ELSEVIER Journal of Econometrics 80 (1997) 269–286

JOURNAL OF Econometrics

Inference in a nearly integrated autoregressive model with nonnormal innovations

Thomas J. Rothenberg[a,*], James H. Stock[b]

[a]*Department of Economics, University of California, 549 Evans Hall #3880, Berkeley CA 94720-3880, USA*

[b]*Kennedy School of Government, Harvard University, USA*

Abstract

Robust tests and estimators based on nonnormal quasi-likelihood functions are developed for autoregressive models with near unit root. Asymptotic power functions and power envelopes are derived for point-optimal tests of a unit root when the likelihood is correctly specified. The shapes of these power functions are found to be sensitive to the extent of nonnormality in the innovations. Power loss resulting from using least-squares unit-root tests in the presence of thick-tailed innovations appears to be greater than in stationary models. © 1997 Elsevier Science S.A.

Keywords: Robust inference; Power envelope; Unit root; Maximum likelihood
JEL classification: C22

1. Introduction

Econometricians have devoted considerable attention to parameter estimation and hypothesis testing in autoregressive time-series models where the largest root of the lag polynomial is near unity. Banerjee et al. (1993) and Stock (1995) survey some of the literature. Most previous studies examine

* Corresponding author.

We are indebted to Jurgen Doornik for providing us with his computer programs and to Anthony Atkinson, Andre Lucas, and Neil Shephard for helpful comments on an earlier draft of this paper.

PII S0304-4076(97)00040-7

least-squares methods, which are likelihood based if the data are normally distributed. If the data are not normal, inference using these methods typically remains valid in large samples, but is inefficient compared to methods which exploit the correct form of the likelihood. Here we develop some alternative estimators and tests based on nonnormal likelihoods for an almost integrated univariate autoregressive time series.

Since unusually large movements in economic time series seem to occur more often than would be implied by normality, inference specifically designed for thick-tailed distributions should be relevant for applied econometric work. In practice, of course, the exact innovation distribution is not known, so we need to construct estimates and tests whose performance is robust to misspecification. We derive point estimates that are approximately unbiased and tests that have approximately the correct size for a large class of distributions possessing second order moments. Although our procedures will be optimal only if based on the actual likelihood, they should have good sampling properties for other data distributions with similar tail behavior.

In addition to providing some practical tools for data analysis, our study may also be of more general methodological interest. In typical parametric inference problems, there exist asymptotically optimal estimates and uniformly most powerful tests for one-sided, one-dimensional alternatives. Furthermore, most common estimates and test statistics are asymptotically normal and the loss from using an inefficient estimator or test can be captured by a single scale parameter. It is well known that the usual asymptotic normal theory does not apply in nearly integrated autoregressive models. However, as we show in later sections, the theory for comparing alternative inference procedures in such models has many similarities with the normal theory. Our analysis follows closely that developed by Elliott et al. (1996) for the case of normal innovations. Related work on robust inference in nearly integrated autoregressive models can be found in Cox and Llatas (1991), Lucas (1995a, b), and Phillips (1993).

2. Likelihood-based inference

A general maintained model often employed by past researchers specifies that an observed univariate time-series $y_1, \ldots, y_n$ is generated as

$$\begin{aligned} y_t &= d_t + u_t \\ & \qquad\qquad t = 1, \ldots, n \\ u_t &= \alpha u_{t-1} + v_t \end{aligned} \tag{1}$$

where $\{d_t\}$ is a deterministic 'trend' typically assumed to be a polynomial function of t and $\{v_t\}$ is an unobserved mean-zero stationary sequence with finite variance. The restriction $\alpha = 1$ is known as the unit-root hypothesis; when α is close to one the process is said to be almost integrated. Typically, there will be unknown nuisance parameters which characterize the trend, the v_t process, and the initial value u_0, and these must be taken into account when conducting inference on α. To simplify the exposition, we avoid these complications and concentrate on the special case where the d_t are zero and the v_t process is white noise. Specifically, we assume the data are generated by the autoregression

$$y_t = \alpha y_{t-1} + \varepsilon_t, \quad t = 1, \ldots, n \tag{2}$$

where $y_0 = 0$ and the ε_t are unobserved iid errors with mean zero and unit variance. Extensions to the more general model (1) are discussed briefly in Section 7.

Since our interest is in inference when the parameter α is close to one, we shall employ local-to-unity asymptotics where the parameter space is assumed to be a shrinking neighborhood of unity as the sample size grows. That is, we reparameterize the model writing $\gamma = n(\alpha - 1)$ and take γ to be a constant when making limiting arguments. Cf. Chan and Wei (1987).

Let $e^{g(\varepsilon)}$ be a density function with mean zero and unit variance. We assume that the log density $g(\varepsilon)$ possesses derivatives $g'(\varepsilon)$ and $g''(\varepsilon)$. If this were the actual error density, the log likelihood function would be

$$L(\gamma) = \sum_{t=1}^{n} g(\Delta y_t - n^{-1}\gamma y_{t-1})$$

and, by Taylor series, we have the quadratic approximation

$$L(\gamma) = \sum g(\Delta y_t) - \gamma n^{-1} \sum y_{t-1} g'(\Delta y_t) + \tfrac{1}{2}\gamma^2 n^{-2} \sum y_{t-1}^2 g''(\Delta y_t) + r_n(\gamma). \tag{3}$$

Under regularity conditions spelled out in Section 3, the pair of random variables

$$A_g \equiv -\frac{1}{n}\sum_{t=1}^{n} y_{t-1} g'(\Delta y_t) \quad \text{and} \quad B_g \equiv -\frac{1}{n^2}\sum_{t=1}^{n} y_{t-1}^2 g''(\Delta y_t) \tag{4}$$

has a nondegenerate limiting distribution and the remainder term $r_n(\gamma)$ is $o_p(1)$ uniformly on compact sets of γ values. Hence A_g and B_g are asymptotically jointly sufficient statistics for γ when $g(\varepsilon)$ is the log density for ε.

Although there is just one unknown parameter in our model, the asymptotic sufficient statistic is two dimensional. Thus we should not expect to find a uniformly best estimate or a uniformly most powerful test even asymptotically. For example, by the Neyman–Pearson lemma, the optimal test of $\gamma = \gamma_0$ against the alternative $\gamma = c$ rejects when $L(\gamma_0) - L(c)$ is small. Using our Taylor series approximation, we see that any test which rejects for small values of a linear combination of A_g and B_g will be asymptotically admissible in the sense that, in large samples, it has highest possible power for some alternative in a neighborhood of unity.

Inference based on A_g and B_g may have good properties even if $g(\varepsilon)$ is not the correct log density. For example, if normality is assumed sc that $g(\varepsilon) = -\frac{1}{2}\varepsilon^2$, then A_g and B_g become the least-squares statistics

$$A_N \equiv \frac{1}{n}\sum_{t=1}^{n} y_{t-1}\Delta y_t \quad \text{and} \quad B_N \equiv \frac{1}{n^2}\sum_{t=1}^{n} y_{t-1}^2.$$

If the errors are actually normal, admissible tests can be constructed from the sufficient statistics A_N and B_N. (For details, see Elliott et al. 1996.) But inference based on A_N and B_N may be good (albeit not optimal) even if the errors are far from normal. This property, of course, is not unique to least-squares statistics. In the next section we show how to construct asymptotically valid estimates and tests based on A_g and B_g whenever the data distribution is known to be symmetric.

3. Asymptotic distribution theory

Following the previous literature, we shall find it convenient to approximate sample statistics by stochastic integrals. Let $[\cdot]$ denote the greatest lesser integer function and let $W(t)$ represent standard Brownian motion defined on $[0,1]$. Suppose, in addition to the assumptions already made, the innovations satisfy the moment condition $E[\varepsilon_t^k] < \infty$ for some $k > 2$. Then, as n tends to infinity, the sequence of random functions $y_n^*(s) \equiv n^{-1/2} y_{[sn]}$ converges weakly on $[0,1]$ to the Ornstein–Uhlenbeck process

$$J_\gamma(s) = \int_0^s e^{\gamma(s-r)} \mathrm{d}W(r)$$

which satisfies the stochastic differential equation

$$\mathrm{d}J_\gamma(s) = \gamma J_\gamma(s)\,\mathrm{d}s + dW(s)$$

with initial condition $J_\gamma(0) = 0$. Since our statistics A_g and B_g are well-behaved functions of the y_t, they can be approximated by functionals of the process J_γ.

The analysis for the least squares statistics A_N and B_N has been worked out by Chan and Wei (1987) and Phillips (1988). It is known that, under local alternatives where γ is fixed as the sample size n tends to infinity,

$$(A_N, B_N) \Rightarrow \left(\tfrac{1}{2}[J_\gamma^2(1) - 1], \int_0^1 J_\gamma^2(t)\right) \tag{5}$$

where $\Rightarrow$ represents weak convergence. Note that, although A_N and B_N are sufficient statistics only under normality, the convergence in (5) holds for arbitrary error distributions satisfying our moment assumptions.

To compute asymptotically valid tests based on A_N and B_N, we need the null distributions of the test statistics. For approximate Neyman–Pearson tests of the unit-root hypothesis that $\gamma = 0$, the distributions of linear combinations of the random variables

$$A_N^* = \tfrac{1}{2}[W^2(1) - 1] \quad \text{and} \quad B_N^* = \int_0^1 W^2(t) \tag{6}$$

are required. These can be obtained from the series representation of the joint distribution of A_N^* and B_N^* given by Abadir (1992, 1995). Alternatively, distributions of functions of A_N^* and B_N^* are readily computed by Monte Carlo simulation, and that is the approach used here.

A similar analysis can be applied to the statistics A_g and B_g based on a general log density g. The relevant asymptotic theory can be found in Cox and Llatas (1991) and Lucas (1995b). Suppose the moments

$$\omega^2 \equiv \operatorname{var}[g'(\varepsilon)], \quad \delta \equiv -\operatorname{E}[g''(\varepsilon)], \quad \rho \equiv -\operatorname{E}[\varepsilon g'(\varepsilon)] \tag{7}$$

exist when expectations are taken with respect to the true error density. Consider the random variables $\eta_t \equiv g'(\varepsilon_t) + \rho\varepsilon_t$ which, by construction, are iid with variance $\omega^2 - \rho^2$ and are uncorrelated with the ε_t. Then, if $g''(\cdot)$ is well behaved, we find by a standard Taylor series argument

$$\begin{aligned} A_g &\equiv -\frac{1}{n}\sum y_{t-1} g'(\Delta y_t) = -\frac{1}{n}\sum y_{t-1} g'(\varepsilon_t) - \frac{\gamma}{n^2}\sum g''(\varepsilon_t) y_{t-1}^2 + o_p(1) \\ &= \frac{\rho}{n}\sum y_{t-1}\Delta y_t - \frac{\gamma\rho}{n^2}\sum y_{t-1}^2 - \frac{1}{n}\sum y_{t-1}\eta_t - \frac{\gamma}{n^2}\sum g''(\varepsilon_t) y_{t-1}^2 + o_p(1), \end{aligned}$$

and

$$\begin{aligned} B_g &\equiv -\frac{1}{n^2}\sum y_{t-1}^2 g''(\Delta y_t) = -\frac{1}{n^2}\sum y_{t-1}^2 g''(\varepsilon_t) + o_p(1) \\ &= \frac{\delta}{n^2}\sum y_{t-1}^2 + o_p(1). \end{aligned}$$

But the random variable

$$-\left[(\omega^2 - \rho^2)\frac{1}{n^2}\sum y_{t-1}^2\right]^{-1/2} \frac{1}{n}\sum y_{t-1}(\eta_t - \operatorname{E}\eta_t) \tag{8}$$

converges to a standard normal variate Z which is independent of the process $J_\gamma(t)$. Using the asymptotic representations for A_N and B_N, we have:

Proposition. Consider the autoregressive model (2) *where the* ε_t *are iid with mean zero and variance one. Suppose* $[\varepsilon, g'(\varepsilon)]$ *possesses kth order moments for some* $k > 2$, $\mathrm{E}[g'(\varepsilon_t)] = 0$, *and* g'' *satisfies a linear Lipschitz condition. Then, under local alternatives where* $\gamma = n(1-\alpha)$ *is fixed as n tends to infinity, the statistics defined in* (3) *have the limiting representations*

$$\begin{aligned} A_g &\Rightarrow \tfrac{1}{2}\rho\,[J_\gamma^2(1) - 1] + Z[(\omega^2 - \rho^2)\textstyle\int J_\gamma^2]^{1/2} + \gamma(\delta - \rho)\textstyle\int J_\gamma^2, \\ B_g &\Rightarrow \delta \textstyle\int J_\gamma^2 \end{aligned} \tag{9}$$

where Z *is standard normal independent of the process* J_γ.

If $e^{g(\varepsilon)}$ is the actual density for ε, we find from integration by parts that $g'(\varepsilon)$ necessarily has mean zero and that $\rho = 1$ and $\delta = \omega^2 \geqslant 1$. Under correct specification, the limiting distribution of (A_g, B_g) depends only on γ and ω^2. Of course, if $g = -\frac{1}{2}\varepsilon^2$, then $\rho = \delta = \omega = 1$ no matter what the actual error density. If both $e^{g(\varepsilon)}$ and the actual error density are symmetric about zero, $g'(\varepsilon)$ will have mean zero even under misspecification.

To construct asymptotically valid tests of the unit-root hypothesis $\gamma = 0$, we need to know the joint distribution of the random variables

$$\begin{aligned} A_g^* &= \tfrac{1}{2}\rho[W^2(1) - 1] + Z[(\omega^2 - \rho^2)\textstyle\int W^2]^{1/2}, \\ B_g^* &= \delta \textstyle\int W^2. \end{aligned} \tag{10}$$

If one has confidence that g is the actual log density, ρ can be set equal to unity and $\delta = \omega^2$ can be computed from the definition

$$\omega^2 = \int [g'(\varepsilon)]^2 e^{g(\varepsilon)}\,\mathrm{d}\varepsilon.$$

The joint distribution of any function of A_g^* and B_g^* can then be obtained analytically or by simulating Z and the Brownian motion $W(t)$.

More commonly, the density e^g chosen by the econometrician will not be the (unknown) true error density so ω^2, ρ, and δ must be treated as unknown parameters. Nevertheless, we can take advantage of the fact that the joint distribution of A_g^* and B_g^* is unaffected if these parameters are replaced in (10) by consistent estimates. Thus a robust estimate of the null distribution of our test statistic can be obtained by simulating (10) after first replacing the unknown moments by estimates such as

$$\hat{\omega}^2 = \frac{1}{n}\sum [g'(\Delta y_t)]^2, \quad \hat{\delta} = -\frac{1}{n}\sum g''(\Delta y_t), \quad \hat{\rho} = -\frac{1}{n}\sum \Delta y_t g'(\Delta y_t). \tag{11}$$

In practice, however, such robust tests could be inconvenient since the critical values will have to be evaluated for each data set. In the next section we show that considerable simplification occurs if inference is based on particular functions of A_g and B_g, namely, the maximum likelihood estimator and t-ratio.

4. Inference based on the MLE

In practice, it is common to estimate unknown parameters by the method of maximum likelihood and to construct test statistics from these estimates. For example, to test $\gamma = \gamma_0$ against the alternative $\gamma < \gamma_0$, one might reject for small values of $\hat{\gamma} - \gamma_0$ where $\hat{\gamma}$ is the maximum likelihood estimate of γ. When the normal likelihood is used, this means rejecting when A_N/B_N is less than some critical value, say d. Since this is equivalent to rejecting when $A_N - dB_N$ is negative, the least-squares test is a member of the Neyman–Pearson family of tests under normality. Since the reciprocal of the standardized Hessian B_N is a common estimate of the asymptotic variance of the least squares estimate, an alternative is to reject the hypothesis that $\gamma = \gamma_0$ if the t-ratio $\sqrt{B_N}(\hat{\gamma} - \gamma_0) = (A_N - \gamma_0 B_N)/\sqrt{B_N}$ is small. Although this rejection region cannot be written in terms of a linear combination of A_N and B_N, numerical calculations suggest that this t-ratio test is admissible asymptotically and has optimal power characteristics for alternatives where power is near one-half.[1] The null distribution of the least squares t-ratio is bell-shaped with modest skewness and kurtosis: it has been tabulated by Dickey and Fuller (1979) for the unit-root case where $\gamma_0 = 0$.

In the case of nonnormal errors, we find that the MLE for γ is

$$\hat{\gamma} = \frac{\frac{1}{n}\sum y_{t-1} g'(\Delta y_t)}{\frac{1}{n^2}\sum y_{t-1}^2 g''(\Delta y_t)} + o_p(1) \approx \frac{A_g}{B_g}.$$

Furthermore, the likelihood ratio statistic for testing $\gamma = \gamma_0$ can be approximated as

$$R^2 = 2[L(\hat{\gamma}) - L(\gamma_0)] = \frac{(A_g - \gamma_0 B_g)^2}{B_g} + o_p(1).$$

[1] See, for example, Elliott et al. (1996).

The square root of the likelihood ratio statistic, appropriately signed, is asymptotically equivalent to the t-statistic based on the MLE and has the asymptotic representation $(A_g - \gamma_0 B_g)/B_g^{1/2}$. Thus the two tests based on the OLS estimator are easily generalized to the case of arbitrary maximum-likelihood estimators.

It follows from Section 3 that the t-statistic based on the MLE has the asymptotic representation

$$R_g \Rightarrow \rho\delta^{-1/2}\left[\frac{J_\gamma^2(1)-1}{2S} - \gamma S\right] + Z\left[\frac{\omega^2-\rho^2}{\delta}\right]^{1/2} + \delta^{1/2}(\gamma-\gamma_0)S. \quad (12)$$

where $S^2 \equiv \int J_\gamma^2$. With a normal likelihood function, we have $\rho = \delta = \omega^2 = 1$ and (12) simplifies to

$$R_N \Rightarrow \frac{J_\gamma^2(1)-1}{2S} - \gamma_0 S.$$

The representation for R_g given in (12) has some interesting consequences. When the null hypothesis is true so $\gamma = \gamma_0$, R_g is asymptotically a linear combination of the least-squares t-ratio R_N and an independent normal. Hence the moments of R_g can easily be computed from the (known) moments of R_N. In particular, if $e^{g(\varepsilon)}$ is the actual density for ε, R_g is equal to $R_N\omega^{-1} + Z(1-\omega^{-2})^{1/2}$ and has cumulants between those of R_N and those of Z. In that case, the null distribution of R_g is closer to that of a standard normal than is the distribution of R_N. Paradoxically, the (asymptotic) null distribution of the ML t-statistic has moments closer to a standard normal the farther the population distribution is from normality!

Table 1 presents simulated cumulants for the asymptotic distribution of the MLE $\hat{\gamma}$ and the t-ratio R in the unit root case where $\gamma = \gamma_0 = 0$ and g is the actual log density of the innovations.[2] For both statistics, the mean becomes closer to zero and the skewness is reduced as ω increases. In standard regression problems, the asymptotic variance of the MLE is proportional to the inverse of the information term $\omega^2 = \mathrm{E}(g')^2$; hence we might expect ω^2 times the asymptotic variance to be constant. In fact, the product decreases with increasing ω; in unit root problems, nonnormality increases precision even more than in standard problems.

Critical values for R_g can be obtained by simulation as before. However, since the null distribution is approximately normal, a simpler approach is available.

[2] Simulations from the asymptotic distribution were performed by computing J_γ as the realization of 200 successive observations from a discrete time Gaussian AR(1) process with parameter $1+\gamma/200$ and drawing Z from an independent standard normal distribution. Unless otherwise stated, all simulation results are based on 20,000 Monte Carlo replications.

Table 1
Cumulants for the asymptotic distribution of the MLE and t-ratio (gamma = 0)

	Standardized estimator $\hat{\gamma}$				t-ratio			
ω^2	mean	std × ω	skew	kur	mean	sd	skew	kur
1.00	− 1.771	3.195	− 2.305	8.904	− 0.419	0.984	0.249	0.086
1.25	− 1.406	3.029	− 2.240	8.357	− 0.371	0.983	0.180	0.067
1.50	− 1.168	2.922	− 2.187	8.092	− 0.337	0.984	0.137	0.050
1.75	− 0.999	2.844	− 2.133	7.861	− 0.311	0.986	0.109	0.040
2.00	− 0.872	2.785	− 2.080	7.649	− 0.290	0.987	0.090	0.032
3.00	− 0.578	2.644	− 1.888	6.953	− 0.235	0.990	0.050	0.018
4.00	− 0.431	2.571	− 1.732	6.445	− 0.202	0.992	0.034	0.013
5.00	− 0.344	2.527	− 1.606	6.067	− 0.180	0.993	0.025	0.010
6.00	− 0.285	2.498	− 1.502	5.777	− 0.163	0.994	0.020	0.009
7.00	− 0.244	2.477	− 1.415	5.549	− 0.150	0.995	0.016	0.008
8.00	− 0.213	2.461	− 1.341	5.365	− 0.140	0.995	0.014	0.007
9.00	− 0.189	2.449	− 1.277	5.214	− 0.132	0.996	0.012	0.006
10.00	− 0.169	2.439	− 1.221	5.087	− 0.124	0.996	0.011	0.006

Denoting the ξ-percentile of the standard normal distribution by τ_ξ and setting $k_3 = \mathrm{E}[R - \mathrm{E}R]^3/[\mathrm{Var}(R)]^{3/2}$, we can approximate the ξ critical value of R_g by the Cornish–Fisher expansion

$$\begin{aligned}\tau_\xi^* &= \mathrm{E}(R_g) + [\mathrm{Var}(R_g)]^{1/2}\left[\tau_\xi + \frac{1}{6}k_3(\tau_\xi^2 - 1)\right] \\ &= \theta\mathrm{E}(R_N) + \left[\frac{\omega^2}{\delta} + \theta^2(\mathrm{Var}(R_N) - 1)\right]^{1/2}\left[\tau_\xi + \frac{1}{6}\theta^3 k_{3N}(\tau_\xi^2 - 1)\right]\end{aligned}$$

where $\theta^2 = \rho^2/\delta$ and k_{3N} is the standardized third cumulant of R_N. If the null hypothesis is that $\gamma = 0$ and $n \geqslant 50$, we have by simulation

$$\mathrm{E}(R_N) \cong -0.42, \quad \mathrm{Var}(R_N) \cong 0.975, \quad k_{3N} \cong 0.25.$$

If one is confident that g is the true log density, ω^2/δ can be set to unity and θ set to the known value ω^{-1}. If one wishes to be robust to misspecification of the error distribution, the unknown parameters can be replaced by estimates such as given in (10).

5. Asymptotic power of unit-root tests under correct specification

If e^g is the true likelihood function for the data, the Neyman–Pearson test of the null hypothesis $\gamma = \gamma_0$ against the point alternative $\gamma = c < \gamma_0$ has an asymptotic local power function given by

$$\pi(\gamma,c) = P_\gamma[A_g - \tfrac{1}{2}(\gamma_0 + c)B_g < d(c)] \tag{13}$$

where $d(c)$ is the critical value ensuring that $\pi(\gamma_0, c)$ equals the significance level. The envelope power function $\pi^*(\gamma) = \pi(\gamma,\gamma)$ is an upper bound for the local asymptotic power of any one-sided test. Since the asymptotic distribution of A_g and B_g under correct specification of the likelihood depends only on γ and ω, the envelope $\pi^*(\gamma)$ depends on the single parameter ω. In Fig. 1, we present graphs of the power envelopes for various values of ω when $\gamma_0 = 0$ and the test size is 5%.

The large increase in power as ω rises is not surprising. In the AR(1) model $y_t = \alpha y_{t-1} + \varepsilon_t$ where the ε_t are iid with density e^g and $|\alpha|$ is considerably less than one, efficient tests of $\alpha = \alpha_0$ against $\alpha < \alpha_0$ have local asymptotic power functions of the form $\Phi[d + \omega\sigma c]$ where Φ is the standard normal distribution function, $\omega^2 = \text{var}[g'(\varepsilon_t)]$, $\sigma = (1 - \alpha_0^2)^{-1/2}$ and $c = n^{1/2}(\alpha - \alpha_0)$. That is, asymptotically, the stable AR model behaves like the location model; power measured in terms of normal quantiles is linear in c with slope proportional to ω. (Cf. Appendix.) One might expect the power functions for unit root tests to

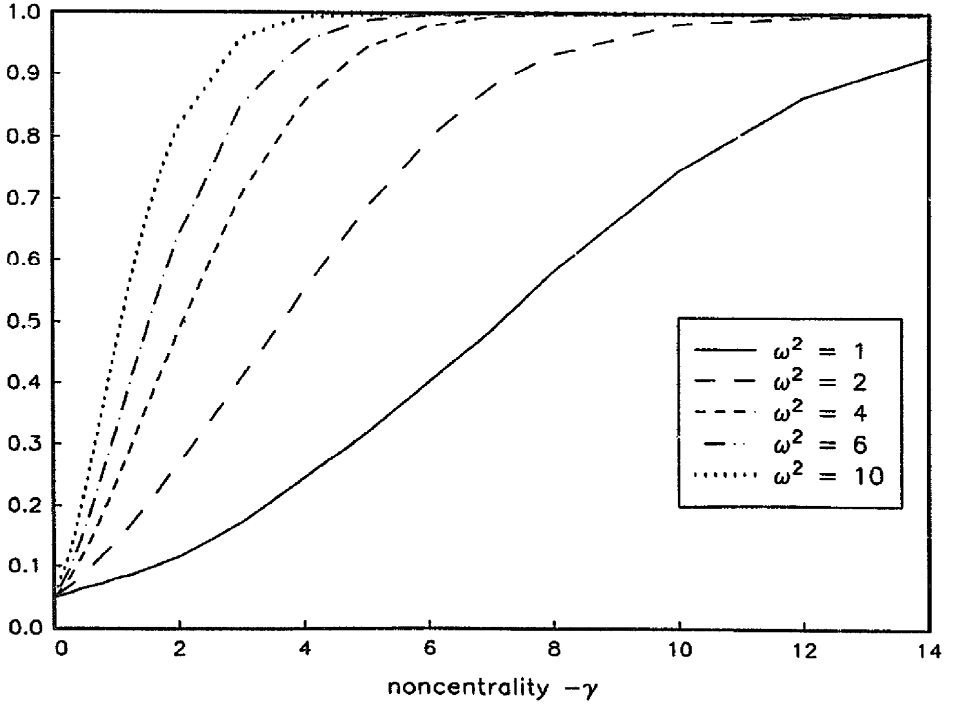


Fig. 1. Asymptotic power envelopes for testing $\gamma = 0$ (5% level).

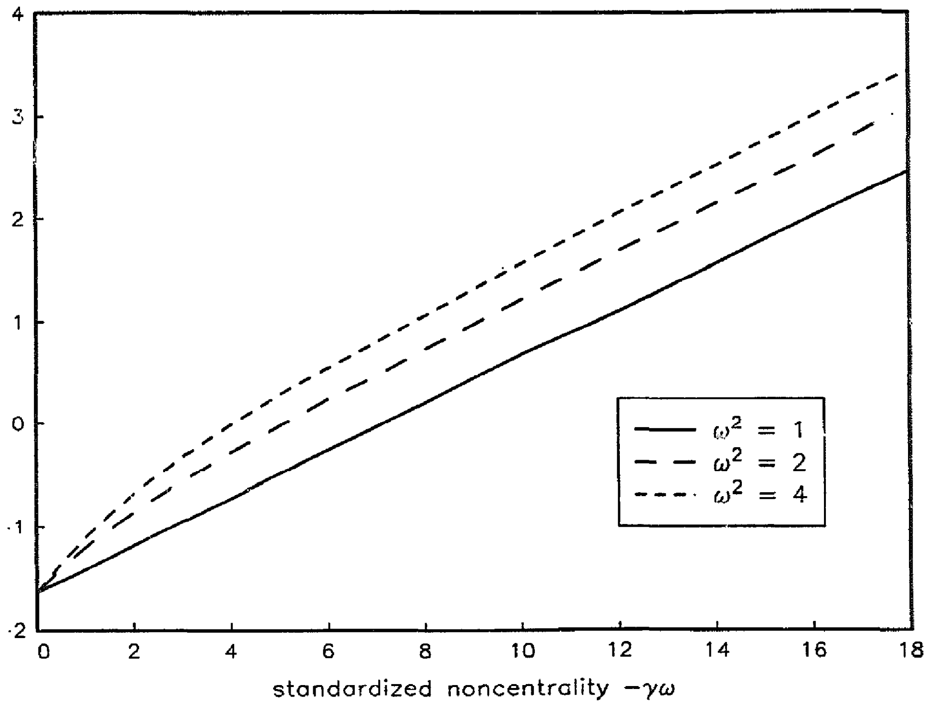
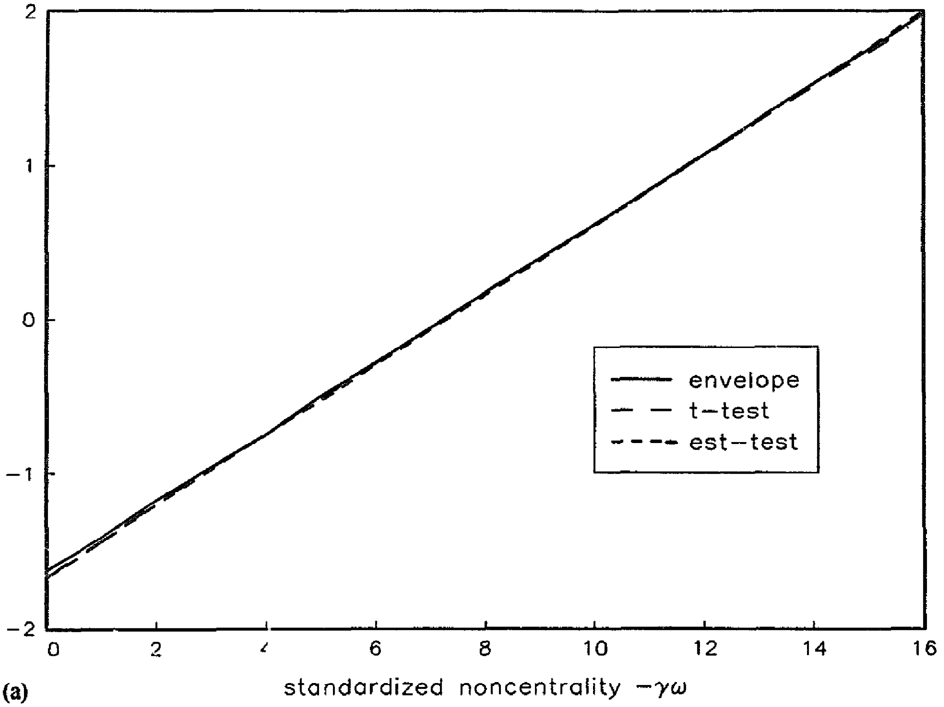


Fig. 2. Asymptotic power envelopes for testing $\gamma = 0$ (5% level; normal quantile scale).

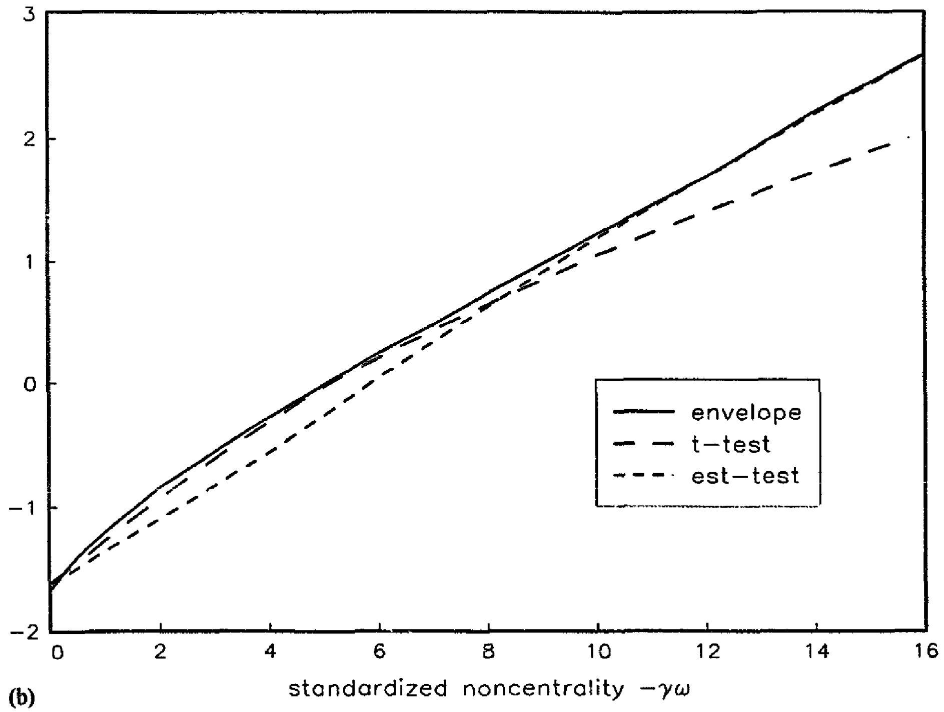
have a similar form. Again using 5% level tests, we plot in Fig. 2 $\Phi^{-1}[\pi^*(\gamma)]$ against $\omega\gamma$ for various values of ω. When $\omega = 1$ (the case for normally distributed ε), transformed power is essentially linear in γ with slope 0.22. As ω rises, the (transformed) power envelope becomes more curved, especially when γ is small, and its slope rises. When testing for a unit root, power is higher the further the errors are from normal and the effect is somewhat greater than in the case where standard asymptotic theory applies.

Fig. 3 plots the asymptotic power (again in terms of normal quantiles) of tests based on the maximum likelihood estimator and t-statistic. When the errors are normal ($\omega = 1$), the power functions for the two tests are essentially identical to the power envelope. For all practical purposes, they are asymptotically equivalent and efficient. This is no longer true when the errors are nonnormal. The power functions for the estimator and t-tests are both tangent to the envelope and hence are asymptotically admissible. But the power functions show considerably greater curvature than the envelope particularly when ω is large. In each case, the tangency for the test based on the MLE occurs at high power and the tangency for the test based on the t-statistic occurs when power is approximately one-half. Curiously, this is exactly the



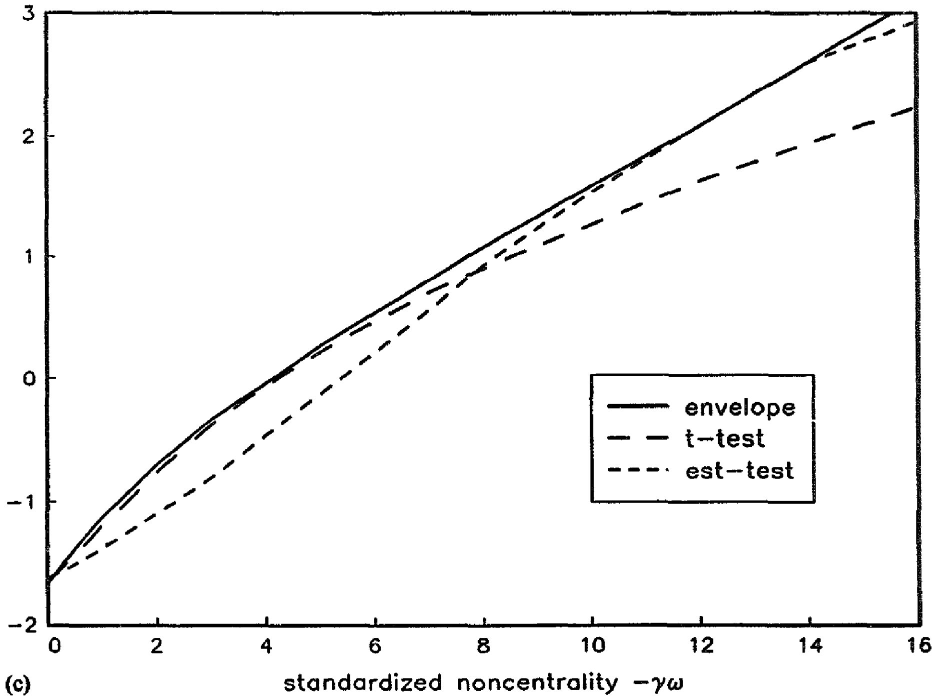
(a)

Fig. 3(a). Asymptotic power functions when $\omega^2 = 1$ (5% level; normal quantile scale).



(b)

Fig. 3(b). Asymptotic power functions when $\omega^2 = 2$ (5% level; normal quantile scale).



(c)

Fig. 3(c). Asymptotic power functions when $\omega^2 = 4$ (5% level; normal quantile scale).

prediction that second-order asymptotic theory makes for standard (nonunit root) problems.[3]

6. Power of robust unit-root tests

Suppose the true error density is e^f but one constructs tests using the density e^g. As discussed in Sections 3 and 4, it is still possible to produce tests with the correct asymptotic size as long as both f and g are symmetric about zero. The cost will be reduced power. The cost, however, is not symmetric in f and g; that is, the loss in power from using g when f is true is not the same as the loss in power from using f when g is true. Some known results for the standard location model are summarized in the appendix. Since similar analytic results are not available in the unit-root case, we conducted a small simulation

[3] See, for example, Rothenberg (1984).

experiment. Data were generated from the first-order autoregressive model (2) using two different error distributions: Case A employed a standard normal density $\phi(\varepsilon)$ whereas Case B employed a mixture of normals having density

$$f(\varepsilon) = \pi\phi(\varepsilon/\sigma_1)/\sigma_1 + (1-\pi)\phi(\varepsilon/\sigma_2)/\sigma_2$$

with $\pi = 0.95$, $\sigma_1^2 = 0.5$ and $\sigma_2^2 = 10.5$. This latter density has zero mean, unit variance, zero skewness, and kurtosis coefficient approximately 13. When this mixture is the actual density, $\rho = 1$ and $\omega^2 = \delta = 1.77$. When the actual density is standard normal, we find by simulation $\omega^2 = 2.65$, $\delta = 1.52$, and $\rho = 1.52$. Table 2 gives rejection probabilities for the t-statistic R_g based on the mixture distribution and for the least squares t-statistic R_N. To illustrate the difference between the unit-root case and the standard case, Table 3 gives corresponding rejection probabilities for testing in an iid location model.

Table 2
Power of unit root test

	Normal errors		Mixture errors	
α	Normal t-test	Mixture t-test	Normal t-test	Mixture t-test
1.00	0.05	0.05	0.05	0.05
0.975	0.15	0.14	0.14	0.28
0.950	0.34	0.30	0.31	0.58
0.925	0.58	0.49	0.51	0.79
0.900	0.78	0.66	0.70	0.91
0.875	0.91	0.80	0.83	0.97

Sample size is 100; significance level is 0.05; 20,000 Monte Carlo trials.

Table 3
Power of location parameter test

	Normal errors		Mixture errors	
$\sqrt{n}(\theta_0 - \theta)$	Normal t-test	Mixture t-test	Normal t-test	Mixture t-test
0	0.05	0.05	0.05	0.05
1.00	0.26	0.24	0.26	0.38
1.64	0.50	0.46	0.50	0.71
2.00	0.64	0.59	0.64	0.84
3.00	0.91	0.87	0.91	0.99

Sample size is 100; significance level is 0.05; 30,000 Monte Carlo trials.

These tables confirm the finding in Section 5 that power losses from using an inefficient test are greater in the unit root case than in the case where standard asymptotics is appropriate. As in the standard case, the power loss from using the (normal-based) least squares t-test when the data really have very thick tails is much larger than the power loss from using the thick-tailed likelihood when in fact data are normal. Finally, other experiments not reported here suggests that the Cornish–Fisher approximate critical values are reasonably accurate. Indeed, the even simpler alternative of using the Dickey–Fuller tables for the 5% one-sided critical value of the least-squares t-statistic works surprisingly well for the general ML t-statistic.

7. The effect of nuisance parameters

For simplicity we have presented the analysis for the special case where there are no nuisance parameters. However, our results are easily modified to cover more realistic situations. Consider, for example, the case where the data are generated by Eq. (1) with d_t equal to the constant β, $\{v_t\}$ is a stationary AR process with known order p, and the initial u_0 has bounded second moment. Suppose further that $C(L)v_t = \varepsilon_t$ where the ε_t are iid with mean zero and variance σ^2 and $C(L)$ is a polynomial in the lag operator L. Defining $\tilde{y}_t \equiv C(L)(y_t - \beta)/\sigma$ for $t > p$, we see that, except for the asymptotically negligible initial condition on $\tilde{y}_p$, the $\tilde{y}_t$ are generated by the simple one-parameter model examined in Sections 3–6. Let $\hat{y}_t$ be the estimate of $\tilde{y}_t$ obtained by replacing β by y_1 and replacing σ and the coefficients in $C(L)$ by estimates consistent under local-to-unity asymptotics.[4] Then, as long as p is small and the roots of $C(L)$ are bounded away from one, A_g and B_g calculated from the $\hat{y}$ data have the same limiting distribution as given in Section 3. Asymptotically, there is no power loss from not knowing β or the parameters describing the AR process $\{v_t\}$. This is the same result found by Elliott et al. (1996) under normality.

The above argument does not work when d_t is a linear time trend of the form $\beta_0 + \beta_1 t$; the use of an estimated β_1 changes the limiting distributions. Furthermore, even in the case where estimating the nuisance parameters has no effect asymptotically, estimates and tests of γ with better small-sample properties might be obtained by using better estimates of the nuisance parameters. In general, suppose the log likelihood function is written as $L(\gamma, \theta)$ where θ is a vector of nuisance parameters. A natural generalization of the Neyman–Pearson tests considered in Section 2 for testing $\gamma = \gamma_0$ against the

[4] For example, one might estimate σ^2 and the parameters of $C(L)$ by a least squares regression (using the last t–p observations) of Δy_t on its p lagged values.

alternative that $\gamma = c$ is to reject for small values of $L(\gamma_0, \hat{\theta}) - L(c, \tilde{\theta})$, where $\hat{\theta}$ is the MLE for θ under the null hypothesis and $\tilde{\theta}$ is the MLE under the alternative. The limiting distribution of this test statistic will be a functional of the J_γ process, but its exact form will depend on the specifics of the problem. When d_t depends on nontrending regressors and v_t is a short memory stationary process, the limiting distribution will typically behave like a linear function of A_g and B_g as in our example.

8. Conclusions

Although standard asymptotic theory does not apply in models for nearly integrated data, robust tests and estimators based on quasi-likelihood functions can be developed just as in the standard case. The loss in efficiency from using least-squares methods when the errors have thick tails can be substantial. Nonnormality appears to increase statistical curvature of the likelihood and affects the limiting distribution of the maximum likelihood estimator and t-ratio. Unlike the normal case, power functions of alternative admissible tests cross and depart substantially from the power envelope. Although our results are based entirely on asymptotic arguments, one expects that they should also be relevant for moderate sample sizes.

In this paper we have only considered robustness with respect to nonnormality of the innovation error distribution. Other forms of misspecification are also possible. In particular, additive measurement error on the observed y_t, which contributes a moving error term to the autoregressive process, is often plausible. This type of misspecification is studied by Franses and Haldrup (1994) and Lucas (1995a) in the context of nearly integrated data. It would be interesting to see if the approach developed here leads to useful results in that model as well.

Appendix

A.1. Robustness theory for the location model

In the location model $y_i = \theta + \varepsilon_i$ (i = 1, ..., n) where the ε_i are iid with zero mean, variance σ^2, and common density function $e^{f(\varepsilon)}$, let $\hat{\theta}_g$ be the pseudo maximum likelihood estimator of θ based on the (possibly) false error density $e^{g(\varepsilon)}$, where both f and g are smooth and symmetric about zero. Denoting derivatives by primes, let $\delta_{gf} = -\mathrm{E}_f[g''(\varepsilon)]$ and $\omega_{gf}^2 = \mathrm{Var}_f[g'(\varepsilon)]$ where expectations are with respect to e^f. Then, under standard regularity conditions, $\sqrt{n}(\hat{\theta}_g - \theta)$ is asymptotically normal with mean zero and variance $\omega_{gf}^2/\delta_{gf}^2$. The maximum likelihood estimator using the correct likelihood is the special case where $f = g$. Since $\delta_{ff} = \omega_{ff}^2$, the asymptotic variance of the actual MLE $\hat{\theta}_f$ is $1/\omega_{ff}^2$ where

$$\omega_{ff}^2 = \mathrm{Var}_f[f'(\varepsilon)] = \int [f'(\varepsilon)]^2 \mathrm{e}^{-f(\varepsilon)} \mathrm{d}\varepsilon \geqslant \sigma^{-2}.$$

Among densities having zero mean and variance σ^2, ω_{ff}^2 takes the minimum value of σ^{-2} for the normal density. By contrast, it equals $2\sigma^{-2}$ for the Student density with 3 degrees of freedom. The efficiency of $\hat{\theta}_g$ compared to $\hat{\theta}_f$ can be measured by the variance ratio

$$\lambda^2 = \frac{\mathrm{var}_f[\hat{\theta}_f]}{\mathrm{var}_f[\hat{\theta}_g]} = \frac{\mathrm{E}_f^2[g'f']}{\mathrm{E}_f(g'g')\mathrm{E}_f(f'f')}$$

which lies between 0 and 1.

When testing $\theta = \theta_0$ against $\theta < \theta_0$, a natural rejection region based on $\hat{\theta}_g$ is $\sqrt{n}(\hat{\theta}_g - \theta_0)\hat{\delta}_{gf}/\hat{\omega}_{gf} < \tau_\xi$, where τ_ξ is the ξ percentage point of a normal distribution and $\hat{\delta}_{gf}$ and $\hat{\omega}_{gf}^2$ are consistent estimates such as

$$\hat{\delta}_{gf} = \frac{1}{n}\sum g''(y_i - \hat{\theta}_g), \qquad \hat{\omega}_{gf}^2 = \frac{1}{n}\sum [g'(y_i - \hat{\theta}_g)]^2.$$

For local alternatives, the test has asymptotic power $\Phi[\sqrt{n}(\theta_0 - \theta)\delta_{gf}/\omega_{gf} + \tau_\xi]$. The corresponding test based on the actual MLE $\hat{\theta}_f$ has asymptotic power function given by $\Phi[\sqrt{n}(\theta_0 - \theta)\omega_{ff} + \tau_\xi]$. The test based on the actual MLE $\hat{\theta}_f$ attains the same power as the test based on $\hat{\theta}_g$ at an alternative λ times as distant.

Because the expectation is taken with respect to the true distribution e^f, λ is not symmetric in f and g. For the distribution used in our simulation experiment in Section 6, we find that, if the errors are normal, but we use the normal mixture density for e^g, $\lambda = 0.93$; if the errors are a normal mixture, but we use the normal density, $\lambda = 0.75$. Hence, we have the well known fact that least-squares methods (which follow from a normal likelihood) perform badly when the errors have thick tails, but robust methods designed for thick-tailed error distributions do not perform badly when the errors are actually normal. Although the above formulas are derived from a simple location model, similar results hold for linear and nonlinear regression and stationary autoregression.

References

Abadir, K.M., 1992. A distribution generating equation for unit-root statistics. Oxford Bulletin of Economics and Statistics 54, 305–323.

Abadir, K.M., 1995. The joint density of two functionals of a brownian motion. Mathematical Methods of Statistics 4, 449–462.

Banerjee, A., Dolado, J.J., Galbraith, J.W., Hendry, D.F., 1993. Co-integration, Error Correction, and the Econometric Analysis of Non-stationary Data, Oxford University Press, Oxford.
Chan, N.H., Wei, C.Z., 1987. Asymptotic inference for nearly nonstationary AR(1) processes. Annals of Statistics 15, 1050–1063.
Cox, D.D., Llatas, I., 1991. Maximum likelihood type estimation for nearly non-stationary autoregressive time series. Annals of Statistics 19, 1109–1128.
Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the Amercian Statistical Association 74, 427–431.
Elliott, G., Rothenberg, T.J., Stock, J.H., 1996, Efficient tests for an autoregressive unit root. Econometrica 64, 813–836.
Franses, P.H., Haldrup, N., 1994. The effects of additive outliers on tests of unit roots and cointegration. Journal of Business and Economic Statistics 12, 471–478.
Lucas, A., 1995a. An outlier robust unit root test with an application to the extended Nelson–Plosser data. Journal of Econometrics 66, 153–173.
Lucas, A., 1995b. Unit root tests based on M estimators. Econometric Theory 11, 331–346.
Phillips, P.C.B., 1988. Regression theory for near-integrated time series. Econometrica 56, 1021–1043.
Phillips, P.C.B., 1993. Robust nonstationary regression, mimeo.
Rothenberg, T.J., 1984. Approximating the distributions of econometric estimators and test statistics. in: Griliches, Z., Intriligator, M. (Eds.), The Handbook of Econometrics, vol. 2, North Holland, Amsterdam.
Stock, J.H., 1995. Unit roots, structural breaks and trends. In: Engle, R., McFadden, D. (Eds.), The Handbook of Econometrics, vol. 4, North Holland, Amsterdam.