Advances in econometrics
Sixth World Congress

Volume I

Edited by

CHRISTOPHER A. SIMS

CAMBRIDGE
UNIVERSITY PRESS

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1994
First paperback edition 1996

Printed in Great Britain at the University Press, Cambridge

A catalogue record for this book is available from the British Library

Library of Congress cataloguing in publication data

Advances in econometrics: Sixth World Congress / edited by
Christopher A. Sims. Vol. I
p. cm.
"Sixth World Congress of the Econometric Society in Barcelona"
ISBN 0 521 40374 X
1. Econometrics – Congresses. I. Sims, Christopher A.
II. Econometric Society. World Congress (6th: Barcelona, Spain)
HB139.A345 1994
330'.01'5195 – dc20 92-42041 CIP

ISBN 0 521 44459 4 hardback
ISBN 0 521 56610 X paperback

SE

Contents

I	**Time series econometrics**	*page* 1
Chapter 1	Testing for the stationarity and the stability of equilibrium *Mototsugu Fukushige, Michio Hatanaka, and Yasuji Koto*	3
Chapter 2	Time series with strong dependence *P.M. Robinson*	47
Chapter 3	Recursive linear models of dynamic economies *Lars Peter Hansen and Thomas J. Sargent*	97
II	**Microeconometric theory**	141
Chapter 4	The selection problem *Charles F. Manski*	143
Chapter 5	Quantile regression, censoring, and the structure of wages *Gary Chamberlain*	171
III	**Seasonality**	211
Chapter 6	The economics of seasonal cycles *Jeffrey A. Miron*	213
	Comment by Svend Hylleberg	252
Chapter 7	On the economics and econometrics of seasonality *Eric Ghysels*	257
	Comment by Denise R. Osborn	317

CHAPTER 1

Testing for the stationarity and the stability of equilibrium: with applications to international capital markets

Mototsugu Fukushige, Michio Hatanaka, and Yasuji Koto

1 A STATISTICAL TEST FOR THE STATIONARITY AND THE STABILITY OF EQUILIBRIUM

1.1 Introduction

One of the most important developments in econometrics in the 1980s is what has been conveniently summarized as the unit root. Concerning the least-squares estimator of an autoregressive parameter, say ρ, its asymptotic distribution when the true ρ is unity is different from that when $|\rho|$ is less than unity, as shown by Dickey and Fuller (1979) and Phillips (1987). The point is important in applied econometrics because the finite sample distribution when $|\rho|$ is less than but near unity resembles the asymptotic distribution for $\rho=1$ more closely than the asymptotic distribution for $|\rho|<1$.[1] One implication is that the power in testing $\rho=1$ against $|\rho|<1$ by the least-squares estimator is bound to be low when the sample size is not very large. Nevertheless we are frequently compelled to discriminate between $\rho=1$ and $|\rho|<1$ by the economic and statistical problems.

However, as judged from the articles published or yet unpublished as of the time of the present writing, the latest research efforts are somewhat counteracting to the previous ones: (1) The Bayesian inference contains no such anomaly as found in the sampling approach (see Zellner, 1971, p.187), but a number of problems arise on the prior. Sims (1988) points out that economic theories do not necessarily justify a sharp point prior placed on $\rho=1$, and Wago and Tsurumi (1991) refer to the inference problems caused by such a prior.[2] Phillips (1991) criticizes the flat prior as a representation of ignorance. (2) The wisdom of taking $\rho=1$ for the null hypothesis has been questioned, and the stationarity for the null hypothesis has been investigated in Park (1990), Fukushige and Hatanaka (1989), Ogaki and Park

(1989), Fisher and Park (1990), and Bierens (1990). Schotman and van Dijk (1990, 1991) treat the stationarity and the unit root symmetrically and derive a posterior odds ratio. (3) As regards Nelson and Plosser (1982), which revealed the importance of the unit root in economic data, Schmidt and Phillips (1992), Choi (1990), and Haldrup (1990) find drawbacks in the method used. Each proposes a revised method within the framework of sampling approach, and Choi (1990) in particular argues that his revision reverses the conclusion of Nelson and Plosser. From the standpoint of the Bayesian inference DeJong and Whiteman (1991) also reverse the conclusion of Nelson and Plosser (1982). It seems that the Box–Jenkins modeling of trends should be considered with a grain of salt.

In the sampling school the null hypothesis is such that one wishes to control the probability of rejecting it mistakenly when it is indeed true. In the Box–Jenkins modeling of time series the failure to difference when differencing is required induces a serious loss in the subsequent inference, and we wish to control the probability of such a failure. In our opinion it makes sense to take $\rho = 1$ for the null hypothesis in the Box–Jenkins type of time series analysis. However, in testing for cointegration, which is the stationarity in the residual of a relationship (see Engle and Granger, 1987), it might make more sense to have the cointegration for the null hypothesis. It would depend upon the purpose of the cointegration analysis.[3]

The present chapter, which is a revised version of Fukushige and Hatanaka (1989), presents a method for testing stationarity against non-stationarity in a scalar autoregressive process, taking stationarity for the null hypothesis. Many methods are now available for such tests. Our method differs from the others in the following two points: (1) we do not assume that the dominating root(s) of the characteristic equation is (are) real; (2) the non-stationarity includes not only the moduli of roots equal to unity but also those exceeding unity. Thus the unity is nothing more than the boundary of the non-stationarity region. In both of these two points our approach is more general than the Box–Jenkins'.

After our test for stationarity is presented, we shall apply it to the *error* of the equilibrium relation in economic theory. The stationarity of the error is equivalent to the stability of the equilibrium in the sense in which the word is used in economic theory, and the Box–Jenkins modeling of trends is irrelevant to this equivalence. For many decades stability has been considered an important property of any economic equilibrium theory, without which the theory is rendered meaningless. Even though the empirical validity of stability has seldom been questioned, important exceptions are, for example, stability in the equilibrium of trade balances in terms of the Marshall–Lerner condition and stability in the markets that are driven by speculation. As for the latter example, as theories of expectation have advanced since the mid 1970s, questions regarding stability are now phrased in terms of convergence of an economy with sunspots to a particular equilibrium unaffected by them (see Azariadis, 1981), and also in terms of the convergence of expectations formation to rational expectations (see Marcet and Sargent, 1989). In testing stability empirically in these cases, economic theorists have a "prior feeling" that stability is likely to hold true, and the suggestion by Cox and Hinkley (1974, p. 65) is that stability should be taken for the null hypothesis in such a case.[4]

1.2 The model and notation

Consider a scalar autoregressive process, $\{y_t\}$, generated by

$$(1-b_1L-\ldots-b_pL^p)y_t=\mu+\epsilon_t, \tag{2.1}$$

where L is the lag operator and $\{\epsilon_t\}$ is an iid Gaussian scalar process with mean zero and variance σ^2. The order, p, is assumed to be known while we develop the stationarity test. The characteristic equation associated with (2.1) is

$$\lambda^p-b_1\lambda^{p-1}-\ldots-b_p=0 \tag{2.2}$$

The stationarity of (2.1) is that the dominating roots of (2.2) are less than unity in moduli, and the non-stationarity is that the dominating roots are equal to or larger than unity in moduli. Our task is to test the former against the latter.

A portion of $\{y_t\}$, $y'=(y_1,\ldots,y_T)$ is observed. Construct a $T\times T$ matrix

$$S=\begin{bmatrix} 0 & 0 & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & 1 & 0 \end{bmatrix}$$

$Sy=(0,y_1,\ldots,y_{T-1})'$ and $S'y=(y_2,\ldots,y_T,0)'$ so that S^τ and S'^τ represent a τ periods lagg and forward respectively. Moreover

$$\begin{cases} S^{T+h}=0 \text{ for } h\geq 0 \\ SS'=I-e_1e_1',\ S'S=I-e_Te_T' \end{cases} \tag{2.3}$$

where $e_1'=(1,0,\ldots,0)$ and $e_T'=(0,\ldots,0,1)$. For any non-stochastic vector, $c'=(c_1,\ldots,c_q)$, we construct

$$S(c)\equiv I-c_1S-\ldots-c_qS^q$$

Let d_0 $(\equiv 1), d_1, d_2,\ldots$ be such that identity with respect to x

$$(1-c_1x-\ldots-c_qx^q)(1+d_1x+d_2x^2+\ldots)=1$$

No matter whether the sequence $\{d_j\}$ converges or not

$$S(c)^{-1} = I + d_1 S + \ldots + d_{T-1} S^{T-1}$$

holds exactly because of (2.3). Moreover $\det[S(c)] = 1 = \det[S(c)^{-1}]$.

1.3 Testing for stationarity in an expository model

Though we shall present later our testing procedure for a general order p, the basic ideas may be explained more simply for the simplest case

$$(1 - \rho L) y_t = \mu + \epsilon_t,\ t = 1, \ldots, T \tag{3.1}$$

where y_0 is unobservable, and $|\rho|$ may exceed unity. The root of the characteristic equation is ρ. We write $S(\rho) = I - \rho S$. No matter whether $|\rho| \geq 1$ or < 1, $S(\rho)^{-1} = I + S + \ldots + \rho^{T-1} S^{T-1}$ holds exactly. If $|\rho| \leq 1$ y_0 does not take an important part in the asymptotic analysis, but if $|\rho| > 1$ it may have a decisively important role.

Let $\iota = (1, \ldots, 1)$, and $\epsilon' = (\epsilon_1, \ldots, \epsilon_T)$. The matrix representation of (3.1) is

$$S(\rho) y - \rho y_0 e_1 = \mu \iota + \epsilon \tag{3.1'}$$

which gives the probability density function of y conditional upon y_0. Assume that y_0 is distributed in the normal distribution with mean $\mu*$ and variance $\sigma*^2$. The joint density of y and y_0 is derived, and from it one can obtain the density of y alone by integrating y_0 out. In fact y is distributed in the normal distribution with mean vector

$$\eta \equiv S(\rho)^{-1}(\mu\iota + \rho\mu* e_1)$$

and covariance matrix

$$V \equiv \sigma^2 S(\rho)^{-1}(1 I_T + \sigma*^2 \sigma^{-2} \rho^2 e_1 e_1') S(\rho)'^{-1}$$

both exactly. From these it follows that

$$-(2T)^{-1} \log \det[V] = -\tfrac{1}{2} \log \sigma^2 - (2T)^{-1} \log(1 + \sigma*^2 \sigma^{-2} \rho^2)$$

and

$$\begin{aligned} -(2T)^{-1}(y - \eta)' V^{-1}(y - \eta) = &-(2T\sigma^2)^{-1}([O, I_{T-1}](S(\rho)y - \mu\iota))' \\ &([O, I_{T-1}](S(\rho)y - \mu\iota)) - (2\sigma^2 T)^{-1}(1 + \sigma^{-2}\sigma*^2\rho^2)(y_1 - \mu - \rho\mu*)^2 \end{aligned} \tag{3.2}$$

No matter whether $|\rho| >, =, < 1$, T^{-1} times the log of the p.d.f. of y is, apart from $-\tfrac{1}{2}\log 2\pi$,

$$\begin{aligned} &-\tfrac{1}{2}\log \sigma^2 - (2\sigma^2 T)^{-1}([O, I_{t-1}](S(\rho)y - \mu\iota))'([O, I_{t-1}](S(\rho) - \mu\iota)) \\ &\quad + O_p(T^{-1}) \end{aligned} \tag{3.3}$$

and $(\mu*, \sigma*^2)$ is hidden in $O_p(T^{-1})$, which we ignore.

Equation (3.3) is simple to deal with, and it will provide a basis for a number of inference problems. Whichever of the maximum likelihood and Bayesian approaches may be chosen the least-squares estimator will be introduced in some context.

We wish to test stationarity, $|\rho| < 1$, against non-stationarity, $|\rho| \geq 1$, without singling out any particular value of ρ as a representation of either region. The Bayesian approach to such problem is to introduce prior probabilities, p_1 and p_2 respectively for $|\rho| < 1$ and $|\rho| \geq 1$ and then distribute p_1 and p_2 somehow among the ρ in each region. The prior distribution within $|\rho| < 1$ may be uniform, but that for $|\rho| \geq 1$ cannot be. Thus some values have to be chosen for several parameters. The maximum likelihood approach is the Cox (1961) test for separate families of hypotheses. To apply the method to our problem the stationarity and non-stationarity regions have to the separated.[5] In so far as the test is applied to the stability of an economic equilibrium, the "demilitarized zone" should be that part of the stationarity region which is adjacent to the non-stationarity region. With a small positive number, δ, the null hypothesis is $|\rho| \leq 1 - \delta$; the alternative hypothesis is $|\rho| \geq 1$; and in developing the testing procedure it is assumed that ρ never falls in the demilitarized zone, $1 > |\rho| > 1 - \delta$. The basic idea of Cox (1961) is to maximize the log-likelihood in each of the regions, $|\rho| \leq 1 - \delta$ and $|\rho| \geq 1$, and then compare the two maximized likelihoods. For the asymptotic theory it is found that δ may be $O(T^{-1+c})$, where c is an arbitrary small positive constant, but we shall assume that δ is a constant in order to simplify the following presentation. The choice of δ in practice and the distribution of the test statistic when ρ does fall in the demilitarized zone will be considered later in section 1.6. Note that μ and σ^2 are nuisance parameters.

If we proceed with Cox's idea, with the log-likelihood function (3.3), we would face a non-standard problem. This is because its maximum over $|\rho| \geq 1$ occurs at the boundary, $|\rho| = 1$, asymptotically with probability 1 provided that y is generated by a value of ρ such that $|\rho| < 1$.[6]

With a hint found in Quenouille (1957, p. 57) we shall demonstrate that the p.d.f. of $(y_1, \ldots, y_{T-1})$ conditional upon y_T leads us to a problem that is closer to a standard one. Using

$$\iota(\rho)' = (\rho^{T-1}, \rho^{T-2}, \ldots, 1) \tag{3.4}$$

which is the last row of $S(\rho)^{-1}$, the marginal distribution of y_T has mean

$$\mu_T \equiv \iota(\rho)'(\mu\iota + \rho\mu_* e_1)$$

and variance,

$$v_T \equiv \sigma^2 \iota(\rho)'\iota(\rho) + \sigma_*^2 \rho^{2T} = \begin{cases} \rho^{2T}(\sigma_*^2 + (\rho^2 - 1)^{-1}\sigma^2) - (\rho^2 - 1)^{-1}\sigma^2 \\ \qquad \text{if} \quad |\rho| \neq 1 \\ \sigma^2 T + \sigma_*^2 \text{ if } \rho = \pm 1 \end{cases}$$

both exactly. It is seen that

$$(2T)^{-1}\log v_T = \begin{cases} \frac{1}{2}\log \rho^2 + O(T^{-1}) & \text{if } |\rho| > 1 \\ (2T)^{-1}\log T + O(T^{-1}) & \text{if } \rho = \pm 1 \\ O(T^{-1}) & \text{if } |\rho| < 1 \end{cases} \tag{3.5}$$

$$(2T)^{-1}(y_T - \mu_T)^2 v_T^{-1} = O_p(T^{-1}) \text{ regardless of } |\rho| >, =, < 1 \tag{3.6}$$

Thus T^{-1} times the log of the p.d.f. of $(y_1, \ldots, y_{T-1})$ conditional upon y_T, i.e., the log-likelihood function divided by T, may be approximated as follows. Let $\theta = (\mu, \sigma^2, \rho)$, and for an arbitrarily small, positive δ define

$\Omega_s: (1-\delta) \geq |\rho|$; μ and σ^2 are unrestricted,
$\Omega_e: |\rho| \geq 1$; μ and σ^2 are unrestricted.

The log-likelihood function (divided by T) associated with Ω_s is

$$L_s(\theta) \equiv -\tfrac{1}{2}\log \sigma^2 - (2\sigma^2 T)^{-1}([O, I_{t-1}](S(\rho) - \mu\iota))' \\ ([O, I_{t-1}](S(\rho) - \mu\iota)) \tag{3.7.S}$$

Though (3.6) should not be dropped from the log-likelihood function for Ω_e at $\rho = \pm 1$ (unless μ is a priori specified as zero),[7] we shall consider

$$L_e(\theta) \equiv -\tfrac{1}{2}\log \sigma^2 + \tfrac{1}{2}\log \rho^2 - (2\sigma^2 T)^{-1}([O, I_{t-1}](S(\rho) - \mu\iota))' \\ ([O, I_{t-1}](S(\rho) - \mu\iota)) \tag{3.7.E}$$

defined over Ω_e. The only difference between (3.7.S) and (3.7.E) is the presence of $\frac{1}{2}\log \rho^2$ in the latter.

Following Cox(1961) $L_s(\theta)$ and $L_e(\theta)$ are maximized over Ω_s and Ω_e respectively. Let $\hat{\theta}_s \equiv (\hat{\mu}_s, \hat{\sigma}_s^2, \hat{\rho}_s)$ and $\hat{\theta}_e \equiv (\hat{\mu}_e, \hat{\sigma}_e^2, \hat{\rho}_e)$ be the values of θ at which the maxima occur. The null hypothesis is that the true value of θ, $\theta_0 \equiv (\mu_0, \sigma_0^2, \rho_0)$, is in the interior of Ω_s. It is well known that (i) plim $\hat{\theta}_s = \theta_0$ and (ii) $T^{1/2}(\hat{\theta}_s - \theta_0)$ converges in distribution to a normal distribution with zero mean vector and covariance matrix $(\text{plim} - \frac{\partial^2 L_s}{\partial\theta\partial\theta'}(\theta_0))^{-1}$ under the null hypothesis.

We are concerned with the asymptotic properties of $\hat{\theta}_e$ under the null hypothesis. Construct

$$\theta_{0e} \equiv (-\rho_0^{-1}\mu_0,\ \rho_0^{-2}\sigma_0^2,\ \rho_0^{-1}) \tag{3.8}$$

Lemma 1
Under the null hypothesis plim $\hat{\theta}_e = \theta_{0e}$.

Lemma 2
Under the null hypothesis $C_e \equiv \text{plim} - \frac{\partial^2 L_e}{\partial\theta\partial\theta'}(\theta_{0e})$ exists.

Both $T^{1/2}\frac{\partial L_e}{\partial\theta}(\theta_{0e})$ and $T^{1/2}(\hat{\theta}_e - \theta_{0e})$ are asymptotically normally distributed with zero mean. Their covariance matrices are C_e and C_e^{-1} respectively. Note that L_e is a log-likelihood divided by T.

Proof of Lemma 1 Let $f(\ )$ be the vector valued function representing the transformation from (μ,σ^2,ρ) to $(-\rho^{-1}\mu,\rho^{-2}\sigma^2,\rho^{-1})$. (i) If (μ,σ^2,ρ) is a θ in Ω_s, $f(\mu,\sigma^2,\rho)$ is in Ω_e. (ii) If (μ,σ^2,ρ) is a θ in Ω_e and if $|\rho|\geq(1-\delta)^{-1}$, then $f(\mu,\sigma^2,\rho)$ is in Ω_s. We wish to extend this latter relation to include $|\rho|$ lying between $(1-\delta)^{-1}$ and 1. In fact this is possible because (a) there is nothing that prevents the extension of the argument ρ up to $|\rho|=1$ in (3.7.S) in so far as $|\rho_0|<1$ and (b) the θ at which (3.7.S) is maximized remains unaffected (with probability as close to 1 as possible) even if Ω_s is thus extended, in so far as T is sufficiently large and $|\rho_0|\leq 1-\delta$. Notice also $f(f(\theta))=\theta$.

From (2.3) we obtain the identities

$$\begin{cases} S(\rho^{-1})'[\mathrm{O},I_{T-1}]'[\mathrm{O},I_{T-1}]S(\rho^{-1})=\rho^{-2}S(\rho)'[\mathrm{O},I_{T-1}]'[\mathrm{O},I_{T-1}]S(\rho) \\ \qquad +(1-\rho^{-2})(e_Te_T'-e_1e_1'), \\ \iota'[\mathrm{O},I_{T-1}]'[\mathrm{O},I_{T-1}]S(\rho^{-1})=-\rho^{-1}\iota'[\mathrm{O},I_{T-1}]'[\mathrm{O},I_{T-1}]S(\rho) \\ \qquad +(1+\rho^{-1})(e_T-e_1)' \end{cases} \tag{3.9}$$

One can verify that (3.7.E) can be rewritten precisely in

$$-\tfrac{1}{2}\log(\rho^{-2}\sigma^2)-(2(\rho^{-2}\sigma^2)T)^{-1}([\mathrm{O},I_{T-1}](S(\rho^{-1})y-(-\rho^{-1}\mu)\iota))' \\ ([\mathrm{O},I_{T-1}](S(\rho^{-1})y-(-\rho^{-1}\mu)\iota))+\hat{\delta} \tag{3.10}$$

where

$$\hat{\delta}=(2\sigma^2T)^{-1}((\rho^2-1)(y_T^2-y_1^2)+2\mu(1+\rho)(y_T-y_1)-\mu^2)$$

Since (3.10) is $L_s(f(\theta))+\hat{\delta}$, we have

$$L_e(\theta)=L_s(f(\theta))+\hat{\delta},\ \theta\epsilon\Omega_e \tag{3.11}$$

Suppose that T is sufficiently large and that the null hypothesis holds. Then $\hat{\delta}$ is negligible and $\hat{\theta}_s$ maximizes L_s not only over Ω_s but also over its extension mentioned above. Because of (3.11) the global maximum of $L_e(\theta)$ within Ω_e must be attained near the θ such that $f(\theta)=\hat{\theta}_s$, i.e., near $f(\hat{\theta}_s)$. Notice that $\theta_{0e}=f(\theta_0)$. The analysis of the local condition is given in appendix 1.

Proof of Lemma 2 Construct a 3×3 matrix

$$A=\begin{bmatrix} -\rho_0 & 0 & 0 \\ 0 & \rho_0^2 & 0 \\ -\rho_0\mu_0 & -2\rho_0\sigma_0^2 & -\rho_0^2 \end{bmatrix}$$

In fact A^{-1} is the Jacobian, $\partial f'/\partial\theta$, evaluated at θ_0. In (3.11) replace θ by $f(\theta)$ on the left-hand side and $f(\theta)$ by θ on the right-hand side with $f(\theta)\epsilon\Omega_s$, and then differentiate both sides by θ. It is seen[8] that

$$T^{1/2}\frac{\partial L_e}{\partial \theta}(\theta_{0e}) = A T^{1/2}\frac{\partial L_s}{\partial \theta}(\theta_0) + O_p(T^{-1/2}) \tag{3.12.E}$$

$$-\text{plim}\frac{\partial^2 L_e}{\partial \theta \partial \theta'}(\theta_{0e}) = A(-\text{plim}\frac{\partial^2 L_s}{\partial \theta \partial \theta'}(\theta_0))A' \tag{3.13.E}$$

where plim is taken under the null hypothesis. The left-hand sides of (3.12.E) and (3.13.E) are well understood in terms of the standard asymptotic theory. Through a Taylor expansion of $\frac{\partial L_e}{\partial \theta}(\hat{\theta}_e) = 0$, which should hold if T is sufficiently large,

$$T^{1/2}(\hat{\theta}_e - \theta_{0e}) \simeq (-\text{plim}\frac{\partial^2 L_e}{\partial \theta \partial \theta'}(\theta_{0e}))^{-1} T^{1/2}\frac{\partial L_e}{\partial \theta}(\theta_{0e}) \tag{3.14.E}$$

Lemma 2 follows from (3.12.E), (3.13.E), (3.14.E) and the well-known fact that $T^{1/2}\frac{\partial L_s}{\partial \theta}(\theta_0)$ is asymptotically distributed in the normal distribution with zero mean and the covariance matrix, $-\text{plim}\frac{\partial^2 L_s}{\partial \theta \partial \theta'}(\theta_0)$. QED

Following Cox (1961) let us try a test statistic,

$$L_s(\hat{\theta}_s) - L_e(\hat{\theta}_e) - \text{an estimate of } E(L_s(\hat{\theta}_s) - L_e(\hat{\theta}_e)) \tag{3.15}$$

where the expectation is under the null hypothesis.

Lemma 3
An estimate of $E(L_s(\hat{\theta}_s) - L_e(\hat{\theta}_e))$ is zero. If y_0 is distributed in the normal distribution with mean $(1-\rho_0)^{-1}\mu_0$ and variance $(1-\rho_0^2)^{-1}\sigma_0^2$, i.e., the stationary mean and variance

$$Q \equiv L_s(\hat{\theta}_s) - L_e(\hat{\theta}_e) \simeq (2T)^{-1}(\xi^2 - \eta^2) \tag{3.16}$$

where ξ^2 and η^2 are both χ^2 variable with 1 degree of freedom and mutually independent.

Proof Consider

$$L_s(\hat{\theta}_s) \simeq L_s(\theta_0) + \frac{\partial L_s}{\partial \theta'}(\theta_0)(\hat{\theta}_s - \theta_0)$$

$$L_e(\hat{\theta}_e) \simeq L_e(\theta_{0e}) + \frac{\partial L_e}{\partial \theta'}(\theta_{0e})(\hat{\theta}_e - \theta_{0e})$$

Using (3.12.E), (3.13.E), and (3.14.E) and

$$T^{1/2}(\hat{\theta}_s - \theta_0) \simeq (\text{plim}\frac{\partial L_s}{\partial \theta \partial \theta'}(\theta_0))^{-1} T^{1/2}\frac{\partial L_s}{\partial \theta}(\theta_0) \tag{3.14.S}$$

we see that

$$\left\| \frac{\partial L_s}{\partial \theta'}(\theta_0)(\hat{\theta}_s - \theta_0) - \frac{\partial L_e}{\partial \theta'}(\theta_{0e})(\hat{\theta}_e - \theta_{0e}) \right\|$$

is in the order of $o_p(T^{-1})$ rather than $O_p(T^{-1})$ so that

$$L_s(\hat{\theta}_s) - L_e(\hat{\theta}_e) = L_s(\theta_0) - L_e(\theta_{0e}) + o_p(T^{-1})$$

It is easily seen[9] that

$$L_s(\theta_0) - L_e(\theta_{0e}) = (2\sigma_0^2 T)^{-1}(1-\rho_0^2)\{(y_1 - \frac{\mu_0}{1-\rho_0})^2 - (y_T - \frac{\mu_0}{1-\rho_0})^2\} \tag{3.17}$$

holds exactly. As $T \to \infty\, y_1$ and y_T become independent. QED

We propose to test $|\rho| \leq 1-\delta$ against $|\rho| \geq 1$ with the statistic, $Q \equiv L_s(\hat{\theta}_s) - L_e(\hat{\theta}_e)$. Let $\zeta = \xi^2 - \eta^2$, where ξ^2 and η^2 are mutually independent χ^2 (1). The c such that Prob[$|\zeta| > c$]=0.05 is 4.364, and the c such that Prob[$|\zeta| > c$]=0.01 is 7.208.[10] Our test is asymptotically similar as the distribution of Q is asymptotically free from μ and σ^2. The proof of Lemma 3 shows that the test examines the probability structure of the first and the last observations while using all the observations to estimate θ_0 and θ_{0e}.

The readers might be perplexed by the fact that the expectation of (3.17) is zero, which is contradictory to the well-known inequality of information theory. The explanation is that (3.17) is $O_p(T^{-1})$, while the terms neglected in arriving at (3.7.S) and (3.7.E) are also $O_p(T^{-1})$; i.e., if we had not attempted any approximations such as in (3.2) and (3.5) and in dropping (3.6), then we should have a strictly positive expectation of (3.17). It would however complicate enormously our computation of $\hat{\theta}_s$ and $\hat{\theta}_e$ especially in the general case that will be considered in the next section. We have decided to proceed along with (3.7.S) and (3.7.E) without being bothered by the inequality of the information theory.

Our next task is to demonstrate the consistency of the above test. The alternative hypothesis is that the true value of θ, $\theta_0 \equiv (\mu_0, \sigma_0^2, \rho_0)$, is in Ω_e, in particular, $|\rho_0| \geq 1$.

Lemma 4
Under the alternative hypothesis plim $\hat{\theta}_e = \theta_0$.

Proof In both (3.7.S) and (3.7.E) we shall replace $-(2\sigma^2 T)^{-1}$ $([O, I_{T-1}]$ $(S(\rho)y - \mu\iota))'([O, I_{T-1}](S(\rho)y - \mu\iota))$ by $-(2\sigma^2 T)^{-1}(S(\rho)y - \mu\iota)'(S(\rho)y - \mu\iota)$. The replacement has no effect upon the result while simplifying the following demonstration. The concentration of the revised (3.7.E) with respect to μ and σ^2 gives

$$\tfrac{1}{2}\log \rho^2 - \tfrac{1}{2}\log\{T^{-1}(S(\rho)y - T^{-1}\iota' S(\rho)y\iota)'(S(\rho)y - T^{-1}\iota' S(\rho)y\iota)\} \tag{3.18.E}$$

which must be maximized over Ω_e.

(a) The case where $\rho_0 = \pm 1$. Let

$$u \equiv (\mu_0\iota + \epsilon + \rho_0 y_0 e_1)$$

Then $y = \sum_{j=0}^{T-1}(\pm 1)^j S^j u$, and the second term of (3.18.E) is

$$-\tfrac{1}{2}\log\{T^{-1}\, u'(I + (1 \mp \rho)\sum_{j=1}^{T-1}(\pm 1)^j S^j)'(I - T^{-1}\iota\iota')(I + (1 \mp \rho)\sum_{j=1}^{T-1}(\pm 1)^j S^j)u\}$$

At $\rho = \pm 1$ the term in $\{\quad\}$ is $O_p(1)$, but at $|\rho| > 1$ or $\rho = 1$ it is $O_p(T)$ if $\mu_0 = 0$, and $O_p(T^2)$ if $\mu_0 \neq 0$. The probability that (3.18.E) attains its maximum at $|\rho| > 1$ or $\rho = \mp 1$ is made arbitrarily small by taking T sufficiently large. Thus plim $\hat{\rho}_e = \pm 1$.

(b) The case where $|\rho_0| > 1$. From (3.1) we have

$$S(\rho_0^{-1})'y = -\rho_0^{-1}S'u + y_T e_T$$

i.e.

$$y = -\rho_0^{-1}S(\rho_0^{-1})^{-1\prime}S'u + y_T\iota(\rho_0^{-1}) \tag{3.19.E}$$

where $\iota(\quad)$ is defined in (3.4). (3.19.E) decomposes y into a stationary process and an explosive term. When (3.19.E) is substituted into (3.18.E) the term inside $\{\quad\}$ in the second log becomes

$$\begin{aligned} &T^{-1}[-\rho_0^{-1}S(\rho_0^{-1})^{-1\prime}S'u + y_T\iota(\rho_0^{-1})]'S(\rho)'(I - T^{-1}\iota\iota') \\ &\quad S(\rho)[-\rho_0^{-1}S(\rho_0^{-1})^{-1\prime}S'u + y_T\iota(\rho_0^{-1})] \end{aligned} \tag{3.20.E}$$

Concerning the maximization of (3.18.E), $\tfrac{1}{2}\log \rho^2$ is dominated by the log of (3.20.E), and within (3.20.E).

$$\begin{aligned} &T^{-1}y_T^2\iota(\rho_0^{-1})'S(\rho)'(I - T^{-1}\iota\iota')S(\rho)\iota(\rho_0^{-1}) \\ &= T^{-1}y_T^2\{(1 - \rho\rho_0^{-1})^2[(1 - \rho_0^{-2})^{-1} - T^{-1}(1 - \rho_0^{-1})^2] + O(|\rho_0|^{-2T})\} \end{aligned} \tag{3.21.E}$$

dominates other terms. In (3.21.E) $\{\quad\}$ becomes $O(|\rho_0|^{-2T})$ only at $\rho = \rho_0$. Asymptotically in probability 1 the maximum of (3.20.E) occurs at $\rho = \rho_0$, where y_T disappears from (3.20.E). It immediately follows that plim $\hat{\sigma}_e^2 = \sigma_0^2$, plim $\hat{\mu}_e = \mu_0$.

Lemma 5
The test statistic Q is unbounded toward minus infinity if $|\rho_0| \geq 1$.

Proof (a) It follows from Lemma 4 that $L_e(\hat{\theta}_e)$ converges in probability to $-\frac{1}{2}\log \sigma_0^2 + \frac{1}{2}\rho_0^2 - \frac{1}{2}$. (b) Consider $L_s(\hat{\theta}_s)$ when $|\rho_0| \geq 1$. As $T \to \infty$ $\hat{\rho}_s$ approaches to $1-\delta$ if $\rho_0 \geq 1$, and to $-(1-\delta)$ if $\rho_0 \leq -1$. Thus plim $\hat{\rho}_s$ exists. It can be shown that $\hat{\mu}_s$ and $\hat{\sigma}_s^2$ are not even bounded in probability. Thus $L_s(\hat{\theta}_s) = -\frac{1}{2}\log \hat{\sigma}_s^2 - \frac{1}{2}$ is unbounded.

Thus the consistency of our test has been established.

The above test assumes that the unobserved initial y_0 is distributed with the stationary mean and variance. An alternative test statistic will be shown in section 1.5. Its asymptotic distribution is free from the mean and variance of y_0, but it has a demerit about the size of the test when T is small and μ^2 is large in comparison with σ^2.

1.4 Testing for stationarity in the general model

For the case where the order of autoregression, p, is not necessarily 1, our testing procedure consists of (a) one root test, (b) two roots test, and (c) combining them. Henceforth a root being non-stationary will mean that the root is equal to or larger than unity in its modulus. The model behind the one root test is a special case of (2.1),

$$(1-\rho L)(1-a_1L-\ldots-a_{p-1}L^{p-1})y_t = \mu + \epsilon_t \tag{4.1}$$

where $|\rho|$ may be non-stationary, but the characteristic equation

$$\lambda^{p-1} - a_1\lambda^{p-2} - \ldots - a_{p-1} = 0 \tag{4.2}$$

has all roots less than unity in moduli. The largest of the absolute values of roots of (4.2) is denoted by α, and if $p=1$, α is set to zero. When $p \geq 2$ we assume that $|\rho|$ exceeds α for the purpose of identification. $S(a) = I - a_1S - \ldots - a_{p-1}S^{p-1}$ for $a' \equiv (a_1, \ldots, a_{p-1})$. Note $S(\rho)S(a) = S(a)S(\rho)$. We write

$$S(a)^{-1} = I + d_1S + \ldots + d_{T-1}S^{T-1}$$

Here $\{d_j\}$ is bounded by a decaying exponential. Denote

$$\tilde{a}' \equiv (-a_{p-1}, \ldots, -a_1, 1)$$

$$R(a) \equiv \begin{bmatrix} 0 & -a_{p-1} & \cdot & -a_1 \\ & & & \cdot \\ & & & -a_{p-1} \\ 0 & \cdots & & 0 \\ \cdot & & & \cdot \\ 0 & \cdots & & 0 \end{bmatrix} \qquad (T \times p)$$

It is assumed that the initials, $y*' \equiv (y_{-p+1}, \ldots, y_0)$ is distributed in a normal distribution with mean $\mu*$ and covariance matrix $\sum*$. The model is

$$S(\rho)S(a)y + (S(\rho)R(a) - \rho e_1 \tilde{a}')y* = \mu\iota + \epsilon$$

Integrating $y*$ out of the joint density of y and $y*$, we have the normal p.d.f. of y alone with mean

$$\eta \equiv S(\rho)^{-1}S(a)^{-1}(\mu\iota - R(\rho, a)\mu*)$$

and covariance matrix

$$V \equiv \sigma^2 S(\rho)^{-1}S(a)^{-1}(I + \sigma^{-2}R(\rho, a)\textstyle\sum*R(\rho, a)')S(a)'^{-1}S(\rho)'^{-1}$$

where $R(\rho, a) \equiv S(\rho)R(a) - \rho e_1 \tilde{a}'$. Note that

$$-\tfrac{1}{2}\log\det[V] = -\frac{T}{2}\log\sigma^2 - \tfrac{1}{2}\log\det[I_p + \sigma^{-2}R(\rho, a)'R(\rho, a)\textstyle\sum*]$$

where the second term on the right-hand side is O(1). Thus T^{-1} times the log of the p.d.f. of y is

$$-\tfrac{1}{2}\log\sigma^2 - (2\sigma^2 T)^{-1}([\mathrm{O}, I_{T-P}](S(\rho)S(a)y - \mu\iota))'$$
$$([\mathrm{O}, I_{T-P}](S(\rho)S(a)y - \mu\iota)) + \mathrm{O}_p(T^{-1})$$

and $\mu*$ and $\sum*$ are hidden in $\mathrm{O}_p(T^{-1}$.

As for the marginal distribution of y_T, its exact variance is

$$v_T \equiv \sigma^2\iota(\rho)'S(a)^{-1}[I + \sigma^{-2}R(\rho, a)\textstyle\sum*R(\rho, a)']S(a)^{-1\prime}\iota(\rho)$$

Thus $(-2T)^{-1}$ log $v_T = (-2T)^{-1}$ log $\iota(\rho)'S(a)^{-1}S(a)^{-1\prime}\iota(\rho)$ apart from $\mathrm{O}(T^{-1})$, and

$$\iota(\rho)'S(a)^{-1}S(a)^{-1\prime}\iota(\rho) = \sum_{i=0}^{T-1}\sum_{j=0}^{T-1} d_i d_j \rho^{|i-j|}(1 + \rho^2 + \ldots + \rho^{2(T-1-\max(i,j))})$$

$$= \begin{cases} \rho^{2T}(\rho^2 - 1)^{-1}(\sum_{i=0}^{\infty}\rho^{-i}d_i)^2(1 + \mathrm{o}(|\rho|^{-T})), & \text{if } |\rho| > 1 \\ T(\sum_{i=0}^{\infty}(\pm 1)^i d_i)^2(1 + \mathrm{O}(T^{-1}), & \text{if } \rho = \pm 1 \\ \mathbf{O}(1) & \text{if } |\rho| < 1 \end{cases}$$

Our parameter is $\theta^{(1)} \equiv (\mu, \sigma^2, \rho, a')$. For non-stationarity we construct

$$L_e^{(1)}(\theta^{(1)}) = -\tfrac{1}{2}\log\sigma^2 + \tfrac{1}{2}\log\rho^2 - (2\sigma^2 T)^{-1}([O, I_{T-P}](S(\rho)S(a)y - \mu\iota))'([O, I_{T-P}](S(\rho)S(a)y - \mu\iota))$$

over the domain

$$\Omega_e^{(1)}: |\rho| \geq 1,\ |\rho|a < 1,\ \mu \text{ and } \sigma^2 \text{ are unrestricted.}$$

If we had eliminated the above condition, $|\rho|a < 1$, there would exist asymptotically more than one isolated point at which the maximum is attained, which involves some inconvenience if not a theoretical difficulty.[11] For stationarity L_s is

$$L_s^{(1)}(\theta^{(1)}) = -\tfrac{1}{2}\log\sigma^2 - (2\sigma^2 T)^{-1}([O, I_{T-P}](S(\rho)S(a)y - \mu\iota))'([O, I_{T-P}](S(\rho)S(a)y - \mu\iota))$$

defined over

$$\Omega_s^{(1)}: 1 - \delta \geq |\rho| > a,\ \mu \text{ and } \sigma^2 \text{ are unrestricted.}$$

The demilitarized zone is $1 > |\rho| > 1 - \delta$. $\hat{\theta}_s^{(1)}$ and $\hat{\theta}_e^{(1)}$ are the $\theta^{(1)}$ where $L_s^{(1)}$ and $L_e^{(1)}$ are maximized over $\Omega_s^{(1)}$ and $\Omega_e^{(1)}$ respectively.

The proofs of the following theorems are quite analogous to those in the previous section.

Theorem 1
If y is generated by $\theta_0^{(1)}$ in the interior of $\Omega_s^{(1)}$ in the model (4.1), plim $\hat{\theta}_e^{(1)} = \theta_{0e}^{(1)} \equiv (-\rho_0^{-1}\mu_0, \sigma_0^{-2}\rho_0^2, \rho_0^{-1}, a_0')$.

Theorem 2
Under the same condition $C^{(1)} \equiv -\text{plim}\dfrac{\partial^2 L_e^{(1)}}{\partial\theta^{(1)}\partial\theta^{(1)\prime}}(\theta_{0e}^{(1)})$ exists. Both $T^{1/2}\dfrac{\partial L_e^{(1)}}{\partial\theta^{(1)}}(\theta_{0e}^{(1)})$ and $T^{1/2}(\hat{\theta}_{0e}^{(1)} - \theta_{0e}^{(1)})$ are asymptotically normally distributed with zero mean. The covariance matrices are $C^{(1)}$ and $C^{(1)-1}$ respectively.

Theorem 3
Under the same condition and under the assumption that y_* is distributed with the stationary mean vector and covariance matrix,

$$Q^{(1)} \equiv L_s^{(1)}(\hat{\theta}_s^{(1)}) - L_e^{(1)}(\hat{\theta}_e^{(1)}) \simeq (2T)^{-1}(\xi_{(1)}^2 - \eta_{(1)}^2)$$

where $\xi_{(1)}^2$ and $\eta_{(1)}^2$ are mutually independent χ^2 (1).

The critical points have been indicated in section 1.3.

Theorem 4
If $\theta_0^{(1)}$ is in $\Omega_e^{(1)}$ in the model (4.1), plim $\hat{\theta}_e^{(1)} = \theta_0^{(1)}$.

The only extension of the proof of Lemma 4 which deserves note is that (3.21.E) is here

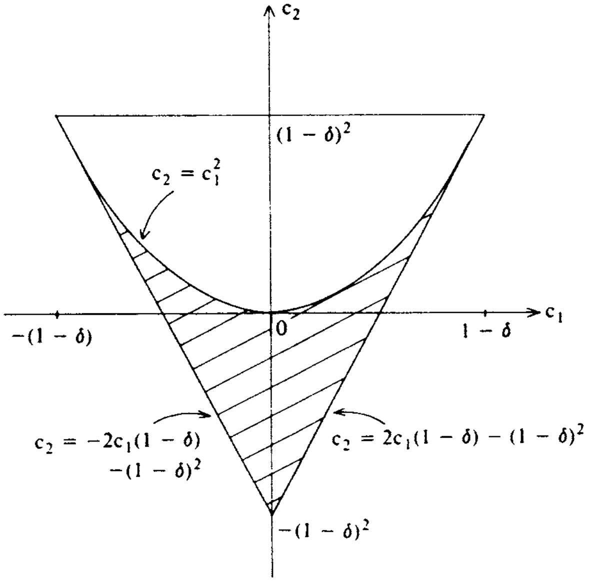

Figure 1.1 Stability region

$$\begin{aligned} &T^{-1}y_T^2\iota(\rho_0^{-1})'S(a)'S(\rho)'(I-T^{-1}\iota\iota')S(\rho)S(a)\iota(\rho_0^{-1}) \\ &\quad = T^{-1}y_T^2[(1-\rho\rho_0^{-1})^2(1-\rho_0^{-1}a_1-\ldots-\rho_0^{-p+1}a_{p-1})^2 \times \\ &\qquad (1-\rho_0^{-2})^{-1}-T^{-1}(1-\rho_0^{-1})^2+\mathrm{O}(|\rho_0|^{-2T})] \end{aligned}$$

Since $a=(a_1,\ldots,a_{p-1})$ is restricted to make the roots of (4.2) less than unity in moduli $(1-\rho_0^{-1}a_1-\ldots-\rho_0^{-p+1}a_{p-1})^2$ can never be zero, thus necessitating $\rho=\rho_0$. Finally

Theorem 5
The test statistic $Q^{(1)}$ is unbounded toward minus infinity if $\theta_0^{(1)}$ is in $\Omega_e^{(1)}$.

The above test will be called the one root test.

The model behind the two roots test is

$$(1-2c_1L+c_2L^2)(1-a_1L-\ldots-a_{p-2}L^{p-2})y_t=\mu+\epsilon_t \tag{4.4}$$

Here we assume that the characteristic equation, $\lambda^{p-2}-a_1\lambda^{p-3}-\ldots-a_{p-2}=0$, has all roots less than $\alpha(<1)$ in absolute values. The equation

$$\lambda^2-2c_1\lambda+c_2=0 \tag{4.5}$$

has a pair of mutually conjugate complex roots or two real roots that may possibly be non-stationary. Any points inside the whole triangle in figure 1.1 represent the case where both of two roots of (4.5) have absolute values

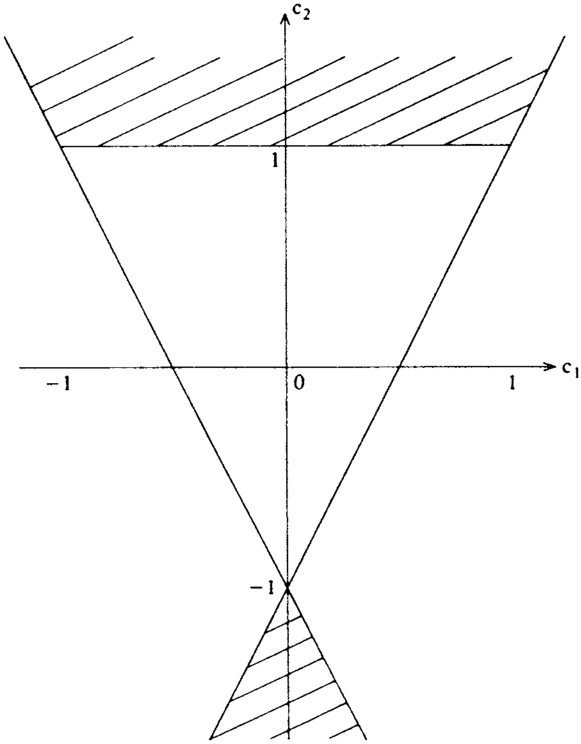

Figure 1.2 Instability region

less than $(1-\delta)$, whereas any points in the shaded area of figure 1.2 represent the case where both roots have absolute values greater than or equal to 1. The two roots test compares the triangle in figure 1.1 and the shaded area in figure 1.2. On the other hand the one root test compares the shaded area in the triangle in figure 1.1 and the unshaded area outside the triangle in figure 1.2.

The distribution of $(y_1, \ldots, y_T)$ is obtained through integrating the initial unobservables out, and the distribution of $(y_1, \ldots, y_{T-2})$ conditional upon (y_{T-1}, y_T) is derived. Let V_T be the covariance matrix of (y_{T-1}, y_T). If (4.5) has complex roots, $\rho\exp(i\omega)$[12] and $\rho\exp(-i\omega)$ and if $|\rho| > 1$, it can be shown that

$$\det[V_T] = \rho^{4T} \Big| \sum_{j=0}^{\infty} \rho^{-j} d_j \exp(ij\omega) \Big|^4 (1-\rho^2)^{-2} |1-\rho^2 \exp(i2\omega)|^{-1} \times (1 + o(\rho^{-2T}))$$

Therefore $(2T)^{-1}\log\det[V_T]\simeq\frac{1}{2}\log\rho^4=\frac{1}{2}\log c_2^2$. If (4.5) has two real roots, ρ_1, ρ_2, it also holds that $(2T)^{-1}\log\det[V_T]\simeq\frac{1}{2}\log\rho_1^2\rho_2^2=\frac{1}{2}\log c_2^2$. Thus no matter whether (4.5) has real or complex roots, we define

$$L_e^{(2)}(\theta^{(2)})\equiv-\tfrac{1}{2}\log\sigma^2+\tfrac{1}{2}\log c_2^2-(2\sigma^2T)^{-1}([O,I_{T-P}](S(c)S(a)y-\mu\iota))'([O,I_{T-P}](S(c)S(a)y-\mu\iota)) \quad (4.6.E)$$

where $S(c)=I-2c_1S+c_2S^2$, $S(a)=I-a_1S-\ldots-a_{p-2}S^{p-2}$, and $\theta^{(2)}\equiv(\mu,\sigma^2,c_1,c_2,a')$. As regards the roots of (4.5) let $\bar{\rho}$ be $\max(|\rho_1|,|\rho_2|)$ in the case of two real roots and $|\rho|$ in the case of complex roots. The domain of (4.6.E) is

$\Omega_e^{(2)}$: both of two roots of (4.5) are non-stationary; $\bar{\rho}\alpha<1$, μ and σ^2 are unrestricted.

As for the stationarity,

$$L_s^{(2)}(\theta^{(2)})\equiv-\tfrac{1}{2}\log\sigma^2-(2\sigma^2T)^{-1}([O,I_{T-P}](S(c)S(a)y-\mu\iota))'([O,I_{T-P}](S(c)S(a)y-\mu\iota)) \quad (4.6.S)$$

$\Omega_s^{(2)}$: (4.5) has both of two roots between $1-\delta$ and α in absolute values; μ and σ^2 are unrestricted.

We define $\theta_0^{(2)}\equiv(\mu_0,\sigma_0^2,c_{01},c_{02},a_0')$, $\hat{\theta}_s^{(2)}$, and $\hat{\theta}_e^{(2)}$ in the same way as above.

Theorem 6
If y is generated with $\theta_0^{(2)}$ in the interior of $\Omega_s^{(2)}$, then plim $\hat{\theta}_e^{(2)}=\theta_{0e}^{(2)}$, which is defined as follows. Two equations, $\lambda^2-2c_{01}\lambda+c_{02}=0$ and $c_{0e2}\lambda^2-2c_{0e1}\lambda+1=0$ have identical roots, which defines (c_{0e1},c_{0e2}). $\mu_{0e}=c_{02}^{-1}\mu_0$, $\sigma_{0e}^2=c_{02}^{-2}\sigma_0^2$, $a_{0e}=a_0$.

A theorem analogous to Theorem 2 holds. Moreover we have

Theorem 7
Under the same condition as in Theorem 6

$$Q^{(2)}\equiv L_s^{(2)}(\hat{\theta}_s^{(2)})-L_e^{(2)}(\hat{\theta}_e^{(2)})\simeq(2T)^{-1}(\xi_{(2)}^2-\eta_{(2)}^2)$$

where $\xi_{(2)}^2$ and $\eta_{(2)}^2$ are mutually independent $\chi_{(2)}^2$, provided that y_* is distributed in the stationary mean vector and covariance matrix.

The p.d.f. of $\zeta_{(2)}\equiv\xi_{(2)}^2-\eta_{(2)}^2$ is symmetric about zero, and the p.d.f. over $\zeta_{(2)}>0$ is identical to $\frac{1}{2}\times$(the p.d.f. of $\chi_{(2)}^2$). Thus c such that $\text{Prob}[|\zeta_{(2)}|>c]=0.05$ is 5.991.

Theorems analogous to Theorems 4 and 5 also hold.

How to compute $\hat{\theta}_s^{(1)}$, $\hat{\theta}_e^{(1)}$, $Q^{(1)}$, $\hat{\theta}_s^{(2)}$, $\hat{\theta}_e^{(2)}$, $Q^{(2)}$ is explained in appendix 3.

Let us consider how the one root and the two roots test can be combined. Since (4.1) and (4.4) are both special cases of (2.1), let the roots of (2.2) be r_1, r_2, $r_3,\ldots,r_p$ in the order descending in absolute values. Three cases may be distinguished: (i) r_1 and r_2 are both real, and r_3 may be real or complex; (ii) r_1

is real, and r_2 and r_3 are complex conjugates; and (iii) r_1 and r_2 are complex conjugates, and r_3 may be real or complex. The model of the one root test consists of (i) and (ii), but does not include (iii). The model of the two roots test consists of (i) and (iii), but does not include (ii). The unconstrained OLS estimate of $(b_1, \ldots, b_p)$ in (2.1) may be used to determine which of (i), (ii), and (iii) has the highest likelihood. If (ii) is found most likely one should accept the result in the one root test and ignore the two roots test. If (iii) is found most likely one should accept the result in the two roots test only. If (i) is found most likely, the stationarity rejected in the one root test but not rejected in the two roots test means a single real root being non-stationary, and the stationarity rejected in the two roots test but not rejected in the one root test means two real roots being non-stationary.

1.5 An alternative group of tests

We may rewrite (3.2) as

$$-(2T)^{-1}(y-\eta)'V^{-1}(y-\eta) = -(2T\sigma^2)^{-1}(S(\rho)y-\mu\iota)'(S(\rho)y-\mu\iota)$$
$$-(2\sigma^2T)^{-1} \text{ (a quadratic function of } (y_1-\mu-\rho\mu*) \text{ with coefficients involving } \mu*,\ \sigma*^2,\ \sigma^2, \text{ and } \rho).$$

A test alternative to the one in section 1.3 may be based on

$$\tilde{L}_s(\theta) \equiv -\tfrac{1}{2}\log\sigma^2 - (2\sigma^2T)^{-1}(S(\rho)y-\mu\iota)'(S(\rho)y-\mu\iota),$$
$$\tilde{L}_s(\theta) \equiv -\tfrac{1}{2}\log\sigma^2 + \tfrac{1}{2}\log\rho^2 - (2\sigma^2T)^{-1}(S(\rho)y-\mu\iota)'(S(\rho)y-\mu\iota)$$

each to be maximized over Ω_s and Ω_e respectively. Let $\tilde{\theta}_s \equiv (\tilde{\mu}_s, \tilde{\sigma}_s^2, \tilde{\rho}_s)$ be the θ at which the maximum of $\tilde{L}_s(\theta)$ occurs. Even if the unobserved initial, y_0, is not distributed with the stationary mean and variance,

$$\tilde{Q} \equiv L_s(\tilde{\theta}_s) - L_e(\tilde{\theta}_e) + (2T)^{-1}(1-\tilde{\mu}_s^2\tilde{\sigma}_s^{-2}(1-\tilde{\rho}_s)^{-1}(1+\tilde{\rho}_s))$$

is distributed as $(-2T)^{-1}(\chi^2(1)-1)$ under the null hypothesis of stationarity. This test is consistent.

Regarding the model (4.1) replace $-(2\sigma^2T)^{-1}([\mathrm{O}, I_{T-p}](S(\rho)S(a)y-\mu\iota))'([\mathrm{O}, I_{T-p}](S(\rho)S(a)y-\mu\iota))$ by $-(2\sigma^2T)^{-1}(S(\rho)S(a)y-\mu\iota)'(S(\rho)S(a)y-\mu\iota)$ in the definitions of $L_e^{(1)}(\ )$ and $L_s^{(1)}(\ )$. With $\tilde{L}_e^{(1)}(\ )$ and $\tilde{L}_s^{(1)}(\ )$ thus constructed,

$$\tilde{Q}^{(1)} \equiv \tilde{L}_s^{(1)}(\tilde{\theta}_s^{(1)}) - \tilde{L}_e^{(1)}(\tilde{\theta}_e^{(1)}) + (2T)^{-1}(1-\tilde{\mu}_s^{(1)2}\tilde{\sigma}_s^{(1)-2}(1-\tilde{\rho}_s^{(1)})^{-1}(1+\tilde{\rho}_s^{(1)}))$$

is asymptotically distributed as $(-2T)^{-1}(\chi^2(1)-1)$. Regarding the model (4.4) the same replacement in $L_s^{(2)}(\ )$ and $L_e^{(2)}(\ )$ yields $\tilde{L}_s^{(2)}(\ )$ and $\tilde{L}_e^{(2)}(\ )$, and

$$\tilde{Q}^{(2)} \equiv \tilde{L}_s^{(2)}(\tilde{\theta}_s^{(2)}) - \tilde{L}_e^{(2)}(\tilde{\theta}_e^{(2)}) + T^{-1}(1-\tilde{\mu}_s^2\tilde{\sigma}_s^{-2}(1-\tilde{c}_{2s})(1-2\tilde{c}_{1s}+\tilde{c}_{2s})^{-1})$$

Table 1.1[13]

$P[2T|Q^{(1)}| > 4.364]$ $p=1$, $\mu_0=0$ a priori specified
$T=150$, 1,000 replications

ρ_0	$1-\delta=0.93$	$1-\delta=0.97$
0.90	0.05*	0.04*
0.95	0.27**	0.06*
0.98	0.69**	0.27**
1.00	0.91***	0.68***
1.02	0.98***	0.95***

Table 1.2a

$P[2T|Q^{(1)}| > 4.364]$ $T=100$, 1,000 replications
$p=2$, $\mu_0=0$ a priori specified
$\rho_0=1$

a_{01}	$1-\delta=0.92$	$1-\delta=0.95$	$1-\delta=0.98$
0	0.76***	0.66***	0.36***
0.5	0.69***	0.62***	0.35***
0.8	0.56***	0.45***	0.30***

Note:
The first column is the true value of a_1 in (4.1). Even in the case where $a_{01}=O$, g_1 is estimated.

is asymptotically distributed as $(-2T)^{-1}(\chi^2(2)-2)$. The asymptotic distributions of both $\tilde{Q}^{(1)}$ and $\tilde{Q}^{(2)}$ are free from the mean vector and covariance matrix of y.

1.6 Simulation studies

So far our simulations have been made only for the cases where a great deal is a priori known; i.e., not only p is known but it is also known which one of the models behind the one root and the two roots test holds. Our major concern here is (1) how small δ can be in the lengths of time series data that are usually used in econometrics, and (2) how the test statistics $Q^{(1)}$ and $Q^{(2)}$ developed in section 1.4 would behave if the true value of dominating roots falls in the demilitarized zone.

Tables 1.1 through 1.4 show the probabilities of the critical regions constructed for the 5 percent significance level. The cases where ρ_0 falls in the region of the null hypothesis, the demilitarized zone, and the region of the alternative hypothesis are indicated by *, **, and *** respectively. In

Table 1.2b

$P[2T|Q^{(1)}|>4.364]$ $\rho_0=1$, $T=200$

a_{01}	$1-\delta=0.95$	$1-\delta=0.98$
0	0.88***	0.53***
0.5	0.86***	0.52***
0.8	0.69***	0.46***

Table 1.3

$P[2T|Q^{(2)}|>5.991]$ $T=150$, $p=2$, 1,000 replications

(a) Complex root,[14] the roots of (4.5) are $\rho\exp(i\omega)$ and $\rho\exp(-i\omega)$

	$1-\delta=0.93$				$1-\delta=0.97$			
ρ_0	0.95		1.00		0.95		1.00	
ω_0	$\pi/4$	$\pi/2$	$\pi/4$	$\pi/2$	$\pi/4$	$\pi/2$	$\pi/4$	$\pi/2$
	0.21**	0.20**	0.99***	0.96***	0.05*	0.06*	0.81***	0.80***

(b) Real root, the roots of (4.5) are ρ_1 and ρ_2

	$1-\delta=0.97$			
ρ_{01}	0.95*	0.80*	0.95*	1.00*
ρ_{02}	0.75*	0.80*	0.95*	1.00*
	0.06*	0.06*	0.08*	0.91***

table 1.1 T is 150, and the model is the simplest of all; $p=1$, and $\mu_0=0$ is a priori specified and utilized in constructing the test statistic. The empirically determined size of the test agrees well with the nominal size. The power of the test indicated by the figures with *** increases as the demilitarized zone is expanded below with a fixed T. The ideal would be that, as ρ_0 moves toward 1 from below, the probability of the critical region remains low while ρ_0 is in the demilitarized zone and then suddenly jumps up at $\rho_0=1$. That is impossible. In many applications of the equilibrium analysis of economics the ρ that is quite close to unity is just as bad as the ρ equal to unity. Assuming that this is generally accepted among economists we suggest selecting $1-\delta$ in the light of the power of the test at $\rho=1$. The high power at $\rho=1$ necessarily accompanies high powers at $\rho>1$.

Tables 1.2a and 1.2b show the power at $\rho_0=1$ in the case where $p=2$ and

Table 1.4a

$P[2T|Q^{(1)}|>4.364]$ $T=150$, 1,000 replications
$p=1$, $\sigma_0^2=1$, $1-\delta=0.97$
$y_0 \sim N(\frac{\mu_0}{1-\rho_0}, \frac{\sigma_0^2}{1-\rho_0^2})$

ρ_0	0.92	0.92	0.95	0.95	0.98	0.98
μ_0	0.50	5.00	0.50	5.00	0.50	5.00
	0.06*	0.06*	0.08*	0.08*	0.13**	0.13**

Table 1.4b

$P[2T|Q^{(1)}|>4.364]$ $y_0 \sim N(0, 1)$, $1-\delta=0.97$

ρ_0	0.90	0.90	0.95	0.95
μ_0	0.05	0.50	0.05	0.50
	0.06*	0.32*	0.08*	0.64*

$\mu_0=0$ is a priori specified in the model of the one root test. They show that adequate power is secured with $1-\delta=0.95$ when T equals 100.

Table 1.3 shows the probability of the critical region in the two roots test, where $p=2$ and $\mu_0=0$ is a priori specified. The empirically determined test size agrees well with the nominal size.

As for the non-zero μ_0, the size of our tests crucially depends upon the moments of the unobservable initial y_0. In table 1.4a y_0 has the stationary mean and variance, and the size is just as expected from the asymptotic theory. In table 1.4b y_0 has zero mean in spite of μ_0 being non-zero, and the size of our test quickly approaches unity as μ_0^2 deviates from zero in relation to σ_0^2.

In our opinion the stationarity, which is our null hypothesis, should be a property of a stochastic process defined over $t=-\infty,\dots,0,\dots,+\infty$ even though the process is observed only over $t=1,\dots,T$. It implies that y_0 should have the stationary mean.

As for tests $\tilde{Q}^{(1)}$ and $\tilde{Q}^{(2)}$ developed in section 1.5 our simulation studies show (1) that the empirically determined test size roughly agrees with the nominal size in so far as $\mu_0=0$ is a priori specified and incorporated in the test statistic, but (2) that the agreement becomes worse as μ_0^2 deviates from zero in relation to σ_0^2 unless T is extremely large. The reason can be explained for the simplest model, (3.1). $(S(\rho)y-\mu\iota)'(S(\rho)y-\mu\iota)$ evaluated at $\rho=\rho_0$ and $\mu=\mu_0$ is