Chapter 46

UNIT ROOTS, STRUCTURAL BREAKS AND TRENDS

JAMES H. STOCK*

Harvard University

Contents

*The author thanks Robert Amano, Donald Andrews, Jushan Bai, Ngai Hang Chan, In Choi, David Dickey, Frank Diebold, Robert Engle, Neil Ericsson, Alastair Hall, James Hamilton, Andrew Harvey, Sastry Pantula, Pierre Perron, Peter Phillips, Thomas Rothenberg, Pentti Saikkonen, Peter Schmidt, Neil Shephard and Mark Watson for helpful discussions and/or coments on a draft of this chapter. Graham Elliott provided outstanding research assistance. This research was supported in part by the National Science Foundation (Grants SES-89-10601 and SES-91-22463).

Handbook of Econometrics, Volume IV, Edited by R.F. Engle and D.L. McFadden

Abstract

This chapter reviews inference about large autoregressive or moving average roots in univariate time series, and structural change in multivariate time series regression. The "problem" of unit roots is cast more broadly as determining the order of integration of a series; estimation, inference, and confidence intervals are discussed. The discussion of structural change focuses on tests for parameter stability. Much emphasis is on asymptotic distributions in these nonstandard settings, and one theme is the general applicability of functional central limit theory. The quality of the asymptotic approximations to finite-sample distributions and implications for empirical work are critically reviewed.

1. Introduction

The past decade has seen a surge of interest in the theoretical and empirical analysis of long-run economic activity and its relation to short-run fluctuations. Early versions of new classical theories of the business cycle (real business cycle models) predicted that many real economic variables would exhibit considerable persistence, more precisely, would contain a unit root in their autoregressive (AR) representations. Hall's (1978) fundamental work on the consumption function showed that, under a simple version of the permanent income hypothesis, future changes in consumption are unpredictable, so consumption follows a random walk or, more generally, a martingale. The efficient markets theory of asset pricing recapitulated by Fama (1970) had the same prediction: if future excess returns were predictable, they would be bid away so that the price (or log price) would follow a martingale. The predictions of these theories often extended to multivariate relations. For example, if labor income has a unit root, then a simple version of the intertemporal permanent income hypothesis implies that consumption will also have a unit root and moreover that income minus consumption (savings) will not have a unit root, so that consumption and income are, in Engle and Granger's (1987) terminology, cointegrated [Campbell (1987)]. Similarly, versions of real business cycle models predict that aggregate consumption, income and investment will be cointegrated.

The empirical evidence on persistence in economic time series was also being refined during the 1980's. The observation that economic time series have high persistence is hardly new. Orcutt (1948) found a high degree of serial correlation in the annual time series data which Tinbergen (1939) used to estimate his econometric model of the U.S. economy. By plotting autocorrelograms and adjusting for their downward bias when the true autocorrelation is large, Orcutt concluded that many of these series – including changes in aggregate output, investment and consumption –

were well characterized as being generated by the first-order autoregression, $\Delta y_t = 0.3\Delta y_{t-1} + \varepsilon_t$ [Orcutt (1948, eq. 50)], where $\Delta y_t = y_t - y_{t-1}$, that is, they contained an autoregressive unit root. During the 1960's and 1970's, conventional time series practice was to model most economic aggregates in first differences, a practice based on simple diagnostic devices rather than formal statistical tests. In their seminal article, Nelson and Plosser (1982) replaced this informal approach with Dickey and Fuller's (1979) formal tests for a unit root, and found that they could not reject the hypothesis of a unit autoregressive root in 13 of 14 U.S. variables using long annual economic time series, in some cases spanning a century. Similarly, Meese and Singleton (1982) applied Dickey–Fuller tests and found that they could not reject the null of a single unit root in various exchange rates. Davidson et al. (1978) found that an error-correction model, later recognized as a cointegrating model, provided stable forecasts of consumption in the U.K. As Campbell and Mankiw (1987a, 1987b) and Cochrane (1988) pointed out, the presence of a unit root in output implies that shocks to output have great persistence through base drift, which can even exceed the magnitude of the original shock if there is positive feedback in the form of positive autocorrelation.

This body of theoretical and empirical evidence drew on and spurred developments in the econometric theory of inference concerning long-run properties of economic time series. This chapter surveys the theoretical econometrics literature on long-run inference in univariate time series. With the exception of Section 5.1 on stability tests, the focus here is strictly univariate; for multivariate extensions see the chapter by Watson in this Handbook. Throughout, we write the observed series y_t as the sum of a deterministic trend d_t and a stochastic term u_t

$$y_t = d_t + u_t, \qquad t = 1, 2, \ldots, T. \tag{1.1}$$

The trend in general depends on unknown parameters, for example, in the leading case of a linear time trend, $d_t = \beta_0 + \beta_1 t$, where β_0 and β_1 are unknown. Unless explicitly stated otherwise, it is assumed that the form of the trend is correctly specified. If u_t has a unit autoregressive root, then u_t is integrated of order one (is I(1)) in the sense of Box and Jenkins (1976). If Δu_t has a unit moving average (MA) root, then u_t is integrated of order zero (is I(0)). In the treatment here, the focus is on these largest roots and the parameters describing the deterministic term are treated as nuisance parameters. The two types of unit roots (AR and MA) introduce obvious ambiguity in the phrase "unit root", so this chapter emphasizes instead the I(0) and I(1) terminology. Precise definitions are given in Section 2.

The specific aim of this chapter is to outline the econometric theory of four areas of inference in time series analysis: unit autoregressive roots and inference for I(1) and nearly I(1) series; unit moving average roots and testing for a series being I(0); inference on d_t, and, in particular, testing for a unit autoregressive root when d_t might have breaks, for example, be piecewise linear; and tests for parameter instability and structural breaks in regression models. Although the analysis of

structural breaks stems from a different literature than unit roots, the mathematics and indeed some test statistics in these two areas are closely related, and this survey emphasizes such links.

There have been four main areas of application of the techniques for inference about long-run dependence discussed in this chapter. The first and perhaps the most straightforward is data description. Does real GNP contain an autoregressive unit root? What is a 95% confidence interval for the largest root? If output has a unit root, then it has a permanent component, in the sense that it can be decomposed into a stochastic trend (a martingale component) plus an I(0) series. What does this permanent component, trend output, look like, and how can it be estimated? This question has led to estimating and testing an unobserved components model. For empirical applications of the unobserved components model see Harvey (1985), Watson (1986), Clark (1987, 1989), Quah (1992) and Harvey and Jaeger (1993); for a technical discussion, see Harvey (1989); for reviews see Stock and Watson (1988a) and Harvey and Shephard (1992). A natural question is whether there is in fact a permanent component. As will be seen in Section 4, this leads to testing for a unit moving average root or, more generally, testing the null hypothesis that the series is I(0) against the I(1) alternative.

A second important application is medium- and long-term forecasting. Suppose one is interested in making projections of a series over a horizon that represents a substantial fraction of the sample at hand. Such long-term forecasts will be dominated by modeling decisions about the deterministic and stochastic trends. Several of the techniques for inference studied in this chapter – for example, tests for unit AR or MA roots and the construction of median-unbiased estimates of autoregressive coefficients – have applications to long-run forecasting and the estimation of forecast error bands.

A third application, perhaps the most common in practice, is to guide subsequent multivariate modeling or inference involving the variable in question. For example, suppose that primary interest is in the coefficients on y_t in a regression in which y_t appears as a regressor. Inference in this regression in general depends on the order of integration of y_t and on its deterministic component [see West (1988a), Park and Phillips (1988), Sims et al. (1990), and the chapter by Watson in this Handbook]. As another example, if multiple series are I(1) then the next step might be to test for and model cointegration. Alternatively, suppose that the objective is to decompose multiple time series into permanent and transitory components, say to study short-run dynamic effects of permanent shocks [Blanchard and Quah (1989), King et al. (1991)]. In each of these cases, how best to proceed hinges on knowing whether the individual series are I(0) or I(1). Although these multivariate applications are beyond the scope of this chapter, inference about univariate AR and MA roots plays a key initial step in these multivariate applications.

Fourth, information on the degree of persistence in a time series and, in particular, on its order of integration can help to guide the construction or testing of economic theories. Indeed, a leading interpretation of Nelson and Plosser's (1982) findings

was that the prevalence of I(1) series in their long annual data set provided support for real theories of the business cycle. Alternatively, knowledge of the order of integration of certain variables can be used to suggest more precise statements (and to guide inference) about certain economic theories, for example, the possibility of a vertical long-run Phillips curve or the neutrality of money [Fisher and Seater (1993), King and Watson (1992)].

In addition to these empirical applications, technical aspects of the econometric theory of unit roots, trend breaks and structural breaks are related to several other problems in econometric theory, such as inference in cointegrated systems. The theory developed here provides an introduction to the more involved multivariate problems.

Several good reviews of this literature are already available and an effort has been made here to complement them. Phillips (1988) surveys the theoretical literature on univariate and multivariate autoregressive unit root distributions, and a less technical introduction to these topics is given in Phillips (1992a). Diebold and Nerlove (1990) provide a broad review of the econometric literature on measures and models of persistence. Campbell and Perron (1991) provide an overview of the literature on unit autoregressive roots, as well as on cointegration, with an eye towards advising applied researchers. Banerjee et al. (1992a) provide a thorough introduction to testing and estimation in the presence of unit autoregressive roots and multivariate modeling of integrated time series, with special attention to empirical applications.

The main approach to inference about long-term properties of time series which is excluded from this survey is fractional integration. In this alternative to the I(0)/I(1) framework, it is supposed that a series is integrated of order d, where d need not be an integer. The econometric theory of inference in fractionally integrated models has seen ongoing important work over the past two decades. This literature is large and the theory is involved, and doing it justice would require a lengthier treatment than possible here. The R/S statistic of Mandelbrot and Van Ness (1968), originally developed to detect fractional integration, is discussed briefly in Section 3.2 in the context of tests for an AR unit root. Otherwise, the reader is referred to recent contributions in this area. Two excellent surveys are Beran (1992) and, at a more rigorous level, Robinson (1993). Important contributions to the theory of inference with fractional integration include Geweke and Porter-Hudak (1983), Fox and Taqqu (1986), Dahlhaus (1989) and Sowell (1990, 1992). Recent empirical work in econometrics includes Lo (1991) (R/S analysis of stock prices), Diebold and Rudebusch (1989, 1991a) and Diebold et al. (1991) (estimation of the fractional differencing parameter for economic data).

The chapter is organized as follows. Section 2 describes the I(0) and I(1) models and reviews some tools for asymptotic analysis. Section 3 examines inference about the largest autoregressive root when this root equals or is close to one. Section 4 studies inference about unit or near-unit moving average roots. Two related topics are covered in Section 5: tests for parameter stability and structural breaks when

the break date is unknown, and tests for AR unit roots when there are broken trends. Section 6 concludes by drawing links between the I(0) and I(1) testing problems and by suggesting some conclusions concerning these techniques that, it is hoped, will be useful in empirical practice. Most of the formal analysis in this chapter is based on asymptotic distribution theory. The treatment of the theory here is self-contained. Readers primarily interested in empirical applications can omit Sections 2.4, 3.2.3 and 4.2.3 with little loss of continuity. Readers primarily interested in tests for parameter stability and structural breaks in time series regression can restrict their attention to Sections 2 and 5.1.

2. Models and preliminary asymptotic theory

This section provides an introduction to the basic models and limit theory which will be used to develop and to characterize the statistical procedures studied in the remainder of this chapter. Section 2.1 introduces basic notation used throughout the chapter, and provides formulations of the I(0) and I(1) hypotheses. This section also introduces a useful tool, Beveridge and Nelson's (1981) decomposition of an I(1) process into I(0) and I(1) components. This leads naturally to a second expression for the I(0) and I(1) hypothesis in a "components" representation.

Section 2.2 summarizes the limit theory which will be used to derive the asymptotic properties of the various test procedures. A variety of techniques have been and continue to be used in the literature to characterize limiting distributions in the unit MA and AR roots problems. However, the most general and the simplest to apply is the approach based on the functional central limit theorem (FCLT, also called the invariance principle or Donsker's theorem) and the continuous mapping theorem (CMT), and that is the approach used in this chapter. [There are a number of excellent texts on the FCLT. The classic text is Billingsley (1968). A more modern treatment, on which this chapter draws, is Hall and Heyde (1980). Ethier and Kurtz (1986) provide more advanced material and applications. Also, see the chapter by Wooldridge in this Handbook.] The version of the FCLT used in this chapter, which applies to the sequence of partial sums of martingale difference sequences, is due to Brown (1971). The main advantage of this approach is that, armed with the FCLT and the CMT, otherwise daunting asymptotic problems are reduced to a series of relatively simple calculations. White (1958) was the first to suggest using the FCLT to analyze "unit root" distributions. Other early applications of the FCLT, with i.i.d. or martingale difference sequence errors, to statistics involving I(1) processes include Bobkoski (1983) and Solo (1984). Phillips' (1987a) influential paper demonstrated the power of this approach by deriving the distribution of the AR(1) estimator and t-statistic in the misspecified case that the process has additional [non-AR(1)] dependence. These were paralleled by important developments in the asymptotics of multivariate unit root models; see the chapter by Watson in this Handbook for a review.

The aim of this chapter is to provide a treatment at a level suitable for graduate students and applied econometricians. To enhance accessibility, we make two main compromises in generality and rigor. The first is to restrict attention to time series which can be written as linear processes with martingale difference errors, subject to some moment restrictions. This class is rich enough to capture the key complications in the theory and practice of inference concerning unit roots and trend breaks, namely the presence of possibly infinitely many nuisance parameters describing the short-run dependence of an I(0) disturbance. However, most of the results hold under some forms of nonstationarity. References to treatments which handle such nonstationarity are given in Section 2.4. The second technical compromise concerns details of proofs of continuity of functionals needed to apply the continuous mapping theorem; these details are typically conceptually straightforward but tedious and notationally cumbersome, and references are given to complete treatments when subtleties are involved.

2.1. *Basic concepts and notation*

Throughout this chapter, v_t denotes a purely stochastic I(0) process and ε_t denotes a serially uncorrelated stochastic process, specifically a martingale difference sequence. The term "I(0)" is vague, so defining a process to be I(0) requires additional technical assumptions. The formulation which shall be used throughout this chapter is that v_t is a linear process with martingale difference sequence errors. That is, the I(0) process v_t has the (possibly infinite) moving average representation,

$$v_t = c(L)\varepsilon_t, \qquad t = 0, \pm 1, \pm 2, \ldots, \tag{2.1}$$

where $c(L) = \sum_{j=0}^{\infty} c_j L^j$ is a one-sided moving average polynomial in the lag operator L which in general has infinite order. The errors are assumed to obey

$$\begin{aligned} &E(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) = 0, \\ &T^{-1} \sum_{t=1}^{T} E(\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) \to_{\text{a.s.}} E\varepsilon_t^2 = \sigma_\varepsilon^2 > 0 \quad \text{as } T \to \infty, \\ &E(\varepsilon_t^4 | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) < K_4 \quad \text{a.s. for all } t. \end{aligned} \tag{2.2}$$

That is, ε_t can exhibit conditional heteroskedasticity but this conditional heteroskedasticity must be stationary in the sense that fourth moments exist and that ε_t is unconditionally homoskedastic. Because ε_t is unconditionally homoskedastic, under (2.1) and (2.2) v_t is covariance stationary.[1] This simplifies the discussion of

[1] A time series y_t is *strictly stationary* if the distribution of $(y_{k+1}, \ldots, y_{k+T})$ does not depend on k. The series is *covariance stationary* (or *second-order stationary*) if Ey_t and $Ey_t y_{t-j}$, $j = 0, \pm 1, \ldots$ exist and are independent of t.

functions of second moments of v_t such as its spectrum, $s_v(\omega)$, or autocovariances, $\gamma_v(j), j = 0, \pm 1, \pm 2, \ldots$ The representation (2.1) is similar to the Wold representation for a covariance stationary series, although the Wold representation only implies that the errors are serially uncorrelated, not martingale difference sequences.

Central to the idea that a process is I(0), is that the dependence between distant observations is limited. In the context of (2.1), this amounts to making specific assumptions on $c(L)$. The assumption which will be maintained throughout is that $c(L)$ has no unit roots and that it is one-summable [e.g. Brillinger (1981, ch. 2.7)]

$$c(1) \neq 0 \quad \text{and} \quad \sum_{j=0}^{\infty} j|c_j| < \infty, \tag{2.3}$$

where $c(1) = \sum_{j=0}^{\infty} c_j$. The conditions (2.3) can alternatively be written as restrictions on the spectrum of v_t, $s_v(\omega)$. Because $s_v(\omega) = (\sigma_\varepsilon^2/2\pi)|\sum_{j=0}^{\infty} c_j e^{i\omega j}|^2$ (where $i = \sqrt{-1}$), $s_v(0) = \sigma_\varepsilon^2 c(1)^2/2\pi$, so $c(1) \neq 0$ implies that the spectral density of v_t at frequency zero is nonzero. Similarly, the one-summability condition implies that $ds_v(\omega)/d\omega$ is finite at $\omega = 0$. Thus, these conditions on $c(L)$ restrict the long-term behavior of v_t. Unless explicitly stated otherwise, throughout this chapter it is assumed that v_t satisfies (2.1)–(2.3).

The definition of general orders of integration rests on this definition of I(0): a process is said to be I(d), $d \geq 1$, if its dth difference, $\Delta^d u_t$, is I(0). Thus u_t is I(1) if $\Delta u_t = v_t$, where v_t satisfies (2.1)–(2.3). In levels, $u_t = \sum_{s=1}^{t} v_s + u_0$, so that the specification of the levels process of u_t must also include an assumption about the initial condition. Unless explicitly stated otherwise, it is assumed that, if u_t is I(1), then the initial condition satisfies $Eu_0^2 < \infty$.

A leading example of processes which satisfy (2.3) are finite-order ARMA models as popularized by Box and Jenkins (1976). If v_t has an ARMA(p, q) representation, then it can be written

$$\rho(L)v_t = \phi(L)\varepsilon_t, \tag{2.4}$$

where $\rho(L)$ and $\phi(L)$, respectively, have finite orders p and q. If the roots of $\rho(L)$ and $\phi(L)$ lie outside the unit circle, then the ARMA process is stationary and invertible and v_t is integrated of order zero. If v_t satisfies (2.4) and is stationary and invertible, then $v_t = c(L)\varepsilon_t$ where $c(L) = \rho(L)^{-1}\phi(L)$ and it is readily verified that (2.3) is satisfied, since $\phi(1) \neq 0$ and eventually $c(L)$ decays exponentially.

ARMA models provide a simple framework for nesting the I(0) and I(1) hypotheses, and are the origin of the "unit root" terminology. Suppose u_t in (1.1) satisfies

$$(1 - \alpha L)u_t = (1 - \theta L)v_t, \tag{2.5}$$

where v_t is I(0). If $|\alpha| < 1, u_t$ is stationary. If $|\theta| < 1$, then $(1 - \theta L)$ is said to be

invertible. If $\alpha = 1$ and $|\theta| < 1$, then u_t is integrated of order one; that is, $u_t - u_0$ can be expressed as the partial sum – loosely, the "integration" – of a stationary process. If $\alpha = 1$ and $\theta = 1$, then $u_t = v_t + (u_0 - v_0)$ and u_t is stationary, or integrated of order zero. [If $\theta = 1$ and $|\alpha| < 1$, then u_t is integrated of order -1, but we will not consider this case since then u_t in (2.2) can be replaced by its accumulation $\sum_{s=1}^{t} u_s$, which in turn is I(0).]

This framework provides an instructive interpretation of the I(1) and I(0) models in terms of the properties of long-run forecasts of the series. As Harvey (1985, 1989) has emphasized, an intuitively appealing definition of the trend component of a series is that its long-run forecast *is* its trend. If u_t is I(1), then its long-run forecast follows a martingale, while if u_t is I(0), its long-run forecast tends to its unconditional mean (here zero). In this sense, if u_t is I(1) then it and y_t can be said to have a stochastic trend.

This correspondence between the order of integration of a series and whether it has a stochastic trend is formally provided by Beveridge and Nelson's (1981) decomposition of u_t into I(1) and I(0) components. Suppose that $\Delta u_t = v_t$. The Beveridge–Nelson (1981) decomposition rests on writing $c(L)$ as $c(L) = c(1) + [c(L) - c(1)] = c(1) + c^*(L)\Delta$, where $\Delta = 1 - L$ and $c_j^* = -\sum_{i=j+1}^{\infty} c_i$ (this identity is readily verified by writing out $c(L) - c(1) = \Delta c^*(L)$ and collecting terms). Thus v_t can be written $v_t = c(1)\varepsilon_t + c^*(L)\Delta\varepsilon_t$. Then, because $u_t = \sum_{s=1}^{t} v_s + u_0$, we get the *Beveridge–Nelson decomposition*

$$u_t = c(1)\sum_{s=1}^{t} \varepsilon_s + c^*(L)\varepsilon_t + \tilde{u}_0, \tag{2.6}$$

where $\tilde{u}_0 = u_0 - c^*(L)\varepsilon_0$. It is readily verified that, under (2.1)–(2.3), $c^*(L)\varepsilon_t$ is covariance stationary. This follows from the one-summability of $c(L)$, which implies that $c^*(L)$ is summable. [Specifically,

$$\sum_{j=0}^{\infty} |c_j^*| = \sum_{j=0}^{\infty} \left| -\sum_{i=j+1}^{\infty} c_i \right| \leqslant \sum_{j=0}^{\infty} \sum_{i=j+1}^{\infty} |c_i| = \sum_{i=0}^{\infty} i|c_i|,$$

which is finite by (2.3).] Thus,

$$E(c^*(L)\varepsilon_t)^2 = \sum_{j=0}^{\infty} (c_j^*)^2\sigma_\varepsilon^2 \leqslant \left(\sum_{j=0}^{\infty} |c_j^*| \right)^2 \sigma_\varepsilon^2,$$

which is finite by (2.2) and (2.3).

The Beveridge–Nelson decomposition (2.6) therefore represents u_t as the sum of a constant times a martingale, a covariance stationary disturbance and an initial condition $\tilde{u}_0$. If u_0 is fixed or drawn from a distribution on the real line, then $\tilde{u}_0$ can be neglected and often is set to zero in statements of the Beveridge–Nelson

decomposition. The martingale term can be interpreted as the long-run forecast of u_t: because $c^*(L)$ is summable, the long-term forecast, $u_{t+k|t}$ for k very large, is $c(1)\sum_{s=1}^{t}\varepsilon_s$. Thus an I(1) series can be thought of as containing a stochastic trend. Equally, if u_t is I(0), then $\text{plim}_{k\to\infty} u_{t+k|t} = 0$, so that u_t does not have a stochastic trend.

2.2. *The functional central limit theorem and related tools*

If u_t is stationary, or more generally has sufficiently many moments and limited dependence on past observations, then averages such as $T^{-1}\sum_{t=1}^{T} u_t^2$ will be consistent for their expectation, and scaled sums like $T^{-1/2}\sum_{t=1}^{T} u_t$ will obey a central limit theorem; see the chapter by Wooldridge in this Handbook for a general treatment. By the nature of the problems being studied, however, conventional limit theory does not apply to many of the statistics covered in this chapter. For example, the null distribution of a test for a unit autoregressive root is derived for u_t being I(1). However, this violates the assumptions upon which conventional asymptotic tools, such as the weak law of large numbers (WLLN), are based. For example, if u_t is I(1), then the sample mean $\bar{u}$ is $O_p(T^{1/2})$ and $T^{-1/2}\bar{u}$ has a limiting normal distribution, in sharp contrast to the I(0) case in which $\bar{u}$ is consistent.[2]

The approach to this and related problems used in this chapter is based on the functional central limit theorem. The FCLT is a generalization of the conventional CLT to function-valued random variables, in the case at hand, the function constructed from the sequence of partial sums of a stationary process. Before discussing the FCLT, we introduce extensions to function spaces of the standard notions of consistency, convergence in distribution, and the continuous mapping theorem. Let $C[0, 1]$ be the space of bounded continuous functions on the unit interval with the sup-norm metric, $d(f, g) = \sup_{s\in[0,1]}|f(s) - g(s)|$, where $f, g \in C[0, 1]$.

Consistency. A random element $\xi_T \in C[0, 1]$ converges in probability to f (that is, $\xi_T \xrightarrow{p} f$) if $\Pr[d(\xi_T, f) > \delta] \to 0$ for all $\delta > 0$.

Convergence in distribution. Let $\{\xi_T, T \geqslant 1\}$ be a sequence of random elements of $C[0, 1]$ with induced probability measures $\{\pi_T\}$. Then π_T converges weakly to π, or equivalently $\xi_T \Rightarrow \xi$ where ξ has the probability measure π, if and only if $\int f\,d\pi_T \to \int f\,d\pi$ for all bounded continuous $f: C[0, 1] \to \mathscr{R}$. The notations $\xi_T \Rightarrow \xi$ and $\xi_T(\cdot) \Rightarrow \xi(\cdot)$, where "$\cdot$" denotes the argument of the functions ξ_T and ξ, are used interchangeably in this chapter.

[2] Suppose $\Delta u_t = \varepsilon_t$ and $u_0 = 0$. Clearly conventional assumptions in the WLLN, such as u_t having a bounded second moment, do not hold. Rather, $T^{-1/2}\bar{u} = T^{-3/2}\sum_{t=1}^{T} u_t = T^{-3/2}\sum_{t=1}^{T}\sum_{s=1}^{t}\varepsilon_s = T^{-1/2}\sum_{s=1}^{T}(1 - (s-1)/T)\varepsilon_s$, so a central limit theorem for weighted sums implies that $T^{-1/2}\bar{u} \xrightarrow{d} N(0, \sigma_\varepsilon^2/3)$.

The continuous mapping theorem (CMT). If h is a continuous functional mapping $C[0,1]$ to some metric space and $\xi_T \Rightarrow \xi$, then $h(\xi_T) \Rightarrow h(\xi)$.

The FCLT generalizes the usual CLT to random functions $\xi_T \in C[0,1]$. Let $[\cdot]$ denote the greatest lesser integer function. Let $\xi_T(\lambda)$ be the function constructed by linearly interpolating between the partial sums of ε_t at the points $\lambda = (0, 1/T, 2/T, \ldots, 1)$, that is,

$$\xi_T(\lambda) = (\sigma_\varepsilon^2 T)^{-1/2} \left\{ \sum_{t=1}^{[T\lambda]} \varepsilon_t + (T\lambda - [T\lambda])\varepsilon_{[T\lambda]+1} \right\},$$

so that ξ_T is a piecewise-linear random element of $C[0,1]$. The CLT for vector-valued processes ensures that, if $\lambda_1, \ldots, \lambda_k$ are fixed constants between zero and one and condition (2.2) holds, then $[\xi_T(\lambda_1), \xi_T(\lambda_2), \ldots, \xi_T(\lambda_k)]$ converges in distribution jointly to a k-dimensional normal random variable. The FCLT extends this result to hold not just for finitely many fixed values of λ, but rather for ξ_T treated as a function of λ. The following FCLT is a special case of Brown's (1971) FCLT [see Hall and Heyde (1980), Theorem 4.1 and discussion].

Theorem 1 (Functional central limit theorem for a martingale)

Suppose that ε_t is a martingale difference sequence which satisfies (2.2). Then $\xi_T \Rightarrow W$, where W is a standard Brownian motion on the unit interval.

An FCLT for processes which satisfy (2.1)–(2.3) can be obtained by verifying that condition (5.24) in Hall and Heyde's (1980) Theorem 5.5 is satisfied if $c(L)$ is one-summable. [One-summability is used because of its prior use in unit root asymptotics [Stock (1987)], although it can be replaced by the weaker condition that $c(L)$ is $\frac{1}{2}$-summable; see Solo (1989) and Phillips and Solo (1992).] However, Hall and Heyde's theorem is more general than needed here and for completeness an FCLT is explicitly derived from Theorem 1 for processes satisfying (2.1)–(2.3). The argument here relies on inequalities in Hall and Heyde (1980) and follows Phillips and Solo (1992), except that the somewhat stronger conditions used here simplify the argument. See Phillips and Solo (1992) for an extensive discussion, based on the Beveridge–Nelson decomposition, of conditions under which the FCLT holds for linear processes.

To show that Theorem 1 and the Beveridge–Nelson decomposition can be used to yield directly an FCLT for partial sums of I(0) processes which satisfy conditions (2.1)–(2.3), let

$$\xi_{vT}(\lambda) = (\sigma_\varepsilon^2 T)^{-1/2} \left\{ \sum_{t=1}^{[T\lambda]} v_t + (T\lambda - [T\lambda])v_{[T\lambda]+1} \right\}.$$

According to the Beveridge–Nelson decomposition (2.6), this scaled partial sum for

fixed λ is $c(1)T^{-1/2}\sum_{s=1}^{[T\lambda]}\varepsilon_s$ plus a term which is $T^{-1/2}$ times an I(0) variable. Because $\xi_T \Rightarrow W$, this suggests that $\xi_{vT} \Rightarrow c(1)W$.

To show this formally, the argument that $\xi_{vT} - c(1)\xi_T \xrightarrow{p} 0$ must be made uniformly in λ, that is, that $\Pr[\sup_\lambda |\xi_{vt}(\lambda) - c(1)\xi_T(\lambda)| > \delta] \longrightarrow 0$ for all $\delta > 0$. Now,

$$
\begin{aligned}
|\xi_{vT}(\lambda) - c(1)\xi_T(\lambda)| &= (\sigma_\varepsilon^2 T)^{-1/2}\Bigg|\sum_{t=1}^{[T\lambda]} v_t + (T\lambda - [T\lambda])v_{[T\lambda]+1} - c(1)\sum_{t=1}^{[T\lambda]}\varepsilon_t \\
&\qquad - (T\lambda - [T\lambda])c(1)\varepsilon_{[T\lambda]+1}\Bigg| \\
&= (\sigma_\varepsilon^2 T)^{-1/2}\Bigg|c(1)\sum_{t=1}^{[T\lambda]}\varepsilon_t + c^*(L)\varepsilon_{[T\lambda]} - c^*(L)\varepsilon_0 \\
&\qquad + (T\lambda - [T\lambda])(c(1)\varepsilon_{[T\lambda]+1} + c^*(L)\Delta\varepsilon_{[T\lambda]+1}) \\
&\qquad - c(1)\sum_{t=1}^{[T\lambda]}\varepsilon_t - (T\lambda - [T\lambda])c(1)\varepsilon_{[T\lambda]+1}\Bigg| \\
&= (\sigma_\varepsilon^2 T)^{-1/2}|c^*(L)\varepsilon_{[T\lambda]} - c^*(L)\varepsilon_0 + (T\lambda - [T\lambda])(c^*(L)\Delta\varepsilon_{[T\lambda]+1})| \\
&\leqslant (\sigma_\varepsilon^2 T)^{-1/2}\{|c^*(L)\varepsilon_{[T\lambda]}| + |c^*(L)\varepsilon_{[T\lambda]+1}| + |c^*(L)\varepsilon_0|\} \\
&\leqslant 2\sigma_\varepsilon^{-1}\max_{t=1,\ldots,T}|T^{-1/2}c^*(L)\varepsilon_t| + \sigma_\varepsilon^{-1}T^{-1/2}|c^*(L)\varepsilon_0| \quad (2.7)
\end{aligned}
$$

where the second equality uses the Beveridge–Nelson decomposition. The term $T^{-1/2}|c^*(L)\varepsilon_0|$ in the final line of (2.7) does not depend on λ and is asymptotically negligible, so we drop it and have

$$
\begin{aligned}
\Pr[\sup_\lambda|\xi_{vT}(\lambda) - c(1)\xi_T(\lambda)| > \delta] &\leqslant \Pr[2\sigma_\varepsilon^{-1}\max_t|T^{-1/2}c^*(L)\varepsilon_t| > \delta] \\
&\leqslant \left(\frac{1}{2}\delta\sigma_\varepsilon\right)^{-3} E\max_t|T^{-1/2}c^*(L)\varepsilon_t|^3 \\
&\leqslant \left(\frac{1}{2}\delta\sigma_\varepsilon\right)^{-3} ET^{-3/2}\sum_{t=1}^{T}|c^*(L)\varepsilon_t|^3 \\
&\leqslant \left(\frac{1}{2}\delta\sigma_\varepsilon\right)^{-3} T^{-1/2}\left(\sum_{j=0}^{\infty}|c_j^*|\right)^3 \max_t E|\varepsilon_t|^3
\end{aligned}
\qquad (2.8)
$$

where the final inequality follows from Minkowski's inequality. Because $\max_t E|\varepsilon_t|^3 < \infty$ [by (2.2)] and $\sum_{j=1}^{\infty}|c_j^*| < \infty$ [by the argument following (2.6)], $\Pr[\sup_\lambda|\xi_{vT}(\lambda) - c(1)\xi_T(\lambda)| > \delta] \to 0$ for all $\delta > 0$ so $\xi_{vT} - c(1)\xi_T \xrightarrow{p} 0$. Combining this asymptotic equivalence with Theorem 1, we have the general result that, if v_t satisfies (2.1)–(2.3), then $\xi_{vT} \Rightarrow c(1)W$.

The continuity correction involved in constructing ξ_T and ξ_{vT} is cumbersome and is asymptotically negligible in the sup-norm sense [this can be shown formally using the method of (2.7) and (2.8)]. We shall therefore drop this correction henceforth and write the result $\xi_{vT} \Rightarrow c(1)W$ as the *FCLT for general I(0) processes*,

$$T^{-1/2} \sum_{s=1}^{[T\cdot]} v_s \Rightarrow \sigma_\varepsilon c(1) W(\cdot) = \omega W(\cdot), \tag{2.9}$$

where $\omega = \sigma_\varepsilon c(1)$.[3]

Suppose u_t is an I(1) process with $\Delta u_t = v_t$ and, as is assumed throughout, $Eu_0^2 < \infty$. Then the levels process of u_t obeys the FCLT (2.9): $T^{-1/2}u_{[T\cdot]} = T^{-1/2}\sum_{s=1}^{[T\cdot]} v_s + T^{-1/2}u_0 \Rightarrow \omega W(\cdot)$, where $T^{-1/2}u_0 \xrightarrow{p} 0$ by Chebyshev's inequality. A special case of this is when u_0 is fixed and finite.

The result (2.9) provides a concrete link between the assumptions (2.2) and (2.3) used to characterize an I(1) process, the Beveridge–Nelson decomposition (2.6) and the limit theory which will be used to analyze statistics based on I(1) processes. Under (2.2) and (2.3), the partial sum process is dominated by a stochastic trend, as in (2.6). In the limit, after scaling by $T^{-1/2}$, this behaves like ω times a Brownian motion, where $\omega^2 = 2\pi s_v(0)$ is the zero-frequency power, or long-run variance, of v_t. Thus the limiting behavior of u_t, where $\Delta u_t = v_t$, is the same (up to a scale factor) for a wide range of $c(L)$ which satisfy (2.2). It is in this sense that we think of processes which satisfy (2.1)–(2.3) as being I(0).

2.3. *Examples and preliminary results*

The FCLT and the CMT provide a powerful set of tools for the analysis of statistics involving I(1) processes. The examples in this section will be of use later but are also of independent interest.

Example 1. Sample moments of I(1) processes

A problem mentioned in Section 2.2 was the surprising behavior of the sample mean of an I(1) process. The limiting properties of this and higher moments are readily characterized using the tools of Section 2.2. Let u_t be I(1), so that $\Delta u_t = v_t$, and let $u_0 = 0$. Then,

$$T^{-1/2}\bar{u} = T^{-3/2}\sum_{t=1}^{T} u_t = T^{-1}\sum_{t=1}^{T}(T^{-1/2}u_t) = \int_0^1 (T^{-1/2}u_{[T\lambda]})\,d\lambda + T^{-3/2}u_T, \tag{2.10}$$

[3] Formally, the process on the left-hand side of (2.9) is an element of $D[0,1]$, the space of functions on $[0,1]$ that are right-continuous and have left-hand limits. However, the discontinuous partial sum process is asymptotically equivalent to $\xi_{vT} \in C[0,1]$, for which Theorem 1 applies. See Billingsley (1968, ch. 3) or Ethier and Kurtz (1986) for a treatment of convergence on $D[0,1]$.

where the final equality follows by definition of the integral. The final expression in (2.10) can be written $T^{-1/2}\bar{u} = h_1(T^{-1/2}u_{[T\cdot]}) + T^{-1}h_2(T^{-1/2}u_{[T\cdot]})$, where h_1 and h_2 are functions from $[0,1] \to \mathscr{R}$, namely $h_1(f) = \int_0^1 f(\lambda)\,d\lambda$ and $h_2(f) = f(1)$. Both functions are readily seen to be continuous with respect to the sup-norm, so by (2.9) and the continuous mapping theorem,

$$h_1(T^{-1/2}u_{[T\cdot]}) \Rightarrow h_1(\omega W) = \omega \int_0^1 W(\lambda)\,d\lambda$$

and

$$h_2(T^{-1/2}u_{[T\cdot]}) \Rightarrow \omega W(1),$$

so $T^{-1}h_2(T^{-1/2}u_{[T\cdot]}) \xrightarrow{p} 0$ and $T^{-1/2}\bar{u} \Rightarrow \omega\int_0^1 W(\lambda)\,d\lambda$, which has a normal distribution (cf. footnote 2).

This approach can be just as easily applied to higher moments, say the kth moment,

$$T^{-(1/2)k}\left(T^{-1}\sum_{t=1}^{T} u_t^k\right) = \int_0^1 (T^{-1/2}u_{[T\lambda]})^k\,d\lambda + o_p(1) \Rightarrow \omega^k \int W^k, \tag{2.11}$$

where the convergence follows from the FCLT and the CMT.

The final expression in (2.11) uses a notational convention which will be used commonly in this chapter: the limits on integrals over the unit interval will be omitted, so for example $\int W^k$ denotes $\int_0^1 (W(\lambda))^k\,d\lambda$. Similarly, the stochastic (Itô) integral $\int_{\lambda=0}^1 W(\lambda)\,dG(\lambda)$ is written $\int W\,dG$ for two continuous-time stochastic processes W and G.

Example 2. Detrended I(1) processes

Because of the presence of the deterministic term d_t in (1.1), many statistics of interest involve detrending. It is therefore useful to have limiting representations of the detrended series. The most common form of detrending is by an ordinary least squares (OLS) regression of y_t on polynomials in time, the leading cases being demeaning and linear detrending. Let

$$y_t^{\mu} = y_t - T^{-1}\sum_{s=1}^{T} y_s, \tag{2.12a}$$

$$y_t^{\tau} = y_t - \hat{\beta}_0 - \hat{\beta}_1 t, \tag{2.12b}$$

where $(\hat{\beta}_0, \hat{\beta}_1)$ are the OLS estimators of the parameters in the regression of y_t onto $(1, t)$. If $d_t = \beta_0$, then (2.12a) applies, while if $d_t = \beta_0 + \beta_1 t$, then (2.12b) applies.

As in the previous example, suppose that u_t is I(1), so that (2.9) applies. Now $y_t^{\mu} = u_t - T^{-1}\sum_{s=1}^{T} u_s$, so $T^{-1/2}y_{[T\cdot]}^{\mu} = T^{-1/2}u_{[T\cdot]} - T^{-1/2}\bar{u}$. The CMT and (2.10) thus imply that $T^{-1/2}y_{[T\cdot]}^{\mu} \Rightarrow \omega\{W(\cdot) - \int W\}$, so that the demeaned I(1) process converges to a "demeaned" Brownian motion. Similar arguments, with a bit more algebra, apply to the detrended case. Summarizing these two results, if u_t is a general I(1) process so that $\Delta u_t = v_t$ where v_t satisfies (2.1)–(2.3) and $Eu_0^2 < \infty$, then we have[4]

$$T^{-1/2}y_{[T\cdot]}^{\mu} \Rightarrow \omega W^{\mu}(\cdot), \text{ where } W^{\mu}(\lambda) = W(\lambda) - \int W, \tag{2.13a}$$

$$T^{-1/2}y_{[T\cdot]}^{\tau} \Rightarrow \omega W^{\tau}(\cdot), \text{ where } W^{\tau}(\lambda) = W(\lambda) - (4-6\lambda)\int W - (12\lambda - 6)\int sW(s)\,ds. \tag{2.13b}$$

Perron (1991c) provides expressions extending (2.13) to the residuals of an I(1) process which has been detrended by an OLS regression onto a pth order polynomial in time.

These results can be used to study the behavior of an I(1) process which has been spuriously detrended, that is, regressed against $(1, t)$ when in fact y_t is purely stochastic. Because the sample R^2 is $1 - \{\sum_{t=1}^{T}(y_t^{\tau})^2 / \sum_{t=1}^{T}(y_t^{\mu})^2\}$, the results (2.13) and the CMT show that R^2 has the limit $R^2 \Rightarrow 1 - \{\int (W^{\tau})^2 / \int (W^{\mu})^2\}$, which is positive with probability one; that is, the regression R^2 is, asymptotically, a positive random variable. It follows that the standard t-test for significance of β_1 rejects with probability one asymptotically even though the true coefficient on time is zero [Durlauf and Phillips (1988)][5]. Next, consider the autocorrelogram of the detrended process, $\rho_T(\lambda) = \hat{\gamma}_{y\tau}([T\lambda])/\hat{\gamma}_{y\tau}(0) = h_{\rho}(T^{-1/2}y_{[T\cdot]}^{\tau})$, say, where $\hat{\gamma}_{y\tau}(j) = (T - |j|)^{-1} \times \Sigma_{t=|j|+1}^{T} y_t^{\tau} y_{t-|j|}^{\tau}$. Because h_{ρ} is a continuous mapping from $D[0,1]$ to $D[0,1]$, the FCLT and CMT imply that $\rho_T \Rightarrow \rho^*$, where $\rho^*(\lambda) = (1-\lambda)^{-1}\int_{s=\lambda}^{1} W^{\tau}(s)W^{\tau}(s-\lambda)\,ds / \int W^{\tau 2}$, $0 \leqslant \lambda < 1$. Thus, in particular, the first k sample autocorrelations converge in probability to one for k fixed, although it eventually declines towards zero. Nelson and Kang (1981) show, using other techniques, that the autocorrelogram dips below zero, suggesting periodicity which spuriously arises from the detrending of the I(1) process. The results here indicate that this is an asymptotic phenomenon, when the autocorrelation is interpreted as a fraction of the sample size.

Example 3. Cumulated detrended I(0) processes

Section 4 considers testing for a unit moving average root when it is maintained that there is a deterministic trend. A statistic which arises in this context is the

[4] Derivations of W^{τ} are given in the proof of Theorem 5.1 of Stock and Watson (1988b) and in Park and Phillips (1988); the result can also be derived from Theorem 2.1 of Durlauf and Phillips (1988). As Park and Phillips (1988) demonstrate, W^{τ} can be thought of as detrended Brownian motion, the residual of the projection of W onto $(1, s)$.

[5] Phillips (1986) gives similar results for regressions with two independent random walks.

cumulation of an I(0) process, which has been detrended by OLS. The asymptotics of this process are also readily analyzed using the FCLT and the CMT.

For this example, suppose that u_t is a general I(0) process and $u_t = v_t$, where v_t satisfies (2.1)–(2.3). Consider the demeaned case, and define the statistic

$$Y_{0T}^{\mu}(\lambda) = T^{-1/2} \sum_{s=1}^{[T\lambda]} y_s^{\mu},$$

so

$$Y_{0T}^{\mu}(\lambda) = T^{-1/2} \sum_{s=1}^{[T\lambda]} (u_s - \bar{u}) = T^{-1/2} \sum_{s=1}^{[T\lambda]} u_s - \left(\frac{[T\lambda]}{T}\right) T^{-1/2} \sum_{t=1}^{T} u_t.$$

Then (2.9) and the CMT yield the limit $Y_{0T}^{\mu} \Rightarrow \omega B^{\mu}$, where $B^{\mu}(\lambda) = W(\lambda) - \lambda W(1)$. The process B^{μ} is a standard Brownian bridge on the unit interval, so called because it is a Brownian motion that is "tied down" to be zero at 0 and 1. Similarly, define $Y_{0T}^{\tau}(\lambda) = T^{-1/2} \sum_{s=1}^{[T\lambda]} y_s^{\tau}$; then $Y_{0T}^{\tau} \Rightarrow \omega B^{\tau}$, where B^{τ} is a second-level Brownian bridge on the unit interval, given by $B^{\tau}(\lambda) = W(\lambda) - \lambda W(1) + 6\lambda(1-\lambda)\{\frac{1}{2}W(1) - \int W\}$ [MacNeill (1978)]. Collecting these results, we have

$$T^{-1/2} \sum_{s=1}^{[T\cdot]} y_s^{\mu} \Rightarrow \omega B^{\mu}(\cdot), \quad B^{\mu}(\lambda) = W(\lambda) - \lambda W(1), \tag{2.14a}$$

$$T^{-1/2} \sum_{s=1}^{[T\cdot]} y_s^{\tau} \Rightarrow \omega B^{\tau}(\cdot), \quad B^{\tau}(\lambda) = W(\lambda) - \lambda W(1) + 6\lambda(1-\lambda)\{\tfrac{1}{2}W(1) - \textstyle\int W\}, \tag{2.14b}$$

MacNeill (1978) extended these results to kth order polynomial detrending. Let $Y_{0T}^{(k)}(\lambda) = T^{-1/2} \sum_{s=1}^{[T\lambda]} y_s^{(k)}$ be the process of cumulated kth order detrended data, where $y_t^{(k)}$ is the residual from the OLS regression of y_t onto $(1, t, \ldots, t^k)$. Then $Y_{0T}^{(k)} \Rightarrow \omega B^{(k-1)}$, where $B^{(k)}$ is a kth level generalized Brownian bridge, expressions for which are given by MacNeill (1978, eq. 8).

Example 4. Processes with an autoregressive root local to unity

One of the issues considered in Section 3 is the asymptotic properties of statistics when the process is nearly I(1), in the sense that the largest root of the process is local to unity. The starting point in these calculations is characterizing the large-sample behavior of the series itself when the root is close to one. Let u_t obey

$$u_t = \alpha u_{t-1} + v_t, \quad \text{where} \quad \alpha = 1 + c/T \quad \text{and} \quad Eu_0^2 < \infty. \tag{2.15}$$

This is the local-to-unity model considered (under various assumptions on the disturbances v_t) by Bobkoski (1983), Cavanagh (1985), Chan and Wei (1987) and

Phillips (1987b). The treatment here follows Bobkoski (1983). In particular, we use the method of proof of Bobkoski's (1983) Lemma 3.4 to generalize his local-to-unity representations from his case of i.i.d. disturbances to general I(0) disturbances which satisfy (2.1)–(2.3). As we shall see, this extension is a straightforward application of the FCLT (2.9) and the CMT.

Use recursive substitution in (2.15) to write u_t as

$$\begin{aligned} T^{-1/2}u_t &= T^{-1/2}\sum_{s=1}^{t}\alpha^{t-s}v_s + T^{-1/2}\alpha^t u_0 \\ &= T^{-1/2}\sum_{s=1}^{t-1}(\alpha^{t-s}-1)v_s + T^{-1/2}\sum_{s=1}^{t}v_s + T^{-1/2}\alpha^t u_0 \\ &= (\alpha-1)\sum_{s=1}^{t-1}\alpha^{t-s-1}\left(T^{-1/2}\sum_{r=1}^{s}v_r\right) + T^{-1/2}\sum_{s=1}^{t}v_s + T^{-1/2}\alpha^t u_0 \\ &= h_4(\sigma_\varepsilon \xi_{vT})(t/T) + o_p(1). \end{aligned} \tag{2.16}$$

The third equality in (2.16) obtains as an identity by noting that $\alpha^r - 1 = (\alpha-1)\sum_{j=0}^{r-1}\alpha^j$ and rearranging summations. The final equality obtains by noting that $(1+c/T)^{[T\lambda]} = \exp(c\lambda) + o(1)$ uniformly in λ, $0 < \lambda < 1$, and by defining $h_4(f)(\lambda) = c\int_0^\lambda e^{c(\lambda-s)}f(s)\,ds + f(\lambda)$, where ξ_{vT} is defined in Section 2.2. The $o_p(1)$ term in the final expression arises from the assumption $Eu_0^2 < \infty$, so $T^{-1/2}u_0 = o_p(1)$, and from the approximation $(1+c/T)^{[T\lambda]} \cong \exp(c\lambda)$.

As in the previous examples, h_4 is a continuous functional, in this case from $C[0,1]$ to $C[0,1]$. Using the FCLT (2.9) and the CMT we have

$$T^{-1/2}u_{[T\cdot]} \Rightarrow h_4(\omega W)(\cdot) \equiv \omega W_c(\cdot), \tag{2.17}$$

where $W_c(\lambda) = c\int_0^\lambda e^{c(\lambda-s)}W(s)\,ds + W(\lambda)$. The stochastic process W_c is the solution to the stochastic differential equation, $dw_c = cW_c(\lambda)\,d\lambda + dW(\lambda)$ with $W_c(0) = 0$. Thus, for α local-to-unity, $T^{-1/2}u_{[T\cdot]}$ converges to ω times a diffusion, or Ornstein–Uhlenbeck, process.

A remark on the interpretation of limiting functionals of Brownian motion. The calculations of this section ended when the random variable of interest was shown to have a limiting representation as a functional of Brownian motion or, in the local-to-unity case of Example 4, a diffusion process. These representations show that the limiting distribution exists; they indicate when a limiting distribution is nonstandard; and, importantly, they show when and how nuisance parameters describing the short-run dependence of v_t enter the limiting distribution. Because W is a Gaussian process, the results occasionally yield simply-evaluated distributions. For example, $W(1)$, $\int W$, and $\int sW$ have normal distributions. However, in most cases the limiting distributions are nonstandard.

This leads to the practical question of how to compute limiting distribution functions, once one has in hand the limiting representation of the process as a functional of Brownian motion. The simplest approach, both conceptually and in terms of computer programming, is to evaluate the functional by Monte Carlo simulation using discretized realizations of the underlying Brownian motions. This is equivalent to generating pseudo-data $\tilde{y}_t$ from a Gaussian random walk with $\tilde{y}_0 = 0$ and with unit innovation variance and replacing W by its discretized realization. For example, $W(1)^2$ would be replaced by $(T^{-1/2}\tilde{y}_T)^2$ and $W^{\mu}(\cdot)$ would be replaced by $T^{-1/2}\{\tilde{y}_{[T\cdot]} - T^{-1}\sum_{t=1}^{T}\tilde{y}_t\}$. For T sufficiently large, the FCLT ensures that the limiting distribution of these pseudo-random variates converges to those of the functionals of Brownian motion. The main disadvantage of this approach is that high numerical accuracy requires many Monte Carlo repetitions. For this reason, considerable effort has been devoted to alternative methods for evaluating some of these limiting distributions. Because these techniques are specialized, they will not be discussed in detail, although selected references are given in Section 2.4.

2.4. *Generalizations and additional references*

The model (2.1)–(2.3) provides a concise characterization of I(0) processes with possibly infinitely many nuisance parameters describing the short-run dependence, but this simplicity comes at the cost of assuming away various types of nonstationarity and heteroskedasticity which might be present in empirical applications. The key result used in Section 2.3 and in the sections to follow is the FCLT, which obtains under weaker conditions than stated here. The condition (2.2) is weakened in Brown's (1971) FCLT, which uses the Lindeberg condition and admits unconditional heteroskedasticity which is asymptotically negligible, in the sense that $T^{-1}\sum_{t=1}^{T}E\varepsilon_t^2 \to \sigma_\varepsilon^2$. The result (2.9) for linear processes can be obtained under Brown's (1971) weaker conditions by modifying the argument in Section 2.2; see Hall and Heyde (1980, Chapter 4) and Phillips and Solo (1992).

An alternative approach is to use mixing conditions, which permit an explicit tradeoff between the number of moments and the degree of temporal dependence in v_t and which admit certain nonstationarities (which are asymptotically negligible in the sense above). This approach was introduced to the unit roots literature by Phillips (1987a), who used Herrndorf's (1984) mixing-condition FCLT, and much of the recent unit roots literature uses these conditions. Phillips (1987b) derives the local-to-unity result (2.17) using Herrndorf's (1984) mixing-condition FCLT.

An elegant approach to defining I(0) is simply to make the high-level assumption that v_t is I(0) if its partial sum process converges weakly to a constant times a standard Brownian motion. Thus (2.9) is taken as the assumption rather than the implication of (2.1)–(2.3). With additional conditions assuring convergence of sample moments, such as sample autocovariances, this "high-level" assumption provides a general definition of I(0), which automatically incorporates v_t which

satisfy Herrndorf's (1984) FCLT's. The gain in elegance of this approach comes at the cost of concreteness. However, the results in this chapter that rely solely on the FCLT and CMT typically can be interpreted as holding under this alternative definition.

The FCLT approach is not the only route to asymptotic results in this literature. The approach used by Fuller (1976), Dickey and Fuller (1979) and Sargan and Bhargava (1983a) was to consider the limiting behavior of quadratic forms such as $\sum_{t=1}^{T} u_t^2$ expressed as $\eta' A_T \eta$, where η is a $T \times 1$ standard normal variate; thus the limiting behavior is characterized by the limiting eigenvalues of A_T. See Chan (1988) and Saikkonen and Luukkonen (1993b) for discussions of computational issues involved with this approach.

There is a growing literature on numerical evaluation of these asymptotic distributions. In some cases, it is possible to obtain explicit expressions for moment generating functions or characteristic functions which can be integrated numerically; see White (1958, 1959), Evans and Savin (1981a), Perron (1989b, 1991a), Nabeya and Tanaka (1990a, 1990b) and Tanaka (1990a). Finally, under normality, exact finite-sample distributions can be computed using the Imhof method; see, for example, Evans and Savin (1981b, 1984).

3. Unit autoregressive roots

This section examines inference concerning α in the model

$$y_t = d_t + u_t, \quad u_t = \alpha u_{t-1} + v_t, \qquad t = 1, 2, \ldots, T \tag{3.1}$$

where α is either close to or equal to one and v_t is I(0) with spectral density at frequency zero of $\omega^2/2\pi$. Unless explicitly stated otherwise, it is assumed that u_0 might be random, with $Eu_0^2 < \infty$, and that v_t is a linear process satisfying (2.1)–(2.3).

The trend term d_t will be specified as known up to a finite-dimensional parameter vector β. The leading cases for the deterministic component are (i) no deterministic term ($d_t = 0$); (ii) a constant ($d_t = \beta_0$); and (iii) a linear time trend ($d_t = \beta_0 + \beta_1 t$). Extensions to higher-order polynomial trends or trends satisfying more general conditions are typically straightforward and are discussed only briefly. Another possibility is a piecewise-linear (or broken) trend [Rappaport and Reichlin (1989), Perron (1989a, 1990b)], a topic taken up in Section 5.

Most of the procedures for inference on α treat the unknown parameters in the trend term d_t as nuisance parameters, so that many of the statistics can be represented generally in terms of detrended data. Throughout, y_t^d denotes a general detrended process with unspecified detrending. For specific types of detrending, we adopt Dickey and Fuller's (1979) notation: y_t^μ denotes demeaned data and y_t^τ denotes linearly detrended data when the detrending is by OLS as in (2.12).

The focus of this section is almost exclusively on the case in which there is at most a single real unit root. This rules out higher orders of integration (two or more real autoregressive unit roots) and seasonal unit roots (complex roots on the unit circle). These topics have been omitted because of space limitations. However, the techniques used here extend to these other cases. References on estimation and testing with seasonal unit roots include Hasza and Fuller (1981), Dickey et al. (1984), Chan and Wei (1988), Ghysels (1990), Jegganathan (1991), Ghysels and Perron (1993), Hylleberg et al. (1990), Diebold (1993) and Beaulieu and Miron (1993). See Banerjee et al. (1992a) for an overview.

In the area of testing when there might be two or more unit roots, an important practical lesson from the theoretical literature is that a "downward" testing procedure (starting with the greatest plausible number of unit roots) is consistent, while an "upward" testing procedure (starting with a test for a single unit root) is not. This was shown for F-type tests in the no-deterministic case by Pantula (1989). Based on simulation evidence, Dickey and Pantula (1987) recommend a downward-testing, sequential t-test procedure. Pantula (1989) proves that the distribution of the relevant t-statistic under each null has the standard Dickey–Fuller (1979) distribution. Also, Hasza and Fuller (1979) provide distribution theory for testing two versus zero unit roots in an autoregression.

3.1. *Point estimation*

The four main qualitative differences between regressions with I(1) and I(0) regressors are that, in contrast to the case of I(0) regressors, inference on certain linear combinations of regression coefficients is nonstandard, with: (i) estimators which are consistent at rate T rather than at the usual rate $\sqrt{T}$; (ii) limiting distributions of estimators and test statistics which are often nonstandard and have nonzero means; (iii) estimators which are consistent even if the regression misspecifies the short-run dynamics, although, in this case, the limiting distributions change; and (iv) limiting distributions which depend on both the true and estimated trend specifications.

These differences between I(0) and I(1) regressors can be seen by examining the OLS estimator of α in (3.1). First, consider the no-deterministic case, so $\hat{\alpha} = \sum_{t=2}^{T} y_t y_{t-1} / \sum_{t=2}^{T} y_{t-1}^2$. When $|\alpha| < 1$ and $v_t = \varepsilon_t$, conventional $\sqrt{T}$ asymptotics apply and $\hat{\alpha}$ has a normal limiting distribution,

$$\text{if } |\alpha| < 1, \quad T^{1/2}(\hat{\alpha} - \alpha) \xrightarrow{\text{d}} N(0, 1 - \alpha^2), \tag{3.2}$$

which was derived by Mann and Wald (1943) under the assumptions that ε_t is i.i.d. and all the moments of ε_t exist.

In contrast, suppose that the true value of α is 1, and let v_t follow the general

linear process (2.1)–(2.3). Then the OLS estimator can be rewritten,

$$T(\hat{\alpha}-1)=\frac{T^{-1}\sum_{t=2}^{T}\Delta y_t y_{t-1}}{T^{-2}\sum_{t=2}^{T}y_{t-1}^2}$$
$$=\frac{\frac{1}{2}\left\{T^{-1}(y_T^2-y_1^2)-T^{-1}\sum_{t=2}^{T}(\Delta y_t)^2\right\}}{T^{-2}\sum_{t=2}^{T}y_{t-1}^2} \tag{3.3}$$

where the second line uses the identity $y_T^2-y_1^2=2\sum_{t=2}^{T}\Delta y_t y_{t-1}+\sum_{t=2}^{T}(\Delta y_t)^2$. Although the conditions for Mann and Wald's result do not apply here because of the unit autoregressive root, an asymptotic result for $T(\hat{\alpha}-1)$ nonetheless can be obtained using the FCLT (2.9) and the CMT. Because $Eu_0^2<\infty$, $T^{-1/2}y_1=T^{-1/2}(u_0+v_1)\xrightarrow{p}0$. Thus, because $y_t=u_0+\sum_{s=1}^{t}v_s$, by (2.9) and the CMT, we have $T^{-1/2}y_T\Rightarrow\omega W(1)$ and $T^{-2}\sum_{t=2}^{T}y_{t-1}^2\Rightarrow\omega^2\int W^2$. Also, $T^{-1}\sum_{t=2}^{T}(\Delta y_t)^2=\hat{\gamma}_{\Delta y}(0)\xrightarrow{p}\gamma_{\Delta y}(0)=\gamma_v(0)$. Thus,

$$\text{if } \alpha=1,\quad T(\hat{\alpha}-1)\Rightarrow\frac{\frac{1}{2}(W(1)^2-\kappa)}{\int W^2,}\quad\text{where}\quad\kappa=\frac{\gamma_v(0)}{\omega^2}. \tag{3.4a}$$

This expression was first obtained by White (1958) in the AR(1) model with $\kappa=1$ (although his result was in error by a factor of $\sqrt{2}$) and by Phillips (1987a) for general κ.

An alternative expression for this limiting result obtains by using the continuous-time analogue of the identity used to obtain the second line of (3.3), namely $\int W\,dW=\frac{1}{2}(W(1)^2-1)$ [Arnold (1973, p. 76)]; thus,

$$\text{if } \alpha=1,\quad T(\hat{\alpha}-1)\Rightarrow\left\{\int W\,dW-\tfrac{1}{2}(\kappa-1)\right\}\Big/\int W^2. \tag{3.4b}$$

This result can also be obtained from the first line in (3.3) by applying Theorem 2.4 of Chan and Wei (1988).

The results (3.2) and (3.4) demonstrate the first three of the four main differences between I(0) and I(1) asymptotics. First, the OLS estimator of α is "superconsistent", converging at rate T rather than $\sqrt{T}$. While initially surprising, this has an intuitive interpretation: if the true value of α is less than one, then in expectation the mean squared error $E(y_t-\alpha y_{t-1})^2$ is minimized at the true value of α but remains finite for other values of α. In contrast, if α is truly 1, then $E(\Delta y_t)^2$ is finite but, for any

fixed value of $\alpha \neq 1$, $(1-\alpha L)y_t = \Delta y_t + (1-\alpha)y_{t-1}$ has an integrated component; thus the OLS objective function $T^{-1}\sum_{t=2}^{T}(y_t - \alpha y_{t-1})^2$ is finite, asymptotically, for $\alpha = 1$ but tends to infinity for fixed $\alpha \neq 1$. An alternative intuitive interpretation of this result is that the variance of the usual OLS estimator depends on the sampling variability of the regressors, here $\sum_{t=2}^{T} y_{t-1}^2$; but, because y_t is I(1), this sum is $O_p(T^2)$ rather than the conventional rate $O_p(T)$.

Second, the limiting distribution in (3.4) is nonstandard. While the marginal distribution of $W(1)^2$ is χ_1^2, the distribution of the ratio in (3.4a) does not have a simple form. This distribution has been extensively studied. In the leading case that v_t is serially uncorrelated, then $\omega^2 = \gamma_v(0)$ so that $\kappa = 1$ and (3.4a) becomes $\frac{1}{2}(W(1)^2 - 1)/\int W^2$ [and (3.4b) becomes $\int W\,dW/\int W^2$]. This distribution was tabulated by Dickey (1976) and reproduced in Fuller (1976, Table 8.5.1). The distribution is skewed, with asymptotic lower and upper 5 percent quantiles of -8.1 and 1.28.

Third, $\hat{\alpha}$ is consistent for α even though the regression of y_t onto y_{t-1} is misspecified, in the sense that the error term v_t is serially correlated and correlated with (differences of) the regressor. This misspecification affects the limiting distribution in an intuitive way. Use the definition of κ to write

$$\frac{1}{2}(\kappa - 1) = \frac{\frac{1}{2}\left[\gamma_{\Delta u}(0) - \sum_{j=-\infty}^{\infty}\gamma_{\Delta u}(j)\right]}{\omega^2} = \frac{-\sum_{j=1}^{\infty} Ev_t v_{t-j}}{\omega^2}.$$

Because ω^2 can be thought of as the long-run variance of v_t, $\frac{1}{2}(\kappa - 1)$ represents the correlation between the error and the regressor, which enters as a shift in the numerator of the limiting representation but does not introduce inconsistency. This term can increase the bias of $\hat{\alpha}$ in finite samples. Although this bias decreases at the rate T^{-1}, when v_t is negatively serially correlated, so that $\frac{1}{2}(\kappa - 1)$ is positive, in sample sizes often encountered in practice this bias can be large. For example, if v_t follows the MA(1) process $v_t = (1-\theta L)\varepsilon_t$, then $\frac{1}{2}(\kappa - 1) = \theta/(1-\theta)^2$, so for $\theta = 0.8$, $\frac{1}{2}(\kappa - 1) = 20$.

To examine the fourth general feature of regressions with I(1) variables, the dependence of limiting distributions on trend specifications, consider the case that $d_t = \beta_0 + \beta_1 t$. Substitute this into (3.1), transform both sides of (3.1) by $(1-\alpha L)$, and thus write

$$y_t = \delta_0 + \delta_1 t + \alpha y_{t-1} + v_t, \qquad t = 1, 2, \ldots, T, \tag{3.5}$$

where $\delta_0 = (1-\alpha)\beta_0 + \alpha\beta_1$ and $\delta_1 = (1-\alpha)\beta_1$. If both β_0 and β_1 are unrestricted, (3.5) suggests estimating α by the regression of y_t onto $(1, t, y_{t-1})$; if β_1 is restricted a priori to be zero, then α can be estimated by regressing y_t onto $(1, y_{t-1})$. Consider, for the moment, the latter case in which $d_t = \beta_0$ where β_0 is unknown. Then, by the algebra of least squares, the OLS estimator of α, $\hat{\alpha}^{\mu}$, can be written (after centering

and scaling) as

$$T(\hat{\alpha}^{\mu}-1)=\frac{T^{-1}\sum_{t=2}^{T}\Delta y_t y_{t-1}^{\mu}}{T^{-2}\sum_{t=2}^{T}(y_{t-1}^{\mu})^2}$$

$$=\frac{\frac{1}{2}\left\{T^{-1}(y_T^{\mu 2}-y_1^{\mu 2})-T^{-1}\sum_{t=2}^{T}(\Delta y_t)^2\right\}}{T^{-2}\sum_{t=2}^{T}(y_{t-1}^{\mu})^2}, \tag{3.6}$$

where $y_{t-1}^{\mu}=y_{t-1}-(T-1)^{-1}\sum_{s=2}^{T}y_{s-1}$.

The method for obtaining a limiting representation of $T(\hat{\alpha}^{\mu}-1)$ under the hypothesis that $\alpha=1$ is analogous to that used for $T(\hat{\alpha}-1)$, namely, to use the FCLT to obtain a limiting representation for $T^{-1/2}y_{t-1}^{\mu}$ and then to apply the continuous mapping theorem. Expression (2.13a) provides the needed limiting result for the demeaned levels process of the data; applying this to (3.6) yields

$$T(\hat{\alpha}^{\mu}-1)\Rightarrow\tfrac{1}{2}\{W^{\mu}(1)^2-W^{\mu}(0)^2-\kappa\}\Big/\int W^{\mu 2}=\left\{\int W^{\mu}\,\mathrm{d}W-\tfrac{1}{2}(\kappa-1)\right\}\Big/\int W^{\mu 2} \tag{3.7}$$

where the second representation is obtained using $W^{\mu}(0)=-\int W$.

The detrended case can be handled the same way. Let $\hat{\alpha}^{\tau}$ denote the estimator of α obtained from estimating (3.5) including both the constant and time as regressors. Then $T(\hat{\alpha}^{\tau}-1)$ can be written in the form (3.6), with y_t^{τ} replacing y_t^{μ}. The application of (2.13b) to this modification of (3.6) yields the limiting representation

$$T(\hat{\alpha}^{\tau}-1)\Rightarrow\tfrac{1}{2}\{W^{\tau}(1)^2-W^{\tau}(0)^2-\kappa\}\Big/\int W^{\tau 2}=\left\{\int W^{\tau}\,\mathrm{d}W-\tfrac{1}{2}(\kappa-1)\right\}\Big/\int W^{\tau 2}. \tag{3.8}$$

Because the distributions of W, W^{μ} and W^{τ} differ, so do the distributions in (3.4), (3.7) and (3.8). When $v_t=\varepsilon_t$ so that $\kappa=1$, the distribution of $T(\hat{\alpha}^{\mu}-1)$ is skewed and sharply shifted to the left, with asymptotic lower and upper 5 percent quantiles of -14.1 and -0.13. With linear detrending, the skewness is even more pronounced, with 5 percent quantiles of -21.8 and -2.66. This imparts a substantial bias to the estimates of α: for example, with $T=100$ and $v_t=\varepsilon_t$, the mean of $\hat{\alpha}^{\tau}$, based on the asymptotic approximation (3.8), is 0.898.

Another feature of regression with I(1) regressors is that, when the regression contains both I(1) and I(0) regressors, estimators of coefficients (and their associated

test statistics) on the I(1) and I(0) regressors, in a suitably transformed regression, are asymptotically independent. This is illustrated here in the AR(p) model with a unit root as analyzed by Fuller (1976). General treatments of regressions with integrated regressors in multiple time series models are given by Chan and Wei (1988), Park and Phillips (1988) and Sims et al. (1990). When v_t has nontrivial short-run dynamics so that $\omega^2 \neq \gamma_{\Delta u}(0)$, an alternative approach to estimating α is to approximate the dynamics of v_t by a pth order autoregression, $a(L)v_t = e_t$. In the time-trend case, this leads to the OLS estimator, $\hat{\alpha}^\tau$, from the regression

$$\Delta y_t = \delta_0 + \delta_1 t + (\alpha - 1)y_{t-1} + \sum_{j=1}^{p} a_j \Delta y_{t-j} + e_t, \qquad t = 1, 2, \ldots, T. \tag{3.9}$$

If v_t in fact follows an AR(p), then $e_t = \varepsilon_t$ and (3.9) is correctly specified. To simplify the calculation, consider the special case of no deterministic terms, conditional homoskedasticity and $p = 1$, so that y_t is regressed on $(y_{t-1}, \Delta y_{t-1})$. Define $\Upsilon_T = \text{diag}(T^{1/2}, T)$, let $a = (a_1, \alpha - 1)'$, let $z_{t-1} = (\Delta y_{t-1}, y_{t-1})$, and let $\hat{a}$ be the OLS estimator of a. Then

$$\Upsilon_T(\hat{a} - a) = \left(\Upsilon_T^{-1} \sum_{t=2}^{T} z_{t-1} z_{t-1}' \Upsilon_T^{-1} \right)^{-1} \left(\Upsilon_T^{-1} \sum_{t=2}^{T} z_{t-1} \varepsilon_t \right).$$

A direct application of the FCLT and the CMT shows that $\Upsilon_T^{-1} \sum_{t=2}^{T} z_{t-1} z_{t-1}' \Upsilon_T^{-1} \Rightarrow \text{diag}(\gamma_{\Delta y}(0), \omega^2 \int W^2)$, where $\omega = \sigma_\varepsilon/(1 - a(1))$ (in the special case $p = 1$, $a(1) = a_1$). Similarly,

$$\Upsilon_T^{-1} \sum_{t=2}^{T} z_{t-1} \varepsilon_t = \left(T^{-1/2} \sum_{t=2}^{T} \Delta y_{t-1} \varepsilon_t, T^{-1} \sum_{t=2}^{T} y_{t-1} \varepsilon_t \right)'.$$

Because $\Delta y_{t-1} \varepsilon_t$ is a martingale difference sequence which satisfies Theorem 1, $T^{-1/2} \sum_{t=2}^{T} \Delta y_{t-1} \varepsilon_t \xrightarrow{d} \eta^*$, where $\eta^* \sim N(0, \sigma_\varepsilon^2 E(\Delta y_t)^2)$. Direct application of Chan and Wei's (1988) Theorem 2.4 (or, alternatively, an algebraic rearrangement of the type leading to (3.4b)) implies that the second term has the limit $\sigma_\varepsilon \omega \int W \, dW$. From Theorem 2.2 of Chan and Wei (1988), this convergence is joint and moreover the (W, η^*) are independent. Upon completing the calculation, one obtains the result

$$\{T^{1/2}(\hat{a}_1 - a_1), T(\hat{\alpha} - 1)\} \Rightarrow \left\{ N(0, 1 - a_1^2), (\sigma_\varepsilon/\omega) \int W \, dW \Big/ \int W^2 \right\}, \tag{3.10}$$

where asymptotically the two terms are independent. The joint asymptotic distribution of $\{T^{1/2}(\hat{a}_1 - a_1), T(\hat{\alpha} - 1)\}$ was originally obtained by Fuller (1976) using different techniques. The result extends to values of $p > 1$ and to more general time trends. For example, in the AR(p) model with a constant and a linear time trend, $T(\hat{\alpha}^\tau - 1) \Rightarrow (\sigma_\varepsilon/\omega) \int W^\tau \, dW / \int (W^\tau)^2$.

Ordinary least squares estimation is not, of course, the only way to estimate α. Interestingly, the asymptotic distribution is sensitive to seemingly minor changes in the estimator. Consider, for example, Dickey's et al. (1984) "symmetric least squares" estimator in the no-deterministic case

$$\bar{\alpha}_s = \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T-1} y_t^2 + \frac{1}{2}(y_1^2 + y_T^2)}. \tag{3.11a}$$

Straightforward algebra and an application of the FCLT reveals that

$$T(\bar{\alpha}_s - 1) = \frac{-\frac{1}{2}T^{-1}\sum_{t=2}^{T}(\Delta y_t)^2}{T^{-2}\left\{\sum_{t=2}^{T-1} y_t^2 + \frac{1}{2}(y_1^2 + y_T^2)\right\}} \Rightarrow -\frac{\kappa}{2}\left(\int W^2\right)^{-1}, \tag{3.11b}$$

so that $T(\bar{\alpha}_s - 1)$ is negative with probability one.

If point estimates are of direct interest, then the bias in the usual OLS estimator can be a problem. For example, if one's object is forecasting, then the use of a biased estimator of α will result in median-biased conditional forecasts of the stochastic component.[6] This has led to the development of median-unbiased estimators of α. This problem is closely related to the construction of confidence intervals for α and is taken up in Section 3.3.

3.2. *Hypothesis tests*

3.2.1. *Test of $\alpha = 1$ in the Gaussian AR(1) model*

The greatest amount of research effort concerning autoregressive unit roots, both empirical and theoretical, has been devoted to testing for a unit root. Because of the large number of tests available, a useful starting point is the no-deterministic i.i.d. Gaussian AR(1) model,

$$y_t = \alpha y_{t-1} + \varepsilon_t, \quad \varepsilon_t \text{ i.i.d. } N(0, \sigma^2), \quad t = 1, 2, \ldots, T, \tag{3.12}$$

[6] When $|\alpha| < 1$ and α is fixed, $\hat{\alpha}$ is also biased towards zero. In the Gaussian AR(1) model with $d_t = 0$, Hurwicz (1950) derives the approximation $E\hat{\alpha} = \{(T^2 - 2T + 3)/(T^2 - 1)\}\alpha$ for α close to zero. When α is close to one, Hurwicz's approximation breaks down but the distribution becomes well-approximated using the local-to-unity approximations discussed in the next section, and the downward bias remains. Approximate biases in the stationary constant model are given by Marriot and Pope (1954). For an application of these bias expressions, see Rudebusch (1993). Also see Magnus and Pesaran (1991) and Stine and Shaman (1989).

where $y_0 = 0$. Suppose further that σ^2 is known, in which case we can set $\sigma^2 = 1$. Because there is only one unknown parameter, α, the Neyman–Pearson lemma can be used to construct the most powerful test of the null hypothesis $\alpha = \alpha_0$ vs. the point alternative $\alpha = \bar{\alpha}$. The likelihood function is proportional to $L(\alpha) = k_T \exp(-\frac{1}{2}(\hat{\alpha} - \alpha)^2 \sum_{t=2}^{T} y_{t-1}^2)$, where k_T does not depend on α. The Neyman–Pearson test of $\alpha = 1$ v.s. $\alpha = \bar{\alpha}$ rejects if $L(\bar{\alpha})/L(1)$ is sufficiently large; after some manipulation, this yields a critical region of the form

$$[T(\bar{\alpha} - 1)]^2 T^{-2} \sum_{t=2}^{T} y_{t-1}^2 - 2T(\bar{\alpha} - 1) T^{-1} \sum_{t=2}^{T} y_{t-1} \Delta y_t < k, \tag{3.13}$$

where k is a constant.

The implication of (3.13) is that the most powerful test of $\alpha = 1$ vs. $\alpha = \bar{\alpha}$ is a linear combination of two statistics, with weights that depend on $\bar{\alpha}$. It follows that, even in this simplified problem, there is no uniformly most powerful (UMP) test of $\alpha = 1$ vs. $|\alpha| < 1$. This difficulty is present even asymptotically: suppose that the alternative of interest is local to one in the sense (2.15), so that $\bar{\alpha} = 1 + \bar{c}/T$, where $\bar{c}$ is a fixed constant. Then, $T(\bar{\alpha} - 1) = \bar{c}$. Under the null

$$\left(T^{-2} \sum_{t=2}^{T} y_{t-1}^2, 2T^{-1} \sum_{t=2}^{T} y_{t-1} \Delta y_t \right) \Rightarrow \sigma_\varepsilon^2 \left(\int W^2, W(1)^2 - 1 \right),$$

so both terms in (3.13) are $O_p(1)$. Thus, there is no single candidate test which dominates on theoretical grounds, either in finite samples [Anderson (1948), Dufour and King (1991)] or asymptotically.[7]

From the perspective of empirical work, the model (3.12) is overly restrictive because of the absence of deterministic components and because the errors are assumed to be i.i.d. The primary objective of the large literature on tests for unit autoregressive roots has, therefore, been to propose tests that have three characteristics: first, the test is asymptotically similar under the general I(1) null, in the sense that the null distribution depends on neither the parameters of the trend process (assuming the trend has been correctly specified) nor the nuisance parameters describing the short-run dynamics of v_t; second, it has good power in large samples; and third, it exhibits small size distortions and good power over a range of empirically plausible models and sample sizes. The next three subsections, respectively, summarize the properties of various unit root tests in terms of these three characteristics.

[7] This draws on Rothenberg (1990). Manipulation of (3.13) shows that the Dickey–Fuller (1979) $\hat{\rho}$ test, which rejects if $T(\hat{\alpha} - 1) < k'$ (where $k' < 0$ for conventional significance levels), is efficient against $\bar{c} = 2k'$, although this does not extend to the demeaned or detrended cases. We thank Thomas Rothenberg and Pentti Saikkonen for pointing this out.

3.2.2. Tests of the general I(1) null

This subsection describes the basic ideas used to generalize tests from the AR(1) model to the general I(1) null by examining four sets of tests in detail. Some other tests of the general I(1) null are then briefly mentioned.

If v_t follows an AR(p) and $d_t = \beta_0 + \beta_1 t$, then the regression (3.9) serves as a basis for two tests proposed by Dickey and Fuller (1979): a test based on the t-statistic testing $\alpha = 1$, $\hat{\tau}^\tau$, and a test based on $\hat{\rho}^\tau = (\hat{\omega}_{AR}/\hat{\sigma}_\varepsilon)T(\hat{\alpha} - 1)$, where $\hat{\omega}^2_{AR}$ is the autoregressive spectral density estimator (the *AR estimator of* ω^2):

$$\hat{\omega}^2_{AR} = \frac{\hat{\sigma}^2_\varepsilon}{\left(1 - \sum_{j=1}^{p} \hat{a}_j\right)^2}, \tag{3.14}$$

where $(\hat{a}_1, \hat{a}_2, \ldots, \hat{\sigma}_\varepsilon)$ are the OLS estimators from (3.9), modified, respectively, to omit t or $(1, t)$ as regressors in the $d_t = \beta_0$ or $d_t = 0$ cases. In the time-trend case, under the null hypothesis $\alpha = 1$ and the maintained AR(p) hypothesis, the limiting representations of these statistics are

$$\{\hat{\rho}^\tau, \hat{\tau}^\tau\} \Rightarrow \left\{ \frac{\int W^\tau \, dW}{\int (W^\tau)^2}, \frac{\int W^\tau \, dW}{\left(\int W^{\tau 2}\right)^{1/2}} \right\}, \tag{3.15}$$

neither of which depends on nuisance parameters. Thus these statistics form the basis for an asymptotically similar test of the unit root hypothesis in the AR(p)/time-trend model. Their distributions have come to be known as the Dickey–Fuller (1979) "$\hat{\rho}^\tau$" and "$\hat{\tau}^\tau$" distributions and are tabulated in Fuller (1976), Tables 8.5.1 and 8.5.2, respectively. In the constant-only case ($d_t = \beta_0$), the only modification is that t is dropped as a regressor from (3.9) and W^μ replaces W^τ in (3.15). In the no-deterministic case, the intercept is also dropped from (3.9) and W replaces W^τ.

In an important extension of Fuller (1976) and Dickey and Fuller (1979), Said and Dickey (1984) used Berk's (1974) results for AR(∞) I(0) autoregressions to analyze the case that v_t follows a general ARMA(p, q) process with unknown p, q. In this case, the true autoregressive order is infinite so the regression (3.9) is misspecified. If, however, the autoregressive order p_T increases with the sample size (specifically, $p_T \to \infty$, $p_T^3/T \to 0$), then Said and Dickey (1984) showed that the results (3.15) continue to hold. Thus the Dickey–Fuller/Said–Dickey tests have a nonparametric interpretation, in the sense that they are valid under a more general I(0) null with weak conditions on the dynamics of v_t.[8]

[8] Berk's (1974) conditions on $c(L)$ (in the notation of (2.1)) are less restrictive than Said and Dickey's (1984) assumption that v_t obeys an ARMA(p, q). For a related discussion and extension to the multivariate case, see Lewis and Reinsel (1985) and especially Saikkonen (1991).

Alternative tests were proposed by Phillips (1987a) and Phillips and Perron (1988). They recognized that if κ in (3.4) were consistently estimable, then $T(\hat{\alpha}-1)$ from the misspecified AR(1) model could be adjusted so that it would be asymptotically similar. This reasoning led Phillips (1987a) to propose the corrected statistics

$$Z_\alpha = T(\hat{\alpha}-1) + \frac{\frac{1}{2}(\hat{\kappa}-1)\hat{\omega}^2}{T^{-2}\sum_{t=2}^{T} y_{t-1}^2} \Rightarrow \frac{\int W\,dW}{\int W^2}, \tag{3.16a}$$

$$Z_\tau = \hat{\tau} + \frac{\frac{1}{2}(\hat{\kappa}-1)\hat{\omega}}{\left(T^{-2}\sum_{t=2}^{T} y_{t-1}^2\right)^{1/2}} \Rightarrow \frac{\int W\,dW}{\left(\int W^2\right)^{1/2}}, \tag{3.16b}$$

where $\hat{\kappa}$ and $\hat{\omega}^2$ are consistent estimators of κ and ω^2, and where τ is the t-statistic testing for a unit root in the OLS regression of y_t onto y_{t-1}. Phillips and Perron (1988) extended these statistics to the constant and time-trend cases by replacing the regression of y_t onto y_{t-1} with a regression of y_t onto $(1, y_{t-1})$ in the constant case or, in the linear-trend case, onto $(1, t, y_{t-1})$. The limiting distributions for these two cases are as in (3.16), with W^μ or W^τ, respectively, replacing W.

Because $\frac{1}{2}(\kappa-1) = \frac{1}{2}(\gamma_v(0)-\omega^2)/\omega^2$, the estimation of the correction entails the estimation of ω^2. Phillips (1987a) and Phillips and Perron (1988) suggested estimating ω^2 using a sum-of-covariances (SC) spectral estimator (the *SC estimator of* ω^2),

$$\hat{\omega}_{\mathrm{SC}}^2 = \sum_{m=-l_T}^{l_T} k\left(\frac{m}{l_T}\right)\hat{\gamma}_{\hat{v}}(m), \tag{3.17}$$

where $\hat{\gamma}_x(m) = (T-m)^{-1}\sum_{t=|m|+1}^{T}(x_t-\bar{x})(x_{t-m}-\bar{x})$, $k(\cdot)$ is a kernel weighting function and $\hat{v}_t$ is the residual from the regression of y_t onto y_{t-1}, $(1, y_{t-1})$ or $(1, t, y_{t-1})$ in the no-deterministic, constant or linear-trend cases, respectively. The appropriate choice of kernel ensures that $\hat{\omega}_{\mathrm{SC}}^2 > 0$ [Newey and West (1987), Andrews (1991)]; Phillips (1987a) and Phillips and Perron (1988) suggested using Bartlett (linearly declining) weights. If l_T increases to infinity at a suitable rate [e.g. $l_T^4/T \to 0$ from Phillips (1987a, Theorem 4.2); see Andrews (1991) for optimal rates], then $\hat{\omega}_{\mathrm{SC}}^2 \xrightarrow{p} \omega^2$ as required. Like the Said–Dickey tests, the Phillips/Phillips–Perron tests thus provide a way to test the general (nonparametric) I(1) null.

A third test for the general I(1) null can be obtained by generalizing a statistic derived by Sargan and Bhargava (1983a) in the no-trend and constant cases and extended to the time-trend case by Bhargava (1986). They used Anderson's (1948) approximation to the inverse of the covariance matrix to derive the locally most

powerful invariant (LMPI) test [although the test is not LMPI if the true inverse is used, as pointed out by Nabeya and Tanaka (1990b)]. The Sargan–Bhargava statistics, $\tilde{R}_T$, $\tilde{R}_T^{\mu}$, and $\tilde{R}_T^{\tau}$, are

$$\tilde{R}_T = \frac{T^{-2}\sum_{t=1}^{T}(y_t)^2}{T^{-1}\sum_{t=2}^{T}(\Delta y_t)^2}, \tag{3.18a}$$

$$\tilde{R}_T^{\mu} = \frac{T^{-2}\sum_{t=1}^{T}(y_t^{\mu})^2}{T^{-1}\sum_{t=2}^{T}(\Delta y_t)^2}, \tag{3.18b}$$

$$\tilde{R}_T^{B} = \frac{T^{-2}\sum_{t=1}^{T}(y_t^{B})^2}{T^{-1}\sum_{t=2}^{T}(\Delta y_t - \overline{\Delta y})^2}, \tag{3.18c}$$

where $y_t^{\mu} = y_t - T^{-1}\sum_{t=1}^{T} y_t$ and $y_t^{B} = y_t - \hat{\beta}_0^{B} - \hat{\beta}_1^{B} t$, where $\hat{\beta}_0^{B} = T^{-1}\sum_{t=1}^{T} y_t - [(T+1)/2(T-1)](y_T - y_1)$ and $\hat{\beta}_1^{B} = (y_T - y_1)/(T-1)$. Note that $\hat{\beta}_1^{B}$ is the maximum likelihood estimator (MLE) of β_1 under the null $\alpha = 1$ when v_t is i.i.d. normal. Also, the statistic $\tilde{R}_T$ is asymptotically equivalent to minus one-half times the inverse of the symmetric least squares estimator $T(\bar{\alpha}_s - 1)$ in (3.11).

These statistics have seen little use in empirical work because of their derivation in the first-order case and because, in their form (3.18), the tests are not similar under the general I(1) null. They are, however, readily extended to the general I(1) null. While of independent interest, this extension is worth explaining here because it demonstrates a simple way that a large class of tests of the unit root model can be extended to the general I(1) case, namely, by replacing y_t by $y_t/\hat{\omega}$. A direct application of the FCLT and the CMT yields $\tilde{R}_T \Rightarrow \kappa^{-2}\int W^2$, $\tilde{R}_T^{\mu} \Rightarrow \kappa^{-2}\int (W^{\mu})^2$ and $\tilde{R}_T^{B} \Rightarrow \kappa^{-2}\int (W^{B})^2$, where $W^{B}(\lambda) = W(\lambda) - (\lambda - \frac{1}{2})W(1) - \int W$. Thus, if $\hat{\kappa}^2$ is consistent, modified Sargan–Bhargava (MSB) statistics are obtained as $R_T = \hat{\kappa}^2 \tilde{R}_T$, $R_T^{\mu} = \hat{\kappa}^2 \tilde{R}_T^{\mu}$ and $R_T^{B} = \hat{\kappa}^2 \tilde{R}_T^{B}$, which are similar under the general I(1) null.

These statistics can be summarized using a compact functional notation. Note that the demeaned (say) MSB statistic with $\hat{\kappa} = \hat{\gamma}_{\Delta y}(0)/\hat{\omega}$ can be written as $R_T^{\mu} = \hat{\omega}^{-2} T^{-1}\sum_{t=1}^{T} Y_T^{\mu}(t/T)^2$, where $Y_T^{\mu}(\lambda) = T^{-1/2} y_{[T\lambda]}^{\mu}$. This suggests the notation

$$\text{MSB} = \int_{\lambda=0}^{1} f(\lambda)^2 \, \mathrm{d}\lambda, \tag{3.19}$$

where $f(\lambda) = \hat{\omega}^{-1} Y_T(\lambda), \hat{\omega}^{-1} Y_T^{\mu}(\lambda)$ and $\hat{\omega}^{-1} Y_T^{B}(\lambda)$, respectively, in the three cases, where $Y_T(\lambda) = T^{-1/2} y_{[Ts]}$ and $Y_T^{B}(\lambda) = T^{-1/2} y_{[T\lambda]}^{B}$.

The approach used to extend the SB statistic to the general I(1) null can be applied to other statistics as well. The Neyman–Pearson test regions (3.13) have the same drawback as the SB critical regions, that is, they depend on ω under the general I(1) null. Consider the no-deterministic case under which (3.13) was derived. Then, $2T^{-1}\sum_{t=2}^{T} y_{t-1}\Delta y_t = (T^{-1/2} y_T)^2 - \gamma_v(0) + o_p(1)$, so the critical regions in (3.13) are asymptotically equivalent to critical regions based on $(T(\bar{\alpha} - 1))^2 T^{-2}\sum_{t=2}^{T} y_{t-1}^2 - T(\bar{\alpha} - 1)(T^{-1/2} y_T)^2$, which has the limiting representation $\omega^2\{\bar{c}^2 \int W^2 - \bar{c} W(1)^2\}$. While this depends on ω, if $\hat{\omega}^2 \xrightarrow{p} \omega^2$ under the null and local alternatives, then an asymptotically equivalent test can be performed using the statistic $P_T = \hat{\omega}^{-2}\{\bar{c}^2 T^{-2}\sum_{t=2}^{T} y_{t-1}^2 - \bar{c} T^{-1} y_T^2\}$. Because $P_T \Rightarrow \bar{c}^2 \int W^2 - \bar{c} W(1)^2$, this test is asymptotically similar and is moreover asymptotically equivalent to the Neyman–Pearson test in the case that v_t is i.i.d. $N(0, \sigma^2)$.

When deterministic terms are present, it is desirable to modify the P_T statistic. A feature of the tests discussed so far is that they are invariant to the values of the parameters describing the deterministic terms (β_0 and β_1 in the linear-trend case), that is, a change in the value of β does not induce a change in the value of the test statistic. This feature is desirable, particularly in the case of β_0. For example, when a test is performed on a series in logarithms, then a change in the units of measurement of the series (from thousands to millions of dollars, say) will appear, after taking logarithms, as an additive shift in β_0. It is natural to require a test for unit roots to be unaffected by the units of measurement, which translates here into requiring that the test be invariant to β_0.

This line of reasoning led Dufour and King (1991), drawing on King (1980), to develop most powerful invariant (MPI) finite-sample tests of $\alpha = \alpha_0$ vs. $\alpha = \alpha_1$, where α_0 is a general value, not necessarily one, in the Gaussian AR(1) model with $Eu_0^2 < \infty$. The finite-sample distribution of these tests hinges on the Gaussian AR(1) assumption and they are not similar under the general I(1) null. These were extended by Elliott et al. (1992) to the general case using the same device as was used to extend the Neyman–Pearson tests to the statistic P_T. The resultant statistics, P_T^{μ} and P_T^{τ}, are asymptotically MPI against the alternative $c = \bar{c}$. These statistics have forms similar to P_T, but are constructed using demeaned and detrended series, where the trend coefficients are estimated by generalized least squares (GLS) under a local alternative ($c = \bar{c}$), rather than under the null. This "local" GLS detrending results in intercept estimators which are asymptotically negligible, so that $P_T^{\mu} \Rightarrow \bar{c}^2 \int W^2 - \bar{c} W(1)^2$. This suggests examining other unit root test statistics using local detrending. One such statistic, proposed by Elliott et al. (1992), is their Dickey–Fuller GLS (DF–GLS) statistic, in which the local GLS-demeaned or local GLS-detrended series is used to compute the t-statistic in the regression (3.9), where the intercept and time trend are suppressed.[9] The construction of the P_T^{μ}, P_T^{τ}, DF–GLS$^{\mu}$ and

[9] The DF–GLS$^{\tau}$ statistic is computed in two steps. Let $z_t = (1, t)$. (1) β_0 and β_1 are estimated by GLS under the assumption that the process is an AR(1) with coefficient $\bar{\alpha} = 1 + \bar{c}/T$ and $u_0 = 0$. That is, β_0

DF–GLS$^{\tau}$ tests requires the user to choose $\bar{c}$. Drawing upon the arguments in King (1988), a case can be made for choosing $\bar{c}$ so that the test achieves the power envelope against stationary alternatives (is asymptotically MPI) at 50 percent power. This turns out to be achieved by setting $\bar{c} = -7$ in the demeaned case and $\bar{c} = -13.5$ in the detrended case.

Another statistic of independent interest is the so-called rescaled range (R/S) statistic which was proposed and originally analyzed by Mandelbrot and Van Ness (1968), Mandelbrot (1975) and Mandelbrot and Taqqu (1979). The statistic is

$$\mathrm{R/S} = \frac{T^{-1/2}(\max_{t=1,\ldots,T} y_t - \min_{t=1,\ldots,T} y_t)}{T^{-1}\sum_{t=2}^{T}(\Delta y_t)^2}. \tag{3.20}$$

Although the R/S statistic was originally proposed as a method for measuring the differencing parameter in a fractionally integrated (fractionally differenced) model, the R/S test also has power against stationary roots in the autoregressive model. In functional notation, the statistic is $\sup_\lambda f(\lambda) - \inf_\lambda f(\lambda)$, which is a continuous functional from $C[0,1]$ to $\mathscr{R}^1$. As Lo (1991) pointed out, this statistic is not similar under the general I(1) null, but if evaluated using $f(\lambda) = T^{-1/2} y_{[T\lambda]}/\hat{\omega}$ it is asymptotically similar (note that this statistic needs no explicit demeaning in the $d_t = \beta_0$ case). Thus, the asymptotic representation of this modified R/S statistic under the general I(1) null is $\sup_\lambda W(\lambda) - \inf_\lambda W(\lambda)$.

Although a large number of unit root tests have been proposed, many fall in the same family, in the sense that they have the same functional representation. It will be shown in the next section that if two tests have the same functional representation then they have the same local asymptotic power functions. However, as will be seen in Section 3.2.4, tests which are asymptotically equivalent under the null and local alternatives can perform quite differently in finite samples.

3.2.3. *Consistency and local asymptotic power*

Consistency. A simple argument proves the consistency of unit root tests which can be written in functional notation such as (3.19). Suppose that a test has the representation $g(f)$, where g: $C[0,1] \to \mathscr{R}$ is continuous, and f is $T^{-1/2} y_{[T\cdot]}/\hat{\omega}$, $T^{-1/2} y^{\mu}_{[T\cdot]}/\hat{\omega}$, or $T^{-1/2} y^{\tau}_{[T\cdot]}/\hat{\omega}$ in the no-deterministic, demeaned or detrended cases, respectively. If $g(0)$ falls in the rejection region, then consistency follows immediately, provided that the process being evaluated is consistent for zero under all fixed

---

and β_1 are estimated by regressing $[y_1, (1-\bar{\alpha}L)y_2, \ldots, (1-\bar{\alpha}L)y_T]$ onto $[z_1, (1-\bar{\alpha}L)z_2, \ldots, (1-\bar{\alpha}L)z_T]$; call the resulting estimator $\bar{\beta}_{\mathrm{GLS}}$. Detrended $\bar{y}_t = y_t - z_t'\bar{\beta}_{\mathrm{GLS}}$ is then computed. (2) The Dickey–Fuller regression (3.9) is run using $\bar{y}_t$ without the intercept and time trend; the t-statistic on $\bar{y}_{t-1}$ is the DF–GLS$^{\tau}$ statistic. The DF–GLS$^{\mu}$ statistic is computed similarly except that the regressor t is omitted in the first step. The DF–GLS$^{\mu}$ statistic has the no-deterministic Dickey–Fuller $\hat{\tau}$ distribution and the distribution of DF–GLS$^{\mu}$ is tabulated in Elliott et al. (1992).

alternatives. As a concrete example, let $d_t = \beta_0$ and consider the demeaned MSB statistic (3.19) with $f = \hat{\omega}^{-1} Y_T^{\mu}$. The test rejects for small values of the statistic, so consistency against the general I(0) alternative follows if $\hat{\omega}^{-1} Y_T^{\mu} \xrightarrow{p} 0$. Now if u_t is I(0) then $\Pr[\sup_\lambda | Y_T^{\mu}(\lambda)| > \delta] \to 0$ for all $\delta > 0$ [the proof is along the lines of (2.8)]. It follows that $\hat{\omega}^{-1} Y_T^{\mu} \xrightarrow{p} 0$ if $\hat{\omega} \xrightarrow{p} k > 0$ for some constant k under the I(0) alternative. Thus, with this additional assumption, $\text{MSB}^{\mu} = \int (\hat{\omega}^{-1} Y_T^{\mu}(\lambda))^2 \, d\lambda + o_p(1) \xrightarrow{p} 0$ under the fixed I(0) alternative, and the test is consistent.

The assumption that $\hat{\omega}^2 \xrightarrow{p} k > 0$ for some constant k under the I(0) alternative is valid for certain variants of both the SC and AR spectral estimators. For the AR spectral estimator, this was shown by Stock (1988, Lemma 1). For the SC spectral estimator, test consistency is an implication of Phillips and Ouliaris' (1990, Theorem 5.1) more general result for tests for cointegration. These results, combined with some additional algebra, demonstrate the consistency for the MSB, Z_α, Z_τ, P_T and R/S statistics.[10]

Local asymptotic power. Power comparisons are a standard way to choose among competing tests. Because finite-sample distribution theory in nearly $I(1)$ models is prohibitively complicated, research has focused on asymptotic approximations to power functions. For consistent tests, this requires computing power against alternatives which are local to (in a decreasing neighborhood of) unity. Applications of asymptotic expansions commonly used in $T^{1/2}$-asymptotic problems, in particular Edgeworth expansions and saddlepoint approximations, provided poor distributional approximations for α near unity [Phillips (1978); also, see Satchell (1984)]. This led to the exploration of the alternative nesting, $\alpha_T = 1 + c/T$; important early work developing this approach includes Bobkoski (1983), Cavanagh (1985), Phillips (1987a, 1987b), Chan and Wei (1987) and Chan (1988, 1989).

The treatment here follows Bobkoski (1983) as generalized in Example 4 of Section 2.3. The key observation is that, under the local-to-unity alternative (2.15), the processes $T^{-1/2} u_{[T\cdot]} \Rightarrow \omega W_c(\cdot)$, where W_c is a diffusion process on the unit interval satisfying $dW_c(\lambda) = cW_c(\lambda) + dW(\lambda)$ with $W_c(0) = 0$. In addition, both the SC and AR spectral estimators have the property that $\hat{\omega}^2 \xrightarrow{p} \omega^2$ under the local alternative.[11] These results directly yield local-to-unity representations of those test statistics with functional representations such as (3.19).

[10] Not all plausible estimators of ω^2 will satisfy this condition. For example, consider the SC estimator constructed using not the quasi-difference $y_t^d - \hat{\alpha} y_{t-1}^d$, as in (3.17), but the first difference Δy_t^d. These two estimators are asymptotically equivalent under the null, but under the alternative y_t is overdifferenced. Thus, the spectrum of Δy_t^d at frequency zero is zero under the I(0) alternative, the SC estimator of the spectrum does not satisfy the positivity condition, and tests constructed using the first-differenced SC estimator are not in general consistent. [Precise statements of this result are given by Stock and Watson (1988b) in the MA(q) case with l_T fixed and by Phillips and Ouliaris (1990, Theorem 5.2) in the general case.] This problem of overdifferencing by imposing $\alpha = 1$ when nuisance parameters are estimated also results in the inconsistency of Solo's (1984) Lagrange multiplier (LM) test for a unit AR root in an ARMA(p, q) model, as demonstrated by Saikkonen (1993).

[11] See Phillips (1987b) for the SC estimator and Stock (1988) for the AR estimator.

As a concrete example, again consider the MSB$^{\mu}$ statistic. Under the local-to-unity alternative, $Y_T^{\mu} \Rightarrow \omega(W_c - \int W_c) \equiv \omega W_c^{\mu}$. Thus, test statistics of the form $g(\hat{\omega}^{-1} Y_T^{\mu})$ have the local-to-unity representation $g(W_c^{\mu})$. An important implication is that the local asymptotic power of these tests does not depend on the nuisance parameter ω, simplifying their comparison.

Phillips (1987b, Theorem 2) showed that this framework bridges the gap between the conventional Gaussian I(0) asymptotics and the nonstandard I(1) asymptotics. Specifically, as $c \to -\infty$ the (suitably normalized) local-to-unity approximations for $T(\hat{\alpha} - 1)$ and the associated t-statistic approached their I(0) Gaussian limits and, as $c \to +\infty$, these distributions, respectively, tend to Cauchy and normal, in accordance with the asymptotic results of White (1958, 1959) and Anderson (1959) for the Gaussian AR(1) model with $|\alpha| > 1$.

Another application of this approach is to derive the asymptotic Gaussian power envelope for unit root tests, that is, the envelope of the power functions of the family of most powerful unit root tests. Because there is no UMP test (or, in the time-trend case, no uniformly most powerful invariant test), this envelope provides a concrete way to judge the absolute asymptotic performance of various unit root tests. In the no-deterministic case, this envelope is readily derived using the local-to-unity limit (2.17) and the Neyman–Pearson critical regions (3.13). Assume that (i) the process is a Gaussian AR(1) so that $\omega^2 = \sigma_\varepsilon^2$; (ii) $Eu_0^2 < \infty$; (iii) the alternative against which the test is most powerful is local-to-unity, so that $\bar{c} = T(\bar{\alpha} - 1)$ is fixed; and (iv) the true process is local-to-unity with $c = T(\alpha - 1)$. Then, the probability of rejecting $\alpha = 1$ against the one-sided alternative $|\bar{\alpha}| \sim 1$ is, asymptotically,

$$\Pr\left[(T(\bar{\alpha} - 1))^2 T^{-2} \sum_{t=2}^{T} y_{t-1}^2 - 2T(\bar{\alpha} - 1) T^{-1} \sum_{t=2}^{T} y_{t-1} \Delta y_t < k \right]$$

$$\to \Pr\left[\bar{c}^2 \int W_c^2 - \bar{c} W_c(1)^2 < k' \right], \tag{3.21}$$

where k and k' are constants which do not depend on c. When $c = \bar{c}$, the second expression in (3.21) provides the envelope of the power functions of the most powerful (Neyman–Pearson) tests and, thus, provides an asymptotic performance bound on all unit root tests in the Gaussian model.

This result is extended in several ways in Elliott et al. (1992). The bound (3.21) is shown to hold if d_t is unknown but changes slowly, in the sense that d_t satisfies

$$T^{-1} \sum_{t=1}^{T} (\Delta d_t)^2 \to 0 \quad \text{and} \quad T^{-1/2} \max_{t=1,\ldots,T} |d_t| \to 0. \tag{3.22}$$

If $d_t = \beta_0 + \beta_1 t$, then the bound (3.21) cannot be achieved uniformly in β_1, but a similar bound can be derived among the class of all invariant tests, and this is achieved by the P_T^{τ} statistic. Although the bound (3.21) was motivated here for v_t i.i.d.

$N(0, \sigma_\varepsilon^2)$, this bound applies under the more general condition that v_t obeys a Gaussian AR(p).

We now turn to numerical results for asymptotic power, computed using the local-to-unity asymptotic representations of various classes of statistics. In addition to the tests discussed so far, we include expressions for the Park (1990) $J(p, q)$ variable-addition test [also, see Park and Choi (1988)] and, in the no-deterministic case, the modified asymptotically LMPI test [invariant under change of scale, this test rejects for small values of $(T^{-1/2}y_T/\hat{\omega})^2$ and is obtained by letting $c \to 0$ in (3.13) and rearranging]. Let W_c^d denote a general OLS-detrended W_c process, that is, $W_c^d = W_c$ in the no-deterministic case, $W_c^d(\lambda) = W_c^\mu(\lambda) = W_c(\lambda) - \int W_c$ in the demeaned case and $W_c^d(\lambda) = W_c^\tau(\lambda) = W_c(\lambda) - (4 - 6\lambda)\int W_c - (12\lambda - 6)\int sW_c(s)\,ds$ in the detrended case [cf. (2.13)]. Let W_c^B denote the asymptotic limit of the Bhargava-detrended process $T^{-1/2}y^B_{[T\cdot]}$ used to construct the Bhargava statistic (3.18c) in the detrended case; specifically, $W_c^B(\lambda) = W_c(\lambda) - (\lambda - \frac{1}{2})W_c(1) - \int W_c$. Finally, let $V_c(\lambda) = W_c(\lambda) - \lambda\{\bar{c}^\dagger W_c(1) + (1 - \bar{c}^\dagger)3\int rW_c(r)\,dr\}$, where $\bar{c}^\dagger = (1 - \bar{c})/(1 - \bar{c} + \bar{c}^2/3)$, denote the limit of the detrended process obtained from local GLS detrending, which is used to construct the P_T^τ and DF–GLS$^\tau$ statistics. The local-to-unity representations for various classes of unit root test statistics are given by the following expressions:

$$\hat{\rho}\text{-class} \qquad \frac{1}{2}\left[\int (W_c^d)^2\right]^{-1}\{W_c^d(1)^2 - W_c^d(0)^2 - 1\}; \tag{3.23a}$$

$$\hat{\tau}\text{-class} \qquad \frac{1}{2}\left[\int (W_c^d)^2\right]^{-1/2}\{W_c^d(1)^2 - W_c^d(0)^2 - 1\}; \tag{3.23b}$$

$$\text{SB-class} \qquad \int (W_c^d)^2 \quad \text{(no-deterministic, demeaned cases);} \tag{3.23c}$$

$$\text{SB-class} \qquad \int (W_c^B)^2 \quad \text{(detrended case);} \tag{3.23d}$$

$$\text{R/S} \qquad \sup_{\lambda\in(0,1)} W_c^d(\lambda) - \inf_{\lambda\in(0,1)} W_c^d(\lambda); \tag{3.23e}$$

$$J(0,1) \qquad \frac{\int (W_c^\mu)^2}{\int (W_c^\tau)^2} - 1 \quad \text{(demeaned case);} \tag{3.23f}$$

$$J(1,2) \qquad \frac{\int (W_c^\tau)^2}{\int (W_c^q)^2} - 1 \quad \text{(detrended case);} \tag{3.23g}$$

$$\text{LMPI} \quad W_c(1)^2 \quad \text{(no-deterministic case only);} \tag{3.23h}$$

$$P_T \quad \bar{c}^2 \int W_c^2 - \bar{c} W_c(1)^2 \quad \text{(no-deterministic, demeaned cases);} \tag{3.23i}$$

$$P_T \quad \bar{c}^2 \int V_c^2 - (\bar{c} - 1) V_c(1)^2 \quad \text{(detrended case);} \tag{3.23j}$$

$$\text{DF–GLS} \quad \frac{1}{2}\left(\int W_c^2 \right)^{-1/2} \{ W_c(1)^2 - W_c(0)^2 - 1 \} \quad \text{(demeaned case);} \tag{3.23k}$$

$$\text{DF–GLS} \quad \frac{1}{2}\left(\int V_c^2 \right)^{-1/2} \{ V_c(1)^2 - V_c(0)^2 - 1 \} \quad \text{(detrended case).} \tag{3.23l}$$

The $\hat{\rho}$ class includes the Dickey–Fuller (1979) $\hat{\rho}$ tests and the Phillips (1987a)/Phillips–Perron (1988) Z_α tests. The $\hat{\tau}$ class includes the Dickey–Fuller (1979) t-tests and the Phillips (1987a)/Phillips–Perron (1988) Z_τ tests. The SB class includes the Schmidt–Phillips (1992) test. Most of these representations can be obtained by directly applying the previous results. For those statistics with functional representations already given and where the statistic is evaluated using OLS detrending [the SB-class (demeaned case) and R/S statistics], the results obtain as a direct application of the continuous mapping theorem. In the cases involving detrending other than OLS in the time trend case (the SB-class and DF–GLS$^\tau$ statistics), an additional calculation must be made to obtain the limit of the detrended processes. The other expressions follow by direct calculation.[12]

Asymptotic power functions for leading classes of unit root tests (5 percent level) are plotted in Figures 1, 2 and 3 in the no-deterministic, constant and trend cases, respectively.[13] The upper line in these figures is the Gaussian power envelope. In the $d_t = 0$ case, the power functions for the $\hat{\tau}, \hat{\rho}$ and SB tests are all very close to the power envelope, so this comparison provides little basis for choosing among them. Also plotted in Figure 1 is the power function of the LMPI test. Although this test has good power against c quite close to zero, its power quickly falls away from the envelope and is quite poor for distant alternatives.

[12] References for these results include: for the $\hat{\rho}, \hat{\tau}$ statistics, Phillips (1987b) (Z_α statistic) and Stock (1991) (Dickey–Fuller AR(p) statistics); for the SB-class statistics, Schmidt and Phillips (1992) and Stock (1988); for the P_T and DF–GLS statistics, Elliott et al. (1992).

[13] The asymptotic power functions were computed using the functional representations in (3.23) evaluated with discrete Gaussian random walks ($T = 500$) replacing the Brownian motions, with 20,000 Monte Carlo replications. Nabeya and Tanaka (1990b) tabulate the power functions for tests including the SB and $\hat{\tau}$ tests, although they do not provide the power envelope. Because they derive and integrate the characteristic function for these statistics in the local-to-unity case, their results presumably have higher numerical accuracy than those reported here. Standard errors of rejection rates in Figures 1–3 are at most 0.004. Some curves in these figures originally appeared in Elliott et al. (1992).

In the empirically more relevant cases of a constant or constant and trend, the asymptotic power functions of the various tests differ sharply. First, consider the case $d_t = \beta_0$. Perhaps the most commonly used test in practice is the Dickey–Fuller/Said–Dickey t-test, $\hat{t}^{\mu}$; however, its power is well below not just the power envelope but the power of the $\hat{\rho}^{\mu}$ (equivalently, $Z_{\hat{\alpha}}$) test. The SB-class statistics have

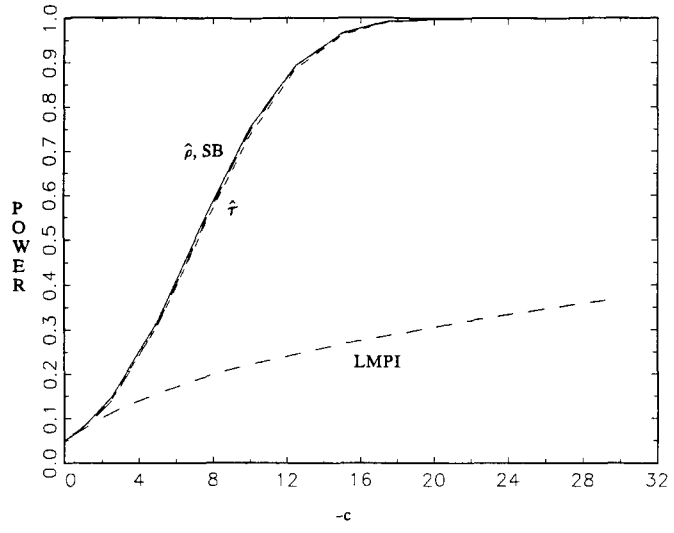

Figure 1

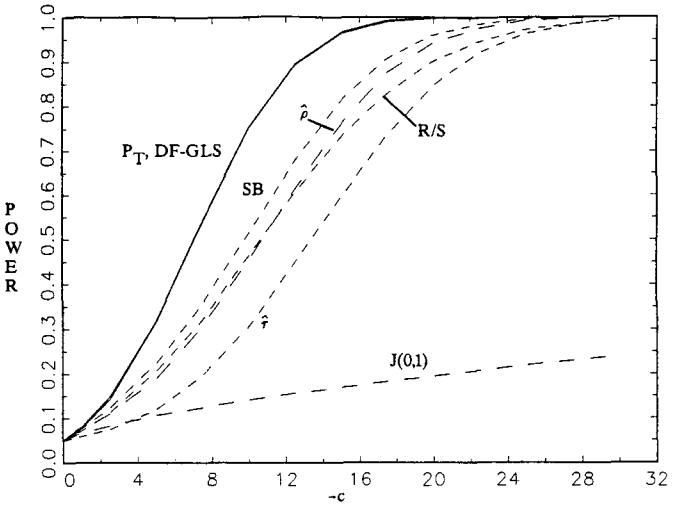

Figure 2

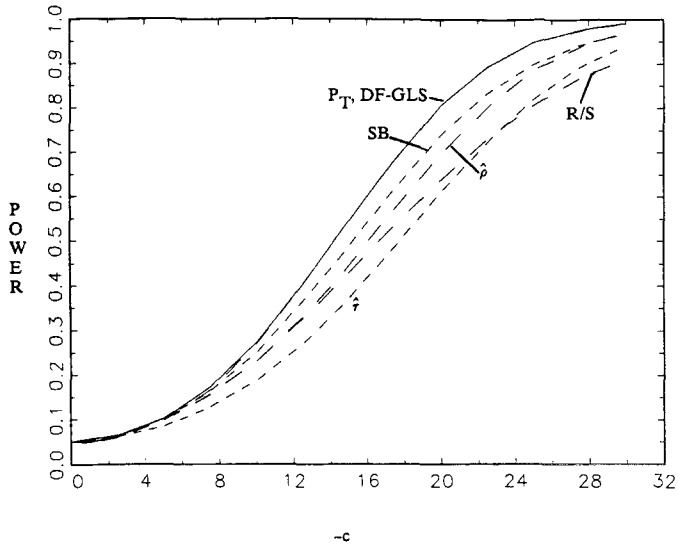

Figure 3

asymptotic power slightly above the $\hat{\rho}^{\mu}$ statistics, particularly for power between 0.3 and 0.8, but remains well below the envelope. In contrast, the asymptotic local power function of the P_T^{μ} test, which is, by construction, tangent to the power envelope at 50 percent power, is effectively on the power envelope for all values of c. Similarly, the DF–GLS$^{\mu}$ power function is effectively on the power envelope.

Pitman efficiency provides a useful way to assess the importance of these power differences. Pitman's proposal was to consider the behavior of two tests of the same hypothesis against a sequence of local alternatives, against which at least one of the tests had nondegenerate power. The Pitman efficiency [or asymptotic relative efficiency (ARE)] is the ratio of the sample sizes giving, asymptotically, the same power for that sequence. In conventional $\sqrt{T}$-normal asymptotics, often the ARE can be computed as a ratio of the variances entering the denominators of the two Studentized test statistics. Although this approach is inapplicable here, the ARE can be calculated using the asymptotic power functions. Suppose that two tests achieve power β against local alternatives $c_1(\beta)$ and $c_2(\beta)$; then the ARE of the first test relative to the second test is $c_1(\beta)/c_2(\beta)$ [Nyblom and Mäkeläinen (1983)]. Using this device, the ARE of the P_T^{μ} test, relative to the optimal test, at power of 50 percent is 1.0, by construction, and the ARE of the DF–GLS$^{\mu}$ test is effectively 1. In contrast, the ARE's of the SB-, $\hat{\rho}$- and $\hat{\tau}$-class tests, relative to the P_T^{μ} test, are 1.40, 1.53 and 1.91. That is, to achieve 50 percent power against a local alternative using the Dickey–Fuller t-statistic asymptotically requires 90 percent more observations than are needed using the asymptotically efficient P_T^{μ} test or the nearly efficient DF–GLS$^{\mu}$ test.

The results in the detrended case are qualitatively similar but, quantitatively, the power differences are less. The $\hat{t}^\tau$-class statistics have low power relative to the envelope and to the SB- and $\hat{\rho}^\tau$-class tests. The SB-class tests have power slightly above the $\hat{\rho}^\tau$-class tests and all power functions are dominated by the P_T^τ test. Some of the other tests, in particular the R/S test, have power that is competitive with the MSB- and $\hat{\rho}^\tau$-class tests. At 50 percent power, the Pitman efficiency of the $\hat{t}$ tests is 1.39 and of the $\hat{\rho}$ tests is 1.25. Interestingly, the power function of P_T^τ actually lies above the power function of $\hat{t}^\mu$, even though P_T^τ involves the additional estimation of the linear-trend coefficient β_1. Comparing the results across the figures highlights a common theme in this literature: including additional trend terms reduces the power of the unit root tests if the trends are unnecessary.[14]

So far, the sampling frequency has been fixed at one observation per period. A natural question is whether power can be increased by sampling more frequently, for example, by moving from annual to quarterly data, while keeping the span of the data fixed. A simple argument, however, shows that it is the span of the data which matters for power, not the frequency of observation. To be concrete, consider the demeaned case and suppose that the true value of α is $1 + c_1/T$, based on T annual observations, where c_1 is fixed. Suppose that the MSB statistic is used with sufficiently many lags for $\hat{\omega}^2$ to be consistent. With the annual data, the test statistic has the limiting representation $\int(W_{c_1}^\mu)^2$. The quarterly test statistic has the limiting representation $\int(W_{4c_4}^\mu)^2$, where c_4 is the local-to-unity parameter at the quarterly frequency and the factor of 4 arises because there are four times as many quarterly as annual observations. Because $\alpha = 1 + c_1/T$ at the annual level, at the quarterly level this root is $\alpha^{1/4} \cong 1 + (c_1/4T) = 1 + (c_4/T)$, so $c_4 = c_1/4$. Thus, $\int(W_{4c_4}^\mu)^2 = \int(W_{c_1}^\mu)^2$ and the quarterly and annual statistics have the same limiting representations and, hence, the same rejection probabilities. Although there are four times as many observations, the quarterly root is four times closer to one than the annual root, and these two effects cancel asymptotically. For theoretical results, see Perron (1991b); for Monte Carlo results, see Shiller and Perron (1985). More frequent observations, however, might improve estimation of the short-run dynamics, and this, apparently, led Choi (1993) to find higher finite-sample power at higher frequencies in a Monte Carlo study.

The case of u_0 drawn from its unconditional distribution. The preceding analysis makes various assumptions about u_0: to derive the finite-sample Neyman–Pearson tests, that $u_0 = 0$ (equivalently, u_0 is fixed and known) and for the asymptotics, that

[14] Asymptotic power was computed for the Dickey–Fuller (1981) and Perron (1990a) F-tests, but this is not plotted in the figures. These statistics test the joint restriction that $\alpha = 1$ and that $\delta_1 = 0$ in (3.5) or (3.9). Unlike the other tests considered here, these F-tests are not invariant to the trend parameter under local and fixed alternatives. The power of the two F-tests depends on β_1 under the alternative, so for drifts sufficiently large their power functions can, in theory, exceed the power envelope for invariant tests. If $\beta_1 = 0$ or is small, the F-tests have very low asymptotic power; well below the $\hat{t}$-class tests. Perron's (1990a) calculations indicate, however, that for β_1 sufficiently large, the F-tests can have high (size-adjusted) power.

$T^{-1/2}u_0 \xrightarrow{p} 0$, as specified after (2.9). Under the null, the tests considered are invariant to β_0 and thus to u_0. Although this finite-u_0 case has received the vast majority of the attention in the literature, some work addresses the alternate model that u_0 is drawn from its unconditional distribution or is large relative to the sample size. In finite samples, this modification is readily handled and leads to different tests [see Dufour and King (1991)]. The maximum likelihood estimator is different from that when u_0 is fixed, being the solution to a cubic equation [Koopmans (1942); for the regression case, Beach and MacKinnon (1978)].

As pointed out by Evans and Savin (1981b, 1984) and further studied by Nankervis and Savin (1988), Perron (1991a), Nabeya and Sorensen (1992), Schmidt and Phillips (1992) and DeJong et al. (1992a), the power of unit tests depends on the assumption about u_0. Analytically, this dependence arises automatically if the asymptotic approximation relies on increasingly finely observed data, in Phillips (1987a) terminology, continuous record asymptotics [see Perron (1991a, 1992), Sorensen (1992) and Nabeya and Sorensen (1992)]. Alternatively, equivalent expressions can be obtained with the local-to-unity asymptotics used here if $T^{-1/2}u_0 = O_p(1)$ [in the stationary AR(1) case, a natural device is to let $T^{-1/2}u_0$ be distributed $N(0, \sigma_\varepsilon^2/T(1-\alpha^2)) \to N(0, -\frac{1}{2}\sigma_\varepsilon^2/c)$, where $c < 0$, so that an additional term appears in (2.17)]. Elliott (1993a) derives the asymptotic power envelope under the unconditional case and shows that tests which are efficient in the unconditional case are not efficient in the conditional case in either the demeaned or detrended cases. The quantitative effect on the most commonly used unit root tests of drawing u_0 from its unconditional distribution is investigated in the Monte Carlo analysis of the next subsection.

3.2.4. *Finite-sample size and power*

There is a large body of Monte Carlo evidence on the performance of tests for a unit AR root. The most influential Monte Carlo study in this literature is Schwert (1989), which found large size distortions in tests which are asymptotically similar under the general I(1) null, especially the Phillips–Perron (1988) Z_α and Z_τ statistics. A partial list of additional papers which report simulation evidence includes Dickey and Fuller (1979), Said and Dickey (1985), Perron (1988, 1989c, 1990a), Diebold and Rudebusch (1991b), Pantula and Hall (1991), Schmidt and Phillips (1992), Elliott et al. (1992), Pantula et al. (1992), Hall (1992a), DeJong et al. (1992b), Ng and Perron (1993a, 1993b) and Bierens (1993).

Taken together, these experiments suggest four general findings. First, all the asymptotically valid tests exhibit finite-sample size distortions for models which are in a sense close to I(0) models. However, the extent of the distortion varies widely across tests and depends on the details of the construction of the spectral estimator $\hat{\omega}^2$. Second, the estimation of nuisance parameters describing the short-run dynamics reduces test power, in some cases dramatically. Third, these two observations lead to the use of data-dependent truncation or AR lag lengths in the

estimation of ω^2 and the resulting tests show considerable improvements in size and power. Fourth, the presence of nonnormality or conditional heteroskedasticity in the errors results in size distortions, but these are much smaller than the distortions arising from the short-run dynamics.

We quantify these findings using a Monte Carlo study with eight designs (data generating processes or DGP's) which reflect some leading cases studied in the literature. In each, $y_t = u_t$, where $u_t = \alpha u_{t-1} + v_t$. Five values of α were considered: 1.0, 0.95, 0.9, 0.8 and 0.7. All results are for $T = 100$. The DGP's are

$$\text{Gaussian MA(1):} \qquad v_t = \varepsilon_t - \theta\varepsilon_{t-1}, \quad u_0 = 0, \quad \theta = 0.8, 0.5, 0, -0.5, -0.8, \tag{3.24a}$$

$$\text{Gaussian MA(1), } u_0 \text{ unconditional:} \qquad v_t = \varepsilon_t - \theta\varepsilon_{t-1}, \quad u_0 \sim N(0, \gamma_u(0)), \quad \theta = 0.5, 0, -0.5, \tag{3.24b}$$

where in each case $\varepsilon_t \sim$ i.i.d. $N(0, 1)$. The Gaussian MA(1) DGP (3.24a) has received the most attention in the literature and was the focus of Schwert's (1989) study. The unconditional variant is identical under the null, but under the alternative u_0 is drawn from its unconditional distribution $N(0, \gamma_u(0))$, where $\gamma_u(0) = (1 + \theta^2 - 2\theta\alpha)/(1 - \alpha^2)$. This affects power but the size is the same as for (3.24a). The unconditional model is of particular interest because the power functions in Section 3.2.3 were for the so-called conditional (u_0 fixed) case.

The tests considered are: the Dickey–Fuller $\hat{\rho}$ statistic $T(\hat{\alpha} - 1)/(1 - \sum_{j=1}^{p} \hat{a}_j)$ [where $\hat{\alpha}, \hat{a}_1, \ldots, \hat{a}_p$ are ordinary least squares estimators (OLSE's) from (3.9)]; the Phillips (1987a)/Phillips–Perron (1988) Z_α statistic (3.16a); the Dickey–Fuller $\hat{t}$ statistic computed from the AR($p + 1$) (3.9); the MSB statistic (3.19) computed using $\hat{\omega}_{\text{AR}}$; the Schmidt–Phillips (1992) statistic, which is essentially (3.19) computed using $\hat{\omega}_{\text{SC}}$; and the DF–GLS of statistic of Elliott et al. (1992).[15]

Various procedures for selecting the truncation parameter l_T in $\hat{\omega}^2_{\text{SC}}$ and the autoregressive order p_T in $\hat{\omega}^2_{\text{AR}}$ are considered. Theoretical and simulation evidence suggest using data-based rules for selecting l_T. Phillips and Perron (1988) and DeJong et al. (1992b) use the Parzen kernel, so this kernel is adopted here.[16] The truncation parameter l_T was chosen using Andrews' (1991) optimal procedure for this kernel as given in his equations (5.2) and (5.4). The AR estimator lag length p_T in (3.14) (with a constant but no time trend in the regression, in both the demeaned

[15] The results here are drawn from the extensive tabulations of 20 tests in 13 data generating processes (DGP's) in Elliott (1993b). Other tests examined include: the Dickey–Fuller (1981) and Perron (1990b) F-tests; the Phillips–Perron Z_τ test; the modified R/S statistic; Hall's (1989) instrumental variable statistic; Stock's (1988) MZ_α statistic; and the Park $J(p, p + 3)$ tests for $p = 1, 2$. In brief, each of these tests had drawbacks – distorted size, low power or both – which, in our view, makes them less attractive than the tests examined here, so, to conserve space, these results are omitted.

[16] The Parzen kernel is given by: $k(x) = 1 - 6x^2 + 6|x|^3$, $0 \leq |x| \leq \frac{1}{2}$; $k(x) = 2(1 - |x|)^3$, $\frac{1}{2} \leq |x| \leq 1$; and $k(x) = 0$, $|x| > 1$.

Table 1
Size and size-adjusted power of selected tests of the I(1) null: Monte Carlo results
(5 percent level tests, demeaned case, $T = 100$, $y_t = u_t, u_t = \alpha u_{t-1} + v_t, v_t = \varepsilon_t - \theta\varepsilon_{t-1}$).[a]

Test Statistic	α	Asymptotic Power	$MA(1)$, $\theta =$					Unconditional: $MA(1)$, $\theta =$		
			−0.8	−0.5	0.0	0.5	0.8	−0.5	0.0	0.5
DF-$\hat{t}^{\mu}$ AR(4)	1.00	*0.05*	*0.04*	*0.05*	*0.05*	*0.06*	*0.30*	*0.05*	*0.05*	*0.06*
	0.95	0.12	0.07	0.09	0.09	0.11	0.13	0.09	0.10	0.12
	0.90	0.31	0.14	0.18	0.20	0.25	0.32	0.19	0.21	0.26
	0.80	0.85	0.37	0.44	0.49	0.66	0.77	0.46	0.51	0.67
	0.70	1.00	0.58	0.66	0.73	0.89	0.95	0.67	0.74	0.90
DF-$\hat{t}^{\mu}$ AR(BIC)	1.00	*0.05*	*0.08*	*0.06*	*0.06*	*0.08*	*0.46*	*0.06*	*0.06*	*0.08*
	0.95	0.12	0.11	0.10	0.10	0.13	0.13	0.11	0.11	0.13
	0.90	0.31	0.23	0.22	0.22	0.31	0.31	0.24	0.24	0.32
	0.80	0.85	0.55	0.56	0.59	0.77	0.78	0.57	0.60	0.77
	0.70	1.00	0.76	0.79	0.83	0.96	0.96	0.80	0.84	0.96
DF-$\hat{t}^{\mu}$ AR(LR)	1.00	*0.05*	*0.08*	*0.08*	*0.07*	*0.14*	*0.43*	*0.08*	*0.07*	*0.14*
	0.95	0.12	0.10	0.11	0.11	0.15	0.13	0.12	0.11	0.14
	0.90	0.31	0.20	0.25	0.26	0.34	0.30	0.26	0.26	0.34
	0.80	0.85	0.40	0.56	0.65	0.72	0.63	0.56	0.66	0.72
	0.70	1.00	0.56	0.73	0.85	0.84	0.72	0.74	0.86	0.85
DF-$\hat{\rho}^{\mu}$ AR(BIC)	1.00	*0.05*	*0.13*	*0.10*	*0.08*	*0.13*	*0.62*	*0.10*	*0.08*	*0.13*
	0.95	0.19	0.17	0.18	0.17	0.17	0.13	0.15	0.16	0.15
	0.90	0.45	0.33	0.35	0.37	0.40	0.31	0.35	0.36	0.38
	0.80	0.94	0.67	0.73	0.76	0.85	0.77	0.73	0.76	0.85
	0.70	1.00	0.85	0.91	0.93	0.98	0.91	0.90	0.93	0.98
Z_{α} SC(auto)	1.00	*0.05*	*0.02*	*0.02*	*0.06*	*0.44*	*0.98*	*0.02*	*0.06*	*0.44*
	0.95	0.19	0.16	0.17	0.18	0.21	0.13	0.15	0.16	0.16
	0.90	0.45	0.38	0.39	0.44	0.50	0.31	0.40	0.42	0.44
	0.80	0.94	0.85	0.87	0.94	0.96	0.71	0.88	0.94	0.95
	0.70	1.00	0.98	0.99	1.00	1.00	0.93	0.99	1.00	1.00
MSB AR(BIC)	1.00	*0.05*	*0.15*	*0.11*	*0.09*	*0.09*	*0.41*	*0.11*	*0.09*	*0.09*
	0.95	0.22	0.19	0.20	0.20	0.18	0.13	0.16	0.16	0.14
	0.90	0.51	0.36	0.40	0.40	0.38	0.28	0.36	0.37	0.34
	0.80	0.96	0.66	0.73	0.76	0.76	0.67	0.71	0.73	0.74
	0.70	1.00	0.82	0.89	0.91	0.92	0.89	0.87	0.90	0.90
MSB SC(auto)	1.00	*0.05*	*0.01*	*0.02*	*0.04*	*0.38*	*0.97*	*0.02*	*0.04*	*0.38*
	0.95	0.22	0.21	0.22	0.22	0.22	0.22	0.16	0.17	0.16
	0.90	0.51	0.47	0.48	0.52	0.52	0.45	0.43	0.47	0.44
	0.80	0.96	0.90	0.92	0.96	0.96	0.74	0.90	0.95	0.95
	0.70	1.00	0.99	0.99	1.00	1.00	0.62	0.99	1.00	1.00
DF-GLS$^{\mu}$ AR(BIC)	1.00	*0.05*	*0.10*	*0.08*	*0.07*	*0.11*	*0.45*	*0.08*	*0.07*	*0.11*
	0.95	0.32	0.26	0.27	0.28	0.30	0.30	0.17	0.16	0.17
	0.90	0.75	0.56	0.59	0.60	0.67	0.68	0.37	0.37	0.37
	0.80	1.00	0.87	0.92	0.93	0.97	0.98	0.66	0.68	0.68
	0.70	1.00	0.96	0.98	0.99	1.00	1.00	0.79	0.80	0.76

[a]For each statistic, the first row of entries (in italics) are the empirical rejection rates under the null, that is, the empirical size of the test. The remaining entries are the size-adjusted power under the model described in the column heading. The column Asymptotic Power gives the local-to-unity approximation to the asymptotic power function for each statistic. The test statistics are discussed in the text. The entry beside the name of each statistic indicates the spectral density estimator used. SC(auto) is the Andrews automatic bandwidth estimator with the Parzen kernel. AR(4) is the AR estimator (3.14) with $p = 4$ in (3.9) [with a constant but no time trend included in the regression (3.9)]; AR(BIC) is the AR estimator with p in (3.9) selected according to the BIC rule, subject to $3 \leqslant p \leqslant 8$; and AR(LR) is the AR estimator with p selected according to sequential downward likelihood ratio tests, with 10 percent critical value at each step, with $1 \leqslant p \leqslant 8$. Based on 5000 Monte Carlo repetitions.

Table 2
Size and size-adjusted power of selected tests of the I(1) null: Monte Carlo results
(5 percent level tests, detrended case, $T = 100$, $y_t = u_t, u_t = \alpha u_{t-1} + v_t, v_t = \varepsilon_t - \theta\varepsilon_{t-1}$).[a]

Test Statistic	α	Asymptotic Power	$MA(1)$, $\theta =$					Unconditional: $MA(1)$, $\theta =$		
			−0.8	−0.5	0.0	0.5	0.8	−0.5	0.0	0.5
DF-$\hat{t}^{\tau}$ AR(4)	1.00	*0.05*	*0.03*	*0.05*	*0.05*	*0.06*	*0.37*	*0.05*	*0.05*	*0.06*
	0.95	0.09	0.07	0.07	0.07	0.08	0.09	0.07	0.08	0.08
	0.90	0.19	0.10	0.12	0.13	0.16	0.18	0.12	0.14	0.17
	0.80	0.61	0.24	0.28	0.32	0.43	0.49	0.28	0.32	0.44
	0.70	0.94	0.40	0.45	0.53	0.71	0.78	0.45	0.52	0.72
DF-$\hat{t}^{\tau}$ AR(BIC)	1.00	*0.05*	*0.10*	*0.07*	*0.05*	*0.09*	*0.58*	*0.07*	*0.05*	*0.09*
	0.95	0.09	0.09	0.08	0.08	0.09	0.08	0.08	0.08	0.09
	0.90	0.19	0.16	0.14	0.15	0.18	0.17	0.15	0.15	0.18
	0.80	0.61	0.36	0.36	0.39	0.51	0.50	0.36	0.39	0.52
	0.70	0.94	0.57	0.58	0.64	0.81	0.80	0.58	0.64	0.81
DF-$\hat{t}^{\tau}$ AR(LR)	1.00	*0.05*	*0.09*	*0.11*	*0.08*	*0.22*	*0.65*	*0.11*	*0.08*	*0.22*
	0.95	0.09	0.08	0.09	0.09	0.10	0.09	0.09	0.09	0.09
	0.90	0.19	0.14	0.16	0.17	0.22	0.17	0.16	0.18	0.20
	0.80	0.61	0.29	0.39	0.46	0.56	0.42	0.39	0.47	0.56
	0.70	0.94	0.42	0.58	0.74	0.76	0.57	0.58	0.74	0.77
DF-$\hat{\rho}^{\tau}$ AR(BIC)	1.00	*0.05*	*0.21*	*0.16*	*0.13*	*0.21*	*0.81*	*0.16*	*0.13*	*0.21*
	0.95	0.10	0.09	0.09	0.10	0.10	0.08	0.09	0.09	0.10
	0.90	0.23	0.18	0.18	0.20	0.22	0.17	0.18	0.19	0.20
	0.80	0.70	0.42	0.45	0.49	0.58	0.47	0.43	0.48	0.57
	0.70	0.97	0.62	0.67	0.74	0.87	0.73	0.66	0.73	0.86
Z_{α} SC(auto)	1.00	*0.05*	*0.00*	*0.01*	*0.05*	*0.65*	*1.00*	*0.01*	*0.05*	*0.65*
	0.95	0.10	0.09	0.09	0.11	0.10	0.09	0.09	0.10	0.09
	0.90	0.23	0.19	0.20	0.25	0.25	0.16	0.20	0.23	0.21
	0.80	0.70	0.56	0.62	0.74	0.73	0.44	0.62	0.73	0.70
	0.70	0.97	0.89	0.92	0.98	0.97	0.72	0.92	0.98	0.97
MSB AR(BIC)	1.00	*0.05*	*0.23*	*0.17*	*0.13*	*0.12*	*0.49*	*0.17*	*0.13*	*0.12*
	0.95	0.10	0.10	0.09	0.10	0.10	0.08	0.09	0.09	0.09
	0.90	0.25	0.19	0.20	0.21	0.21	0.16	0.18	0.19	0.19
	0.80	0.73	0.42	0.45	0.48	0.50	0.42	0.41	0.44	0.46
	0.70	0.97	0.60	0.65	0.69	0.74	0.69	0.61	0.64	0.69
MSB SC(auto)	1.00	*0.05*	*0.00*	*0.01*	*0.03*	*0.46*	*0.99*	*0.01*	*0.03*	*0.46*
	0.95	0.10	0.10	0.10	0.11	0.11	0.11	0.09	0.10	0.10
	0.90	0.25	0.24	0.24	0.28	0.27	0.22	0.21	0.24	0.23
	0.80	0.73	0.63	0.66	0.75	0.74	0.42	0.61	0.70	0.65
	0.70	0.97	0.89	0.91	0.97	0.94	0.42	0.86	0.93	0.89
DF-GLS$^{\tau}$ AR(BIC)	1.00	*0.05*	*0.11*	*0.08*	*0.07*	*0.11*	*0.58*	*0.08*	*0.07*	*0.11*
	0.95	0.10	0.11	0.10	0.10	0.11	0.12	0.09	0.09	0.09
	0.90	0.27	0.23	0.23	0.24	0.28	0.27	0.19	0.19	0.21
	0.80	0.81	0.53	0.57	0.61	0.72	0.70	0.46	0.49	0.54
	0.70	0.99	0.75	0.80	0.84	0.94	0.91	0.67	0.71	0.76

[a]AR(BIC) indicates that the AR spectral estimator based on (3.9) with the time trend suppressed was used. See the notes to Table 1.

and the detrended cases) was selected using the Schwartz (1977) Bayesian information criterion (BIC), with a minimum lag of 3 and a maximum of 8. For comparison purposes a sequential likelihood ratio (LR) downward-testing procedure with 10 percent critical values, as suggested by Ng and Perron (1993b), was also applied to the Dickey–Fuller t-statistic.[17]

The results for tests of asymptotic level 5 percent are summarized in Table 1 for the demeaned case and in Table 2 for the detrended case. For each statistic, the first column provides the asymptotic approximation to the size (which is always 5 percent) and to the local-to-unity power. The remaining entries for $\alpha = 1$ are the empirical size, that is, the Monte Carlo rejection rate based on asymptotic critical values. The entries for $|\alpha| < 1$ are the size-adjusted power, that is, the Monte Carlo rejection rates when the actual 5 percent critical value computed for that model with $\alpha = 1$ is used to compute the rejections. Of course, in practice the model and this correct critical value are unknown, so the size-adjusted powers do not reflect the empirical rejections based on the asymptotic critical values. However, it is the size-adjusted powers, not the empirical rejection rates, which permit examining the quality of the local-to-unity asymptotic approximations reported in the first column.

These results illustrate common features of other simulations. Test performance, both size and power, varies greatly across the statistics, the models generating the data and the methods used to estimate the long-run variance. The most commonly used test in practice is the Dickey–Fuller $\hat{\tau}$ statistic. Looking across designs, this statistic has size closer to its level than any other statistic considered here, with size in the range 5–10 percent in both the demeaned and detrended cases with $\theta \leqslant 0.5$, for both the AR(4) and BIC choices of lag length. However, as the asymptotic comparisons of the previous subsection suggest, this ability to control size in a variety of models comes at a high cost in power. For example, consider the case $\theta = -0.5$. In the demeaned case with $\alpha = 0.9$, the DF-$\hat{\tau}$ test has power of 0.22 (BIC case) while the DF–GLS test has power of 0.59. In the detrended case, as the asymptotic results suggest, the power loss from using the DF-$\hat{\tau}$ statistic is less. Again in the $\theta = -0.5$, $\alpha = 0.9$ case, the powers of the DF-$\hat{\tau}$ and the DF–GLS statistics are 0.14 and 0.23. Typically, the $\hat{\rho}$- and SB-class tests also have better size-adjusted power than the DF-$\hat{\tau}$ statistics.

Three lag length selection procedures are compared for the DF $\hat{\tau}^{\mu}$ and $\hat{\tau}^{\tau}$ statistics, and the choice has important effects on both size and power. In the $\theta = 0$ case, for example, using 4 lags results in substantial power declines against distant alternatives, relative to either data-dependent procedure. DeJong et al. (1992b) show that increasing p typically results in a modest decrease in power but a substantial reduction in size distortions. The results here favor the BIC over the LR selector;

[17] Alternative strategies, both data-based and not, were also studied, but they, typically, did not perform as well as the procedures reported here and thus are not reported here to save space. In general, among SC estimators, the Andrews (1991) procedure studied here performed substantially better (in terms of size distortions and size-adjusted power) than non-data-based procedures with $l_T = k(T/100)^{0.2}$ with $k = 4$ or 12. See Elliott (1993b).

a finding congruent with Hall's (1992b) proof that the asymptotic null distribution of the DF statistic is the same using the BIC as if the true order were known (assuming the maximum possible lag is known and fixed). However, Ng and Perron (1993b) provide evidence supporting the sequential LR procedure. In any event, currently available information suggests using one of these two lag selection procedures.

Although the size distortions are slight for the cases with positive serial correlation in v_t, the introduction of moderate negative serial correlation results in very large size distortions for several of the statistics. This is the key finding of Schwert's (1989) influential Monte Carlo study and is one of the main lessons for practitioners of this experiment. For several statistics, these size distortions are extreme. For example, for the Gaussian MA(1) process with $\theta = 0.5$, which corresponds to a first autocorrelation of v_t of -0.4, the detrended Phillips–Perron Z_α statistic has a rejection rate of 65 percent. These large size distortions are partially but not exclusively associated with the use of the SC spectral estimator. For example, the sizes of the MSB$^\mu$/AR(BIC) test and Schmidt and Phillips' (1992) version of this test implemented with the Parzen kernel, the MSB$^\mu$/SC(auto) statistic, are respectively 9 percent and 38 percent in the $\theta = 0.5$ case. Similarly, the Z_α test can be modified using an AR estimator to reduce the distortions substantially, although they remain well above the distortions of the DF-$\hat{\tau}$ or DF–GLS statistics. Ng and Perron (1993a) give theoretical reasons for the improvement of the AR over SC estimators. Part of the problem is that the SC estimators are computed using the estimated quasidifference of y_t^μ or y_t^τ, where the quasidifference is based on $\hat{\alpha}$, which in turn is badly biased in the very cases where the correction factor is most important [see the discussion following (3.4b)].[18]

Looking across the statistics, the asymptotic power rankings provide a good guide to finite-sample size-adjusted power rankings, although the finite-sample power typically falls short of the asymptotic power. As predicted by the asymptotic analysis, the differences in size-adjusted powers is dramatic. For example, in the demeaned $\theta = 0$ case with $\alpha = 0.9$, the Dickey–Fuller t-statistic (BIC case) has power of 22 percent, Z_α has power of 44 percent, and DF–GLS has power of 60 percent.

There is some tradeoff between power and size. The DF-τ statistic exhibits the smallest deviation from nominal size, but it has low power. Other tests, such as the Z_α and MSB/SC (auto) statistics, have high size-adjusted power but very large size distortions. The DF–GLS statistic appears to represent a compromise, in the sense that its power is high – based on results in Elliott et al. (1992), typically as high as the asymptotic point-optimal test P_T – but its size distortions are low,

[18] Consistent with the asymptotic theory, introducing generalized autoregressive conditional heteroskedasticity [GARCH, Bollerslev (1986)] has only a small effect on the empirical size or power of any of the statistics. Elliott (1993b) reports simulations with MA(1) GARCH(1, 1) errors and coefficients which add to 0.9. For example, for the DF–GLS statistic, demeaned case, $\theta = 0$ or -0.5, size and power ($T = 100$) differ at most by 0.03 from those in Table 1 for $\alpha = 1$ to 0.7.

although not as low as the DF-τ statistic. In the demeaned results, DF–GLS has sizes of 0.07–0.11, compared to the DF-τ (BIC) which has sizes 0.06–0.08 (except in the extreme, $\theta = 0.8$, case). In the detrended case, the DF–GLS has sizes of 0.07–0.11, while DF-τ has sizes in the range 0.05–0.10.

Drawing the initial value from its unconditional distribution changes the rankings of size-adjusted power; in particular the size-adjusted power of DF–GLS drops, particularly for distant alternatives. However, the DF–GLS power remains above the DF-$\hat{t}$ (BIC) power in both demeaned and detrended cases, and, of course, the large size distortions of the other tests are not mitigated in this DGP. Recent Monte Carlo evidence by Pantula et al. (1992) suggests that, in the correctly specified demeaned AR(1) model, better power can be achieved against the unconditional alternative by a test based on a weighted symmetric least squares estimator. However, the unconditional case has been less completely studied than the conditional case and it seems premature to draw conclusions about which tests perform best in this setting.

3.2.5. *Effects of misspecifying the deterministic trend*

The discussion so far has assumed that the order of the trend has been correctly specified. If the trend is misspecified, however, then the estimators of α and the tests of $\alpha = 1$ can be inconsistent [Perron and Phillips (1987), West (1987, 1988a)].

This argument can be made simply in the case where the true trend is $d_t = \beta_0 + \beta_1 t$ with β_1 a nonzero constant, but the econometrician incorrectly uses the constant-only model. Because y_t contains a linear time trend, asymptotically the OLS objective function will be minimized by first-differencing y_t, whether or not v_t is I(1), and a straightforward calculation shows that $\hat{\alpha} \xrightarrow{p} 1$.[19] It follows that $\hat{\rho}$ and $\hat{t}$ tests will not be consistent. This inconsistency is transparent if one works with the functional representation of the tests: $T^{-1} Y_T^{\mu} \Rightarrow h^{\mu}$, where $h^{\mu}(\lambda) = \beta_1(\lambda - \frac{1}{2})$. In finite samples, the importance of this omitted variable effect depends on the magnitude of the incorrectly omitted time-trend slope relative to ω [West (1987) provides Monte Carlo evidence on this effect]. This problem extends to other types of trends as well, and in particular to misspecification of a piecewise-linear trend (a "broken" trend) as a single linear trend [see Perron (1989a, 1990b), Rappoport and Reichlin (1989) and the discussion in Section 5.2 of this chapter].

The analogy to the usual regression problem of omitted variable bias is useful here: if the trend is underspecified, unit root tests (and root estimators) are inconsistent, while if the trend is overspecified, power is reduced, even asymptotically. This contrasts with the case of mean-zero I(0) regressors, in which the reduction in power, resulting from unnecessarily including polynomials in time, vanishes asymptotically. The difference is that while I(0) regressors are asymptotically

[19] See West (1988a), Park and Phillips (1988) and Sims et al. (1990) for extensions of this result to multiple time series models.

uncorrelated with the included time polynomials, I(1) regressors are asymptotically correlated (with a random correlation coefficient). This asymptotic collinearity reduces the power of the unit root tests when a time trend is included. A procedure of sequential testing of the order of the trend specification prior to inference on α will result in pretest bias arising from the possibility of making a type I error in the tests for the trend order. This problem is further complicated by the dependence of the distributions of the trend coefficients and test statistics on the order of integration of the stochastic component.[20]

3.2.6. *Summary and implications for empirical practice*

If one is interested in testing the null hypothesis that a series is I(1) – as opposed to testing the null that the series is I(0) or to using a consistent decision- or information-theoretic procedure to select between the I(0) and I(1) hypotheses – then the presumption must be that there is a reason that the researcher wishes to control type I error with respect to the I(1) null. If so, then a key criterion in the selection of a unit root test for practical purposes is that the finite-sample size be approximately the level of the test.

Taking this criterion as primary, we can see from Tables 1 and 2 that only a few of the proposed tests effectively control size for a range of nuisance parameters. In the demeaned case, only the Dickey–Fuller $\hat{\tau}^{\mu}$ and DF–GLS$^{\mu}$ tests have sizes of 12 percent or under (excluding the extreme $\theta = 0.8$ case). However, the $\hat{\tau}^{\mu}$ statistic has much lower size-adjusted power than the DF–GLS$^{\mu}$ statistic. Moreover, asymptotically, the DF–GLS$^{\mu}$ statistic can be thought of as approximately UMP since its power function nearly lies on the Neyman–Pearson power envelope in Figure 2, even though, strictly, no UMP test exists. When u_0 is drawn from its unconditional distribution, the power of the DF–GLS$^{\mu}$ statistic exceeds that of $\hat{\tau}^{\mu}$ except against distant alternatives. These results suggest that, of the tests studied here, the DF–GLS$^{\mu}$ statistic is to be preferred in the $d_t = \beta_0$ case.

In the detrended case, only $\hat{\tau}^{\tau}$ and DF–GLS$^{\tau}$ have sizes less than 12 percent (excepting $\theta = 0.8$). The size-adjusted power of the DF–GLS$^{\tau}$ (BIC) test exceeds that of the $\hat{\tau}^{\tau}$ (BIC) test in all cases except u_0 unconditional, $\theta = 0.5$ and $\alpha = 0.7$. Because the differences in size distortions between the $\hat{\tau}^{\tau}$ and DF–GLS$^{\tau}$ tests is minimal, this suggests that again the DF–GLS$^{\tau}$ test is preferred in the detrended case.

In both the demeaned and the detrended cases, an important implication of the Monte Carlo results here and in the literature is that the choice of lag length or truncation parameter can strongly influence test performance. The LR and BIC

[20] In theory, this can be addressed by casting the trend order/integration order decision as a model selection problem and using Bayesian model selection techniques, an approach investigated by Phillips and Ploberger (1992). See the discussion in Section 6 of this chapter.

rules have the twin advantages of relieving the researcher from making an arbitrary decision about lag length and of providing reasonable tradeoffs between controlling size with longer lags and gaining size-adjusted power with shorter lags.

One could reasonably object to the emphasis on controlling size in drawing these conclusions. In many applications, particularly when the unit root test is used as a pretest, it is not clear that controlling type I error is as important as achieving desirable statistical properties in the subsequent analysis. This suggests adopting alternative strategies: perhaps testing the I(0) null or implementing a consistent classification scheme. These strategies are respectively taken up in Sections 4 and 6.

3.3. *Interval estimation*

Confidence intervals are a mainstay of empirical econometrics and provide more information than point estimates or hypothesis tests alone. For example, it is more informative to estimate a range of persistence measures for a given series than simply to report whether or not the persistence is consistent with there being a unit root in its autoregressive representation [see, for example, Cochrane (1988) and Durlauf (1989)]. This suggests constructing classical confidence intervals for the largest autoregressive root α, for the sum of the coefficients in the autoregressive approximation to u_t, or for the cumulative impulse response function. Alternatively, if one is interested in forecasting, then it might be desirable to use a median-unbiased estimator of α, so that forecasts (in the first-order model) would be median-unbiased. Because a median-unbiased estimator of α corresponds to a 0 percent equal-tailed confidence interval [e.g. Lehmann (1959, p. 174)], this again suggests considering the construction of classical confidence intervals for α. Moreover, a confidence interval for α would facilitate computing forecast prediction intervals which take into account the sampling uncertainty inherent in estimates of α.

The construction of classical confidence intervals for α, however, involves technical and computational complications. Only recently has this been the subject of active research. Because of the nonstandard limiting distribution at $\alpha = 1$, it is evident that the usual approach of constructing confidence intervals, as, say, ± 1.96 times the standard error of $\hat{\alpha}$, has neither a finite-sample nor an asymptotic justification. This approach does not produce confidence intervals with the correct coverage probabilities, even asymptotically, when α is large. To see this, suppose that α is estimated in the regression (3.9) and that the true value of α is one. Then the usual "asymptotic 95 percent" confidence interval will contain the true value of α when the absolute value of the t-ratio testing $\alpha = 1$, constructed using α^{τ}, is less than 1.96. When $\alpha = 1$, however, this t-ratio has the Dickey–Fuller $\hat{t}^{\tau}$ distribution, for which $\Pr[|\hat{t}^{\tau}| > 1.96] \to 0.61$. That is, the purported 95 percent confidence interval actually has an asymptotic coverage rate of only 39 percent!

It is, therefore, useful to return to first principles to develop a theory of classical

interval estimation for α. A 95 percent confidence set for α, $S(y_1,\ldots,y_T)$, is a set-valued function of the data with the property that $\Pr[\alpha \in S] = 0.95$ for all values of α and for all values of the nuisance parameters. In general, a confidence set can be constructed by "inverting" the acceptance region of a test statistic that has a distribution which depends on α but not on any nuisance parameters. Were there a UMP test of $\alpha = \alpha_0$ available for all null values α_0, then this test could be inverted to obtain a uniformly most accurate confidence set for α. However, as was shown in Section 3.2.1, no such UMP test exists, even in the special case of no nuisance parameters, so uniformly most accurate (or uniformly most accurate invariant or invariant-unbiased) confidence sets cannot be constructed by inverting such tests. Thus, as in the testing problem, even in the finite-sample Gaussian AR(1) model the choice of which test to invert is somewhat arbitrary.

As the discussion of Section 3 revealed, a variety of statistics for testing $\alpha = \alpha_0$ are available for the construction of confidence intervals. Dufour (1990) and Kiviet and Phillips (1992) proposed techniques for constructing exact confidence regions in Gaussian AR(1) regression with exogenous regressors and Andrews (1993a) develops confidence sets for the Gaussian AR(1) model in the no-deterministic, demeaned and detrended cases with no additional regressors. Dufour's (1990) confidence interval is based on inverting the Durbin–Watson statistic, Kiviet and Phillips (1992) inverted the t-statistic from an augmented OLS regression, and Andrews (1993a) inverted $\hat{\alpha}^{\mu} - \alpha$ (in the detrended case $\hat{\alpha}^{\tau} - \alpha$). In practice, the inversion of these test statistics is readily performed using a graph of the confidence belt for the respective statistics, which plots the critical values of the test statistic as a function of the true parameter. Inverting this graph yields those parameters which cannot be rejected for a given realization of the test statistic, providing the desired confidence interval [see Kendall and Stuart (1967, Chapter 20)].[21]

In practice one rarely, if ever, knows a priori that the true autoregressive order is one and that the errors are Gaussian, so a natural question is how to construct confidence intervals for α in the more general model (3.1). If treated in finite samples, even if one maintains the Gaussianity assumption this problem is quite difficult because of the additional nuisance parameters describing the short-run dependence. However, as first pointed out by Cavanagh (1985), the local-to-unity asymptotics of Section 3.2.3 can be used to construct asymptotically valid confidence intervals for α when α is close to one.

[21] Dufour studied linear regression with Gaussian AR(1) disturbances, of which the constant and constant/time-trend regression problems considered here are special cases. Both Dufour (1990) and Andrews (1993a) computed the exact distributions of these statistics using the Imhof method. In earlier work, Ahtola and Tiao (1984) proposed a method for constructing confidence intervals in the Gaussian AR(1) model with no intercept. Ahtola and Tiao's approach can be interpreted as inverting the score test of the null $\rho = \rho_0$, with two important approximations. First, they use a normal-F approximation to the distribution of the score test, which seems to work well over their tabulated range of α. Second (and more importantly), their proposed procedure for inverting the confidence belt requires the belt to be linear and parallel, which is not the case over a suitably large range of α, even at the scale of the local-to-unity model $\alpha = 1 + c/T$.

If the local-to-unity asymptotic representation of the statistic in the general I(1) case has a distribution which depends only on c [a condition satisfied by any statistic with the limiting representations in (3.23)] then this test can be inverted to construct confidence intervals for c and, thus, for α. In the finite-sample case, α cannot be determined from the data with certainty, and similarly in the asymptotic case c cannot be known with certainty even though α is consistently estimated. However, the nesting $\alpha = 1 + c/T$ provides confidence intervals that shorten at the rate T^{-1} rather than the usual $T^{-1/2}$.

To be concrete, the Dickey–Fuller t-statistic from the pth order autoregression (3.9) (interpreted in the Said–Dickey sense of p increasing with the sample size) has the local-to-unity distribution (3.23b) which depends only on c and, so, can be used to test the hypothesis $c = c_0$ against the two-sided alternative for any finite value of c_0. The critical values for this test depend on c_0. The plot of these values constitutes an asymptotic confidence belt for the local-to-unity parameter c, based on the Dickey–Fuller t-statistic. Inverting the test based on this belt provides an asymptotic local-to-unity confidence interval for c. Asymptotic confidence belts based on the Dickey–Fuller t-statistic in (3.9) and, alternatively, the modified Sargan–Bhargava (MSB) statistic are provided by Stock (1991) in both the demeaned and the detrended cases. Stock's (1991) Monte Carlo evidence suggests that the finite-sample coverage rates of the interval based on the Dickey–Fuller t-statistic are close to their asymptotic confidence levels in the presence of MA(1) disturbances, but the finite-sample coverage rates of the MSB-based statistics exhibit substantial distortions relative to their asymptotic confidence levels.

Both the finite-sample AR(1) and asymptotic AR(p) confidence intervals yield, as special cases, median-unbiased estimators of α. The OLS estimates of α are biased downwards, and both the finite-sample and asymptotic approaches typically produce median-unbiased estimates of α larger than the OLS point estimates.

While this approach produces confidence intervals and median-unbiased estimators of α, the researcher might not be interested in the largest root per se but rather in some function of this root, such as the sum of the AR coefficients in the autoregressive representation. To this end, Rudebusch (1992) proposed a numerical technique based on simulation to construct median-unbiased estimators of each of the $p + 1$ autoregressive parameters; his algorithm searches for those autoregressive parameters for which the median of each of the AR parameters equals the observed OLS estimate for that parameter. Andrews and Chen (1992) propose a similar algorithm, except that their emphasis is the sum of the autoregressive parameters rather than the individual autoregressive parameters themselves and their calculations are done using the asymptotic local-to-unity approximations.

A completely different approach to interval estimation which has been the subject of considerable recent research and controversy is the construction of Bayesian regions of highest posterior probability and an associated set of Bayesian tests of the unit AR root hypothesis. (References are given in Section 6.2.) Although these procedures examine the same substantive issue, they are not competitors of

the classical methods in the sense that, when interpreted from a frequentist perspective, many of the proposed Bayesian intervals have coverage rates that differ from the stated confidence levels, even in large samples. A simple example of this occurs in the Gaussian AR(1) model with a constant and a time trend when there are flat priors on the coefficients. Then, in large samples, the Bayesian 95 percent coverage region is constructed as those values of α which are within 1.96 standard errors (conventional OLS formula) of the point estimate [Zellner (1987)]. As pointed out earlier in this subsection, if $\alpha = 1$ this interval will contain the true value of α only 39 percent of the time in the detrended case. Of course these Bayesian regions have well-defined interpretations in terms of thought-experiments in which α is itself random and have optimality properties, given the priors. However, given the lack of congruence between the classical and the Bayesian intervals in this problem, and the sensitivity of the results to the choice of priors [see Phillips (1991a) and his discussants], applied researchers should be careful in interpreting these results.

4. Unit moving average roots

This section examines inference in two related models, the moving average model and the unobserved components model. The moving average model is

$$y_t = d_t + u_t, \quad \Delta u_t = (1 - \theta L)v_t \tag{4.1}$$

where v_t is, in general, an I(0) process satisfying (2.1)–(2.3). If $\theta = 1, u_t = v_t + (u_0 - v_0)$, so that with the initial condition that $u_0 = v_0, u_t = v_t$ is a purely stochastic I(0) process. If $|\theta| < 1$, then $(1 - \theta L)^{-1}$ yields a convergent series and $(1 - \theta L)$ is invertible, so u_t is I(1). The convention in the literature is to refer to $|\theta| = 1$ as the noninvertible case.

The unobserved components (UC) model considered here can be written

$$y_t = d_t + u_t, \quad u_t = \mu_t + \zeta_t, \quad \mu_t = \mu_{t-1} + v_t, \quad t = 1, 2, \ldots, T, \tag{4.2}$$

where ζ_t and v_t are I(0) with variances σ_ζ^2 and σ_v^2 and where d_t is a trend term as in (1.1). If ζ_t and v_t have a nondegenerate joint distribution, then, in general, the I(1) component μ_t and the I(0) component ζ_t cannot be extracted from the observed series without error, even with known additional parametric structure; hence the "unobserved components" terminology.

It should be observed at the outset that the unit MA root/UC models are a mirror image of the unit AR root model, in the sense that the unit AR root model parameterizes the I(1) model as a point ($\alpha = 1$) and the I(0) model as a continuum ($|\alpha| < 1$), while in the unit MA root/UC models the reverse is true. In the latter two models, the I(0) case is parameterized as a point ($\theta = 1$ in the unit MA root model, $\sigma_v^2 = 0$ in the UC model), while the I(1) case is parameterized as a

continuum ($|\theta| < 1$ in the MA model, $\sigma_v^2 > 0$ in the UC model). This suggests that, at least qualitatively, some of the general lessons from the AR problem will carry over to the MA/UC problems. In particular, because the points $\theta = 1$ and $\sigma_v^2 = 0$ represent discontinuities in the long-run behavior of the process, it is perhaps not surprising that, as in the special case of a unit AR root, the first-order asymptotic distributions of estimators of θ and σ_v^2 exhibit discontinuities at these points. In addition, just as the unit AR root model lends itself to constructing tests of the general I(1) null, the unit MA root model lends itself to constructing tests of the general I(0) null.

Although the MA model (4.1) and UC model (4.2) appear rather different, they are closely related. To see this, consider only the stochastic components of the models. In general, for suitable choices of initial conditions, all MA models (4.1) have UC representations (4.2) and all UC models (4.2) have MA representations of the form (4.1). To show the first of these statements, we need only write $\Delta u_t = (1 - \theta L)v_t = (1 - \theta)v_t + \theta \Delta v_t$; then, cumulating Δu_t with the initial condition $u_0 = v_0$ yields

$$u_t = (1 - \theta) \sum_{s=0}^{t} v_s + \theta v_t. \tag{4.3}$$

By construction $(1 - \theta)\sum_{s=0}^{t} v_s$ is I(1) and θv_t is I(0). Therefore, the MA model (4.1) has the UC representation (4.3) with $v_t = (1 - \theta)v_t$ and $\zeta_t = \theta v_t$. If $\theta = 1$, then the I(1) term in (4.3) vanishes and u_t is I(0). To argue that all UC models have MA representations of the form (4.1), it is enough to consider the two cases, $\sigma_v^2 = 0$ and $\sigma_v^2 > 0$. If $\sigma_v^2 = 0$ and $\mu_0 = 0$, then $u_t = \zeta_t$, which is I(0), so (4.1) obtains with $u_0 = v_0$ and $\theta = 1$. If $\sigma_v^2 > 0$, then u_t is I(1), so $\Delta u_t = v_t + \Delta \zeta_t$ is I(0), and it follows that Δu_t has a Wold decomposition and hence an MA representation of the form (4.1) where $|\theta| < 1$.

A leading special case of the UC model which is helpful in developing intuition and which will be studied below is when $(\Delta \zeta_t, v_t)$ are serially uncorrelated and are mutually uncorrelated. Then, Δu_t in the UC model has MA(1) autocovariances: $\Delta u_t = \Delta \zeta_t + v_t$, so that $\gamma_{\Delta u}(0) = 2\sigma_\zeta^2 + \sigma_v^2, \gamma_{\Delta u}(1) = -\sigma_\zeta^2$ and $\gamma_{\Delta u}(j) = 0, |j| > 1$. Thus Δu_t has the MA(1) representation (4.1), $\Delta u_t = (1 - \theta L)e_t$, where e_t is serially uncorrelated, θ solves $\theta + \theta^{-1} - 2 = \sigma_v^2/\sigma_\zeta^2$ and $\sigma_e^2 = \sigma_\zeta^2/\theta$. Because $E\Delta \zeta_t v_t = 0$ by assumption, this UC model is incapable of producing positive autocorrelations of Δu_t, so, while all uncorrelated UC models have an MA(1) representation, the converse in not true. [The MA(1) model will, however, have a correlated UC representation and a UC representation with independent permanent and transitory components which themselves have complicated short-run dynamics; see Quah (1992).]

The UC model can equivalently be thought of as having I(1) and I(0) components, or as being a regression equation with deterministic regressor(s) d_t (which in general has unknown parameter(s) β), an I(0) error and a constant which is time-varying

and follows an I(1) process. Thus the problem of testing for a unit moving average root and testing for time variation in the intercept in an otherwise standard time series regression model are closely related.

4.1. Point estimation

When the MA process is noninvertible, or nearly noninvertible, estimators of θ fail to have the standard Gaussian limiting distributions. The task of characterizing the limiting properties of estimators of θ when θ is one, or nearly one, is difficult, and the theory is less complete than in the case of nearly-unit autoregressive roots. Most of the literature has focused on the Gaussian MA(1) model with $d_t = 0$ and $v_t = \varepsilon_t$, and this model is adopted in this subsection, except as explicitly noted. One complication is that the limiting distribution depends on the specific maximand and the treatment of initial conditions. Because the objective here is pedagogical rather than to present a comprehensive review, our discussion of point estimation focuses on two specific estimators of θ, the unconditional and conditional (on $\varepsilon_1 = 0$) MLE.

Suppose that the data have been transformed so that $x_t = \Delta y_t, t = 2, \ldots, T$. Then θ can be estimated by maximizing the Gaussian likelihood for $(x_2, \ldots, x_T)$. The exact form of the likelihood depends on the treatment of the initial condition. If x_2 is treated as being drawn from its stationary distribution, so that $x_2 = \varepsilon_2 - \theta\varepsilon_1$, then $X \equiv (x_2, \ldots, x_T)'$ has covariance matrix $\sigma_\varepsilon^2 \Omega_u(\theta)$, where $\Omega_{u,ii} = 1 + \theta^2$ and $\Omega_{u,ii\pm 1} = -\theta$. This is the "unconditional" case, and the Gaussian likelihood is

$$\Lambda(\theta, \sigma_\varepsilon^2) = -\tfrac{1}{2} T \ln 2\pi\sigma_\varepsilon^2 - \tfrac{1}{2} \ln \det(\Omega_u) - \frac{\tfrac{1}{2} X' \Omega_u^{-1} X}{\sigma_\varepsilon^2}, \tag{4.4}$$

where $\det(\Omega_u)$ denotes the determinant of Ω_u. Estimation proceeds by maximization of Λ in (4.4). Numerical issues associated with this maximization are discussed at the end of the subsection.

The "conditional" case sets $\varepsilon_1 = 0$, so that $x_2 = \varepsilon_2$. The conditional likelihood is given by (4.4) with Ω_u replaced by Ω_c, where $\Omega_c = \Omega_u$ except that $\Omega_{c,11} = 1$. A principal advantage of maximizing the conditional Gaussian likelihood is that the determinant of the covariance matrix does not depend on θ, so maximization can proceed by minimizing the quadratic form $X' \Omega_c^{-1} X$. Because $\varepsilon_t = x_t + \theta\varepsilon_{t-1}$, with the additional assumption that $\varepsilon_1 = 0$, the residuals $e_t(\theta)$ can be constructed recursively as $(1 - \theta L) e_t(\theta) = x_t$ and estimation reduces to the nonlinear least squares problem of minimizing $\sum_{t=2}^{T} e_t(\theta)^2$.

If $|\theta| < 1$, so that the process is invertible, then standard $\sqrt{T}$ asymptotic theory applies. More generally, if an ARMA(p, q) process is stationary and invertible and has no common roots, then the Gaussian maximum likelihood estimator of the ARMA parameters is $\sqrt{T}$-consistent and has a normal asymptotic distribution; see Brockwell and Davis (1987, Chapter 10.8). In the MA(1) model, $\sqrt{T}(\hat{\theta}_{\text{MLE}} - \theta)$

is asymptotically distributed $N(0, 1-\theta^2)$. This provides a simple way to construct tests of whether θ equals some particular value. Alternatively, confidence intervals for θ can be constructed as $\hat{\theta} \pm 1.96(1-\hat{\theta}^2)^{1/2}$ (for a 95 percent two-sided confidence interval).

These simple results fail to hold in the noninvertible case. This is readily seen by noting that the asymptotic normal approximation to the distribution of the MLE is degenerate when $\theta = 1$. The most dramatic and initially surprising feature of this failure is the "pileup" phenomenon. In a series of Monte Carlo experiments, investigators found that the unconditional MLE took on the value of exactly one with positive probability when the true value of θ was near one, a surprising finding at the time since θ can take on any value in a continuum. Shephard (1992) and Davis and Dunsmuir (1992) attribute the initial discovery of the pileup effect to unpublished work by Kang (1975); early published simulation studies documenting this phenomenon include Ansley and Newbold (1980), Cooper and Thompson (1977), Davidson (1979, 1981), Dent and Min (1978) and Harvey (1981, pp. 136–9); also see Plosser and Schwert (1977), Dunsmuir (1981) and Cryer and Ledolter (1981).

The intuition concerning the source of the pileup effect is straightforward, and concerns the lack of identification of (θ, σ^2) in the unconditional model. Note that $\Omega_u(\theta) = \theta^2 \Omega_u(\theta^{-1})$; upon substituting this expression into the unconditional likelihood (4.4), one obtains $\Lambda(\theta, \sigma^2) = \Lambda(\theta^{-1}, \theta^2\sigma^2)$ and $\tilde{\Lambda}(\theta) = \tilde{\Lambda}(\theta^{-1})$, where $\tilde{\Lambda}$ denotes the likelihood concentrated to be an argument only of θ. Because $\tilde{\Lambda}$ is symmetric in $\ln\theta$ for $\theta > 0$, it follows immediately that $\tilde{\Lambda}$ will have a local maximum at $\theta = 1$ if $\partial^2\tilde{\Lambda}/\partial\theta^2|_{\theta=1} < 0$, so the probability of a local maximum at $\theta = 1$ is $\Pr[\partial^2\tilde{\Lambda}/\partial\theta^2|_{\theta=1} < 0]$. Sargan and Bhargava (1983b, Corollary 1) [also see Pesaran (1983) and Anderson and Takemura (1986, Theorem 4.1)] provide expressions for this limiting probability in the noninvertible case, which can be calculated by interpolation of Table I in Anderson and Darling (1952) and is 0.657.

These results were extended to the case of higher-order MA and ARMA models by Anderson and Takemura (1986), where the estimation is by Gaussian MLE when the order of the ARMA process is correctly specified. Tanaka (1990b) considered a different problem, in which v_t is a linear process which is I(0) but otherwise is only weakly restricted, but θ is estimated by using the misspecified Gaussian MA(1) likelihood. Tanaka (1990b) found that, despite the misspecification of the model order, the unconditional MLE continues to exhibit the pileup effect, in the sense that the probability of a local (but not necessarily global) maximum at $\hat{\theta} = 1$ is nonzero if the true value of θ is one. Also see Tanaka and Satchell (1989) and Pötscher (1991).

Because of the close link between the UC and MA models, not surprisingly the pileup phenomenon occurs in those models as well. In this model, if the "signal-to-noise ratio" $\sigma_v^2/\sigma_\zeta^2$ is zero or is in a T^{-2} neighborhood of zero, then σ_v^2 is estimated to be precisely zero with finite probability. However, the value of this point probability depends on the precise choice of maximand (e.g. maximum marginal likelihood or maximum profile likelihood) and the treatment of the initial condition

μ_0 (as fixed or alternatively as random with a variance which tends to infinity, or equivalently as being drawn from a diffuse prior). Various versions of this problem have been studied by Nabeya and Tanaka (1988), Shephard and Harvey (1990) and Shephard (1992, 1993).

Research on the limiting distribution of estimators of θ when θ is close to one is incomplete. Davis and Dunsmuir (1992) derive asymptotic distributions of the local maximizer closest to one of the unconditional likelihood, when the true value is in a $1/T$ neighborhood of $\theta = 1$. Their numerical results indicate that their distributions provide good approximations, even for θ as small as 0.6 with $T = 50$. Their approach is to obtain representations of the first and second derivatives of the likelihoods as stochastic processes in $T(1-\theta)$. They do not (explicitly) use the FCLT, and working through the details here would go beyond the scope of this chapter.

A remark on computation. The main technical complication that arises in the estimation of stationary and invertible ARMA(p, q) models is the numerical evaluation of the likelihood when $q \geqslant 1$. If the sample size is small, then $\boldsymbol{\Omega}_u^{-1}$ can be computed and inverted directly. In sample sizes typically encountered in econometric applications, however, the direct computation of $\boldsymbol{\Omega}_u^{-1}$ is time-consuming and can introduce numerical errors. One elegant and general solution is to use the Kalman filter, which is a general device for computing the Gaussian log likelihood, $\mathscr{L}(y_1, \ldots, y_T)$ with the factorization $\mathscr{L}(y_1, \ldots, y_T) = \mathscr{L}(y_1) + \sum_{t=2}^{T} \mathscr{L}(y_t | y_{t-1}, \ldots, y_1)$, when the model can be represented in state space form (as can general ARMA models). The Kalman filter operates by recursively computing the conditional mean and variance of y_t, which, in turn, specifies the conditional likelihood $\mathscr{L}(y_t | y_{t-1}, \ldots, y_1)$. The Gaussian MLE is then computed by finding the parameter vector that maximizes the likelihood. The chapter by Hamilton in this Handbook describes the particulars of the standard Kalman filter and provides a state space representation for ARMA models which can be used to compute their Gaussian MLE. The model (4.2) is a special case of unobserved components time series models, which in general can be written in state space form so that they, too, can be estimated using the Kalman filter; see Harvey (1989) and Harvey and Shephard (1992) for discussions and related examples.

The literature on the estimation of stationary and invertible ARMA models is vast and it will not be covered further in this chapter. See Brockwell and Davis (1987, Chapter 8) for a discussion and references. For additional discussion of the Kalman filter with applications and a bibliography, see the chapter by Hamilton in this Handbook.

4.2. *Hypothesis tests*

4.2.1. *Tests of $\theta = 1$ in the conditional Gaussian MA(1) model*

As Sargan and Bhargava (1983b) pointed out, the pileup phenomenon means that likelihood ratio tests cannot be used for hypothesis testing at conventional

significance levels, at least using the unconditional Gaussian MLE. Given this difficulty, it is perhaps not surprising that research into testing the null of a unit MA root has been limited and has largely focused on the MA(1) conditional Gaussian model. This model, therefore, provides a natural starting point for our discussion of tests of the general I(0) null.

The conditional Gaussian MA(1) model with a general pth order polynomial time trend is

$$y_t = z_t'\beta + u_t, \quad u_1 = \varepsilon_1, \quad \Delta u_t = \varepsilon_t - \theta\varepsilon_{t-1}, \quad t > 1 \tag{4.5}$$

where ε_t is i.i.d. $N(0, \sigma_\varepsilon^2)$ and $z_t = (1, t, \ldots, t^p)'$. Let $z = (z_1, \ldots, z_T)'$, and similarly define the $T \times 1$ vectors y and u. Then y is distributed $N(z\beta, \Sigma_c)$, where $\Sigma_{c,11} = \sigma_\varepsilon^2$ and the remaining elements can be calculated directly from the moving average representation $\Delta u_t = (1 - \theta L)\varepsilon_t$.

The problem of testing values of θ is invariant to transformations of the form $y_t \to ay_t + z_t'b$, $\beta \to a\beta + b$ and $\sigma_\varepsilon^2 \to a^2\sigma_\varepsilon^2$. It is therefore reasonable to restrict attention to the family of tests which are invariant to this transformation, and among that family to find the most powerful tests of $\theta = 1$ against the fixed alternative $\theta = \bar{\theta}$. An implication of the general results of King (1980, 1988) is that the MPI test of $\theta = 1$ vs. $\theta = \bar{\theta}$ rejects for small values of the statistic

$$\tilde{U}'\Sigma_c(\bar{\theta})^{-1}\tilde{U}/(\hat{U}'\hat{U}), \tag{4.6}$$

where $\hat{U} = (\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_T)'$, where $\{\hat{u}_t\}$ are the residuals from the OLS regression of y_t onto z_t, $\tilde{U}$ are the GLS residuals from the estimation of (4.5) under the alternative, and $EUU' = \Sigma_c(\bar{\theta})$ is the covariance matrix of $U = (u_1, \ldots, u_T)$ under the alternative $\bar{\theta}$. In the MA(1) model, the GLS transformation can be written explicitly, and GLS simplifies to the OLS regression of $Y_t(\bar{\theta})$ onto $Z_t(\bar{\theta})$, where $Y_1(\bar{\theta}) = y_1$ and $(1 - \bar{\theta}L)Y_t(\bar{\theta}) = \Delta y_t, t > 1$, and similarly for $Z_t(\bar{\theta})$.

As in the case of MP tests for an autoregressive unit root discussed in Section 3.2.1, the dependence of the MPI test statistic (4.6) on the alternative $\bar{\theta}$ cannot be eliminated, so there does not exist a UMPI test of $\theta = 1$ vs. $|\theta| < 1$. This has led researchers to propose alternative tests. A natural approach is to consider tests which have maximal power for local alternatives, that is, to consider the locally most powerful invariant test. In the special case that d_t is zero, Saikkonen and Luukkonen (1993a) show that the LMPI test has the form, $T\bar{y}^2/\hat{\sigma}_\varepsilon^2$, where $\bar{y} = T^{-1}\sum_{t=1}^T y_t$ and $\hat{\sigma}_\varepsilon^2 = T^{-1}\sum_{t=1}^T y_t^2$ (which is the MLE of σ_ε^2 under the null). In the case $d_t = \beta_0$, Saikkonen and Luukkonen (1993a) use results in King and Hillier (1985) to derive the locally most powerful invariant unbiased test, which is based on the statistic

$$L^\mu = \frac{T^{-2}\sum_{t=1}^T \left(\sum_{s=1}^t y_s^\mu\right)^2}{(\hat{\sigma}_\varepsilon^\mu)^2} = \int_0^1 \left\{\frac{Y_{0T}^\mu(\lambda)}{\hat{\sigma}_\varepsilon^\mu}\right\}^2 \mathrm{d}\lambda + \mathrm{o}_p(1) \tag{4.7}$$

where $Y^{\mu}_{0T}(\lambda) = T^{-1/2}\sum_{s=1}^{[T\lambda]} y^{\mu}_s$ and $(\hat{\sigma}^{\mu}_{\varepsilon})^2 = T^{-1}\sum_{s=1}^{T}(y^{\mu}_s)^2$, where $y^{\mu}_s = y_t - \bar{y}$ [also see Tanaka (1990b)]. Note that $(\hat{\sigma}^{\mu}_{\varepsilon})^2$ is the (conditional) MLE of σ^2_{ε} under the null hypothesis. Because the statistic was derived for arbitrarily small deviations from the null, the parameter θ does not need to be estimated to construct L.[22]

A natural generalization of (4.7) to linear time trends is to replace the demeaned process y^{μ}_t by the detrended process y^{τ}_t:

$$L^{\tau} = \frac{T^{-2}\sum_{t=1}^{T}\left(\sum_{s=1}^{t} y^{\tau}_s\right)^2}{(\hat{\sigma}^{\tau}_{\varepsilon})^2} = \int_0^1 \left\{\frac{Y^{\tau}_{0T}(\lambda)}{\hat{\sigma}^{\tau}_{\varepsilon}}\right\}^2 d\lambda + o_p, \tag{4.8}$$

where $Y^{\tau}_{0T}(\lambda) = T^{-1/2}\sum_{s=1}^{[T\lambda]} y^{\tau}_s$ and $(\hat{\sigma}^{\tau}_{\varepsilon})^2 = T^{-1}\sum_{s=1}^{T}(y^{\tau}_s)^2$, where $y^{\tau}_t = y_t - z_t'\hat{\beta}$ and $\hat{\beta}$ is the OLS estimator from the regression of y_t onto $(1, t)$.

The asymptotic null distributions of L^{μ} and L^{τ} are readily obtained using the FCLT and CMT, under the maintained assumption that the order of the estimated deterministic trend is at least as great as the order of the true trend. First consider L^{μ}. As the second expression in (4.7) reveals, L^{μ} can be written as a continuous functional of $Y^{\mu}_{0T}/\hat{\sigma}^{\mu}_{\varepsilon}$. To obtain limiting representations for L^{μ} it therefore suffices to obtain limiting results for the stochastic process $Y^{\mu}_{0T}/\hat{\sigma}^{\mu}_{\varepsilon}$. The limit of the numerator of this process was derived in Section 2.3 (Example 3) and is given in (2.14a) for u_t being a general I(0) process; because it is assumed in this subsection that $u_t = \varepsilon_t$ under the null, this result applies here with $\omega = \sigma_{\varepsilon}$. In addition, the maintained assumption that the trend is correctly specified ensures that $\hat{\sigma}^{\mu}_{\varepsilon} \xrightarrow{p} \sigma_{\varepsilon}$. It follows that $Y^{\mu}_{0T}/\hat{\sigma}^{\mu}_{\varepsilon} \Rightarrow B^{\mu}$ and that $L^{\mu} \Rightarrow \int (B^{\mu})^2$, where $B^{\mu}(\lambda) = W(\lambda) - \lambda W(1)$ is a standard Brownian bridge. An identical argument applies to the linearly detrended case and yields the limit $L^{\tau} \Rightarrow \int (B^{\tau})^2$, where B^{τ} is the second-level Brownian bridge in (2.14b). In the leading case that d_t is a constant, L^{μ} has the asymptotic distribution of the Cramer–von Mises statistic derived by Anderson and Darling (1952). Nyblom and Mäkeläinen (1983, Table 1) provide critical values of the finite-sample distribution of L^{μ}, computed using the Imhof method for Gaussian errors. Kwiatkowski et al. (1992, Table 1) provide a table of critical values of $\int (B^{\mu})^2$ and $\int (B^{\tau})^2$ which agrees closely with earlier computations, e.g. MacNeill (1978, Table 2). Although the motivation for the L-statistic comes from considering the Gaussian MA(1) model, it is evident from the preceding derivation that the same asymptotic distribution obtains for MA(1) models with errors which satisfy the weaker assumptions, such as being martingale difference sequences which satisfy (2.2).

The L-statistics (4.7) and (4.8) have intuitive interpretations. To be concrete, consider L^{μ}. Under the null hypothesis, $y_t - \beta_0$ is serially uncorrelated and the partial sum process of the demeaned data, $\sum_{s=1}^{[T\lambda]} y^{\mu}_s$, is I(1). The statistic L^{μ} thus can be seen to test the null hypothesis that y_t is I(0) by testing its implication that the

[22] Nabeya and Tanaka (1990b) showed that the statistic (4.7) is also locally MPI unbiased for the unconditional Gaussian MA(1) model with known d_t.

process of accumulated (demeaned) y_t's is I(1). Rejection occurs if L^μ is large, so the statistic tests the null that the accumulation of y_t is I(1) against the alternative that it is I(2). Comparison of (4.7) to the expression (3.18b) for the Sargan–Bhargava statistic testing the null of a unit autoregressive root in the $d_t = \beta_0$ case shows that the two statistics are closely related: the Sargan–Bhargava statistic rejects the I(1) null against the I(0) alternative when the sum of squared y_t's is small, while the LMPI statistic L^μ rejects the I(0) null against the I(1) alternative when the sum of squared accumulated y_t's is large.

Because of the similarities between the UC and MA models, not surprisingly the L-statistics can be alternatively derived as tests of $\sigma_v^2 = 0$ in the UC model. In this formulation, the tests have the interpretation that they are testing the null that the regression intercept in (4.2) is constant, versus the alternative that it is time-varying, evolving as a martingale. To be concrete, suppose that y_t is generated by (4.2) with (ζ_t, v_t) i.i.d. $N(0, \sigma_\zeta^2 \operatorname{diag}(1, q))$, $\mu_0 = 0$, and set $q = \sigma_v^2/\sigma_\zeta^2$. Then $(y_1, \ldots, y_T)$ is distributed $N(z\beta, \sigma_\zeta^2 \Omega_{\mathrm{UC}}(q))$, where $\Omega_{\mathrm{UC}}(q) = I + q\Omega^*$, where $\Omega_{ij}^* = \min(i, j)$. Again, the results of King (1980) imply that the most powerful invariant test of $q = 0$ against $q = \bar{q} > 0$ is a ratio of quadratic forms similar to (4.6) but involving $\Omega_{\mathrm{UC}}(\bar{q})$. The resulting statistic depends on $\bar{q}$, so no uniformly MPI test exists.

Because there is no UMPI test, it is reasonable to examine the locally MPI test in the UC model. In the case $d_t = \beta_0$, Nyblom and Mäkeläinen (1983) derived the LMPI test statistic and showed it to be L^μ. Nyblom (1986) extended this analysis to the case $d_t = \beta_0 + \beta_1 t$ and showed the LMPI statistic to be L^τ. Nabeya and Tanaka (1988) extended these results to the general Gaussian regression problem in which coefficients on some of the variables follow a random walk while others are constant. The special case of Nabeya and Tanaka's (1988) results, of interest here, is when $d_t = z_t'\beta$, where $z_t = (1, t, \ldots, t^p)'$ and the intercept term, μ_t in (4.2), follows a random walk. Then Nabeya and Tanaka's (1988) LM test statistic simplifies to (4.7), except that y_t^d, the residual from the OLS regression of y_t onto the time polynomials z_t, replaces y_t^μ and $Y_t^d = T^{-1/2}\sum_{s=1}^t y_s^d$ replaces Y_t^μ.[23]

Despite the differences in the derivations in the MA and UC cases, the fact that the same local test statistics arise has a simple explanation. As argued above, Δu_t generated by the UC model has an MA(1) representation with parameters $(\theta, \sigma_\varepsilon^2)$ which solve $q = \theta^{-1} + \theta - 2$ and $\theta\sigma_\varepsilon^2 = \sigma_\zeta^2$. Thus, the distribution of $(y_1, \ldots, y_T)$ can be written as $N(z\beta, \sigma^2 \Sigma_{\mathrm{UC}}(\theta))$, where $z = (z_1, \ldots, z_T)$, $\Sigma_{\mathrm{UC},11} = (1 + q)\sigma_\zeta^2/\sigma_\varepsilon^2 = (1 - \theta + \theta^2)$ and where the remaining elements of $\Sigma_{\mathrm{UC}}(\theta)$ equal those of $\Sigma_c(\theta)$. Thus the UC and conditional MA models are the same except for their treatment of the initial value y_1. But $\Sigma_{\mathrm{UC},11}(1) = \Sigma_{c,11}(1)$, so, when $\theta = 1$ (equivalently, when $q = 0$), the two models are identical.

A third interpretation of the L-tests arises from recognizing that the UC model

[23] This simplification obtains from Nabeya and Tanaka's (1988) equation (2.5) by noting that the tth element of their My is y_t^d, by recognizing that, in our application, their D_x is the $T \times T$ identity matrix, and by carrying out the summation explicitly. See Kwiatkowski et al. (1992).

is a special case of time-varying parameter models, so that the tests can be viewed as a test for a time-varying intercept. This interpretation was emphasized by Nabeya and Tanaka (1988) and by Nyblom (1989). We return to this link in Section 5.

Local optimality is not the only testing principle which can be fruitfully exploited here, and other tests of the hypothesis $q = 0$ in the UC model have been proposed. LaMotte and McWhorter (1978) proposed a family of exact tests for random walk coefficients, which contains the i.i.d. UC model (4.1) with $d_t = \beta_0$ and $d_t = \beta_0 + \beta_1 t$ as special cases, under the translation group $y \to y + zb$, $\beta \to \beta + b$. Powers of the LaMotte and McWhorter tests are tabulated by Nyblom and Mäkeläinen (1983) (constant case) and by Nyblom (1986) (time-trend case). Franzini and Harvey (1983) considered tests in the Gaussian UC model with the maintained hypothesis of nonzero drift in μ_t, which is equivalent to (4.2) with (ζ_t, v_t) i.i.d. Gaussian and $d_t = \beta_0 + \beta_1 t$. Franzini and Harvey (1983) suggested using a point-optimal test, that is, choosing an MPI test of the form where their recommendation corresponds to $\bar{q} \cong 0.75$ for $T = 20$. Shively (1988) also examined point-optimal tests in the UC model with an intercept and suggested using the MPI tests tangent to the power envelope at, alternatively, powers of 50 percent and 80 percent, respectively, corresponding to $\bar{q} = 0.023$ and 0.079 for $T = 51$.

4.2.2. *Tests of the general I(0) null*

Because economic theory rarely suggests that an error term is i.i.d. Gaussian, the Gaussian MA(1) and i.i.d. UC models analyzed in the previous subsection are too special to be of interest in most empirical applications. While the asymptotic null distributions of the L^μ and L^τ statistics obtain under weaker conditions than Gaussianity, such as $v_t = \varepsilon_t$ where ε_t satisfies (2.2), these statistics are not asymptotically similar under the general I(0) null in which v_t is weakly dependent and satisfies (2.1)–(2.3). A task of potential practical importance, therefore, is to relax this assumption and to develop tests which are valid under the more general assumption that v_t is I(0).

The two main techniques which have been used to develop tests of the general I(0) null parallel those used to extend autoregressive unit root tests from the AR(1) model to the general I(1) model. The first, motivated by analogy to the way that Phillips and Perron (1988) handled the nuisance parameter ω in their unit root tests, is to replace the estimator of the variance of u_t in statistics such as L^μ and L^τ with an estimator of the spectral density of u_t at frequency zero; this produces "modified" L^μ and L^τ statistics.[24] The second, used by Saikkonen and Luukkonen (1993a, 1993b), is to transform the series using an estimated ARMA process for u_t.

The device of Section 3, in which unit root tests were represented as functionals of

[24] This approach was proposed by Park and Choi (1988) to generalize their variable addition tests, discussed in the subsequent paragraphs, to the general I(0) null. It was used by Tanaka (1990b) to extend the L^μ statistic to the general I(0) null. [Tanaka's (1990b) expression (7) is asymptotically equivalent to (4.7).] Kwiatkowski et al. (1992) used this approach to extend the L^τ statistic to the general I(0) null.

the levels process of the data, can be applied in this problem to provide a general treatment of those tests of the I(0) null which involve an explicit correction using an estimated spectral density. The main modification is that, in the I(0) case, the tests are represented as functionals of the accumulated levels process rather than the levels process itself. This general treatment produces, as special cases, the extended L^{μ} and L^{τ} statistics and the "variable addition" test statistics, $G(p, q)$, proposed by Park and Choi (1988).

Park and Choi's (1988) $G(p, q)$ statistic arises from supposing that the true trend is (at most) a pth order polynomial. The detrending regression is then intentionally overspecified, including polynomials of order q where $q > p$. If u_t is I(0), then the OLS estimators of the coefficients on these additional $q - p$ trends are consistent for zero. If, however, u_t is I(1), then the regression of y_t on the full set of trends introduces "spurious detrending", as discussed in Example 2 of Section 2.3, and the LR test will reject the null hypothesis that the true coefficients on $(t^{p+1}, \ldots, t^q)$ are zero. These two observations suggest considering the modified LR statistic, $G(p, q) = T(\hat{\sigma}^2 - \tilde{\sigma}^2)/\hat{\omega}^2$, where $\hat{\sigma}^2$ and $\tilde{\sigma}^2$, respectively, are the mean squared residuals from the regression of y_t onto $(1, t, \ldots, t^p)$ and of y_t onto $(1, t, \ldots, t^q)$.

In functional notation, the L- and $G(p, q)$-tests have the representations

$$L: \qquad L = g_L(f), \quad g_L(f) = \int f^2, \tag{4.9a}$$

$$G(p, q): \quad G(p, q) = g_G(f), \quad g_G(f) = \sum_{j=p+1}^{q} \left\{ \int h_j f \right\}^2, \tag{4.9b}$$

where f denotes the random function being evaluated and h_j is the jth Legendre polynomial on the unit interval.

As the representations (4.9) make clear, to study the limiting behavior of the statistics it suffices to study the behavior of the function being evaluated and then to apply the CMT. Under the general I(0) null, $T^{-1/2}\sum_{s=1}^{[T\cdot]} u_s \Rightarrow \omega W$, so that the general detrended process Y_{0T}^d (defined in Section 2.3, Example 3) has the limit $Y_{0T}^d/\hat{\sigma}_{\varepsilon}^d \Rightarrow (\omega^2/\sigma_{\varepsilon}^2)B^{(p)}$. This suggests modifying these statistics by evaluating the functionals using V_{0T}^d, where

$$V_T^d(\lambda) = \hat{\omega}_{\mathrm{SC}}^{-1} T^{-1/2} \sum_{s=1}^{[T\lambda]} y_s^d. \tag{4.10}$$

If $\hat{\omega}_{\mathrm{SC}}^2$ is consistent for ω^2 then $V_T^d \Rightarrow B^{(p)}$ and the asymptotic distributions of the statistics (4.9) will not depend on any nuisance parameters under the general I(0) null. The SC estimator $\hat{\omega}_{\mathrm{SC}}^2$ is used for this purpose by Park and Choi (1988), Tanaka (1990b), Kwiatkowski et al. (1992) and Stock (1992). The rate conditions $l_T \to \infty$ and $l_T = o(T^{1/2})$ are sufficient to ensure consistent estimation of ω under the null and, as is discussed in the next subsection, test consistency under a (fixed) alternative.

A second approach to extending the MA(1) tests to the general I(0) null, developed by Saikkonen and Luukkonen (1993a, 1993b), involves modifying the test statistic by, in effect, filtering the data. Saikkonen and Luukkonen (1993a) consider the Gaussian LMPI unbiased (LMPIU) test under an ARMA(p, q) model for the I(0) errors $v = (v_1, \ldots, v_T)$. Let the covariance matrix of v_t be Σ_v and assume that Σ_v were known. With Σ_v known, under the null hypothesis β_0 can be estimated by GLS yielding the estimator $\hat{\beta}_0$, and let $\tilde{v}_t$ denote residuals $y_t - \hat{\beta}_0$. The transformed GLS residuals are then given by $\tilde{e} = \Sigma_v^{-1/2\prime}\tilde{v}$. The Gaussian LMPIU test in this model is a ratio of quadratic forms in $\tilde{e}$ analogous to (4.6), where the covariance matrix in the numerator is evaluated under the $\theta = 1$ null. In practice, the parameters of the short-run ARMA process used to construct Σ_v must be estimated; see Saikkonen and Luukkonen (1993a) for the details. Saikkonen and Luukkonen (1993b) apply this approach to extend the finite-sample point-optimal invariant tests of the form (4.6) to general I(0) errors in the $d_t = \beta_0$ case and to derive the asymptotic distribution of these tests under the null and local alternatives.[25]

4.2.3. *Consistency and local asymptotic power*

Consistency. The statistics with the functional representations (4.9) reject for large values of the statistic. It follows that the tests based on the modified L^μ and L^τ statistics and on the $G(p, q)$ statistics are consistent if $V_T^d \xrightarrow{p} \infty$ under the I(1) alternative. Consider, first, the case $d_t = 0$, so that the numerator of V_T^d in (4.10) is $T^{-1/2}\sum_{s=1}^{[T\cdot]} u_s$. Under the I(1) alternative this cumulation is I(2) and $T^{-3/2}\sum_{s=1}^{[T\cdot]} u_s \Rightarrow \omega\int_{s=0}^{\cdot} W(s)\,ds$. Similarly, if $l_T \to \infty$ but $l_T = o(T^{1/2})$, then $\hat{\omega}^2_{SC}$ has the limit $\hat{\omega}^2_{SC}/[T\sum_{m=-l_T}^{l_T} k(m/l_T)] \Rightarrow \omega^2\int W^2$.[26] Combining these two results, we have

$$N_T^{-1/2} V_T \Rightarrow V^*, \quad \text{where} \quad V^*(\lambda) = \frac{\int_0^\lambda W(s)\,ds}{\left\{\int W^2\right\}^{1/2}} \tag{4.11}$$

and $N_T = T/\sum_{m=-l_T}^{l_T} k(m/l_T)$. Because the kernel k is bounded and $l_T = o(T^{1/2})$,

[25] Bierens and Guo (1993) used a different approach to develop a test of the general I(0) null against the I(1) alternative, in which the distribution under the null is made free of nuisance parameters not by explicit filtering or estimation of the spectral density at frequency zero, but rather by using a weighted average of statistics in which the weights are sensitive to whether the I(0) or I(1) hypothesis is true.

[26] Suppose that $d_t = 0$ and that the SC estimator is constructed using a fixed number l autocovariances of y_t, rather than letting $l_T \to \infty$; this would be appropriate were the MA order of v_t finite and known a priori. If y_t is I(1), $T^{-2}\sum_{t=i+1}^{T} y_t y_{t-i} - T^{-2}\sum_{t=1}^{T} y_t^2 \xrightarrow{p} 0, i = 1, \ldots, k$, and moreover $T^{-2}\sum_{t=1}^{T} y_t^2 \Rightarrow \omega^2\int W^2$. It follows by direct calculation that $\hat{\omega}^2_{SC}/[T\sum_{m=-l}^{l} k(m/l)] \Rightarrow \omega^2\int W^2$. The proof for the general SC estimator entails extending this result from fixed l to a sequence of l_T increasing sufficiently slowly. For details see Phillips (1991b, Appendix) for the $d_t = 0$ case; Kwiatkowski et al. (1992) for OLS detrending with a constant or a linear time trend; Perron (1991c) for general polynomial trends estimated by OLS; and Stock (1992) under general trend conditions including an estimated broken trend.

$N_T \to \infty$ as $T \to \infty$, so under the fixed I(1) alternative $V_T \xrightarrow{p} \infty$. Thus, tests which are continuous functionals of V, and which reject I(0) in favor of I(1) for large realizations of V, will be consistent against the fixed I(1) alternative.

This is readily extended to general trends. For example, in the case $d_t = \beta_0$, $N_T^{-1/2} V_T^{\mu}(\cdot) \Rightarrow V^{*\mu}$, where $V^{*\mu}(\lambda) = \int_0^{\lambda} W^{\mu}(s)\,ds / \{\int W^{\mu 2}\}^{1/2}$. In the detrended case, $N_T^{-1/2} V_T^{\tau}(\cdot) \Rightarrow V^{*\tau}$, where $V^{*\tau}(\lambda) = \int_0^{\lambda} W^{\tau}(s)\,ds / \{\int W^{\tau 2}\}^{1/2}$, where W^{τ} is the OLS-detrended Brownian motion in (2.13b). This, in turn, implies that, under the fixed I(1) alternative, V_T^{μ} and $V_T^{\tau} \xrightarrow{p} \infty$, and consistency of the test statistics in (4.9) follows directly.

Local asymptotic power. We examine local asymptotic power using a local version of the UC model (4.2):

$$y_t = d_t + u_t, \quad u_t = u_{0t} + H_T u_{1t} \tag{4.12}$$

where u_{0t} and u_{1t} are respectively I(0) and I(1) in the sense that $(u_{0t}, \Delta u_{1t})$ satisfy (2.1)–(2.3) and where $H_T = h/T$, where h is a constant. Because $H_T \to 0$, the I(1) component of y_t in (4.12) vanishes asymptotically, so that (4.12) provides a model in which y_t is a local-to-I(0) process.

For concreteness and to make the link between the UC and MA models precise, we will work with the special case of (4.12) in which u_{0t} and Δu_{1t} are mutually and serially uncorrelated and have the same variance. Then the local-to-I(0) UC model (4.12) has the MA(1) representation, $\Delta u_t = \varepsilon_t - \theta_T \varepsilon_{t-1}$, where $\theta_T = 1 - h/T + o(T^{-1})$ and where ε_t is serially uncorrelated. Thus, the local-to-I(0) parameterization $H_T = h/T$ is asymptotically equivalent to a local-to-unit MA root with the nesting $\theta_T = 1 - h/T$.[27]

To analyze the local power properties of the tests in (4.9), we obtain a limiting representation of V_T^d under the local-to-I(0) model. First consider the case $d_t = 0$. The behavior of the numerator follows from the FCLT and CMT. Define the independent Brownian motions W_0 and W_1 respectively as the limits $T^{-1/2} \sum_{s=1}^{[T\cdot]} u_{0s} \Rightarrow \omega_0 W_0(\cdot)$ and $T^{-1/2} u_{1[T\cdot]} \Rightarrow \omega_1 W_1(\cdot)$. By assumption, $\omega_0 = \omega_1$, so $T^{-1/2} \sum_{s=1}^{[T\cdot]} u_s = T^{-1/2} \sum_{s=1}^{[T\cdot]} u_{0s} + hT^{-3/2} \sum_{s=1}^{[T\cdot]} u_{1s} \Rightarrow \omega_0 U_h(\cdot)$, where $U_h(\lambda) = W_0(\lambda) + h\int_0^{\lambda} W_1(s)\,ds$. It can additionally be shown that, in the local-to-I(0) model (4.12), the SC estimator has the limit $\hat{\omega}_{SC}^2 \xrightarrow{p} \omega_0^2$ [Elliott and Stock (1994, Theorem 2)]. Thus $V_T \Rightarrow U_h$, from which it follows that the statistics in (4.9) have the local asymptotic representation $g(U_h)$ for their respective g functionals.

[27] The rate T for this local nesting is consistent with the asymptotic results in the unit MA root and UC test literatures, which in general find that this nesting is an appropriate one for studying rates of convergence of the MA estimators and/or the local asymptotic power of tests. In the MA unit root literature, see Sargan and Bhargava (1983b), Anderson and Takemura (1986), Tanaka and Satchell (1989), Tanaka (1990b) and Saikkonen and Luukkonen (1993b); in the UC literature, see Nyblom and Mäkeläinen (1983), Nyblom (1986, 1989) and Nabeya and Tanaka (1988).

Table 3
Power of MA unit root tests
[5 percent level tests, demeaned case ($d_t = \beta_0$)].[a]

h	PE	L^μ	POI(0.5)	$G(0,1)$	$G(0,2)$
A. $T = 50$					
1	0.064	0.064	0.061	0.058	0.059
2	0.103	0.101	0.095	0.087	0.089
5	0.304	0.299	0.298	0.279	0.268
10	0.631	0.570	0.633	0.583	0.596
15	0.823	0.717	0.815	0.745	0.776
20	0.909	0.803	0.898	0.823	0.864
30	0.976	0.878	0.960	0.890	0.932
40	0.992	0.914	0.979	0.912	0.954
B. $T = 100$					
1	0.064	0.064	0.066	0.065	0.060
2	0.106	0.107	0.101	0.103	0.090
5	0.332	0.319	0.321	0.311	0.289
10	0.659	0.605	0.664	0.623	0.629
15	0.845	0.765	0.841	0.779	0.807
20	0.931	0.852	0.919	0.857	0.890
30	0.985	0.923	0.974	0.924	0.952
40	0.996	0.958	0.991	0.944	0.974
C. $T = 200$					
1	0.062	0.063	0.064	0.062	0.063
2	0.102	0.104	0.099	0.097	0.095
5	0.314	0.309	0.316	0.305	0.299
10	0.669	0.605	0.667	0.621	0.640
15	0.851	0.758	0.841	0.779	0.811
20	0.937	0.847	0.922	0.854	0.894
30	0.988	0.934	0.980	0.924	0.956
40	0.998	0.965	0.995	0.950	0.976
D. $T = 1000$					
1	0.061	0.062	0.057	0.055	0.057
2	0.099	0.102	0.087	0.086	0.081
5	0.329	0.321	0.310	0.296	0.283
10	0.661	0.613	0.663	0.624	0.632
15	0.853	0.777	0.843	0.789	0.811
20	0.944	0.866	0.929	0.871	0.900
30	0.992	0.948	0.985	0.937	0.963
40	0.999	0.978	0.996	0.962	0.981

[a]Data were generated according to the unobserved components model, $y_t = u_t$, where $u_t = u_{0t} + H_T u_{1t}$, $(u_{0t}, \Delta u_{1t})$ are i.i.d. $N(0, I)$ and $H_T = h/T$. PE denotes the power envelope. The remaining tests are based on the indicated statistics as defined in the text. Based on 5000 Monte Carlo repetitions.

The power functions for the statistics in (4.9) along with the power envelope are summarized in Tables 3 (for the case $d_t = \beta_0$) and 4 (for the case $d_t = \beta_0 + \beta_1 t$). The power functions were computed by Monte Carlo simulation for various values of T, so technically all the power functions are finite-sample although the simulations

Table 4
Power of MA unit root tests
[5 percent level tests, detrended case ($d_t = \beta_0 + \beta_1 t$)].[a]

h	PE	L^c	POI(0.5)	$G(1,2)$	$G(1,3)$
A. $T = 50$					
1	0.053	0.052	0.052	0.055	0.055
2	0.062	0.062	0.061	0.062	0.062
5	0.139	0.132	0.131	0.117	0.118
10	0.349	0.322	0.348	0.259	0.299
15	0.585	0.513	0.578	0.379	0.477
20	0.744	0.659	0.746	0.469	0.603
30	0.905	0.810	0.899	0.569	0.733
40	0.958	0.880	0.954	0.618	0.795
B. $T = 100$					
1	0.055	0.055	0.053	0.052	0.051
2	0.064	0.063	0.063	0.060	0.060
5	0.127	0.136	0.130	0.121	0.113
10	0.359	0.349	0.359	0.277	0.298
15	0.610	0.560	0.609	0.414	0.491
20	0.775	0.704	0.777	0.507	0.632
30	0.939	0.864	0.928	0.629	0.777
40	0.984	0.928	0.975	0.689	0.842
C. $T = 200$					
1	0.055	0.054	0.049	0.053	0.050
2	0.063	0.064	0.056	0.062	0.062
5	0.136	0.136	0.122	0.111	0.125
10	0.369	0.357	0.362	0.269	0.323
15	0.613	0.569	0.610	0.415	0.517
20	0.785	0.718	0.776	0.521	0.655
30	0.948	0.880	0.935	0.640	0.792
40	0.989	0.950	0.980	0.708	0.867
D. $T = 1000$					
1	0.051	0.052	0.054	0.052	0.052
2	0.060	0.059	0.063	0.062	0.060
5	0.131	0.127	0.123	0.120	0.116
10	0.370	0.353	0.366	0.323	0.335
15	0.629	0.576	0.624	0.521	0.554
20	0.815	0.743	0.805	0.671	0.712
30	0.963	0.905	0.953	0.825	0.868
40	0.994	0.969	0.990	0.893	0.942

[a]See the notes to Table 3.

suggest that the $T = 1000$ power is effectively the asymptotic local power.[28] In addition, the power of the point-optimal tests which are tangent to the power

[28]Tabulations of exact power functions and the finite-sample power envelope under the Gaussian model appear in several places in the literature. Those tabulations are based on the Imhof algorithm. When results in the literature are directly comparable to those in Tables 3 and 4, they agree to within two decimals. For results in the demeaned UC model, see Nyblom and Mäkeläinen (1983) and Shively (1988); for tabulations in the detrended UC model, see Nyblom (1986). Tanaka (1990b) tabulates both finite-sample and limiting powers of the L^μ statistic, where the latter is computed by inverting numerically its limiting characteristic function [Tanaka (1990b, Theorem 2)]. Tanaka's limiting power for L^μ agrees with the $T = 1000$ powers in Table 3 to within the Monte Carlo error.

envelope at power of 50 percent are reported. In the demeaned case, this test was suggested by Shively (1988), and the test is the MPI test against the local alternative $\bar{h} = 7.74$. In the detrended case, calculations suggest that the power envelope attains 50 percent at approximately $h = 13$, so the MPI test against the local alternative $\bar{h} = 13$ is reported. This test, the POI(0.5) statistic in Table 4, is almost the same test as was proposed by Franzini and Harvey (1983): if the local-to-I(0) asymptotics are used to interpret their recommendations (which were based on a Monte Carlo experiment with $T = 20$), then the Franzini–Harvey statistic is the point-optimal invariant test which is point-optimal against the local alternative $\bar{c} \cong 17$. (Interpreted thus, the Franzini–Harvey statistic is asymptotically MPI at a power of approximately 70 percent.) These tables summarize the power findings of Nyblom and Mäkeläinen (1983), Nyblom (1986), Shively (1988), Tanaka (1990b) and Saikkonen and Luukkonen (1993a, 1993b).

Five main conclusions emerge from these tables. First, the convergence of the finite-sample powers to the asymptotic limits appears to be relatively fast, in the sense that the $T = 100$ powers and $T = 1000$ powers typically differ by less than 0.02. Second, as was the case with tests for a unit autoregressive root, the powers deteriorate as the order of detrending increases from demeaning to linear detrending, particularly for alternatives of h near zero. For example, the L^{μ} statistic has a limiting power of 0.61 against $h = 10$, while the corresponding power for the L^{τ} statistic is 0.35. Third, the point-optimal tests perform better than the LMPIU test against all but the closest alternatives. Fourth, although the Park–Choi $G(p, p+1)$ and $G(p, p+2)$ tests are strictly below the power envelope, they nonetheless perform rather well and in particular have power curves only slightly below the L^{μ} and L^{τ} statistics. Fifth, it is important to emphasize that all these differences are rather modest in comparison to the large differences in powers found among the various tests for a unit AR root. For example, the Pitman efficiency of the L^{μ} statistic relative to the MPI test at power = 50 percent is approximately 1.1, indicating a loss of the equivalent of only 10 percent of the sample if the L^{μ} statistic is used in this case rather than the MPI test.

4.2.4. *Finite-sample size and power*

A small Monte Carlo experiment was performed to examine the finite-sample size and power of tests of the I(0) null. Unlike for tests of the I(1) null, as of this writing, there have been few Monte Carlo investigations of tests of the general I(0) null; exceptions include Amano and van Norden (1992) and Kwiatkowski et al. (1992). The simulation here summarizes the results of these two studies for the L^{τ} statistic by using a similar design (autoregressive errors) and extends them to include the Park–Choi $G(p, p+2)$ statistics and to examine the effect of kernel choice on test performance.

In the $d_t = \beta_0$ case, the experiment considers the modified L^{μ} and $G(0, 2)$ statistics (based on V_T^{μ}); in the $d_t = \beta_0 + \beta_1 t$ case, the statistics are the modified L^{τ} and $G(1, 3)$

statistics (based on V_T^{τ}). The spectral density was estimated using two SC spectral estimators with a truncated automatic bandwidth selector. The automatic bandwidth is $l_T = \min[\hat{l}_T, 12(T/100)^{0.2}]$, where $\hat{l}_T$ is Andrews' (1991) automatic selector based on an estimated AR(1) model. The two kernels are the Parzen kernel and the QS kernel, the latter being Andrews' (1991) optimal kernel, and the appropriate selector for each kernel is used. [The automatic bandwidth selector is truncated because unless $\hat{l}_T$ is bounded in the I(1) case it does not satisfy the $o(T^{-1/2})$ rate condition needed for consistency as described in Section 4.2.3.]

The pseudo-data were generated so that v_t followed the AR(1),

$$y_t = u_t, \quad \Delta u_t = (1 - \theta L)v_t,$$

$$\text{where} \quad v_t = \rho v_{t-1} + \varepsilon_t, \quad \varepsilon_t \text{ i.i.d. } N(0,1), \tag{4.13}$$

where $u_0 = 0$ and v_0 is drawn from its unconditional distribution. When $|\rho| < 1$ and $\theta = 1$, y_t is I(0) and the experiment examines the size of the test. When $|\rho| < 1$ and $|\theta| < 1$, y_t is I(1). When $\rho = 0$, this is the MA(1) model and corresponds to the local-to-unity model (4.12) with $(u_{0t}, \Delta u_{1t})$ mutually and serially uncorrelated with the same variance, in which case $\theta = 1 - h/T + o(T^{-1})$.

Empirical size (in italics) and size-adjusted power are presented for $T = 100$ in Table 5 (the demeaned case) and Table 6 (the detrended case). Size-adjusted power in a (ρ, θ) design, $|\theta| < 1$, is computed using the 5 percent empirical quantile for $(\rho, \theta = 1)$ for each value of ρ.

These results suggest three conclusions. First, the choice of spectral estimator matters for size, less so for size-adjusted power. For example, if the Parzen kernel is used, the size deteriorates substantially when the serial correlation is large ($\rho = 0.9$). [If the Bartlett kernel is used, as suggested by Tanaka (1990b) and Kwiatkowski et al. (1992), similar size distortions arise (results not shown in these tables).] In contrast, the size is much better controlled using the QS kernel. This is true for both of the statistics examined, in both the demeaned and detrended cases. On the other hand, the size-adjusted powers for both statistics in both cases are comparable for the two spectral estimators. Interestingly, for distant alternatives the size-adjusted power declines in the $\rho = 0$ case for the demeaned statistics, and the decline is more pronounced for the QS statistics.

Second, a comparison of the results in Tables 3 and 4 with those in Tables 5 and 6, respectively, reveals that when $\rho = 0$ the finite-sample size-adjusted power is fairly close to the power predicted by the local-to-I(0) asymptotics of Section 4.2.3, at least for close and moderately close alternatives. At least in the $\rho = 0$ case, the use of the SC estimator seems to have little impact on either size or power. However, size-adjusted power deteriorates sharply as the autoregressive nuisance parameter increases towards one. Interestingly, detrending makes little difference in terms of size. This is noteworthy, given the large impact of detrending in the I(1) test situations.

Table 5
Size and size-adjusted power of selected tests of the I(0) null: Monte Carlo results
[5 percent level tests, demeaned case ($d_t = \beta_0$), $T = 100$]
[Data generating process: $(1 - \rho L)\Delta y_t = (1 - \theta L)\varepsilon_t$, ε_t i.i.d. $N(0, 1)$].[a]

Test Statistic	θ	Asymptotic Power	$\rho =$ 0.0	0.9	0.75	0.5	−0.5
L^μ	1.00	*0.05*	*0.05*	*0.26*	*0.10*	*0.06*	*0.04*
P(auto)	0.95	0.32	0.29	0.26	0.25	0.25	0.29
	0.90	0.61	0.55	0.43	0.46	0.47	0.56
	0.80	0.87	0.69	0.53	0.58	0.60	0.79
	0.70	0.95	0.68	0.55	0.61	0.64	0.87
L^μ	1.00	*0.05*	*0.05*	*0.11*	*0.05*	*0.06*	*0.04*
QS(auto)	0.95	0.32	0.30	0.24	0.21	0.24	0.29
	0.90	0.61	0.57	0.39	0.36	0.39	0.56
	0.80	0.87	0.72	0.46	0.47	0.43	0.80
	0.70	0.95	0.67	0.49	0.49	0.44	0.88
$G(0, 2)$	1.00	*0.05*	*0.05*	*0.29*	*0.10*	*0.05*	*0.03*
P(auto)	0.95	0.28	0.28	0.26	0.24	0.23	0.26
	0.90	0.63	0.58	0.48	0.49	0.49	0.57
	0.80	0.90	0.76	0.62	0.66	0.67	0.81
	0.70	0.96	0.76	0.65	0.70	0.72	0.88
$G(0, 2)$	1.00	*0.05*	*0.05*	*0.07*	*0.04*	*0.06*	*0.04*
QS(auto)	0.95	0.28	0.29	0.19	0.16	0.21	0.26
	0.90	0.63	0.60	0.37	0.33	0.34	0.57
	0.80	0.90	0.78	0.49	0.47	0.37	0.81
	0.70	0.96	0.72	0.52	0.52	0.38	0.88

[a]For each statistic, the first row of entries are the empirical rejection rates under the null, that is, the empirical size of the test, based on the asymptotic critical value. The remaining entries are the size-adjusted power for the model given in the column heading. The column Asymptotic Power gives the $T = 1000$ rejection rate for that statistic from Table 3 using $\theta_T = 1 - h/T$. The entry below the name of each statistic indicates the spectral density estimator used. P(auto) and QS(auto) refer to the SC estimator, computed respectively using the Parzen and QS kernels, each with lag lengths chosen by the respective truncated automatic selector in Andrews (1991). Based on 5000 Monte Carlo repetitions.

Third, the differences in size-adjusted power across test statistics are modest. Because of its better size performance, we restrict the discussion to the results for the QS kernel. In the demeaned case, $G(0, 2)$ has somewhat better size-adjusted power than the modified L^μ statistic for distant alternatives when v_t is positively correlated; for θ near one, the modified L^μ statistic is more powerful. In the detrended case, $G(1, 3)$ and modified L^τ have essentially the same size-adjusted powers.

4.2.5. *Summary and implications for empirical practice*

The literature on tests of the general I(0) null against the I(1) alternative is still young. Subject to this caveat, the results here suggest several observations. The asymptotic power analysis of Section 4.2.3 suggests that there is little room for improvement on the performance of the currently proposed tests, at least in terms of local

Table 6
Size and size-adjusted power of selected tests of the I(0) null: Monte Carlo results
[5 percent level tests, detrended case ($d_t = \beta_0 + \beta_1 t$), $T = 100$]
[Data generating process: $(1 - \rho L)\Delta y_t = (1 - \theta L)\varepsilon_t$, ε_t i.i.d. $N(0, 1)$].[a]

Test Statistic	θ	Asymptotic Power	$\rho =$ 0.0	0.9	0.75	0.5	−0.5
L^z	1.00	*0.05*	*0.05*	*0.29*	*0.11*	*0.06*	*0.04*
P(auto)	0.95	0.13	0.13	0.12	0.12	0.12	0.12
	0.90	0.35	0.34	0.23	0.25	0.25	0.32
	0.80	0.74	0.62	0.36	0.41	0.43	0.65
	0.70	0.91	0.64	0.40	0.47	0.50	0.81
L^z	1.00	*0.05*	*0.05*	*0.10*	*0.05*	*0.06*	*0.04*
QS(auto)	0.95	0.13	0.13	0.10	0.09	0.12	0.13
	0.90	0.35	0.35	0.19	0.16	0.22	0.33
	0.80	0.74	0.65	0.28	0.25	0.25	0.67
	0.70	0.91	0.68	0.30	0.28	0.23	0.83
$G(1, 3)$	1.00	*0.05*	*0.04*	*0.30*	*0.12*	*0.07*	*0.04*
P(auto)	0.95	0.12	0.12	0.12	0.11	0.11	0.11
	0.90	0.34	0.31	0.25	0.24	0.23	0.28
	0.80	0.71	0.58	0.40	0.43	0.43	0.59
	0.70	0.87	0.62	0.46	0.49	0.51	0.73
$G(1, 3)$	1.00	*0.05*	*0.04*	*0.13*	*0.07*	*0.07*	*0.04*
QS(auto)	0.95	0.12	0.12	0.10	0.08	0.10	0.11
	0.90	0.34	0.32	0.18	0.15	0.20	0.29
	0.80	0.71	0.61	0.28	0.24	0.25	0.60
	0.70	0.87	0.64	0.32	0.28	0.24	0.74

[a]See the notes to Table 5.

asymptotic power. The various tests have asymptotic relative efficiencies fairly close to one, and the point-optimal tests (the Shively and Franzini–Harvey tests), interpreted in the local-to-I(0) asymptotic framework, have power functions that are close to the power envelope for a large range of local alternatives.

The Monte Carlo results suggest, however, that there remains room for improvement in the finite-sample performance of these tests. With the Parzen kernel, the tests exhibit large size distortions; with the QS kernel, the size distortions are reduced but the finite-sample power can be well below its asymptotic limit. For autoregressive parameters not exceeding 0.75, both the $G(p, p + 2)$ and L statistics, evaluated using the QS(auto) kernel, have Monte Carlo sizes near their asymptotic levels and have comparable power.

5. Structural breaks and broken trends

This section examines two topics: structural breaks and parameter instability in time series regression; and tests for a unit root when there are kinks or jumps in the

deterministic trend (the "broken-trend" model). At first glance these problems seem quite different. However, there are close mathematical and conceptual links which this section aims to emphasize. Mathematically, a multidimensional version of the FCLT plus CMT approach of Section 2 is readily applied to provide asymptotic representations for a variety of tests of parameter stability. [An early and sophisticated application of the FCLT to the change-point problem can be found in MacNeill (1974).] Conceptually, the unobserved components model with a small independent random walk component is in fact a special case of the more general time-varying-parameter model. Also, these topics recently have become intertwined in empirical investigations into unit roots when one maintains the possibility that the deterministic component has a single break, for example is a piecewise-linear time trend.

Section 5.1 addresses testing for and, briefly, estimation of parameter instability in time series regression with I(0) regressors, including the case when there are lagged dependent I(0) variables and, in particular, stationary autoregressions. The main empirical application of these tests is as regression diagnostics and, as an example in Section 5.1.4, the tests are used to assess the stability of the link between various monetary aggregates and output in the U.S. from 1960 to 1992. The literature on parameter instability and structural breaks is vast, and the treatment here provides an introduction to the main applications in econometric time series regression from a classical perspective. The distribution theory for the tests is nonstandard. Here, the alternatives of interest have parameters which are unidentified under the null hypothesis; for example, in the case of a one-time change in a coefficient, under the null of "no break" the magnitude of the change is zero and the break date is unidentified. Davies (1977) showed that, if parameters are unidentified under the null, standard χ^2 inference does not obtain, and many of the results in Section 5.1 can be seen as special cases of this more general problem. For further references on parameter instability and breaks, the reader is referred to the reviews and bibliographies in Hackl and Westlund (1989), Krishnaiah and Miao (1988), Krämer and Sonnberger (1986) and, for Bayesian work in this area, Zacks (1983) and Barry and Hartigan (1993).

Section 5.2 turns to inference about the largest root in univariate autoregression under the maintained hypothesis that there might be one-time breaks or jumps in the deterministic component. In innovative papers, Perron (1989a, 1990b) and Rappoport and Reichlin (1989) independently suggested that the broken-trend model provides a useful description of a wide variety of economic time series. Perron (1989a) argued, inter alia, that U.S. postwar real GNP is best modeled as being I(0) around a piecewise-linear trend with a break in 1973, and Rappoport and Reichlin (1989) argued that U.S. real GNP from 1909–1970 [the Nelson–Plosser (1982) data] was stationary around a broken trend with a break in 1940. These results seem to suggest that the long-term properties of output are determined not by unit-root dynamics, but rather by rare events with lasting implications for mean long-term growth, such as World War II and the subsequent shift to more activist governmental

economic policy, or the oil shock and productivity slowdown of the mid-1970's. Whether this view is upheld statistically is a topic of ongoing debate in which the tests of Section 5.2 play a central role.

5.1. *Breaks in coefficients in time series regression*

5.1.1. *Tests for a single break date*

Suppose y_t obeys the time series regression model

$$y_t = \beta_t' X_{t-1} + \varepsilon_t, \tag{5.1}$$

where under the null hypothesis $\beta_t = \beta$ for all t. Throughout Section 5.1, unless explicitly stated otherwise, it is maintained that ε_t is a martingale difference sequence with respect to the σ-fields generated by $\{\varepsilon_{t-1}, X_{t-1}, \varepsilon_{t-2}, X_{t-2}, \ldots\}$, where X_t is a $k \times 1$ vector of regressors, which are here assumed to be constant and/or I(0) with $EX_t X_t' = \Sigma_X$ and, possibly, a nonzero mean. For convenience, further assume that ε_t is conditionally (on lagged ε_t and X_t) homoskedastic. Also, assume that $T^{-1}\sum_{s=1}^{[T\lambda]} X_s X_s' \xrightarrow{p} \lambda \Sigma_X$ uniformly in λ for $\lambda \in [0, 1]$. Note, in particular, that X_{t-1} can include lagged dependent variables as long as they are I(0) under the null.

The alternative hypothesis of a single break in some or all of the coefficients is

$$\beta_t = \beta, \quad t \leqslant r \quad \text{and} \quad \beta_t = \beta + \gamma, \quad t > r, \tag{5.2}$$

where $r, k+1 < r < T$, is the "break date" (or "change point") and $\gamma \neq 0$.

When the potential break date is known, a natural test for a change in β is the Chow (1960) test, which can be implemented in asymptotically equivalent Wald, Lagrange multiplier (LM), and LR forms. In the Wald form, the test for a break at a fraction r/T through the sample is

$$F_T\left(\frac{r}{T}\right) = \frac{\mathrm{SSR}_{1,T} - (\mathrm{SSR}_{1,r} + \mathrm{SSR}_{r+1,T})}{(\mathrm{SSR}_{1,r} + \mathrm{SSR}_{r+1,T})/(T-2k)}, \tag{5.3}$$

where $SSR_{1,r}$ is the sum of squared residuals from the estimation of (5.1) on observations $1, \ldots, r$, etc. For fixed r/T, $F_T(r/T)$ has an asymptotic χ_k^2 distribution under the null. When the break date is unknown, the situation is more complicated. One approach might be to estimate the break date, then compute (5.3) for that break. However, because the change point is selected by virtue of an apparent break at that point, the null distribution of the resulting test is not the same as if the break date were chosen without regard to the data. The means of determining r/T must be further specified before the distribution of the resulting test can be obtained.

A natural solution, proposed by Quandt (1960) for time series regression and extended by Davies (1977) to general models with parameters unidentified under the null, is to base inference on the LR statistic, which is the maximal F_T statistic over a range of break dates $r_0, \ldots, r_1$. This yields the Quandt likelihood ratio (QLR) statistic,

$$\text{QLR} = \max_{r=r_0,\ldots,r_1} F_T\left(\frac{r}{T}\right). \tag{5.4}$$

Intuition suggests that this statistic will have power against a change in β even though the break date is unknown. The null asymptotic distribution of the QLR statistic remained unknown for many years. The FCLT and CMT, however, provide ready tools for obtaining this limit. The argument is sketched here; for details, see Kim and Siegmund (1989) and, for a quite general treatment of "sup tests" in nonlinear models, Andrews (1993b).

To obtain the limiting null distribution of the QLR statistic, let $\tilde{F}_T(r/T) = \text{SSR}_{1,T} - (\text{SSR}_{1,r} + \text{SSR}_{r+1,T})$ and use (5.1) to write

$$\begin{aligned}
\tilde{F}_T\left(\frac{r}{T}\right) &= -\left(\sum_{t=2}^{T} X_{t-1}\varepsilon_t\right)'\left(\sum_{t=2}^{T} X_{t-1}X_{t-1}'\right)^{-1}\left(\sum_{t=2}^{T} X_{t-1}\varepsilon_t\right) \\
&\quad + \left(\sum_{t=2}^{r} X_{t-1}\varepsilon_t\right)'\left(\sum_{t=2}^{r} X_{t-1}X_{t-1}'\right)^{-1}\left(\sum_{t=2}^{r} X_{t-1}\varepsilon_t\right) \\
&\quad + \left(\sum_{t=r+1}^{T} X_{t-1}\varepsilon_t\right)'\left(\sum_{t=r+1}^{T} X_{t-1}X_{t-1}'\right)^{-1}\left(\sum_{t=r+1}^{T} X_{t-1}\varepsilon_t\right) \\
&= -v_T(1)'V_T(1)^{-1}v_T(1) + v_T\left(\frac{r}{T}\right)'V_T\left(\frac{r}{T}\right)^{-1}v_T\left(\frac{r}{T}\right) \\
&\quad + \left[v_T(1) - v_T\left(\frac{r}{T}\right)\right]'\left[V_T(1) - V_T\left(\frac{r}{T}\right)\right]^{-1}\left[v_T(1) - v_T\left(\frac{r}{T}\right)\right],
\end{aligned} \tag{5.5}$$

where $v_T(\lambda) = T^{-1/2}\sum_{t=2}^{[T\lambda]} X_{t-1}\varepsilon_t$ and $V_T(\lambda) = T^{-1}\sum_{t=2}^{[T\lambda]} X_{t-1}X_{t-1}'$. Because ε_t is a martingale difference sequence, $X_{t-1}\varepsilon_t$ is a martingale difference sequence. Additionally, assume throughout Section 5.1 that X_{t-1} has sufficiently limited dependence and enough moments for $X_{t-1}\varepsilon_t$ to satisfy a multivariate martingale difference sequence FCLT, so $v_T(\cdot) \Rightarrow \sigma_\varepsilon \Sigma_X^{1/2} W_k(\cdot)$, where W_k is a k-dimensional standard Brownian motion. Also, recall that by assumption $V_T(\lambda) \xrightarrow{p} \lambda\Sigma_X$ uniformly in λ. By applying these two limits to the second expression in (5.5), one obtains

$$\tilde{F}_T(\cdot) \Rightarrow \sigma_\varepsilon^2 F^*(\cdot), \tag{5.6}$$

where

$$F^*(\lambda) = -W_k(1)'W_k(1) + \frac{W_k(\lambda)'W_k(\lambda)}{\lambda} + \frac{[W_k(1) - W_k(\lambda)]'[W_k(1) - W_k(\lambda)]}{1-\lambda}$$

$$= \frac{B_k^{\mu}(\lambda)'B_k^{\mu}(\lambda)}{\lambda(1-\lambda)},$$

where $B_k^{\mu}(\lambda) = W_k(\lambda) - \lambda W_k(1)$, where $W_k(1)$ is a k-dimensional Brownian bridge. Because $\tilde{F}_T \Rightarrow \sigma_\varepsilon^2 F^*$ and $\mathrm{SSR}_{1,T}/(T-k) \xrightarrow{\mathrm{p}} \sigma_\varepsilon^2$ under the null, $(\mathrm{SSR}_{1,r} + \mathrm{SSR}_{r+1,T})/(T-2k) \xrightarrow{\mathrm{p}} \sigma_\varepsilon^2$ uniformly in r. Thus $F_T \Rightarrow F^*$. It follows from the continuous mapping theorem that the QLR statistic has the limiting representation,

$$\mathrm{QLR} \Rightarrow \sup_{\lambda \in [\lambda_0, \lambda_1]} \left\{ \frac{B_k^{\mu}(\lambda)'B_k^{\mu}(\lambda)}{\lambda(1-\lambda)} \right\}, \tag{5.7}$$

where $\lambda_i = \lim_{T\to\infty} r_i/T$, $i = 0, 1$. For fixed $\bar{\lambda}$, $F^*(\bar{\lambda})$ has a χ_k^2 distribution.

Andrews (1993b, Table I) reports asymptotic critical values of the functional in (5.7), computed by Monte Carlo simulation for a range of trimming parameters and $k = 1, \ldots, 20$. The critical values are much larger than the conventional fixed-break χ_k^2 critical values. For example, consider 5 percent critical values with truncation fractions $(\lambda_0, \lambda_1) = (0.15, 0.85)$: for $k = 1$, the QLR critical value is 8.85, while the χ_1^2 value is 3.84; for $k = 10$, the QLR critical value is 27.03, while the χ_{10}^2 critical value is 18.3. In practice the researcher must choose the trimming parameters r_0 and r_1. In some applications the approximate break date might be known and used to choose r_0 and r_1. Also, with nonnormal errors and small r_0 the fixed-r distribution of the $F_T(r/T)$ statistic can be far from χ_k^2, so one way to control size is to choose r_0 sufficiently large, say $r_0/T = 0.15$ and $r_1/T = 0.85$.[29]

The error process has been assumed to be serially uncorrelated. If it is serially correlated but uncorrelated with the regressors, then the distribution of the change-point test differs. In the case of a known break date, this problem is well studied and the Wald test statistic should be computed using an autocorrelation-consistent estimator of the covariance matrix; for recent work and discussion of the literature, see Andrews (1991) and Andrews and Monahan (1992). For the extension to break tests with unknown break dates, see Tang and MacNeill (1993).

[29] Functionals of $F_T(\lambda)$ other than the supremum are possible. Examples include the average of F_T, perhaps over a restricted range, as studied by Andrews and Ploberger (1992) and Hansen (1990) [see Chernoff and Zacks (1964) and Gardner (1969) for historical precedents]. Andrews and Ploberger (1992) consider tests which maximize weighted average local asymptotic power, averaged over the unidentified nuisance parameters (here, the break date). The resulting family of optimal tests are weighted averages of exponentials, with the simple weighted average as the limit for nearby alternatives. The Andrews–Ploberger (1992) tests are reviewed in the chapter by Andrews in this Handbook.

The derivation of (5.6) assumes that $T^{-1/2}\sum_{s=1}^{[T\cdot]} X_{s-1}\varepsilon_s$ obeys an FCLT and $T^{-1}\sum_{s=1}^{[T\lambda]} X_s X_s' \xrightarrow{p} \lambda\Sigma_X$ uniformly in λ. These assumptions hold if X_t contains a constant and/or I(0) regressors, but not if X_t is I(1). A sufficient condition for (5.6) *not* to hold is that the standard Chow test for fixed r/T does not have an asymptotic χ^2 distribution, since $F^*(\bar{\lambda})$ has a χ^2 distribution for any fixed $\bar{\lambda}$. This will occur, in general, for I(1) regressors (although there are exceptions in cointegrating relations; see Watson's chapter in this Handbook) and, in these cases, the derivations must be modified; see Banerjee et al. (1992b), Chu and White (1992) and Hansen (1992) for examples.

In principle this approach can be extended to more than one break. A practical difficulty is that the computational demands increase exponentially in the number of breaks (all values of the two-break F-statistic need to be computed for break dates (r,s) over the range of r and s), which makes evaluating the limiting distributions currently difficult for more than two or three break dates. More importantly, positing multiple exogenous breaks raises the modeling question of whether the breaks are better thought of as stochastic or as the result of a continuous process. Indeed, this line of reasoning leads to a formulation in which the parameters change stochastically in each period by random amounts, which is the time-varying parameter model discussed in Section 5.1.3.

A related problem is the construction of confidence intervals for the break date. A natural estimator of the break date is the Gaussian MLE $\hat{\lambda}$, which is the value of $\lambda\in(\lambda_0,\lambda_1)$ which maximizes the LR test statistic (5.3). The literature on inference about the break date is large and beyond the scope of this chapter, and we make only two observations. First, $\hat{\lambda}$ is consistent for λ when the break magnitude is indexed to the sample size ($\gamma=\gamma_T$) and $\gamma_T\to 0, T^{1/2}\gamma_T\to\infty$ [Picard (1985), Yao (1987), Bai (1992)], although $\hat{r}$ itself is not consistent. Second, it is possible to construct asymptotic confidence intervals for λ, but this is not as straightforward as inverting the LR statistic using the QLR critical values because the null for the LR statistic is no break, while the maintained hypothesis for the construction of a confidence interval is that a break exists. Picard's (1985) results can be used to construct confidence intervals for the break date by inverting a Wald-type statistic, an approach extended to time series regression with dependent errors by Bai (1992). Alternatively, finite-sample intervals can be constructed with sufficiently strong conditions on ε_t and strong conditions on X_t; see Siegmund (1988) and Kim and Siegmund (1989) for results and discussion.

5.1.2. *Recursive coefficient estimates and recursive residuals*

Another approach to the detection of breaks is to examine the sequence of regression coefficients estimated with increasingly large data sets, that is, to examine $\tilde{\beta}(\lambda)$, the OLS estimator of β computed using observations $1,\ldots,[T\lambda]$. These tests typically have been proposed without reference to a specific alternative, although the most commonly studied alternative is a single structural break. Related is Brown's et al.

(1975) CUSUM statistic, which rejects when the time series model systematically over- or under-forecasts y_t, more precisely, when the cumulative one-step-ahead forecast errors, computed recursively, are either too positive or negative. The recursive coefficients and Brown's et al. (1975) recursive residuals and CUSUM statistic are, respectively, given by

$$\tilde{\beta}\left(\frac{r}{T}\right)=\left(\sum_{t=2}^{r} X_{t-1}X'_{t-1}\right)^{-1}\left(\sum_{t=2}^{r} X_{t-1}y_t\right), \tag{5.8}$$

$$w_t=\frac{y_t-\tilde{\beta}((t-1)/T)'X_{t-1}}{f_t}, \tag{5.9}$$

$$\text{CUSUM}(\lambda)=\frac{T^{-1/2}\sum_{s=k+1}^{[T\lambda]} w_s}{\tilde{\sigma}_\varepsilon}, \tag{5.10}$$

where $\tilde{\sigma}_\varepsilon=\{T^{-1}\sum_{t=k+1}^{T}(w_t-\bar{w})^2\}^{1/2}$ and $f_t=\{1+X'_{t-1}(\sum_{s=2}^{t-1}X_{s-1}X'_{s-1})^{-1}X_{t-1}\}^{1/2}$ (this comes from noting that the variance of the one-step-ahead forecast error is $\sigma_\varepsilon^2 f_t^2$). The CUSUM test rejects for large values of $\sup_{0\leq\lambda\leq 1}|\text{CUSUM}(\lambda)/(1+2\lambda)|$.

Because the recursive coefficients are evaluated at each point r, the distribution of the recursive coefficients differs from the usual distribution of the OLS estimator. The asymptotics readily obtain using the tools of Section 2. Under the null hypothesis $\beta_t=\beta$, the arguments leading to (5.6), applied here, yield

$$T^{1/2}(\tilde{\beta}(\cdot)-\beta)=V_T(\cdot)^{-1}v_T(\cdot)\Rightarrow\beta^*(\cdot),\quad \beta^*(\lambda)=\frac{\sigma_\varepsilon\Sigma_X^{-1/2}W_k(\lambda)}{\lambda} \tag{5.11}$$

(Ploberger et al. (1989), Lemma A.1). For fixed λ, $\beta^*(\lambda)$ has the usual OLS asymptotic distribution. An immediate implication of (5.11) is that conventional "95 percent" confidence intervals, plotted as bands around the path of recursive coefficient estimates, are inappropriate since those bands fail to handle simultaneous inferences on the full plot of recursive coefficients.

Combined with the CMT, (5.11) can be used to construct a formal test for parameter constancy based on recursive coefficients. An example is Ploberger's et al. (1989) "fluctuations" test [also see Sen (1980)], which rejects for large changes in the recursive coefficients, specifically when $\tilde{\beta}(\lambda)-\tilde{\beta}(1)$ is large. From (5.11), note that $T^{1/2}(\tilde{\beta}(\lambda)-\tilde{\beta}(1))\Rightarrow\sigma_\varepsilon\Sigma_X^{-1/2}(W_k(\lambda)/\lambda-W_k(1))$ uniformly in λ. Because the full-sample OLS estimator $\hat{\sigma}_\varepsilon^2$ is consistent under the null, $B^{(T)}(\lambda)=\hat{\sigma}_\varepsilon^{-1}(T^{-1}\sum^T X_{t-1}X'_{t-1})^{1/2}\times T^{1/2}(\tilde{\beta}(\lambda)-\tilde{\beta}(1))\Rightarrow W_k(\lambda)/\lambda-W_k(1)$, uniformly in λ. This leads to Ploberger's et al. (1989) "fluctuations" test and its limiting representation under the null of parameter

constancy,

$$\max_{r_0 \leq r \leq r_1} \max_{1 \leq i \leq k} \left| \left(\frac{r}{T}\right) B_i^{(T)}\left(\frac{r}{T}\right) \right| \Rightarrow \sup_{\lambda_0 \leq \lambda \leq \lambda_1} \max_{1 \leq i \leq k} |B_{ki}(\lambda)|, \tag{5.12}$$

where $B_i^{(T)}$ is the ith element of $B^{(T)}$ and B_{ki} is the ith element of the k-dimensional Brownian bridge B_k.

The null distribution of the CUSUM test is also obtained by FCLT and CMT arguments. If X_{t-1} is strictly exogenous and ε_t is i.i.d. $N(0, \sigma_\varepsilon^2)$, then w_t is i.i.d. $N(0, \sigma_\varepsilon^2)$, so the FCLT and CMT imply

$$\text{CUSUM}(\cdot) \Rightarrow W(\cdot), \tag{5.13}$$

where W is a 1-dimensional Brownian motion. The same limit obtains with general I(0) regressors and a constant, but the calculation is complicated and is omitted here; for the details, see Krämer et al. (1988), who prove (5.13) for time series regressions possibly including lagged dependent variables and for general i.i.d. errors (their i.i.d. assumption can be relaxed to the martingale difference assumption used here). Critical values for $\sup_\lambda |\text{CUSUM}(\lambda)/(1 + 2\lambda)|$ are obtained from results in Brownian motion theory; see Brown et al. (1975).

An important feature of the CUSUM statistic is that, as shown by Krämer et al. (1988), it has local asymptotic power only in the direction of the mean regressors: coefficient breaks of order $T^{-1/2}$ on mean-zero stationary regressors will not be detected. This has an intuitive explanation. The cumulation of a mean-zero regressor will remain mean-zero (and will obey an FCLT) whether or not its true coefficient changes, while the nonzero mean of the cumulation of the constant implies that breaks in the intercept will result in systematically biased forecast errors.[30] This is both a limitation and an advantage, for rejection suggests a particular alternative (instability in the intercept or the direction of the mean regressors).

Several variants of the CUSUM statistic have been proposed. Ploberger and Krämer's (1992a) version, in which full-sample OLS residuals $\hat{e}_t$ replace the recursive residuals $\hat{w}_t$, is attractive because of its computational simplicity. Again, the distribution is obtained using the FCLT and CMT. Their test statistic and its

[30] Consider the simplest case, in which $y_t = \varepsilon_t$ under the null while under the local alternative $y_t = T^{-1/2}\gamma \mathbf{1}(t > r)X_{t-1} + \varepsilon_t$. Since β is known to equal zero, under the null (with $\gamma = 0$ imposed) the cumulated residuals process is just $T^{-1/2}\sum_{s=1}^{t} y_s$. Under the local alternative, $T^{-1/2}\sum_{s=1}^{[T\lambda]} y_s = T^{-1/2}\sum_{s=1}^{[T\lambda]} \varepsilon_s + \gamma T^{-1}\sum_{s=r+1}^{[T\lambda]} X_{s-1} \Rightarrow W(\lambda) + \gamma \max(0, r/T - \lambda) EX_t$. If EX_t is zero, the distribution is the same under the local alternative and the null; the test only has power in the direction of the mean vector EX_t. Estimation of β, as is of course done in practice, does not affect this conclusion qualitatively because the alternative is local. Also see Ploberger and Krämer (1990).

limiting null representation are

$$\max_{k\in[1,T]}\left|\frac{T^{-1/2}\sum_{s=1}^{k}\hat{e}_s}{\hat{\sigma}_\varepsilon}\right| \Rightarrow \sup_{\lambda\in[0,1]}|B_1^{\mu}(\lambda)|, \tag{5.14}$$

where B_1 is the one-dimensional Brownian bridge and the limit obtains using the FCLT and CMT. Other variants include Brown's et al. (1975) CUSUM-of-Squares test based on w_t^2, and McCabe and Harrison's (1980) CUSUM-of-Squares test based on OLS residuals. See Ploberger and Krämer (1990) for a discussion of the low asymptotic power of the CUSUM-of-Squares test. See Deshayes and Picard (1986) and the bibliography by Hackl and Westlund (1989) for additional references.

If the regressors are I(1), the distribution theory for rolling and recursive tests changes, although it still can be obtained using the FCLT and CMT as it was throughout this chapter. See Banerjee et al. (1992b) for rolling and recursive tests with a single I(1) regressor, Chu and White (1992) for fluctuations tests in models with stochastic and deterministic trends, and Hansen (1992) for Chow-type (e.g. QLR) and LM-type [e.g. Nyblom's (1989) statistic] tests with multiple I(1) regressors in cointegrating equations. Also, the distribution of the CUSUM statistic changes if stochastically or deterministically trending regressors are included; see MacNeill (1978) and Ploberger and Krämer (1992b).

5.1.3. *Tests against the time-varying-parameter model*

A flexible extension of the standard regression model is to suppose that the regression coefficients evolve over time, specifically

$$y_t = \beta_t' X_{t-1} + \varepsilon_t, \quad \beta_t = \beta_{t-1} + v_t, \quad Ev_t v_t' = \tau^2 G, \tag{5.15}$$

where ε_t and v_t are uncorrelated and v_t is serially uncorrelated. The formulation (5.15), of course, nests the standard linear regression model when $\tau^2 = 0$. By setting $v_t = \gamma$, $t = r+1$ and $v_t = 0, t \neq r+1$, (5.15) nests the single-break model (5.2). The alternative of specific interest here, however, is when v_t is i.i.d. $N(0, \tau^2 G)$ (where G is assumed to be known) so that the coefficient β_t follows a multivariate stochastic trend and thus evolves smoothly but randomly over the sample period. When combined with the additional assumption that ε_t is i.i.d. $N(0, \sigma_\varepsilon^2)$, this is referred to as the "time-varying-parameter" (TVP) model [see Cooley and Prescott (1976), Sarris (1973) and the reviews by Chow (1984) and Nicholls and Pagan (1985)]. Maximum likelihood estimation of the TVP model is a direct application of the Kalman filter (β_t is the unobserved state vector and $y_t = \beta_t' X_{t-1} + \varepsilon_t$ is the measurement equation) and the estimation of β_t and its standard error under the alternative is well understood; see the chapter in this Handbook by Hamilton. We therefore focus on the problem of testing the null that $\tau^2 = 0$.

The TVP model (5.15) nests, as a special case, the MA(1) model considered in Section 4. Setting $X_t = 1$ yields the unobserved components model (4.2), $y_t = \beta_0 + u_t$, where $u_t = (\beta_t - \beta_0) + \varepsilon_t = \sum_{s=1}^{t} v_s + \varepsilon_t$. Thus the testing problem in the general TVP model can be seen as an extension of the unit MA root testing problem. Starting with Nyblom and Mäkeläinen (1983), several authors have studied the properties of locally most powerful tests of $\tau^2 = 0$ against $\tau^2 > 0$ in (5.15) or in models where only some of the coefficients are assumed to evolve over time (that is, where G has reduced rank); see for example King and Hillier (1985), King (1988), Nyblom (1989), Nabeya and Tanaka (1988), Leybourne and McCabe (1989), Hansen (1990), Jandhyala and MacNeill (1992) and Andrews and Ploberger (1992). [Also see Watson and Engle (1985) who consider tests against β_t following a stationary AR(1).] The treatment here follows Nyblom and Mäkeläinen (1983) and builds on the discussion in Section 4 of tests of the UC model.

To derive the LMPI test of $\tau^2 = 0$ versus $\tau^2 > 0$, suppose that X_{t-1} is strictly exogenous (although the asymptotics hold more generally). Under the TVP model, (5.15) can be rewritten as $y_t = \beta_0' X_{t-1} + \{(\sum_{s=1}^{t} v_s)' X_{t-1} + \varepsilon_t\}$, where the term in curly brackets is an unobserved error. In standard matrix notation [Y denotes $(y_1, \ldots, y_T)'$, X denotes $(X_0, \ldots, X_{T-1})'$, etc.], the conditional distribution of Y, given X, is

$$Y \sim N\left\{X\beta_0, \sigma_\varepsilon^2\left[I_T + \left(\frac{\tau^2}{\sigma_\varepsilon^2}\right)\bar{V}_T\right]\right\}, \quad \text{where} \quad \bar{V}_T = \Omega^* \odot (XGX') \tag{5.16}$$

where $\Omega_{ij}^* = \min(i, j)$ and $\odot$ denotes the Hadamard (elementwise) product. The testing problem is invariant to scale/translation shifts of the form $y \to ay + b'X$, so the most powerful invariant test against an alternative $\bar{\tau}^2$ will be a ratio of quadratic forms involving $I_T + (\bar{\tau}^2/\sigma_\varepsilon^2)\bar{V}_T$. However, this depends on the alternative, so no uniformly most powerful test exists. One solution is to consider the LMPI test, which rejects for large values of $\hat{e}'\bar{V}_T\hat{e}/\hat{e}'\hat{e}$, where $\{\hat{e}_t\}$ are the full-sample OLS residuals. Straightforward algebra shows that $T^{-2}\hat{e}'\bar{V}_T\hat{e} = T^{-1}\sum_{s=1}^{T} S_T(s/T)'GS_T(s/T)$, where $S_T(\lambda) = T^{-1/2}\sum_{r=[T\lambda]+1}^{T}\hat{e}_t X_{t-1}$, which provides a simpler form for the test: reject if $T^{-1}\sum_{s=1}^{T} S_T(s/T)'GS_T(s/T)$ is large. Because this test and its limiting distribution depend on G, Nyblom (1989) suggested the simplification $G = (T^{-1}\sum_{t=1}^{T} X_{t-1}X_{t-1}')^{-1}$. Accordingly, the test rejects for large values of

$$L = T^{-1}\sum_{s=1}^{T} S_T\left(\frac{s}{T}\right)'\left(\hat{\sigma}_\varepsilon^2 T^{-1}\sum_{t=1}^{T} X_{t-1}X_{t-1}'\right)^{-1} S_T\left(\frac{s}{T}\right). \tag{5.17}$$

Conditional on $\{X_t\}$ the TVP model induces a heteroskedastic random walk into the error term which is detected by L using the cumulated product of the OLS residuals and the regressors.

Nyblom (1989) derived the statistic (5.17) by applying local arguments to a likelihood for generally nonlinear, nonnormal models, and his general statistic

simplifies to (5.17) in the Gaussian linear regression model. If $X_t = 1$, (5.17) reduces to the LMPI test (4.7) of the i.i.d. null against the random walk alternative, or equivalently the test of the null of a unit MA root.[31] Henceforth, we refer to (5.17) as the Nyblom statistic.

The asymptotics of the Nyblom statistic follow from the FCLT and the CMT. As usual, ε_t need not be i.i.d. normal and X_t need not be strictly exogenous; rather the weaker conditions following (5.1) are sufficient for the asymptotics. Under those weaker conditions, $\varepsilon_t X_{t-1}$ is a martingale difference sequence and, by the FCLT and CMT, $S_T(\cdot) \Rightarrow \sigma_\varepsilon \Sigma_X^{1/2} B_k^\mu(\cdot)$, where B_k^μ is a k-dimensional standard Brownian bridge. Because $T^{-1}\sum_{t=1}^T X_{t-1}X_{t-1}' \xrightarrow{p} \Sigma_X$ and $\hat{\sigma}_\varepsilon^2 \xrightarrow{p} \sigma_\varepsilon^2$, under the null hypothesis,

$$L \Rightarrow \int B_k^\mu(\lambda)' B_k^\mu(\lambda)\, d\lambda. \tag{5.18}$$

The literature contains Monte Carlo results on the finite-sample power of the tests in Sections 5.1.1–5.1.3 against various alternatives. The range of alternatives considered is broad and some preliminary conclusions have firmly emerged. Many of the tests overreject in moderately large samples ($T = 100$) when asymptotic critical values are used. This is exacerbated if errors are nonnormal and, especially, if autoregressions have large autoregressive parameters [QLR and related tests; see Diebold and Chen (1992)]. In their Monte Carlo study of the QLR test, exponentially averaged F-tests, the CUSUM test and several other tests against alternatives of one-time breaks and random walk coefficients, Andrews et al. (1992) found that in general the weighted exponential tests performed well and often the QLR and Nyblom tests performed nearly as well. For additional results, see Garbade (1977) and the references in Hackl and Westlund (1989, 1991).

5.1.4. *Empirical application: stability of the money–output relation*

At least since the work of Friedman and Mieselman (1963), one of the long-standing empirical problems in macroeconomics has been whether money has a strong and stable link to aggregate output; for a discussion and recent references, see Friedman and Kuttner (1992). Since their introduction to this literature by Sims (1972, 1980), Granger-causality tests and vector autoregressions have provided the workhorse machinery for quantifying the strength and direction of these relations in nonstructural time series models (see the chapter by Watson in this Handbook for a discussion of vector autoregressions). But for such empirical models to be useful guides for monetary policy they must be stable, and the tests of this section can play a useful role in assessing their stability. Of particular importance is whether one of the several monetary aggregates is arguably most stably related to output.

[31] To show this, rewrite $S_T(s/T)$ using the identity that the mean OLS residual is zero.

Table 7
Tests for structural breaks and time-varying parameters in the money–output relation
(Dependent variable: Nominal GDP growth)
(Estimation period: quarterly, 1960:2 to 1992:2).[a]

	M	r	$\bar{R}^2$	F-tests on coefficients on: M	r	QLR	P–K CUSUM	Nyblom L
1	Base	—	0.153	3.85**	—	40.23***	1.43**	2.83**
2	Base	R-90	0.178	1.43	2.50*	54.75***	1.36**	2.80
3	M1	—	0.140	3.17**	—	32.57***	1.22*	1.90
4	M1	R-90	0.179	1.50	2.87**	51.02***	1.35**	2.85
5	M2	—	0.221	7.60***	—	20.82	0.73	1.31
6	M2	R-90	0.332	8.40***	3.19**	25.44	1.05	1.53

[a]All regressions include 3 lags each of the nominal GDP growth rate, GDP inflation and the growth rate of the monetary aggregate. The M column specifies the monetary aggregate. The r column indicates whether the 90-day U.S. Treasury bill rate is included in the regression. If the interest rate is included, it is included in differences (3 lags) and one lag of an error-correction term from a long-run money demand equation is also included. The $\bar{R}^2$ is the usual OLS adjusted $\bar{R}^2$. The F-tests are Wald tests of the hypothesis that the coefficients on the indicated variable are zero; the restriction that the error-correction term (when present) has a zero coefficient is included in the Wald test on the monetary aggregate. QLR is the Quandt (1960) likelihood ratio statistic (5.4) with symmetric 15 percent trimming; P–K CUSUM is the Ploberger–Krämer (1992a) CUSUM statistic (5.14); and the final column reports the Nyblom (1989) L statistic (5.17). Break test critical values were taken from published tables and/or were computed by Monte Carlo simulation of the limiting functionals of Brownian motion, as described in Section 2.3. Tests are significant at the *10 percent; **5 percent; ***1 percent level.

Table 7 presents regression summary statistics and three tests for parameter stability in typical money–output regressions for three monetary aggregates, the monetary base, M1, and M2, over the period 1960:2–1992:2. The results are taken from Feldstein and Stock (1994), to which the reader is referred for additional detail. Based on preliminary unit root analysis, log GDP, the log GDP deflator, log money and the 90-day U.S. Treasury bill rate are specified as having a single unit root so that the GDP growth rate, GDP inflation, the money growth rate and the first difference of the interest rate are used in the regressions. Drawing on the cointegration evidence in Hoffman and Rasche's (1991) and Stock and Watson's (1993) studies of long-run money demand, in the models including the interest rate we model log money, log output and the interest rate as being cointegrated so that the equations include an error-correction term, the cointegrating residual. The long-run cointegrating equation was estimated by imposing a unit long-run income elasticity and estimating the interest semi-elasticity using the Saikkonen (1991)/Phillips–Loretan (1991)/Stock–Watson (1993) "dynamic OLS" efficient estimator.[32] The main

[32]All data were taken from the Citibase data base. The hypothesis of two unit roots was rejected at the 5 percent level for each series (demeaned case) using DF–GLS tests with AR(BIC), $1 \leq p \leq 8$ except that a unit root in inflation is rejected at the 10 percent but not 5 percent level. For each series, DF–GLS

conclusions are insensitive to empirically plausible changes in the unit root specifications of interest rates and money; in particular see Konishi et al. (1993) for F-statistics in specifications where the interest rate is assumed stationary.

The Granger-causality test results indicate that including the interest rate makes base money and M1 insignificant, although M2 remains significant (this is partly due to the error-correction term). The QLR test rejects the null hypothesis of parameter stability at the 1 percent level in all specifications including base money or M1; the L-statistic rejects in the base-only specification; and the Ploberger–Krämer (1992a) CUSUM based on OLS residuals rejects in the base and M1 specifications. The hypothesis of stability is thus strongly rejected for the base–output and M1–output relations. The evidence against stability is much weaker for the M2–output relation; none of the stability tests reject at the 10 percent level. Once changes in velocity are controlled for by including the error-correction term in regression 6, both M2 and the interest rate enter significantly and there is no evidence of instability. As with any empirical investigation, some caveats are necessary: these results are based on only a few specifications, and stability in this sample is no guarantee of stability in the future. Still, these results suggest that, of the base, M1, and M2, only M2 had a stable reduced-form relationship with output over this period.

5.2. *Trend breaks and tests for autoregressive unit roots*

5.2.1. *The trend-break model and effects of misspecifying the trend*

Rappoport and Reichlin (1989) and Perron (1989a, 1990b) argued that a plausible model for many economic variables is stationarity around a time trend with a break, and that autoregressive unit root tests based on linear detrending as discussed in Section 3 have low power against this alternative. Two such broken-trend

failed to reject a single unit root at the 10 percent level (detrending for each variable except interest rates, for which the demeaned statistics were used), except for the interest rate, which rejected at the 10 percent but not 5 percent level. The 95 percent asymptotic confidence intervals, computed as in Stock (1991) by inverting the Dickey–Fuller $\hat{t}^{\tau}$ statistic ($\hat{t}^{\mu}$ for interest rates) as described in Section 3.3, for the largest autoregressive roots are: log M1, (0.821, 1.026); log M2, (0.998, 1.039); log base, (0.603, 0.882); 90-day T-bill rate, (0.838, 1.015); log GDP, (0.950, 1.037); GDP inflation, (0.876, 1.032). The results are robust to using the AR(BIC) selector with $3 \leqslant p \leqslant 8$, as in the Monte Carlo simulations, except that the M2 confidence interval rises to (1.011, 1.040). For consistency, all monetary aggregates are specified in growth rates [but see Christiano and Ljungqvist (1988) and Stock and Watson (1989)]. These results leave room to argue that inflation should be entered in changes, but for comparability with other specifications in the literature inflation itself is used. There is some ambiguity about the treatment of interest rates, but to be consistent with recent investigations of long-run money demand they are treated here as I(1). The evidence on cointegration involves statistics not covered in this chapter and the reader is instead referred to Hoffman and Rasche (1991) and Stock and Watson (1993).

specifications are

$$\text{(Shift in mean)} \qquad d_t = \beta_0 + \beta_1 \mathbf{1}(t > r), \tag{5.19}$$

$$\text{(Shift in trend)} \qquad d_t = \beta_0 + \beta_1 t + \beta_2 (t - r)\mathbf{1}(t > r), \tag{5.20}$$

where $\mathbf{1}(t > r)$ is a dummy variable which equals one for $t > r$ and zero otherwise.[33] For conciseness, attention here is restricted to the trend-shift model (5.20), a model suggested by Perron (1989a) and Rappoport and Reichlin (1989) for real GNP.

It was emphasized in Section 3.2.5 that if the trend is misspecified then unit root tests can be misleading. This conclusion applies here as well. Suppose that (5.20) is correct and $r/T \to \delta, \delta$ fixed, $0 < \delta < 1$, but that statistics are computed by linear detrending. Then the power of the unit root test tends to zero against fixed alternatives. The intuition is simple: if a linear time trend is fitted to an I(0) process around a piecewise-linear trend, then the residuals will be I(0) around a mean-zero "V"-shaped trend. These residuals have variances growing large (with T) at the start and end of the sample and standard tests will classify the residuals as having a unit root.[34] In the mean-shift case, Dickey–Fuller unit root tests are consistent but have low power if the mean shift is large [Perron (1989a); for Monte Carlo evidence, Hendry and Neale (1991)]. See Campbell and Perron (1991) for further discussion. This troubling effect of trend misspecification raises the question of how to test for AR unit roots in the presence of a possibly broken trend.

5.2.2. *Unit root tests with broken trends*

If the break date is known a priori, as Perron (1989a, 1990b) and Rappoport and Reichlin (1989) assumed, then detrending can be done by correctly specified OLS, and the asymptotic distribution theory is obtained using a straightforward extension of Sections 2 and 3. However, as Christiano (1992) and, subsequently, Banerjee et al. (1992b), Perron and Vogelsang (1992) and Zivot and Andrews (1992) pointed out, the assumption that the break date is data-independent is hardly credible in macroeconomic applications. For example, in Perron's applications to

[33] Under the null hypothesis of a unit root, the mean-shift model is equivalent to assuming that there is a single additive outlier in v_t at time $r+1$, since, under the null hypothesis, (3.1) and (5.19) imply $\Delta y_t = v_t + \beta_1 \mathbf{1}(t = r+1)$. A third trend model is Perron's (1989a) "model C," with both a mean and a trend shift.

[34] To show this, consider the AR(1) case, so that $\gamma_v(0) = \omega^2$, and the Dickey–Fuller root test, $T(\hat{\alpha}^\tau - 1)$. If the trend is, in fact, given by (5.20), then the detrended process is $y_t^\tau = u_t + \hat{d}_t$, where $\hat{d}_t = (\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)t + \beta_2(t - r)\mathbf{1}(t > r)$. For $r/T \to \delta$, δ fixed, if u_t is I(0) then straightforward but tedious calculations show that the scaled detrended process has the deterministic limit, $T^{-1} y^\tau_{[T\lambda]} \xrightarrow{p} \tilde{\beta}_0 + \tilde{\beta}_1 \lambda + \tilde{\beta}_2(\lambda - \delta)\mathbf{1}(\lambda > \delta)$, uniformly in λ, where $\tilde{\beta}_1$ and $\tilde{\beta}_2$ are nonrandom functions of $\beta_0, \beta_1, \beta_2$ and δ. It follows that $T(\hat{\alpha}^\tau - 1) \Rightarrow g(\delta)$, where g is nonrandom. An explicit expression for $g(\delta)$ is given in Perron (1989a, Theorem 1(b)). Perron (1989a) shows that $g(\delta)$ is in the acceptance region of the detrended DF root test, so, asymptotically, the null is incorrectly accepted with probability one.

GNP, the break dates were chosen to be in the Great Depression and the 1973 oil price shock, both of which are widely recognized as having important and lasting effects on economic activity. Thus the problem becomes testing for a unit root when the break dates are unknown and determined from the data.

Two issues arise here: devising tests which control for the nuisance parameters, in particular the unknown break date, and, among such tests, finding the most powerful. To date research has focused on the first of these topics. There is little work which addresses this problem starting from the theory of optimal tests, and this is not pursued here.[35]

The procedures in the literature for handling the unknown date of the trend break are based on a modified Dickey–Fuller test. To simplify the argument, consider the AR(1) case, so that $\omega^2 = \gamma_v(0)$. Then, as suggested by Christiano (1992), Banerjee et al. (1992b) and Zivot and Andrews (1992), one could test for a unit root by examining the minimum of the sequence of Dickey–Fuller t-statistics, constructed by first detrending the series by OLS using (5.20) for r over the range $r_0, \ldots, r_1$,

$$t_{\mathrm{DF}}^{\min} = \min_{\delta \in [\delta_0, \delta_1]} \hat{\tau}^d(\delta), \tag{5.21}$$

where

$$\hat{\tau}^d(\delta) = \frac{T^{-1} \sum_{t=2}^{T} \Delta y_t^d(\delta) y_{t-1}^d(\delta)}{\left\{ (\hat{\sigma}^d)^2(\delta) T^{-2} \sum_{t=2}^{T} (y_{t-1}^d(\delta))^2 \right\}^{1/2}},$$

where $y_t^d(\delta) = y_t - z_t(\delta)'\hat{\beta}(\delta)$, where $\hat{\beta}(\delta) = [\sum_{t=1}^{T} z_t(\delta) z_t(\delta)']^{-1} [\sum_{t=1}^{T} z_t(\delta) y_t]$, $z_t(\delta) = [1, t, (t - [T\delta])\mathbf{1}(t > [T\delta])]$ and $(\hat{\sigma}^d)^2(\lambda)$ is the sample variance of the residual from the regression of $y_t^d(\delta)$ onto $y_{t-1}^d(\delta)$.

Just as the null distribution of the QLR statistic differs from the distribution of the fixed-date F-statistics, the null distribution of $t_{\mathrm{DF}}^{\min}$ differs from the distribution of $\hat{\tau}^d$ for a fixed break point. The approach to obtaining the null distribution is similar; namely to obtain a limiting representation for the sequence of statistics $\hat{\tau}^d(\delta)$, uniformly in δ. Relative to the QLR statistic, this entails an additional complication, because under the null the broken-trend detrended process will be I(1). This leads to limit results for elements of $D[0, 1] \times D[0, 1]$. While no new tools are needed for these calculations, they are tedious and notationally cumbersome and the reader is referred to the articles by Banerjee et al. (1992b) and Zivot and Andrews (1992) for different derivations of the same limiting representation. Not surprisingly, the critical values of the minimal DF statistic are well below the critical values of

[35] Elliott et al. (1992) show that the asymptotic Gaussian power envelope in the mean-shift model (5.19) with β_1 fixed equals the no-detrending power envelope plotted in Figure 1.

the usual linearly detrended statistic; for example, with symmetric 15 percent trimming, the one-sided 10 percent asymptotic critical value is approximately -4.13 [Banerjee et al. (1992b, Table 2)], compared with -3.12 in the linearly detrended case.

5.2.3. *Finite-sample size and power*

There are fewer Monte Carlo studies of the broken-trend and broken-mean unit root statistics than of the linearly detrended case, perhaps in part because the additional minimization dramatically increases the computational demands. Nonetheless, the results of Hendry and Neale (1991), Perron and Vogelsang (1992), Zivot and Andrews (1992) and Banerjee et al. (1992b) provide insights into the performance of the tests. The finite-sample distributions are sensitive to the procedures used to determine the lag length in the augmented DF regression, and the null distributions depend on the nuisance parameters even though the tests are asymptotically similar. Typically, the asymptotic critical values are too small, that is, the sizes of the tests exceed their nominal level. The extent of the distortion depends on the actual values of the nuisance parameters. Zivot and Andrews (1992) examined size distortions by Monte Carlo study of ARIMA models estimated using the Nelson–Plosser (1982) U.S. data set; for the mean-shift model (5.19), the finite-sample 10 percent critical values were found to fall in the range -4.85 to -5.05, while the corresponding asymptotic value is -4.58; for each series, tests of asymptotic level 2.5 percent rejected between 5 percent and 10 percent of the time. Perron and Vogelsang (1992) found larger rejection rates under the null when there is more negative serial correlation than present in the Zivot–Andrews simulations.

The Monte Carlo evidence confirms the view that the finite-sample power of the unit root tests is reduced by trend- or mean-shift detrending, in the sense that if the true trend is linear then introducing the additional break-point reduces power. The extent of this power reduction, however, depends on the nuisance parameters and, in any event, if the broken-trend specification is correct then broken-trend detrending is necessary. The more relevant comparison is across different procedures which entail broken-trend detrending, but only limited results are available [see Perron and Vogelsang (1992) for some conclusions comparing four Dickey–Fuller-type tests with different lag length selection procedures].

5.2.4. *Conclusions and practical implications*

Although the research on trend-break unit root tests is incomplete, it is possible to draw some initial conclusions. On a practical level, the size distortions found in the demeaned and linear detrended cases in Section 3 appear, if anything, to be more severe in the broken-trend case, and the power of the tests also deteriorates. One can speculate that this reflects a dwindling division between the I(1) model and other competing representations; were the trend-shifts I(0) and occurring every period, then the extension of (5.20) would deliver an I(2) model for y_t.

A useful way to summarize the broken-trends literature is to return to our original four motivating objectives for analyzing unit roots. As a matter of data description, Perron's (1989a, 1990b) and Rappoport and Reichlin's (1989) analyses demonstrate that the broken-trend models deliver very different interpretations from conventional unit root models, emphasizing the importance of a few irregularly occurring events in determining the long-run path of aggregate variables; this warrants continued research in this area. The practical implications concerning the remaining three objectives remain largely unexplored. From a forecasting perspective, if the single-break model is taken as a metaphor for multiple irregular breaks, then one must be skeptical that out-of-sample forecasts will be particularly reliable, since another break could occur. Equally importantly, for this reason treating the break as a one-time nonrandom event presumably leads to understating the uncertainty of multistep forecasts. Little is currently known about the practical effect of misspecifying trend breaks in subsequent multivariate modeling, although the asymptotic theory of inference in vector autoregressions (VAR) with unit roots and cointegration analysis discussed in Watson's chapter in this Handbook must be modified if there are broken trends. Finally, the link between these trend-break models and economic theory is undeveloped. In any event, the statistical difficulties with inference in this area does not make one optimistic that trend-break models will parse economic theories, however capable they are of producing suggestive stylized facts.

6. Tests of the I(1) and I(0) hypotheses: links and practical limitations

Sections 3, 4, and 5.2 focused on inference in the I(1) and I(0) models. When inference is needed about the order of integration of a series, sometimes there is no compelling a priori reason to think that one or other of these models is the best starting point; rather, the models might best be treated symmetrically. In this light, this section addresses three topics. Section 6.1 examines some formal links between the I(1) and I(0) models. Section 6.2 summarizes some recent work taking a different approach to these issues, in which the determination of whether a series is I(0) or I(1) is recast as a classification problem, so that the tools of Bayesian analysis and statistical decision theory can be applied. Section 6.3 then raises several practical difficulties which arise in the interpretation of both these Bayesian classification schemes and classical unit-root hypothesis tests in light of the size distortions coupled with low power of the tests studied in the Monte Carlo experiments of Sections 3 and 4.

6.1. *Parallels between the I(0) and I(1) testing problems*

The historical development of tests of the I(0) and I(1) hypotheses treated the issues as conceptually and technically quite different. To a large extent, these differences

are artificial, arising from their ARMA parameterizations. Since an integrated I(0) process is I(1), a test of the I(0) null against the I(1) alternative is, up to the handling of initial conditions, equivalent to a test of the I(1) null against the I(2) alternative. In this sense, the tests of the previous sections can both be seen as tests of the I(1) null, on the one hand, against I(0) and, on the other hand, against I(2). What is interesting is that this reinterpretation is valid not just on a heuristic level but also on a technical level.

To make this precise, consider the case $d_t = \beta_0$, $v_t = \varepsilon_t$. The LMPIU test of the unit MA root in (4.1) rejects for large values of the Nyblom–Mäkeläinen (1983) statistic

$$L^{\mu} = \frac{T^{-2}\sum_{t=1}^{T}\left(\sum_{s=1}^{t} y_s^{\mu}\right)^2}{T^{-1}\sum_{t=1}^{T}(y_t^{\mu})^2}, \tag{6.1}$$

where $y_s^{\mu} = y_s - \bar{y}$. If instead the null hypothesis is that u_t is a Gaussian random walk and the alternative is that u_t is an AR(1) with $|\alpha| < 1$, then one could test this hypothesis by rejecting for small values of the demeaned Sargan–Bhargava statistic

$$\tilde{R}_T^{\mu} = \frac{T^{-2}\sum_{t=1}^{T}(y_t^{\mu})^2}{T^{-1}\sum_{t=2}^{T}(\Delta y_t^{\mu})^2}. \tag{6.2}$$

The L^{μ} statistic rejects if the mean square of the I(1) process, the cumulation of y_t^{μ}, is large, while $\tilde{R}_T^{\mu}$ rejects if the mean square of the I(1) process, y_t^{μ}, is small. Both tests can be seen as tests of the I(1) null but, respectively, against the I(2) and I(0) alternatives.

6.2. *Decision-theoretic classification schemes*

A standard argument for using conventional hypothesis tests is that the researcher has a particular reason for wishing to control the Type I error rate. While this might be appropriate in some of the applications listed in Section 1, in others, such as forecasting, the ultimate objective of the empirical analysis is different and classical hypothesis tests are not necessarily the best tools to achieve those objectives. In such cases, the researcher might rather be interested in having a procedure which will deliver consistent inference, in the sense that the probability of correctly classifying a process as I(1) or I(0) asymptotically tends to one; that is, the probabilities of both Type I and Type II errors tend to zero.

In theory, this can be achieved by using a sequence of critical values which tend to $-\infty$ as an appropriate function of the sample size. To be concrete, suppose that the researcher computed the Dickey–Fuller $\hat{\tau}$ statistic, and evaluated it using critical values, b_T. If the null is true, then $\Pr[\hat{\tau} < b_T | \mathrm{I}(1)] \to 0$ for any sequence $b_T \to -\infty$, so that the probability of correctly concluding that the process is I(1) tends to one. Similarly, it is plausible that for a suitable choice of b_T, if the process is truly I(0) then $\Pr[\hat{\tau} < b_T | \mathrm{I}(0)] \to 1$ and the Type II error rate tends to zero. For such a choice of b_T, this would be a consistent classification scheme. Because the Dickey–Fuller t-statistic tends to $-\infty$ at the rate $T^{1/2}$ under a fixed alternative, one candidate for b_T is $b_T = -k_0 - k_1 \ln T$ for some positive constants (k_0, k_1). Thus the rule is

$$\text{Classify } y_t \text{ as I(0) if } \quad \hat{\tau} < -k_0 - k_1 \ln T \tag{6.3}$$

and, otherwise, classify y_t as I(1). The problem with this scheme is that, in practice, the researcher is left to choose k_0 and k_1. Because the sample size is, of course, fixed in an actual data set, the conceptual device of choosing this sequence is artificial and the researcher is left with little practical guidance.

One solution is to frame this as a classification or decision-theoretic problem and to apply Bayesian techniques. In this context, an observed series is classified as I(0) or I(1) based on the posterior odds ratio Π_T, which we write heuristically as

$$\Pi_T = \left(\frac{\pi_1}{\pi_0}\right) B_T, \quad \text{where} \quad B_T = \frac{\Pr[(y_1, \ldots, y_T) | \mathrm{I}(1)]}{\Pr[(y_1, \ldots, y_T) | \mathrm{I}(0)]}, \tag{6.4}$$

where π_1 and π_0 are prior weights that the series is I(1) and I(0) and where B_T is the Bayes ratio. If $\Pi_T > 1$, then the posterior odds favor the I(1) model and the series is classified as I(1).

Although (6.4) appears simple, numerous subtleties are involved in its evaluation and addressing these subtleties has spawned a large literature on Bayesian approaches to autoregressive unit roots; see in particular Sims (1988), Schotman and van Dijk (1990), Sims and Uhlig (1991), DeJong and Whiteman (1991a, 1991b), Diebold (1990), Sowell (1991) and the papers by Phillips (1991a) and his discussants in the special issue of the *Journal of Applied Econometrics* (October–December, 1991). In most cases, implementations of (6.4) have worked within specifications which require placing explicit priors over key continuous parameters, such as the largest autoregressive root. The proposed priors differ considerably and can imply substantial differences in empirical inferences [see the review by Uhlig (1992)]. Because of this dependence on priors, and given space limitations, no attempt will be made here to summarize this literature. Instead, we briefly discuss two recent approaches, by Phillips and Ploberger (1991) and Stock (1992), which provide simple ways to evaluate the posterior odds ratio (6.4) and which avoid explicit integration over priors on continuous parameters. These procedures require only

that the researcher place priors π_0 and $\pi_1 = 1 - \pi_0$ on the respective point hypotheses "I(0)" and "I(1)".

Phillips and Ploberger (1991) derive their procedure from a consideration of the likelihood ratio statistic in the AR(1) model, and obtain the rule

$$\text{Classify } y_t \text{ as I(0) if } \quad \hat{\tau}^2 > \ln\left(\frac{\pi_1}{\pi_0}\right) + \ln\left(\frac{\sum_{t=2}^{T} y_{t-1}^2}{\hat{\sigma}_\varepsilon^2}\right) \tag{6.5}$$

and, otherwise, classify y_t as I(1), where $\hat{\tau}$ is the Dickey–Fuller t-statistic.[36] The expression (6.5) bears considerable similarity to (6.3): a unit AR root is rejected based on the Dickey–Fuller t-statistic, with a critical value that depends on the sample size. The difference here is that the critical value is data-dependent; if y_t is I(1), the "critical value" will be $2 \ln T + O_p(1)$, while if y_t is I(0), it will be $\ln T + O_p(1)$. As Phillips and Ploberger (1992) point out, this procedure can be viewed as an extension to the I(1) case of the BIC model selection procedure, where the issue is whether to include or to exclude y_{t-1} as a regressor in the DF regression (3.9). The procedure is also closely related to the predictive least squares principle, see Wei (1992).

Another approach is to evaluate the Bayes factor in (6.4) directly, using a reduced-dimensional statistic rather than the full data set. Suppose that ϕ_T is a statistic which is informative about the order of integration, such as a unit root test statistic; then the expression for the Bayes factor in (6.4) could be replaced with

$$B_T^* = \frac{\Pr(\phi_T | \text{I(1)})}{\Pr(\phi_T | \text{I(0)})}. \tag{6.6}$$

The approach in Stock (1992) is to construct a family of statistics which have limiting distributions which, on the one hand, do not depend on nuisance parameters under either the I(1) or I(0) hypothesis but, on the other hand, diverge, depending on which hypothesis is true. The results in the previous sections can, in fact, be used to construct such statistics. Consider the process V_T^d defined in (4.10), and consider the no-deterministic case. If y_t is a general I(0) process then $V_T \Rightarrow W$. On the other hand, if y_t is a general I(1) process then $N_T^{-1/2} V_T \Rightarrow V^*$, where V^* is defined in (4.11) and $N_T = T / \sum_{m=-l_T}^{l_T} k(m/l_T)$. In either case the limiting representation of V_T does not depend on any nuisance parameters. To make this concrete, consider the statistic $\phi_T = \ln L = \ln\{T^{-1} \sum_{t=1}^{T} V_{t-1}^2\}$. Then, for v_t a general I(0) process, from (2.9) (in the I(0) case) and (4.11) (in the I(1) case), ϕ_T has the limiting representations

$$\text{if I(0)} \quad \phi_T \Rightarrow \ln\left(\int W^2\right), \tag{6.7a}$$

[36] Phillips and Ploberger's (1991) formula has been modified for an estimated variance as in Phillips (1992b).

$$\text{if I(1)} \quad \phi_T - \ln N_T \Rightarrow \ln\left(\int V^{*2}\right). \tag{6.7b}$$

The limiting distributions under the I(0) and I(1) models can be computed numerically from (6.7a) and (6.7b), respectively, which in turn permits the numerical evaluation of the Bayes factor (6.6) based on this statistic.

It must be stressed that, although consistent decision-theoretic procedures such as these have both theoretical and intuitive appeal, they have properties which empirical researchers might find undesirable. One is that these procedures will consistently classify local-to-I(1) processes as I(1) rather than I(0), and local-to-I(0) processes as I(0) rather than as I(1). That is, if y_t is local-to-I(1) with local parameter $\alpha = 1 + c/T$, then, as the sample size increases, this process will be classified as I(1) with probability increasing to one, even though along the sequence it is always an I(0) process [see Elliott and Stock (1994) for details]. More generally, because these procedures can have large misclassification rates in finite samples (loosely, their size can be quite large), care must be taken in interpreting the results.

Initial empirical applications [Phillips (1992b)] and Monte Carlo simulations [Elliott and Stock (1994), Stock (1992)] suggest that, for some applications such as forecasting and pretesting, these approaches are promising. To date, however, the investigation of the sampling properties of these and alternative procedures, and in particular the effect of their use in second-stage procedures, is incomplete. It would be premature to make concrete recommendations for empirical practice.

6.3. *Practical and theoretical limitations in the ability to distinguish I(0) and I(1) processes*

6.3.1. *Theory*

The evidence on tests of the I(1) null yields two troubling conclusions. On the one hand, the tests have relatively low power against I(0) alternatives that might be of interest; for example, with 100 observations in the detrended case, the local-to-unity asymptotics indicate that the 5 percent one-sided MPI test has a power of 0.27 against $\alpha = 0.9$ and that the Dickey–Fuller t-test has a power of only 0.19. On the other hand, processes which are I(1) but which have moderate negative autocorrelation in first differences are incorrectly rejected with high probability, that is, the unit AR root tests exhibit substantial size distortions, although the extent of these distortions varies widely across test statistics. The same general conclusions were found for tests of the general I(0) null: the power against interesting alternatives can be low and, depending on the choice of spectral estimator, the rejection rate for null values that have substantial positive autocorrelation can be well above the asymptotic level.

A natural question is how one should interpret these finite-sample size distortions. In this regard, it is useful to develop some results concerning the source of these

size distortions and whether they will persist in large samples. Section 4 examined the behavior of tests of the I(0) null in the event that y_t was I(1) but local-to-I(0) in the sense (4.12), and found that the I(0) tests with functional representations had nondegenerate asymptotic power functions against these alternatives. It is natural to wonder, then, what is the behavior of tests of the I(1) null, if the true process is I(1) but is local to I(0)?

As a starting point, consider the $d_t = 0$ case and the sequence of models

$$u_t = bT\zeta_t + \sum_{s=1}^{t} \eta_s, \quad (\zeta_t, \eta_t) \text{ i.i.d. } N(0, \sigma^2 I). \tag{6.8}$$

This is just the local-to-I(0) model (4.12), rescaled by multiplication by $h^{-1}T$ where $b = h^{-1}$, with $(\Delta u_{1t}, u_{0t})$ Gaussian and mutually and serially uncorrelated.

If in fact $b = 0$, then u_t is a Gaussian random walk, so one might consider using the Sargan–Bhargava statistic $\tilde{R}_T$. A direct calculation indicates that, for $b > 0$, under this nesting $T\tilde{R}_T \xrightarrow{p} \frac{1}{2}$. It follows that $\Pr[\tilde{R}_T < k] = \Pr[T\tilde{R}_T < Tk] \to 1$ for any constant k, so that the rejection probability of the test tends to one. The implication is that (unmodified) Sargan–Bhargava tests of sequences which are local to I(0) in the sense (6.8) will incorrectly reject with asymptotic probability 1. The implication of this result is perhaps clearer when u_t is cast in its ARIMA form, $\Delta u_t = (1 - \theta_T L)\varepsilon_t$. For finite T, $|\theta_T| < 1$, but the limiting result indicates that the rejection probability approaches one and so can be quite large for finite T.

A similar set of calculations can be made for tests of the I(0) null. Here, the relevant sequence of null models to consider are those models against which the AR unit root tests have nondegenerate local asymptotic power, namely the local-to-unity models studied in Section 3.2.3. Again let $d_t = 0$ and suppose that the I(0) null is tested using the L^μ statistic. A straightforward calculation shows that

$$T^{-1}L^\mu \Rightarrow \frac{\left\{\int_{s=0}^{1}\left(\int_{r=0}^{s} W_c^\mu(r)\,dr\right)^2 ds\right\}}{\int (W_c^\mu)^2},$$

where W_c^μ is the demeaned local-to-unity diffusion process defined in Section 3.2.3. It follows that $\Pr[L^\mu > k] = \Pr[T^{-1}L^\mu > T^{-1}k] \to 1$ for any constant k, so that the rejection probability of the test tends to one. For these processes, which are local to I(1), the L^μ test rejects with probability approaching one even though, for fixed T, u_t has the AR(1) representation $(1 - \rho_T L)u_t = \varepsilon_t$ with $|\rho_T| < 1$.

These results elucidate the Monte Carlo findings in Sections 3 and 4. In the AR case, the implication is that there are I(1) models which are local to I(0) for which the I(1) null will be incorrectly rejected with high probability. In the MA case, there are I(0) models which are local to I(1) for which the I(0) null will be incorrectly

rejected with high probability. Thus the high false rejection rates (the size distortions) found in the Monte Carlo analysis can be expected to persist asymptotically, but the range of models suffering size distortions will decrease.

The foregoing analysis is limited both because of the tests considered and because it does not address the question of the size of the neighborhoods of these incorrect rejections; it was shown only that the neighborhoods are at least $O(T^{-1})$. In the case of I(1) tests, Pantula (1991) has provided results on the sizes of these neighborhoods for several tests of the unit AR root hypothesis in the MA(1) model. He found that these neighborhoods vanish at different rates for different tests, with the slowest rate being the Phillips–Perron (1988) statistic. This finding explains the particularly large size distortions of this statistic with negative MA roots, even with very large samples [e.g. $T = 500$; Pantula (1991, Table 2)]. In related work, Perron (1991d) and Nabeya and Perron (1991) provide approximations to the distribution of the OLS root estimator with sequences of negative MA and negative AR roots approaching their respective boundaries.

Because tests for the general I(0) null have only recently been developed, as of this writing there have been few empirical analyses in which both I(0) and I(1) tests are used [exceptions include Fisher and Park (1991) and Ogaki (1992)]. The foregoing theoretical results suggest, however, that there will be a range of models for which the I(1) test will reject with high probability and the I(0) test will not, although the process is I(1); for which the I(0) test will reject and the I(1) test will not, although the process is I(0); and for which both tests will reject and the process is I(1). It also seems plausible that there are models that are truly I(0) but for which both tests reject with high probability, but this has not been investigated formally. There is currently little evidence on the volume of these regions of contradictory results, although Amano and van Norden's (1992) Monte Carlo evidence suggests that they may well be large in moderate sample sizes.

In summary, tests of the general I(0) null and tests of the general I(1) null are neither similar nor unbiased. Asymptotically, the tests have size equal to their stated level for fixed null models; but problems arise when we consider sequences of null and alternative models for which the I(0) and I(1) models become increasingly close. On the one hand, there are null models which will be rejected with arbitrarily high probability; on the other hand, there are alternative models against which the tests will have power approaching the nominal level. Although these regions diminish asymptotically, in finite samples this implies that there is a range of I(0) and I(1) models amongst which the unit MA and AR root tests are unable to distinguish.

6.3.2. Practical implications

The asymptotic inability to distinguish certain I(0) and I(1) models raises the question of how these tests are to be interpreted, and this has generated great controversy in the applied literature on the practical value of unit root tests. Some

of the earliest criticisms of Nelson and Plosser's (1982) stylized fact that many economic time series contain a unit root came from Bayesian analyses [Sims (1988), DeJong and Whiteman (1991a, 1991b); see the references in Section 6.2 following equation (6.4)], although the discussion here follows the debate from a classical perspective in Blough (1992), Christiano and Eichenbaum (1990), Cochrane (1991a), Rudebusch (1992, 1993) and Stock (1990). In particular, the reader is referred to Campbell and Perron's (1991) thoughtful discussion and the comments on it by Cochrane (1991b). There has, however, been little systematic research on the practical implications of this problem, so one's view of the importance of this lack of unbiasedness remains largely a matter of judgment. Because of the prominence of this issue it nonetheless seems appropriate to organize the ways in which such judgment can be exercised. This discussion focuses exclusively on AR unit root tests, although several of the remarks have parallels to MA unit root tests.

It is useful to return to the reasons, listed in the introduction, why one might wish to perform inference concerning orders of integration: as data description; for forecasting; as pretesting for subsequent specification analysis or testing; or for testing or distinguishing among economic theories. Although this discussion proceeds in general terms, it must be emphasized that the size and power problems vary greatly across test statistics, so that the difficulties discussed here are worse for some tests than others.

Data description. The size distortions and low power of even the best-performing tests imply that the literal interpretation of unit AR root tests as similar and unbiased tests of the I(1) null against the I(0) alternative is inappropriate. However, the Monte Carlo evidence provides considerable guidance in the interpretation of unit root test results. For some tests, such as the Dickey–Fuller t-statistic and the DF–GLS statistic, the size is well controlled over a wide range of null models so rejection can be associated rather closely with the absence of a unit root. In contrast, the severe size distortions of the Phillips–Perron tests [or other tests with the SC spectral estimator, such as the Schmidt–Phillips (1992) MSB statistic] in the presence of moderate negative MA roots and their low empirical rejection rates in the stationary case with moderate positive MA or second AR roots indicates that rejection by this statistic is only secondarily associated with the presence or absence of a unit root, and instead is indicative of the extent of positive serial correlation in the process. Interpretation of results based on extant versions of these statistics using SC estimators is, thus, problematic. In any event, confidence intervals for measures of long-run persistence are arguably more informative than unit root tests themselves; constructing these confidence intervals entails testing a range of possible values of α, not just the unit root hypothesis.

An important caveat is that the unit root tests and thus confidence intervals require that the trend order be correctly specified; depending on the type of misspecification, the tests might otherwise be inconsistent. We agree with Campbell and Perron's (1991) emphasis on the importance of properly specifying the trend order

before proceeding with the classical tests, and this is an area in which one should bring economic theory to bear to the maximum extent possible. For example, a priori reasoning might suggest using a constant or shift-in-mean specification in modeling interest rates, rather than including a linear time trend. We speculate that while one could develop a consistent downward-testing procedure, starting with the highest possible trend order and letting the test level decline with the sample size, such an approach would have high misclassification rates in moderate samples (size distortions and low power). The Bayesian approach in Phillips and Ploberger (1992) to joint selection of the trend and order of integration is theoretically appealing for fixed models but the finite-sample performance of this approach has not yet been fully investigated.

Forecasting. Campbell and Perron (1991) and Cochrane (1991b) examined the effect of unit root pretests on forecasting performance. In their Monte Carlo experiment, data were generated by an ARMA(1, 1) and forecasts were made using an autoregression. Their most striking finding was that, in models with a unit AR root and large negative correlation in first differences, the out-of-sample forecast error was substantially lower with the unit root pretest than if the true differences specification was used. This finding appeared both at short and long horizons (1 and 20 periods with a sample of size 100). In cases with less severe negative correlation or with a stationary process, little was lost by pretesting relative to using the true model. Because economic forecasting is largely done using multivariate models, these initial results do not bear directly on most professional forecasting activities. Still, they suggest that for forecasting the size distortions might be an advantage, not a problem. A promising alternative to pretesting is to forecast using median-unbiased estimates of α as discussed in Section 3.3. To date, however, there has been no thorough examination of whether this delivers finite-sample improvements in forecasts and forecast standard errors.

Pretests for second-stage inference. Perhaps the most common use of unit root tests is as pretests for second-stage inference: as a preliminary stage for developing a forecasting model, for formulating a cointegrated system, or for determining subsequent distribution theory. In the final of these applications, the existing distribution theory for inference in linear time series regressions conditions upon the number and location of unit roots in the system, in the sense that the orders of integration and cointegration are assumed known. In empirical work, these orders are typically unknown, so one way to proceed is to pretest for integration or cointegration and then to condition on the results of these pretests in performing second-stage inference. In practice, this can mean using a unit root pretest to decide whether a variable should enter a second-stage regression in levels or differences, as was done in the empirical application in Section 5.1.4. Alternatively, if the relationship of interest involves the level of the variable in a second-stage regression, a unit

root pretest could be used to ascertain whether standard or nonstandard distribution theory should be used to compute second-stage tests.

There has, however, been little research on the implications of this use of unit root tests. Some evidence addressing this is provided by Elliott and Stock (1994), who consider a bivariate problem in which there is uncertainty about whether the regressor has a unit root. In their Monte Carlo simulation, they find that unit root pretests can induce substantial size distortions in the second-stage test. If the innovations of the regressor and the second-stage regression error are correlated, the first-stage Dickey–Fuller t-statistic and the second-stage t-statistic will be dependent so the size of the second stage in this two-stage procedure cannot be controlled effectively, even asymptotically. Although this problem is important when this error correlation is high, in applications with more modest correlations the problem is less severe.

Formulating and testing economic theories. This is arguably the application most damaged by the problems of poor size and low power. In special cases – the martingale theories of consumption and stock prices being the leading examples – simple theories predict that the series is not only I(1) but is a martingale. In this case, the null models are circumscribed and the problems of size distortions do not arise. However, the initial appeal of unit root tests to economists was that they seemed to provide a way to distinguish between broad classes of models: on the one hand, dynamic stochastic equilibrium models (real business cycle models) in which fluctuations were optimal adjustments to supply shocks, on the other hand, Keynesian models in which fluctuations arose in large part from demand disturbances. Indeed, this was the original context in which they were interpreted in Nelson and Plosser's (1982) seminal paper.

Unfortunately, there are two problems, either of which alone is fatal to such an interpretation. The first is a matter of economic theory: as argued by Christiano and Eichenbaum (1990), stochastic equilibrium models need not generate unit roots in observed output, and as argued by West (1988b), Keynesian models can generate autoregressive roots that are very close to unity. Thus a rejection by an ideal unit root test (that is, one with no size distortions) need not invalidate a real business cycle model and a failure to reject should not be interpreted as evidence against Keynesian models. The second is the lack of unbiasedness outlined above: even if the match between classes of macroeconomic theories and whether macroeconomic series are I(1) were exact, the size distortions and low power would mean that the outcomes of unit root tests would not discriminate among theories. In this light the idea, however appealing, that a univariate unit root test could distinguish which class of models best describes the macroeconomy seems in retrospect overly ambitious.

This said, inference about the order of integration of a time series can usefully guide the specification and empirical analysis of relations of theoretical interest in economics. For example, King and Watson (1992) and Fisher and Seater (1993) use

these techniques to provide evidence on which versions of money neutrality (long-run neutrality, superneutrality) can be investigated empirically. They show that long-run neutrality can be tested without specifying a complete model of short-run dynamics, as long as money and income are I(1). Similarly, investigations into whether there are unit roots in exchange rates have proven central to inferences about such matters as long-run purchasing power parity and the behavior of exchange rates in the presence of target zones [see Johnson (1993) and Svensson (1992) for reviews]. Finally, quantitative conclusions about the persistence in univariate series have proven to be a key starting point for modeling the long-run properties of multiple time series and cointegration analysis, an area which has seen an explosion of exciting empirical and theoretical research and is the topic of the next chapter in this Handbook.

References

Ahtola, J. and G.C. Tiao (1984) "Parameter Inference for a Nearly Nonstationary First Order Autoregressive Model", *Biometrika*, 71, 263–72.

Amano, R.A. and S. van Norden (1992) Unit Root Tests and the Burden of Proof, manuscript, Bank of Canada.

Anderson, T.W. (1948) "On the Theory of Testing Serial Correlation", *Skandinavisk Aktuariedskrift*, 31, 88–116.

Anderson, T.W. (1959) "On Asymptotic Distributions of Estimates of Parameters of Stochastic Difference Equations", *Annals of Mathematical Statistics*, 30, 676–687.

Anderson, T.W. and D. Darling (1952) "Asymptotic Theory of Certain 'Goodness of Fit' Criteria Based on Stochastic Processes", *Annals of Mathematical Statistics*, 23, 193–212.

Anderson, T.W. and A. Takemura (1986) "Why Do Noninvertible Moving Averages Occur?", *Journal of Time Series Analysis*, 7, 235–254.

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

Andrews, D.W.K. (1993a) "Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Model", *Econometrica*, 61, 139–166.

Andrews, D.W.K. (1993b) "Tests for Parameter Instability and Structural Change with Unknown Change Point", *Econometrica*, 821–856.

Andrews, D.W.K. and H.-Y. Chen (1992) "Approximately Median-Unbiased Estimation of Autoregressive Models with Applications to U.S. Macroeconomic and Financial Time Series", Cowles Foundation Discussion Paper no. 1026, Yale University.

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

Andrews, D.W.K. and W. Ploberger (1992) "Optimal Tests When a Nuisance Parameter is Identified Only Under the Alternative", Cowles Foundation Discussion Paper no. 1015, Yale University.

Andrews, D.W.K., I. Lee and W. Ploberger (1992) "Optimal Changepoint Tests for Normal Linear Regression", Cowles Foundation Discussion Paper no. 1016, Yale University.

Ansley, C.F. and P. Newbold (1980) "Finite Sample Properties of Estimators for Autoregressive Moving Average Models", *Journal of Econometrics*, 13, 159–183.

Arnold, L. (1973) *Stochastic Differential Equations: Theory and Applications*. Wiley:New York.

Bai, J. (1992) Econometric Estimation of Structural Change, unpublished Ph.D. Dissertation, Department of Economics, University of California, Berkeley.

Banerjee, A., J. Dolado, J.W. Galbraith and D.F. Hendry (1992a) *Cointegration, Error Correction and the Econometric Analysis of Non-Stationary Data*. Oxford:Oxford University Press.

Banerjee, A., R.L. Lumsdaine and J.H. Stock (1992b) "Recursive and Sequential Tests of the Unit Root

and Trend Break Hypotheses: Theory and International Evidence", *Journal of Business and Economic Statistics*, 10, 271–288.
Barry, D. and J.A. Hartigan (1993) "A Bayesian Analysis for Change Point Problems", *Journal of the American Statistical Association*, 88, 309–319.
Beach, C.M. and J.G. MacKinnon (1978) "A Maximum Likelihood Procedure for Regression with Autocorrelated Errors", *Econometrica*, 46, 51–58.
Beaulieu, J.J. and J.A. Miron (1993) "Seasonal Unit Roots in Aggregate U.S. Data", *Journal of Econometrics*, 55, 305–328.
Beran, J. (1992) "Statistical Models for Data with Long-Range Dependence (with discussion)", *Statistical Science*, 4, 404–427.
Berk, K.N. (1974) "Consistent Autoregressive Spectral Estimates", *Annals of Statistics*, 2, 489–502.
Beveridge, S. and C.R. Nelson (1981) "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle' ", *Journal of Monetary Economics*, 7, 151–174.
Bhargava, A. (1986) "On the Theory of Testing for Unit Roots in Observed Time Series", *Review of Economic Studies*, 53, 369–384.
Bierens, H.J. (1993) "Higher-Order Sample Autocorrelations and the Unit Root Hypothesis", *Journal of Econometrics*, 57, 137–160.
Bierens, H.J. and S. Guo (1993) "Testing Stationarity and Trend Stationarity Against the Unit Root Hypothesis", *Econometric Reviews*, 12, 1–32.
Billingsley, P. (1968) *Convergence of Probability Measure*. Wiley: New York.
Blanchard, O.J. and D. Quah (1989) "The Dynamic Effects of Aggregate Demand and Supply Disturbances", *American Economic Review*, 79, 655–73.
Blough, S.R. (1992) "The Relationship Between Power and Level for Generic Unit Root Tests in Finite Samples", *Journal of Applied Econometrics*, 7, 295–308.
Bobkoski, M.J. (1983) Hypothesis Testing in Nonstationary Time Series, unpublished Ph.D. thesis, Department of Statistics, University of Wisconsin.
Bollerslev, T. (1986) "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, 31, 307–327.
Box, G.E.P. and G.M. Jenkins (1976) *Time Series Analysis: Forecasting and Control, Revised Edition*. San Francisco: Holden–Day.
Brillinger, D.R. (1981) *Time Series: Data Analysis and Theory*. San Francisco: Holden–Day.
Brockwell, P.J. and R.A. Davis (1987) *Time Series: Theory and Methods*. New York: Springer–Verlag.
Brown, B.M. (1971) "Martingale Central Limit Theorems", *Annals of Mathematical Statistics*, 42, 59–66.
Brown, R.L., J. Durbin and J.M. Evans (1975) "Techniques for Testing the Constancy of Regression Relationships over Time with Comments", *Journal of the Royal Statistical Society*, Series B, 37, 149–192.
Campbell, J.Y. (1987) "Does Saving Anticipate Declining Labor Income? An Alternative Test of the Permanent Income Hypothesis", *Econometrica*, 55, 1249–1274.
Campbell, J.Y. and N.G. Mankiw (1987a) "Are Output Fluctuations Transitory", *Quarterly Journal of Economics*, 102, 857–880.
Campbell, J.Y. and N.G. Mankiw (1987b) "Permanent and Transitory Components in Macroeconomic Fluctuations", *The American Economic Review*, 77, 111–117.
Campbell, J.Y. and P. Perron (1991) "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots", *NBER Macroeconomics Annual*, 141–200.
Cavanagh, C (1985) Roots Local To Unity, manuscript, Department of Economics, Harvard University.
Chan, N.H. (1988) "On The Parameter Inference for Nearly Nonstationary Time Series", *Journal of the American Statistical Association*, 83, 857–862.
Chan, N.H. (1989) "Asymptotic Inference for Unstable Autoregressive Time Series With Drifts", *Journal of Statistical Planning and Inference*, 23, 301–312.
Chan, N.H. and C.Z. Wei (1987) "Asymptotic Inference For Nearly Nonstationary AR(1) Processes", *Annals of Statistics*, 15, 1050–1063.
Chan, N.H. and C.Z. Wei (1988) "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes", *Annals of Statistics*, 16, 367–401.
Chernoff, H. and S. Zacks (1964) "Estimating the Current Mean of a Normal Distribution Which Is Subject to Changes in Time", *Annals of Mathematical Statistics*, 35, 999–1028.
Choi, I. (1993) "Effects of Data Aggregation on the Power of Tests for a Unit Root: A Simulation Study", *Economic Letters*, forthcoming.

Chow, G.C. (1960) "Tests of Equality Between Sets of Coefficients in Two Linear Regressions", *Econometrica*, 28, 591–605.
Chow, G.C. (1984) "Random and Changing Coefficient Models", in: Z. Griliches and M. Intriligator, eds., *Handbook of Econometrics*. Vol. 2, North-Holland: Amsterdam.
Christiano, L.J. (1992) "Searching for a Break in GNP", *Journal of Business and Economic Statistics*, 10, 237–250.
Christiano, L.C. and M. Eichenbaum (1990) "Unit Roots in GNP: Do We Know and Do We Care?" *Carnegie-Rochester Conference Series on Public Policy*, 32, 7–62.
Christiano, L.J. and L. Ljungqvist (1988) "Money Does Granger-Cause Output in the Bivariate Money-Output Relation", *Journal of Monetary Economics*, 22, 217–236.
Chu, C.-S., and H. White (1992) "A Direct Test for Changing Trend", *Journal of Business and Economic Statistics*, 10, 289–299.
Clark, P.K. (1987) "The Cyclical Component of U.S. Economic Activity", *The Quarterly Journal of Economics*, 102, 798–814.
Clark, P.K. (1989) "Trend Reversion in Real Output and Unemployment", *Journal of Econometrics*, 40, 15–32.
Cochrane, J.H. (1988) "How Big is the Random Walk in GNP?", *Journal of Political Economy*, 96, 893–920.
Cochrane, J.H. (1991a) "A Critique of the Application of Unit Root Tests", *Journal of Economic Dynamics and Control*, 15, 275–284.
Cochrane, J.H. (1991b) "Comment on Campbell and Perron", *NBER Macroeconomics Annual 1991*, 5, 201–210.
Cooley, T.F. and E.C. Prescott (1976) "Estimation in the Presence of Stochastic Parameter Variation", *Econometrica*, 44, 167–184.
Cooper, D.M. and R. Thompson (1977) "A Note on the Estimation of the Parameters of the Autoregressive-Moving Average Process", *Biometrika*, 64, 625–628.
Cryer, J.D. and J. Ledolter (1981) "Small-Sample Properties of the Maximum Likelihood Estimator in the First-Order Moving Average Model", *Biometrika*, 68, 691–694.
Dahlhaus, R. (1989) "Efficient Parameter Estimation for Self-Similar Processes", *Annals of Statistics*, 17, 1749–1766.
Davidson, J.E.H. (1979) "Small Sample Properties of Estimators of the Moving Average Process", in: E.G. Charatsis, ed., *Proceedings of the Econometric Society Meeting in 1979: Selected Econometric* Papers in Memory of Stephan Valananis. North-Holland: Amsterdam.
Davidson, J.E.H. (1981) "Problems With the Estimation of Moving Average Processes", *Journal of Econometrics*, 16, 295–310.
Davidson, J.E.H., D.F. Hendry, F. Srba and S. Yeo (1978) "Econometric Modelling of the Aggregate Time-Series Relationship Between Consumer's Expenditure and Income in the United Kingdom", *Economic Journal*, 86, 661–692.
Davies, R.B. (1977) "Hypothesis Testing When a Nuisance Parameter is Present only Under the Alternative", *Biometrika*, 64, 247–254.
Davis, R.A. and W.T.M. Dunsmuir (1992) Inference for MA(1) Processes with a Root On or Near the Unit Circle, Discussion Paper no. 36, Bond University School of Business.
DeJong, D.N. and C.H. Whiteman (1991a) "Reconsidering Trends and Random Walks in Macroeconomic Time Series", *Journal of Monetary Economics*, 28, 221–254.
DeJong, D.N. and C.H. Whiteman (1991b) "The Temporal Stability of Dividends and Stock Prices: Evidence from the Likelihood Function", *American Economic Review*, 81, 600–617.
DeJong, D.N., J.C. Nankervis, N.E. Savin and C.H. Whiteman (1992a) "Integration versus Trend-Stationarity in Macroeconomic Time Series", *Econometrica*, 60, 423–434.
DeJong, D.N., J.C. Nankervis, N.E. Savin and C.H. Whiteman (1992b) "The Power Problems of Unit Root Tests for Time Series with Autoregressive Errors", *Journal of Econometrics*, 53, 323–43.
Dent, W. and A.-S. Min (1978) "A Monte Carlo Study of Autoregressive Integrated Moving Average Processes", *Journal of Econometrics*, 7, 23–55.
Deshayes, J. and D. Picard (1986) "Off-line Statistical Analysis of Change-Point Models Using Non Parametric and Likelihood Methods", in: M. Basseville and A. Benveniste, eds., *Detection of Abrupt Changes in Signals and Dynamical Systems*. Springer Verlag Lecture Notes in Control and Information Sciences, no. 77, 103–168.
Dickey, D.A. (1976) Estimation and Hypothesis Testing in Nonstationary Time Series, Ph.D. dissertation, Iowa State University.

Dickey, D.A. and W.A. Fuller (1979) "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association*, 74, 427–431.
Dickey, D.A. and W.A. Fuller (1981) "Likelihood Ratio Tests for Autoregressive Time Series with a Unit Root", *Econometrica*, 49, 1057–1072.
Dickey, D.A. and S.G. Pantula (1987) "Determining the Order of Differencing in Autoregressive Processes", *Journal of Business and Economic Statistics*, 5, 455–462.
Dickey, D.A., D.P. Hasza, and W.A. Fuller (1984) "Testing for Unit Roots in Seasonal Time Series", *Journal of the American Statistical Association*, 79, 355–367.
Diebold, F.X. (1990) "On 'Unit Root Econometrics': Discussion of Geweke, Sims, and DeJong and Whiteman", in: *Proceedings of the American Statistical Association Business and Economics Statistics Session*. Washington, D.C.: American Statistical Association.
Diebold, F.X. (1993) "The Effect of Seasonal Adjustment Filters on Tests for a Unit Root: Discussion", *Journal of Econometrics*, 55, 99–103.
Diebold, F.X. and C. Chen (1992) Testing Structural Stability with Endogenous Break Point: A Size Comparison of Analytic and Bootstrap Procedures, manuscript, Department of Economics, University of Pennsylvania.
Diebold, F.X. and M. Nerlove (1990) "Unit Roots in Economic Time Series: A Selective Survey", in: T.B. Fomby and G.F. Rhodes, eds., *Advances in Econometrics: Co-Integration, Spurious Regressions, and Unit Roots*. Greenwich, CT: JAI Press.
Diebold, F.X. and G.D. Rudebusch (1989) "Long Memory and Persistence in Aggregate Output", *Journal of Monetary Economics*, 24, 189–209.
Diebold, F.X. and G.D. Rudebusch (1991a) "Is Consumption Too Smooth? Long Memory and the Deaton Paradox", *Review of Economics and Statistics*, 71, 1–9.
Diebold, F.X. and G.D. Rudebusch (1991b) "On the Power of Dickey–Fuller Tests Against Fractional Alternatives", *Economics Letters*, 35, 155–160.
Diebold, F.X., S. Husted and M. Rush (1991) "Real Exchange Rates Under the Gold Standard", *Journal of Political Economy*, 99, 1252–1271.
Dufour, J.-M. (1990) "Exact Tests and Confidence Sets in Linear Regressions with Autocorrelated Errors", *Econometrica*, 58, 475–494.
Dufour, J.-M. and M.L. King (1991) "Optimal Invariant Tests for the Autocorrelation Coefficient in Linear Regressions with Stationary or Nonstationary AR(1) Errors", *Journal of Econometrics*, 47, 115–143.
Dunsmuir, W.T.M. (1981) "Estimation for Stationary Time Series When Data Are Irregularly Spaced or Missing", in: D.F. Findley, ed., *Applied Time Series II*, Academic Press: New York.
Durlauf, S.N. (1989) "Output Persistence, Economic Structure and the Choice of Stabilization Policy", *Brookings Papers on Economic Activity*, 2, 69–136.
Durlauf, S.N. and P.C.B. Phillips (1988) "Trends Versus Random Walks in Time Series Analysis", *Econometrica*, 56, 1333–1354.
Elliott, G. (1993a) Efficient Tests for a Unit Root When the Initial Observation Is Drawn From Its Unconditional Distribution, manuscript, Department of Economics, Harvard University.
Elliott, G. (1993b) Monte Carlo Performance of Twenty Unit Root Tests, manuscript, Department of Economics, Harvard University.
Elliott, G. and J.H. Stock (1994) "Inference in Time Series Regression When the Order of Integration of a Regressor is Unknown", *Econometric Theory*, forthcoming.
Elliott, G., T.J. Rothenberg and J.H. Stock (1992) Efficient Tests for an Autoregressive Unit Root, NBER Technical Working Paper no. 130.
Engle, R.F. and C.W.J. Granger (1987) "Cointegration and Error Correction: Representation, Estimation and Testing", *Econometrica*, 55, 251–276.
Ethier, S.N. and T.G. Kurtz (1986) *Markov Processes: Characterization and Convergence*. New York: Wiley.
Evans, G.B.A. and N.E. Savin (1981a) "The Calculation of the Limiting Distribution of the Least Squares Estimator of the Parameter in a Random Walk Model", *Annals of Statistics*, 9, 1114–1118.
Evans, G.B.A. and N.E. Savin (1981b) "Testing for Unit Roots: I", *Econometrica*, 49, 753–779.
Evans, G.B.A. and N.E. Savin (1984) "Testing for Unit Roots: II", *Econometrica*, 52, 1241–1260.
Fama, E.F. (1970) "Efficient Capital Markets: A Review of Theory and Empirical Work", *Journal of Finance*, 25, 383–417.
Feldstein, M. and J.H. Stock (1994) "The Use of a Monetary Aggregate to Target Nominal GDP", in: N.G. Mankiw, ed., *Monetary Policy*. Chicago: University of Chicago Press.

Fisher, E.O. and J.Y. Park (1991) "Testing Purchasing Power Parity Under the Null Hypothesis of Cointegration", *Economic Journal*, 101, 1476–84.
Fisher, M.E. and J.J. Seater (1993) "Long-Run Neutrality and Superneutrality in an ARIMA Framework", *American Economic Review*, 83, 402–415.
Fox, R. and M.S. Taqqu (1986) "Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series", *Annals of Statistics*, 14, 517–532.
Franzini, L. and A.C. Harvey (1983) "Testing for deterministic trend and seasonal components in time series models", *Biometrika*, 70, 673–682.
Friedman, B.M. and K.N. Kuttner (1992) "Money, Income, Prices and Interest Rates", *American Economic Review*, 82, 472–492.
Friedman, M. and D. Meiselman (1963) "The Relative Stability of the Investment Multiplier and Monetary Velocity in the United States, 1897–1958", in: *Stabilization Policies*. Commission on Money and Credit, Englewood Cliffs, NJ: Prentice–Hall, 165–268.
Fuller, W.A. (1976) *Introduction to Statistical Time Series*. New York:Wiley.
Garbade, K. (1977) "Two Methods of Examining the Stability of Regression Coefficients", *Journal of The American Statistical Association*, 72, 54–63.
Gardner, L.A. (1969) "On Detecting Changes in the Mean of Normal Variates", *Annals of Mathematical Statistics*, 40, 116–126.
Geweke, J. and S. Porter-Hudak (1983) "The Estimation and Application of Long Memory Time Series Models", *Journal of Time Series Analysis*, 4, 221–238.
Ghysels, E. (1990) "Unit-Root Tests and the Statistical Pitfalls of Seasonal Adjustment: The Case of U.S. Postwar Real Gross National Product", *Journal of Business and Economic Statistics*, 8, 145–152.
Ghysels, E. and P. Perron (1993) "The Effect of Seasonal Adjustment Filters on Tests for a Unit Root", *Journal of Econometrics*, 55, 57–98.
Hackl, P. and A. Westlund (1989) "Statistical Analysis of 'Structural Change': An Annotated Bibliography", *Empirical Economics*, 143, 167–192.
Hackl, P. and A.H. Westlund (1991) eds., *Economic Structural Change: Analysis and Forecasting*. Springer–Verlag: Berlin, 95–119.
Hall, A. (1989) "Testing for a Unit Root in the Presence of Moving Average Errors", *Biometrika*, 76, 49–56.
Hall, A. (1992a) "Testing for a Unit Root in Time Series Using Instrumental Variable Estimation with Pretest Data-Based Model Selection", *Journal of Econometrics*, 54, 223–250.
Hall, A. (1992b) "Unit Root Tests After Data Based Model Selection", *Proceedings of the 1992 Summer Meetings of the American Statistical Association: Business and Economics Section*.
Hall, P. and C.C. Heyde (1980) *Martingale Limit Theory and its Applications*. New York: Academic Press.
Hall, R.E. (1978) "Stochastic Implications of the Life Cycle–Permanent Income Hypothesis: Theory and Evidence", *Journal of Political Economy*, 86, 971–87.
Hansen, B.E. (1990) Lagrange Multiplier Tests for Parameter Instability in Non-Linear Models, manuscript, Economics Department, University of Rochester.
Hansen, B.E. (1992) "Tests for Parameter Instability in Regressions with I(1) Processes", *Journal of Business and Economic Statistics*, 10, 321–336.
Harvey, A.C. (1981) *Time Series Models*. Oxford: Philip Allan.
Harvey, A.C. (1985) "Trends and Cycles in Macroeconomic Time Series", *Journal of Business and Economic Statistics*, 3, 216–227.
Harvey, A.C. (1989) *Forecasting, Structural Models and the Kalman Filter*. Cambridge, U.K.: Cambridge University Press.
Harvey, A.C. and A. Jaeger (1993) "Detrending, Stylized Facts and the Business Cycle", *Journal of Applied Econometrics*, 8, 231–248.
Harvey, A.C. and N.G. Shephard (1992) "Structural Time Series Models", in: G.S. Maddala, C.R. Rao and H.D. Vinod, eds., *Handbook of Statistics*, Vol. 2, Elsevier: Amsterdam.
Hasza, D.P. and W.A. Fuller (1979) "Estimation for Autoregressive Processes with Unit Roots", *Annals of Statistics*, 7, 1106–1120.
Hasza, D.P. and W.A. Fuller (1981) "Testing for Nonstationary Parameter Specifications in Seasonal Time Series Models", *Annals of Statistics*, 10, 1209–1216.
Hendry, D.F. and A.J. Neale (1991) "A Monte Carlo Study of the Effects of Structural Breaks on Tests for Unit Roots", in. P. Hackl and A.H. Westlund, eds., *Economic Structural Change: Analysis and Forecasting*. Springer–Verlag: Berlin, 95–119.

Herrndorf, N.A. (1984) "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables", *Annals of Probability*, 12, 141–153.
Hoffman, D. and R.H. Rasche (1991) "Long-Run Income and Interest Elasticities of Money Demand in the United States", *Review of Economics and Statistics*, 73, 665–674.
Hurwicz, L. (1950) "Least-Squares Bias in Time Series", in: T. Koopmans, ed., *Statistical Inference in Dynamic Economic Models*. New York: Wiley.
Hylleberg, S., R.F. Engle, C.W.J. Granger and S. Yoo (1990) "Seasonal Integration and Cointegration", *Journal of Econometrics*, 44, 215–238.
Jandhyala, V.K. and I.B. MacNeill (1992) "On Testing for the Constancy of Regression Coefficients Under Random Walk and Change-Point Alternatives", *Econometric Theory*, 8, 501–517.
Jegganathan, P. (1991) "On the Asymptotic Behavior of Least-Squares Estimators in AR Time Series with Roots Near the Unit Circle", *Econometric Theory*, 7, 269–306.
Johnson, D. (1993) "Unit Roots, Cointegration and Purchasing Power Parity: Canda and the United States, 1870–1991", in: B. O'Reilly and J. Murray, eds., *The Exchange Rate and the Economy*. Ottawa: Bank of Canada.
Kang, K.M. (1975) A Comparison of Estimators of Moving Average Processes, manuscript, Australian Bureau of Statistics.
Kendall, M.G. and A. Stuart (1967) *The Advanced Theory of Statistics*. Vol. 2, Second edition. Charles Griffin: London.
Kim, H.-J. and D. Siegmund (1989) "The Likelihood Ratio Test for a Change-Point in Simple Linear Regression", *Biometrika*, 76(3), 409–23.
King, M.L. (1980) "Robust Tests for Spherical Symmetry and their Application to Least Squares Regression", *Annals of Statistics*, 8, 1265–1271.
King, M.L. (1988) "Towards a Theory of Point Optimal Testing", *Econometric Reviews*, 6, 169–218.
King, M.L. and G.H. Hillier (1985) "Locally Best Invariant Tests of the Error Covariance Matrix of the Linear Regression Model", *Journal of the Royal Statistical Society*, Series B, 47, 98–102.
King, R.G. and M.W. Watson (1992) Testing Long-Run Neutrality, NBER Working Paper no. 4156.
King, R.G., C.I. Plosser, J.H. Stock and M.W. Watson (1991) "Stochastic Trends and Economic Fluctuations", *American Economic Review*, 81, 819–840.
Kiviet, J.F. and G.D.A. Phillips (1992) "Exact Similar Tests for Unit Roots and Cointegration", *Oxford Bulletin of Economics and Statistics*, 54, 349–368.
Konishi, T., V.A. Ramey and C.W.J. Granger (1993) Stochastic Trends and Short-Run Relationships Between Financial Variables and Real Activity, NBER Working Paper no. 4275.
Koopmans, T. (1942) "Serial Correlation and Quadratic Forms in Normal Variables", *Annals of Mathematical Statistics*, 13, 14–33.
Krämer, W. and H. Sonnberger (1986) *The Linear Regression Model Under Test*. Physica-Verlag: Heidelberg.
Krämer, W., W. Ploberger and R. Alt (1988) "Testing for Structural Change in Dynamic Models", *Econometrica*, 56, 1355–1369.
Krishnaiah, P.R. and B.Q. Miao (1988) "Review about Estimation of Change Points", in: P.R. Krishnaiah and C.R. Rao, eds., *Handbook of Statistics*. 7, New York: Elsevier.
Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992) "Testing the Null Hypothesis of Stationarity Against the Alternatives of a Unit Root: How Sure Are We that Economic Time Series Have a Unit Root?", *Journal of Econometrics*, 54, 159–178.
Lamotte, L.R. and A. McWhorter (1978) "An Exact Test for the Presence of Random Walk Coefficients in a Linear Regression Model", *Journal of the American Statistical Association*, 73, 816–820.
Lehmann, E.L. (1959) *Testing Statistical Hypotheses*. John Wiley and Sons: New York.
Lewis, R. and G.C. Reinsel (1985) "Prediction of Multivariate Time Series by Autoregressive Model Fitting", *Journal of Multivariate Analysis*, 16, 393–411.
Leybourne, S.J. and B.P.M. McCabe (1989) "On the Distribution of Some Test Statistics for Coefficient Constancy", *Biometrika*, 76, 169–177.
Lo, A.W. (1991) "Long-Term Memory in Stock Market Prices", *Econometrica*, 59, 1279–1314.
MacNeill, I.B. (1974) "Tests for Change of Parameter at Unknown Times and Distributions of Some Related Functionals on Brownian Motion", *Annals of Statistics*, 2, 950–962.
MacNeill, I.B. (1978) "Properties of Sequences of Partial Sums of Polynomial Regression Residuals with Applications to Tests for Change of Regression at Unknown Times", *Annals of Statistics*, 6, 422–433.
Magnus, J.R. and B. Pesaran (1991) "The Bias of Forecasts from a First-Order Autoregression", *Econometric Theory*, 7, 222–235.

Mandelbrot, B.B. (1975) "Limit Theorems on the Self-normalized Range for Weakly and Strongly Dependent Processes", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 31, 271–285.
Mandelbrot, B.B. and M.S. Taqqu (1979) Robust R/S Analysis of Long Run Serial Correlation, Invited paper, 42nd Session of the International Institute of Statistics.
Mandelbrot, B.B. and J.W. Van Ness (1968) "Fractional Brownian Motions, Fractional Noise and Applications", *SIAM Review*, 10, 422–437.
Mann, H.B. and A. Wald (1943) "On the Statistical Treatment of Linear Stochastic Difference Equations", *Econometrica*, 11, 173–220.
Marriot, F.H.C. and J.A. Pope (1954) "Bias in the Estimation of Autocorrelations", *Biometrika*, 41, 393–403.
McCabe, B.P.M. and M.J. Harrison (1980) "Testing the Constancy of Regression Relationships over Time Using Least Squares Residuals", *Applied Statistics*, 29, 142–148.
Meese, R.A. and K.J. Singleton (1982) "On Unit Roots and the Empirical Modeling of Exchange Rates", *Journal of Finance*, 37, 1029–1035.
Nabeya, S. and P. Perron (1991) Local Asymptotic Distributions Related to the AR(1) Model with Dependent Errors, Econometric Research Program Memorandum No. 362, Princeton University.
Nabeya, S. and B.E. Sorensen (1992) 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, manuscript, Department of Economics, Brown University.
Nabeya, S. and K. Tanaka (1988) "Asymptotic Theory of a Test for the Constancy of Regression Coefficients Against the Random Walk Alternative", *Annals of Statistics*, 16, 218–235.
Nabeya, S. and K. Tanaka (1990a) "A General Approach to the Limiting Distribution for Estimators in Time Series Regression with Nonstable Autoregressive Errors", *Econometrica*, 58, 145–163.
Nabeya, S. and K. Tanaka (1990b) "Limiting Powers of Unit Root Tests in Time Series Regression", *Journal of Econometrics*, 46, 247–271.
Nankervis, J.C. and N.E. Savin (1988) "The Exact Moments of the Least-Squares Estimator for the Autoregressive Model: Corrections and Extensions", *Journal of Econometrics*, 37, 381–388.
Nelson, C.R. and H. Kang (1981) "Spurious Periodicity in Inappropriately Detrended Time Series", *Econometrica*, 49, 741–751.
Nelson, C.R. and C.I. Plosser (1982) "Trends and Random Walks in Macro-economic Time Series: Some Evidence and Implications", *Journal of Monetary Economics*, 10, 139–162.
Newey, W. and K.D. West (1987) "A Simple Positive Semi Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, 55, 703–708.
Ng, S. and P. Perron (1993a) Useful Modifications to Some Unit Root Tests with Dependent Errors and their Local Asymptotic Properties, manuscript, C.R.D.E., University of Montreal, Quebec.
Ng, S. and P. Perron (1993b) Unit Root Tests in ARMA Models With Data-Dependent Methods for the Selection of the Truncation Lag, manuscript, C.R.D.E., University of Montreal, Quebec.
Nicholls, S. and A. Pagan (1985) "Varying Coefficient Regression", in: *Handbook of Statistics*, 413–449.
Nyblom, J. (1986) "Testing for Deterministic Linear Trend in Time Series", *Journal of the American Statistical Association*, 81, 545–549.
Nyblom, J. (1989) "Testing for the Constancy of Parameters Over Time", *Journal of the American Statistical Association*, 84, 223–30.
Nyblom, J. and T. Mäkeläinen (1983) "Comparisons of Tests for the Presence of Random Walk Coefficients in a Simple Linear Model", *Journal of the American Statistical Association*, 78, 856–864.
Ogaki, M. (1992) "Engel's Law and Cointegration", *Journal of Political Economy*, 100, 1027–1046.
Orcutt, G.H. (1948) "A Study of the Autoregressive Nature of the Time Series Used for Tinbergen's Model of the Economic System of the United States, 1919–1932", *Journal of the Royal Statistical Society B*, 10, 1–45.
Pantula, S.G. (1989) "Testing for Unit Roots in Time Series Data", *Econometric Theory*, 5, 256–271.
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.
Pantula, S.G. and A. Hall (1991) "Testing for Unit Roots in Autoregressive Moving Average Models: An Instrumental Variable Approach", *Journal of Econometrics*, 48, 325–353.
Pantula, S.G., G. Gonzalez-Farias and W.A. Fuller (1992) A Comparison of Unit Root Test Criteria, manuscript, North Carolina State University.
Park, J. (1990) "Testing for Unit Roots and Cointegration by Variable Addition", in: T.B. Fomby and G.F. Rhodes, eds., *Advances in Econometrics: Co-Integration, Spurious Regressions and Unit Roots*. Greenwich, CT: JAI Press.

Park, J. and B. Choi (1988) A New Approach to Testing for a Unit Root, Working Paper no. 88–23, Center for Analytical Economics, Cornell University.
Park, J.Y. and P.C.B. Phillips (1988) "Statistical Inference in Regressions with Integrated Processes: Part I", *Econometric Theory*, 4, 468–497.
Perron, P. (1988) "Trends and Random Walks in Macroeconomic Time Series: Further Evidence from a New Approach", *Journal of Economic Dynamics and Control*, 12, 297–332.
Perron, P. (1989a) "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis", *Econometrica*, 57, 1361–1401.
Perron, P. (1989b) "The Calculation of the Limiting Distribution of the Least-Squares Estimator in a Near-Integrated Model", *Econometric Theory*, 5, 241–255.
Perron, P. (1989c) "Testing for a Random Walk: A Simulation Experiment of Power When the Sampling Interval is Varied", in: B. Raj, ed., *Advances in Econometrics and Modelling*. Kluwer Academic Publishers: Boston.
Perron, P. (1990a) "Tests of Joint Hypothesis for Time Series Regression with a Unit Root", in: T.B. Fomby and G.F. Rhodes, eds., *Advances in Econometrics: Co-Integration, Spurious Regressions and Unit Roots*. Greenwich, CT: JAI Press.
Perron, P. (1990b) "Testing for a Unit Root in a Time Series with a Changing Mean", *Journal of Business and Economic Statistics*, 8, 153–162.
Perron, P. (1991a) "A Continuous Time Approximation to the Stationary First-Order Autoregressive Model", *Econometric Theory*, 7, 236–252.
Perron, P. (1991b) "Test Consistency with Varying Sampling Frequency", *Econometric Theory*, 7, 341–368.
Perron, P. (1991c) A Test for Changes in a Polynomial Trend Function for a Dynamic Time Series, manuscript, Princeton University.
Perron, P. (1991d) The Adequacy of Asymptotic Approximations in the Near-Integrated Autoregressive Model with Dependent Errors, manuscript, Princeton University.
Perron, P. (1992) "A Continuous Time Approximation to the Unstable First Order Autoregressive Process: The Case Without an Intercept", *Econometrica*, 59, 211–236.
Perron, P. and P.C.B. Phillips (1987) "Does GNP Have a Unit Root? A Reevaluation", *Economics Letters*, 23, 139–145.
Perron, P. and T.S. Vogelsang (1992) "Nonstationarity and Level Shifts With an Application to Purchasing Power Parity", *Journal of Business and Economic Statistics*, 10, 301–320.
Pesaran, M.H. (1983) "A Note on the Maximum Likelihood Estimation of Regression Models with First-Order Moving Average Errors with Roots on the Unit Circle", *Australian Journal of Statistics*, 25, 442–448.
Phillips, P.C.B. (1978) "Edgeworth and Saddlepoint Approximations in the First-Order Noncircular Autoregression", *Biometrika*, 65, 91–98.
Phillips, P.C.B. (1986) "Understanding Spurious Regression in Econometrics", *Journal of Econometrics*, 33, 311–340.
Phillips, P.C.B. (1987a) "Time Series Regression with a Unit Root", *Econometrica*, 55, 277–302.
Phillips, P.C.B. (1987b) "Towards a Unified Asymptotic Theory for Autoregression", *Biometrika*, 74, 535–547.
Phillips, P.C.B. (1988) "Multiple Regression with Integrated Time Series", *Contemporary Mathematics*, 80, 79–105.
Phillips, P.C.B. (1991a) "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends", *Journal of Applied Econometrics*, 6, 333–364.
Phillips, P.C.B. (1991b) "Spectral Regression for Cointegrated Time Series", in: W. Barnett, J. Powell and G. Tauchen, eds., *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge University Press: Cambridge, UK.
Phillips, P.C.B. (1992a) "Unit Roots", in: P. Newman, M. Milgate and J. Eatwell, eds., *The New Palgrave Dictionary of Money and Finance*. London: Macmillan, 726–730.
Phillips, P.C.B. (1992b) Bayes Methods for Trending Multiple Time Series with An Empirical Application to the U.S. Economy, manuscript, Cowles Foundation, Yale University.
Phillips, P.C.B. and M. Loretan (1991) "Estimating Long Run Economic Equilibria", *Review of Economic Studies*, 58, 407–436.
Phillips, P.C.B. and S. Ouliaris (1990) "Asymptotic Properties of Residual Based Tests for Cointegration", *Econometrica*, 58, 165–94.

Phillips, P.C.B. and P. Perron (1988) "Testing for a Unit Roots in a Time Series Regression", *Biometrika*, 75, 335–346.
Phillips, P.C.B. and W. Ploberger (1991) Time Series Modeling with a Bayesian Frame of Reference: I. Concepts and Illustrations, manuscript, Cowles Foundation, Yale University.
Phillips, P.C.B. and W. Ploberger (1992) Posterior Odds Testing for a Unit Root with Data-Based Model Selection, manuscript, Cowles Foundation for Economic Research, Yale University.
Phillips P.C.B. and V. Solo (1992) "Asymptotics for Linear Processes", *Annals of Statistics*, 20, 971–1001.
Picard, D. (1985) "Testing and Estimating Change-Points in Time Series", *Advances in Applied Probability*, 176, 841–867.
Ploberger, W. and W. Krämer (1990) "The Local Power of the CUSUM and CUSUM of Squares Tests", *Econometric Theory*, 6, 335–347.
Plobeger, W. and W. Krämer (1992a) "The CUSUM test with OLS Residuals", *Econometrica*, 60, 271–286.
Ploberger, W. and W. Krämer (1992b) A Trend-Resistant Test for Structural Change Based on OLS-Residuals, manuscript, Department of Statistics, University of Dortmund.
Ploberger, W., W. Krämer and K. Kontrus (1989) "A New Test for Structural Stability in the Linear Regression Model", *Journal of Econometrics*, 40, 307–318.
Plosser, C.I. and C.W. Schwert (1977) "Estimation of a Noninvertible Moving-Average Process: The Case of Overdifferencing", *Journal of Econometrics*, 6, 199–224.
Pötscher, B.M. (1991) "Noninvertibility and Pseudo-Maximum Likelihood Estimation of Misspecified ARMA Models", *Econometric Theory*, 7, 435–449.
Quah, D. (1992) "The Relative Importance of Permanent and Transitory Components: Identification and some Theoretical Bounds", *Econometrica*, 60, 107–118.
Quandt, R.E. (1960) "Tests of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes", *Journal of the American Statistical Association*, 55, 324–330.
Rappoport, P. and L. Reichlin (1989) "Segmented Trends and Non-Stationary Time Series", *Economic Journal*, 99, 168–177.
Robinson, P.M. (1993) "Nonparametric Function Estimation for Long-Memory Time Series", in: C. Sims, ed., *Advances in Econometrics, Sixth World Congress of the Econometric Society*. Cambridge: Cambridge University Press.
Rothenberg, T.J. (1990) Inference in Autoregressive Models with Near-Unit Root, seminar notes, Department of Economics, University of California, Berkeley.
Rudebusch, G. (1992) "Trends and Random Walks in Macroeconomic Time Series: A Reexamination", *International Economic Review*, 33, 661–680.
Rudebusch, G.D. (1993) "The Uncertain Unit Root in Real GNP", *American Economic Review*, 83, 264–272.
Said, S.E. and D.A. Dickey (1984) "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order", *Biometrika*, 71, 599–608.
Said, S.E. and D.A. Dickey (1985) "Hypothesis Testing in ARIMA $(p, 1, q)$ Models", *Journal of the American Statistical Association*, 80, 369–374.
Saikkonen, P. (1991) "Asymptotically Efficient Estimation of Cointegrating Regressions", *Econometric Theory*, 7, 1–21.
Saikkonen, P. (1993) "A Note on a Lagrange Multiplier Test for Testing an Autoregressive Unit Root", *Econometric Theory*, 9, 343–362.
Saikkonen, P. and R. Luukkonen (1993a) "Testing for Moving Average Unit Root in Autoregressive Integrated Moving Average Models", *Journal of the American Statistical Association*, 88, 596–601.
Saikkonen, P. and R. Luukkonen (1993b) "Point Optimal Tests for Testing the Order of Differencing in ARIMA Models", *Econometric Theory*, 9, 343–362.
Sargan, J.D. and A. Bhargava (1983a) "Testing Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk", *Econometrica*, 51, 153–174.
Sargan, J.D. and A. Bhargava (1983b) "Maximum Likelihood Estimation of Regression Models with First Order Moving Average Errors when the Root Lies on the Unit Circle", *Econometrica*, 51, 799–820.
Sarris, A.H. (1973) "A Bayesian Approach to Estimation of Time Varying Regression Coefficients", *Annals of Economic and Social Measurement*, 2, 501–523.
Satchell, S.E. (1984) "Approximation to the Finite Sample Distribution for Nonstable First Order Stochastic Difference Equations", *Econometrica*, 52, 1271–1289.

Schmidt, P. and P.C.B. Phillips (1992) "LM Tests for a Unit Root in the Presence of Deterministic Trends", *Oxford Bulletin of Economics and Statistics*, 54, 257–288.
Schotman, P. and H.K. van Dijk (1990) Posterior Analysis of Possibly Integrated Time Series with an Application to Real GNP, manuscript, Erasmus University Econometric Institute, Netherlands.
Schwarz, G. (1978) "Estimating the Dimension of a Model", *Annals of Statistics*, 6, 461–464.
Schwert, G.W. (1989) "Tests for Unit Roots: A Monte Carlo Investigation", *Journal of Business and Economic Statistics*, 7, 147–159.
Sen, P.K. (1980) "Asymptotic Theory of Some Tests for a Possible Change in the Regression Slope Occurring at an Unknown Time Point", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 52, 203–218.
Shephard, N.G. (1992) Distribution of the ML Estimator of a MA(1) and a Local Level Model, manuscript, Nuffield College, Oxford.
Shephard, N.G. (1993) "Maximum Likelihood Estimation of Regression Models with Stochastic Trend Components", *Journal of the American Statistical Association*, 88, 590–595.
Shephard, N.G. and A.C. Harvey (1990) "On the Probability of Estimating a Deterministic Component in the Local Level Model", *Journal of Time Series Analysis*, 11, 339–347.
Shiller, R.J. and P. Perron (1985) "Testing the Random Walk Hypothesis: Power versus Frequency of Observation", *Economics Letters*, 18, 1271–1289.
Shively, T.S. (1988) "An Exact Test for a Stochastic Coefficient in a Time Series Regression Model", *Journal of Time Series Analysis*, 9, 81–88.
Siegmund, D. (1988) "Confidence Sets in Change-point Problems", *International Statistical Review*, 56, 31–47.
Sims, C.A. (1972) "Money, Income and Causality", *American Economic Review*, 62, 540–552.
Sims, C.A. (1980) "Macroeconomics and Reality", *Econometrica*, 48, 1–48.
Sims, C.A. (1988) "Bayesian Skepticism on Unit Root Econometrics", *Journal of Economic Dynamics and Control*, 12, 463–474.
Sims, C.A. and H. Uhlig (1991) "Understanding Unit Rooters: A Helicopter Tour", *Econometrica*, 59, 1591–1600.
Sims, C.A., J.H. Stock, and M.W. Watson (1990) "Inference in Linear Time Series Models with Some Unit Roots", *Econometrica*, 58, 113–144.
Solo, V. (1984) "The Order of Differencing in ARIMA Models", *Journal of the American Statistical Association*, 79, 916–921.
Solo, V. (1989) A Simple Derivation of the Granger Representation Theorem, manuscript, Macquarie University, Sydney, Australia.
Sorensen, B. (1992) "Continuous Record Asymptotics in Systems of Stochastic Differential Equations", *Econometric Theory*, 8, 28–51.
Sowell, F. (1990) "Fractional Unit Root Distribution", *Econometrica*, 58, 495–506.
Sowell, F. (1991) "On DeJong and Whiteman's Bayesian Inference for the Unit Root Model", *Journal of Monetary Economics*, 28, 255–264.
Sowell, F.B. (1992) "Maximum Likelihood Estimation of Stationary Univariate Fractionally-Integrated Time-Series Models", *Journal of Econometrics*, 53, 165–188.
Stine, R.A. and P. Shaman (1989) "A Fixed Point Characterization for Bias of Autoregressive Estimates", *Annals of Statistics*, 17, 1275–1284.
Stock, J.H. (1987) "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors", *Econometrica*, 55, 1035–1056.
Stock, J.H. (1988) A Class of Tests for Integration and Cointegration, manuscript, Kennedy School of Government, Harvard University.
Stock, J.H. (1990) "Unit Roots in GNP: Do We Know and Do We Care?: A Comment", *Carnegie-Rochester Conference on Public Policy*, 26, 63–82.
Stock, J.H. (1991) "Confidence Intervals for the Largest Autoregressive Root in U.S. Economic Time Series", *Journal of Monetary Economics*, 28, 435–460.
Stock, J.H. (1992) "Deciding Between I(0) and I(1)", NBER Technical Working Paper no. 121; *Journal of Econometrics*, forthcoming.
Stock, J.H. and M.W. Watson (1988a) "Variable Trends in Economic Time Series", *Journal of Economic Perspectives*, 2, 147–174.
Stock, J.H. and M.W. Watson (1988b) "Testing for Common Trends", *Journal of the American Statistical Association*, 83, 1097–1107.

Stock, J.H. and M.W. Watson (1989) "Interpreting the Evidence on Money-Income Causality", *Journal of Econometrics*, 40, 161–182.

Stock, J.H. and M.W. Watson (1993) "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems", *Econometrica*, 61, 783–820.

Svensson, L.E.O. (1992) "An Interpretation of Recent Research on Exchange Rate Target Zones", *Journal of economic Perspectives*, 6, 119–144.

Tanaka, K. (1990a) "The Fredholm Approach to Asymptotic Inference in Nonstationary and Non-invertible Time Series Models", *Econometric Theory*, 6, 411–432.

Tanaka, K. (1990b) "Testing for a Moving Average Unit Root", *Econometric Theory*, 6, 433–444.

Tanaka, K. and S.E. Satchell (1989) "Asymptotic Properties of the Maximum Likelihood and Non-Linear Least-Squares Estimators for Noninvertable Moving Average Models", *Econometric Theory*, 5, 333–353.

Tang, S.M. and I.B. MacNeill (1993) "The Effect of Serial Correlation on Tests for Parameter Change at Unknown Time", *Annals of Statistics*, 21, 552–575.

Tinbergen, J. (1939) *Statistical Testing of Business-Cycle Theories. Vol. II: Business Cycles in the United States of America, 1919–1932*. League of Nations: Geneva.

Uhlig, H. (1992) What Macroeconomists Should Know About Unit Roots As Well: The Bayesian Perspective, manuscript, Department of Economics, Princeton University.

Watson, M.W. (1986) "Univariate Detrending Methods with Stochastic Trends", *Journal of Monetary Economics*, 18, 49–75.

Watson, M.W. and R.F. Engle (1985) "Testing for Regression Coefficient Stability with a Stationary AR(1) Alternative", *Review of Economics and Statistics*, LXVII, 341–345.

Wei, C.Z. (1992) "On Predictive Least Squares Principles", *Annals of Statistics*, 20, 1–42.

West, K.D. (1987) "A Note on the Power of Least Squares Tests for a Unit Root", *Economics Letters*, 24, 1397–1418.

West, K.D. (1988a) "Asymptotic Normality when Regressors have a Unit Root", *Econometrica*, 56, 1397–1418.

West, K.D. (1988b) "On the Interpretation of Near Random Walk Behavior in GNP", *American Economic Review*, 78, 202–208.

White, J.S. (1958) "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case", *Annals of Mathematical Statistics*, 29, 1188–1197.

White, J.S. (1959) "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case II", *Annals of Mathematical Statistics*, 30, 831–834.

Yao, Y.-C. (1987) "Approximating the Distribution of the Maximum Likelihood Estimate of the Change-Point in a Sequence of Independent Random Variables", *Annals of Statistics*, 15, 1321–1328.

Zacks, S. (1983) "Survey of Classical and Bayesian Approaches to the Change-Point Problem: Fixed Sample and Sequential Procedures of Testing and Estimation", in: M.H. Rizvi et al., eds., *Recent Advances in Statistics*. New York: Academic Press, 245–269.

Zellner (1987) "Bayesian Inference", in: J. Eatwell, M. Milgate and P. Newman, eds., *The New Palgrave: A Dictionary of Economics*, Macmillan Press: London, 208–219.

Zivot, E. and D.W.K. Andrews (1992) "Further Evidence on the Great Crash, the Oil Price Shock, and the Unit Root Hypothesis", *Journal of Business and Economic Statistics*, 10, 251–270.