) and (1.3), where the lag length in (1.3) was chosen by the Bayes information criterion with a maximum of four lags. The tests were implemented using an estimated value of δ, $\hat{\delta}=\mathrm{corr}(\hat{\epsilon}_{1t},\hat{\epsilon}_{2t})$; given $\hat{\delta}$, the relevant critical values were interpolated from a lookup table of critical values as a function of δ and, for the Bonferroni tests, c. Results are reported for $\theta=-0.5$, 0, 0.5, $\alpha=1$, 0.95, 0.90, 0.80, and $\delta=0.5$, 0.9.

TABLE 4. Monte Carlo results: Finite sample rejection rates, δ estimated, demeaned case

		$\delta=0.5$			$\delta=0.9$		
	c	$\theta=-0.5$	$\theta=0$	$\theta=0.5$	$\theta=-0.5$	$\theta=0$	$\theta=0.5$
Sup-bound	0	0.095	0.101	0.100	0.086	0.099	0.106
	−5	0.073	0.076	0.077	0.044	0.044	0.045
	−10	0.081	0.072	0.077	0.048	0.047	0.047
	−20	0.091	0.082	0.079	0.073	0.063	0.063
Bonferroni	0	0.099	0.094	0.102	0.110	0.082	0.100
	−5	0.094	0.092	0.091	0.092	0.093	0.096
	−10	0.100	0.088	0.091	0.097	0.106	0.104
	−20	0.104	0.092	0.099	0.102	0.103	0.108
Scheffe	0	0.042	0.031	0.038	0.076	0.066	0.082
	−5	0.034	0.030	0.030	0.044	0.044	0.050
	−10	0.052	0.032	0.036	0.052	0.040	0.041
	−20	0.281	0.035	0.053	0.411	0.039	0.050

Notes: Results are based on 2,000 replications with $T=100$. The design is described in the text.

The Monte Carlo results suggest that the asymptotic results in Table 3 provide a good guide to finite sample rejection rates in almost all cases. The Bonferroni and sup-bound procedures have Monte Carlo sizes close to 10%. The Scheffe procedure is somewhat less conservative in this finite sample experiment than it is asymptotically and has rejection rates less than 10% in all cases except $\theta = -0.5$, $c = -20$. Because $T = 100$, this case corresponds to $\alpha = 0.8$, $\theta = -0.5$, so the AR and MA roots are approaching cancellation. This is a case in which it is known that the asymptotics provide a poor approximation in the univariate model (cf. Pantula,1991), and those difficulties evidently carry over to (2.3), particularly as the univariate case is approached for $|\delta|$ large.

5. DISCUSSION AND EXTENSIONS

This paper has investigated several procedures for handling the dependence of the distribution of tests of $\gamma = \gamma_0$ on c. The Monte Carlo simulations suggest that these procedures control size in finite samples with δ unknown, even though they are based on asymptotic analysis in which δ is consistently estimated. When δ is small or moderate, the cost of using these procedures is small, relative to infeasible tests that use knowledge of c. However, for δ large, the relative cost of not knowing c can be large.

The model considered here is simple and stylized. One extension is to include lags of y_t and additional lags of x_t in (1.2). The asymptotic distribution theory for this extension is straightforward under the null that x_t does not enter; the calculations use the techniques in Park and Phillips (1988) and Sims et al. (1990), as adapted in Stock (1991) for the local-to-unity case. The qualitative feature of the current results—that the test statistics have nonstandard distributions that depend on c—will continue to hold under this generalization, although the critical values for the F-statistic testing the coefficients on x_{t-1} and its lags will depend on the number of lags of x_t. Another extension is to nonrecursive models in which (1.2) continues to hold, but in which there is feedback from y to x in (1.1) and (1.3). After suitable modification of δ and the covariance matrix in the W-test, the distributions of the sup-bound and $W(\gamma_0, c_0)$ statistics obtained for the current model also hold for this extension under the null $\gamma = 0$. A third extension is to inference about cointegrating vectors. Although the focus here has been on the null $\gamma_0 = 0$, if γ_0 is nonzero then y_t and x_t are cointegrated, except both x_t and y_t have local-to-unit roots in their univariate representation. This extension is pursued by Elliott (1994), who also considers the behavior of efficient estimators of cointegrating vectors and their test statistics in this model. Even though these extensions are possible, however, considerable work remains to generalize this approach to higher dimensional models with possibly multiple unit roots and cointegrated regressors.

REFERENCES

Bobkoski, M.J. (1983) *Hypothesis Testing in Nonstationary Time Series.* Unpublished Ph.D. Thesis, University of Wisconsin.

Campbell, J.Y. & R.J. Shiller (1988) Stock prices, earnings and expected dividends. *Journal of Finance* 43, 661–676.

Cavanagh, C. (1985) Roots Local to Unity. Manuscript, Harvard University.

Chan, N.H. (1988) On the parameter inference for nearly nonstationary time series. *Journal of the American Statistical Association* 83 (403), 857–862.

Chan, N.H. & C.Z. Wei (1987) Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics* 15, 1050–1063.

Dickey, D.A. & W.A. Fuller (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74 (366), 427–431.

Dufour, J.M. (1990) Exact tests and confidence sets in linear regression with autocorrelated errors. *Econometrica* 58, 475–494.

Elliott, G. (1994) *Application of Local to Unity Asymptotic Theory to Time Series Regression.* Ph.D. Dissertation, Harvard University.

Elliott, G. & J.H. Stock (1994) Inference in time series regression when the order of integration of a regressor is unknown. *Econometric Theory* 10, 672–700.

Fama, E.F. (1991) Efficient capital markets II. *Journal of Finance* 46(5), 1575–1617.

Kitamura, Y. & P.C.B. Phillips (1992) Fully Modified IV, GIVE, and GMM Estimation with Possibly Nonstationary Regressors and Instruments. Manuscript, Cowles Foundation, Yale University.

Lehmann, E.L. (1959) *Testing Statistical Hypotheses.* New York: Wiley.

Nabeya, S. & B.E. Sørensen (1994) Asymptotic distributions of the least-squares estimators and test statistics in the near unit root model with non-zero initial value and local drift and trend. *Econometric Theory* 10, 937–966.

Pantula, S.G. (1991) Asymptotic distributions of the unit-root tests when the process is nearly stationary. *Journal of Business and Economic Statistics* 9, 63–71.

Park, J.Y. & P.C.B. Phillips (1988) Statistical inference in regressions with integrated processes: Part I. *Econometric Theory* 4, 468–497.

Phillips, P.C.B. (1987) Toward a unified asymptotic theory for autoregression. *Biometrika* 74, 535–547.

Phillips, P.C.B. (1995) Fully modified least squares and vector autoregression. *Econometrica* 63, 1023–1078.

Phillips, P.C.B. & W. Ploberger (1991) Time Series Modeling with a Bayesian Frame of Reference: I. Concepts and Illustrations. Manuscript, Cowles Foundation, Yale University.

Sims, C.A., J.H. Stock, & M.W. Watson (1990) Inference in Linear Time Series Models with Some Unit Roots. *Econometrica* 58, 113–144.

Stock, J.H. (1991) Confidence intervals for the largest autoregressive root in U.S. economic time series. *Journal of Monetary Economics* 28 (3), 435–460.

Stock, J.H. (1994) Deciding between I(0) and I(1). *Journal of Econometrics* 63, 105–131.