

7

Improving Size and Power in Unit Root Testing

Niels Haldrup and Michael Jansson

Abstract

A frequent criticism of unit root tests concerns the poor power and size properties that many such tests exhibit. However, during the past decade or so intensive research has been conducted to alleviate these problems and great advances have been made. The present paper provides a selective survey of recent contributions to improve upon both the size and power of unit root tests and, in so doing, the approach of using rigorous statistical optimality criteria in the development of such tests is stressed. In addition to presenting tests where improved size can be achieved by modifying the standard Dickey–Fuller class of tests, the paper presents the theory of optimal testing and the construction of power envelopes for unit root tests under different conditions allowing for serial correlation, deterministic components, assumptions regarding the initial condition, non-Gaussian errors, and the use of covariates.

7.1	Introduction	253
7.2	Unit root testing	254
7.2.1	The augmented Dickey–Fuller and Phillips–Perron classes of tests	256
7.2.2	Size distortions of unit root tests	257
7.2.3	Modified unit root tests with good size	259
7.2.4	Deterministics	260
7.3	Power envelopes for unit root tests	262
7.3.1	The leading special case	262
7.3.2	Serial correlation	265
7.3.3	Deterministics	266
7.3.4	The initial condition	267
7.3.5	Non-Gaussian errors	269
7.3.6	Covariates	271
7.4	Conclusion	272

7.1 Introduction

Since the mid-1980s there has been a veritable explosion of research on the importance of unit roots in the analysis of economic and other time series data. The reasons for this are manifold, but perhaps the most important motivation for this work is the fact that the development of the notion of cointegration by Granger (1981) and Engle and Granger (1987) has stressed the significance of unit roots and the importance of making valid statistical inference in the presence of nonstationary time series data. There is a vast literature on developing statistical theory for unit root (integrated) processes and the list of empirical applications using unit root testing is even more impressive. There is also a tremendous literature examining the power and size of unit root tests, not least by adopting numerical simulation (Monte Carlo) techniques under a multitude of different designs of the underlying process. Referring to exhaustive lists of contributions to the theoretical, numerical and empirical literature in this field goes far beyond the space limitation we have here, but surveys referring to many of these contributions can be found in, *inter alia*, Stock (1994), Maddala and Kim (1998), and Phillips and Xiao (1998).

The present chapter reviews some of the results and findings in the unit roots literature that we believe are the most important contributions made over the past decade or so. The review is deliberately chosen to be selective and focuses on those contributions we believe are most likely to be fruitful for future developments of theory as well as in applications. Historically, the criticisms of unit root testing have concerned both the power and size properties of conventional unit root tests (e.g., Schwert, 1989; Agiakloglou and Newbold, 1992; and DeJong, Nankervis, Savin, and Whiteman, 1992a, 1992b). Stimulated in part by these influential Monte Carlo studies, a considerable amount of effort has been devoted to improving the size and/or power properties of unit root tests and much progress has been made. Most of these advances have been made with the help of rigorous statistical theory and are therefore potentially of general methodological interest. Here we survey (a subset of) these recent contributions.

We focus narrowly on the problem of testing for the presence of autoregressive unit roots, partly for concreteness and partly because this branch of nonstationary time series analysis appears to be the one in which the pertinent problems are best understood at this point. Consequently, several important recent advances in the analysis of nonstationary data are abstracted from in the present exposition. These advances include models of higher orders of integration (e.g., Dickey and Pantula, 1987; Haldrup, 1998), tests of seasonally integrated processes (e.g., Hylleberg, Engle, Granger, and Yoo, 1990; Franses, 1996; Ghysels and Osborn, 2001), tests of unit roots against structural breaks (e.g., Perron, 1989, 2005), fractionally integrated processes (e.g., Granger and Joyeux, 1980; Baillie, 1996; Velasco, 2005), testing stationarity against nonstationarity (e.g., Kwiatkowski, Phillips, Schmidt, and Shin, 1992; Saikkonen and Luukkonen, 1993a, 1993b; Jansson, 2004), and panel data unit root tests (e.g., Levin, Lin, and Chu, 2002; Im, Pesaran, and Shin, 2003; and Choi, 2005). Also, a recent literature has developed bootstrap methods

for unit root tests (e.g., Park, 2002, 2003; Paparoditis and Politis, 2003; Davidson and MacKinnon, 2005).

The chapter proceeds as follows. Section 7.2 is concerned with the Dickey–Fuller class of tests (Dickey and Fuller, 1979) and the modifications of this testing framework that have been suggested in order to accommodate (particularly) serially dependent processes and the size distortions that this complication generally implies. The modifications discussed herein include the proposal of Said and Dickey (1984) to use long autoregressions to approximate general dependent processes and the nonparametric corrections of Phillips (1987a) and Phillips and Perron (1988). Special emphasis will be put on the further improvements towards alleviating size distortions made in a series of papers by Ng and Perron, culminating in Ng and Perron (2001). The focus of this recent work is on the importance of appropriately estimating the long-run variance by use of an autoregressive-based spectral density estimator, whilst simultaneously accounting for the serious bias that the estimate of the least squares estimator of the autoregressive coefficient implies even in large samples. Extending the approach to allow for deterministic components by exploiting the GLS detrending procedure of Elliott, Rothenberg, and Stock (1996), the class of tests suggested by Ng and Perron (2001) are argued to exhibit excellent power and size performance. In fact, the tests of Ng and Perron are “nearly efficient,” in the sense that they almost achieve the asymptotic power envelope for unit root tests.

The concepts of test efficiency and power envelopes in the construction of tests for unit roots are discussed in depth in section 7.3. To highlight the nonstandard aspects of the unit root testing problem, the discussion of efficiency starts from the benchmark case of a zero mean Gaussian AR(1) model, a fully parametric model in which there are no nuisance parameters. Subsequent subsections discuss how Elliott, Rothenberg, and Stock’s (1996) efficiency results for the benchmark change under modifications of that model. The discussion emphasizes the role of nuisance parameters caused by the accommodation of serial correlation (Elliott, Rothenberg, and Stock, 1996), deterministic components (Elliott, Rothenberg, and Stock, 1996), and/or a nonzero initial condition (Müller and Elliott, 2003), but also touches upon two sources of potentially important power gains in unit root testing, namely non-Gaussian errors and stationary covariates.

7.2 Unit root testing

Suppose the observed data $\{y_t : t = 1, \dots, T\}$ is generated by the AR(1) model

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad (7.1)$$

where $y_0 = 0$ and the $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$ are unobserved errors. In this model, the unit root testing problem is that of testing

$$H_0 : \rho = 1 \quad \text{vs} \quad H_1 : \rho < 1.$$

Let

$$\hat{\rho} = \frac{\sum_{t=2}^T y_{t-1} y_t}{\sum_{t=2}^T y_{t-1}^2}$$

denote the OLS estimator of ρ and let

$$t_{\hat{\rho}} = \frac{\hat{\rho} - 1}{s / \sqrt{\sum_{t=2}^T y_{t-1}^2}}, \quad s^2 = T^{-1} \sum_{t=2}^T (y_t - \hat{\rho} y_{t-1})^2,$$

be the t -statistic associated with the unit root hypothesis. As is well known, $\hat{\rho}$ and $t_{\hat{\rho}}$ exhibit nonstandard large sample behavior under the unit root hypothesis. Indeed, when $\rho = 1$,

$$T(\hat{\rho} - 1) \rightarrow_d \frac{\int_0^1 W(r) dW(r)}{\int_0^1 W(r)^2 dr} \quad (7.2)$$

and

$$t_{\hat{\rho}} \rightarrow_d \frac{\int_0^1 W(r) dW(r)}{\sqrt{\int_0^1 W(r)^2 dr}} \quad (7.3)$$

where W is a standard Brownian motion (i.e., a Wiener process) and “ \rightarrow_d ” signifies convergence in distribution. Although the results in the preceding equations can be traced back to White (1958), the limiting distributions in (7.2) and (7.3) are frequently referred to as the Dickey–Fuller distributions, in recognition of the contribution of Dickey, Fuller (1979).

7.2.1 The augmented Dickey–Fuller and Phillips–Perron classes of tests

An important implication of (7.2) and (7.3) is that both $T(\hat{\rho} - 1)$ and $t_{\hat{\rho}}$ are asymptotically pivotal under the null hypothesis; that is, the limiting null distributions of these statistics do not depend on unknown nuisance parameters. However, the practical usefulness of these results is limited by the implausibility of the assumptions regarding the errors ε_t . The impact of serially correlated errors can be illustrated by means of the model

$$y_t = \rho y_{t-1} + u_t, \quad u_t = \psi(L) \varepsilon_t, \quad (7.4)$$

where $y_0 = 0$, $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$, and $\psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$ is a lag polynomial whose coefficients $\{\psi_j\}$ satisfy $\sum_{j=1}^{\infty} j |\psi_j| < \infty$.

The model (7.4) imposes more general assumptions on the serial correlation pattern of $y_t - \rho y_{t-1}$ than does the AR(1) model in (7.1). As in the AR(1) model, the parameter of interest is ρ , the unit root testing problem being $H_0: \rho = 1$ vs $H_1: \rho < 1$.

Relative to the AR(1) model (7.1), the unit root testing problem in model (7.4) is complicated by the presence of the nuisance parameters $\{\psi_j\}$.

Phillips (1987a) shows that in this case (7.2) and (7.3) are modified as follows:

$$T(\hat{\rho} - 1) \rightarrow_d \frac{\int_0^1 W(r)dW(r) + \lambda}{\int_0^1 W(r)^2 dr} \quad (7.5)$$

and

$$t_{\hat{\rho}} \rightarrow_d \frac{\omega \int_0^1 W(r)dW(r) + \lambda}{\sigma \sqrt{\int_0^1 W(r)^2 dr}}, \quad (7.6)$$

where $\lambda = (\omega^2 - \sigma^2)/(2\omega^2)$, $\sigma^2 = E[u_t^2] = \sigma_\varepsilon^2 \left(\sum_{j=0}^{\infty} \psi_j^2 \right)$ is the variance of u_t , and $\omega^2 = \lim_{T \rightarrow \infty} T^{-1} E \left[\left(\sum_{t=1}^T u_t \right)^2 \right] = \sigma_\varepsilon^2 \left(\sum_{j=0}^{\infty} \psi_j \right)^2$ is the long-run variance of u_t .¹

When the innovations u_t are *i.i.d.* (i.e., when $\psi_j = 0$ for $j \geq 1$), $\omega^2 = \sigma^2$ and the limiting distributions in (7.5) and (7.6) simplify to the nuisance parameter free limiting distributions in (7.2) and (7.3).

Different routes have been followed in the literature to account for the presence of nuisance parameters in the limiting distributions of $T(\hat{\rho} - 1)$ and $t_{\hat{\rho}}$. It was shown by Dickey and Fuller (1979) that when u_t is an AR process of (finite) order k , $T(\hat{\rho} - 1)$ and $t_{\hat{\rho}}$ calculated from the regression

$$y_t = \tilde{\rho} y_{t-1} + \sum_{j=1}^{k-1} \tilde{\gamma}_j \Delta y_{t-j} + \tilde{v}_{tk} \quad (t = k+1, \dots, T) \quad (7.7)$$

will indeed have the limiting null distributions in (7.2) and (7.3). However, if u_t is an ARMA(p, q) process (with $q \geq 1$), then the auxiliary regression model (7.7) will inadequately solve the nuisance parameter problem, at least if k is held fixed. On the other hand, utilizing results of the Berk (1974) variety and generalizing the celebrated results of Said and Dickey (1984), it has been shown by Chang and Park (2002) that, if u_t is an ARMA(p, q) process, then the limiting null distributions of $T(\hat{\rho} - 1)$ and $t_{\hat{\rho}}$ coincide with the limiting distributions in (7.2) and (7.3) provided $E(\varepsilon_t^4) < \infty$, $k \rightarrow \infty$, and $k = o(T^{1/2-\delta})$ for some $\delta > 0$.²

Rather than solving the nuisance parameter problem by employing an autoregressive sieve, Phillips (1987a) and Phillips and Perron (1988) use consistent estimators of ω^2 and σ^2 to transform the statistics $T(\hat{\rho} - 1)$ and $t_{\hat{\rho}}$ in a manner that eliminates the influence of nuisance parameters. More specifically, they suggest the statistics

$$Z_\rho = T(\hat{\rho} - 1) - \frac{\hat{\omega}^2 - \hat{\sigma}^2}{2T^{-2} \sum_{t=2}^T y_{t-1}^2} \quad (7.8)$$

and

$$Z_t = \frac{\hat{\sigma}}{\hat{\omega}} t_{\hat{\rho}} - \frac{\hat{\omega}^2 - \hat{\sigma}^2}{2\sqrt{\hat{\omega}^2 T^{-2} \sum_{t=2}^T y_{t-1}^2}}, \quad (7.9)$$

where $\hat{\omega}^2$ and $\hat{\sigma}^2$ are consistent estimators of ω^2 and σ^2 . The limiting null distributions of Z_{ρ} and Z_t coincide with the limiting distributions in (7.2) and (7.3).

7.2.2 Size distortions of unit root tests

Several studies (for example, Schwert, 1989; Agiakloglou and Newbold, 1992) have documented that the tests discussed in the previous subsection generally exhibit significant size distortions in finite samples when errors are serially correlated, especially when the errors are of the moving average type with a root approaching minus one. Because a near cancellation of roots occurs when the MA root is very close to minus one, it is not surprising that any unit root test will suffer from severe size distortions in that case. However, it has been found that size can be seriously inflated even for negative roots of moderate magnitude. Schwert (1989) found that the test with the least size distortion is the (Said–Dickey) t -test based on $t_{\hat{\rho}}$ obtained from a high order autoregression, but even for this test the size problem is not negligible. Moreover, even though long autoregressions may moderate the size problems, the use of a long autoregression typically leads to a nontrivial loss of power (for example, DeJong, Nankervis, Savin, and Whiteman, 1992a, 1992b).

Ng and Perron (1995, 2001) further scrutinized rules for truncating long autoregressions when performing unit root tests based on (7.7). Consider the information criterion

$$IC(k) = \log \tilde{\sigma}_k^2 + kC_T/T, \quad \tilde{\sigma}_k^2 = (T-k)^1 \sum_{t=k+1}^T \tilde{v}_{tk}^2, \quad (7.10)$$

where $\{C_T\}$ is a positive sequence satisfying $C_T = o(T)$. The Akaike Information Criterion (AIC) sets $C_T = 2$, whereas the Schwartz or Bayesian Information Criterion (BIC) puts $C_T = \log T$. Ng and Perron (1995) found that, although these information criteria satisfy the requirement that $k = o(T^{1/3})$, generally models with too low a value of k are selected, with size distortions as a consequence. Ng and Perron (1995) also demonstrate that using a sequential data dependent procedure, where the significance of coefficients on additional lags are sequentially tested, will yield a test with improved size. However, a problem with the latter procedure is that in other cases the sequential test procedure tends to overparameterize, thereby resulting in a loss of power. More recently, Ng and Perron (2001) have developed an information criterion with a penalty function adequate for integrated time series. The idea is to select some lag order k in the interval between 0 and a preselected value k_{\max} , where the upper bound k_{\max} satisfies $k_{\max} = o(T)$. Their preferred criterion, which can be interpreted as a modified form of the AIC,

is given by

$$MAIC(k) = \log \hat{\sigma}_k^2 + 2(\tau_T(k) + k)/(T - k_{\max}), \quad (7.11)$$

where $\hat{\sigma}_k^2 = (T - k_{\max})^{-1} \sum_{t=k_{\max}+1}^T \hat{v}_{tk}^2$ and $\tau_T(k) = \hat{\sigma}_k^2(\hat{\rho} - 1)^2 \sum_{t=k_{\max}+1}^T \gamma_{t-1}^2$.³

It is interesting to observe that the penalty function of the modified criteria is data dependent, which is a way to account for the fact that the bias in the sum of the autoregressive coefficients (i.e., $\hat{\rho} - 1$) is highly dependent upon the selected k . Even though Ng and Perron (2001) did not examine directly the effect on the size of the Said–Dickey test by using the above rules, simulation results in other contexts demonstrate that modified information criteria are superior to conventional information criteria in truncating long autoregressions with integrated variables when moving average errors are present.

In the implementation of the (Phillips–Perron) tests based on Z_ρ and Z_t , the estimator

$$\hat{\sigma}^2 = T^{-1} \sum_{t=2}^T \hat{u}_t^2, \quad \hat{u}_t = \gamma_t - \hat{\rho}\gamma_{t-1},$$

serves as a consistent estimator of σ^2 . With respect to estimation of ω^2 , a wide range of kernel estimators have been considered. These kernel estimators are of the form

$$\hat{\omega}_{KER}^2 = T^{-1} \sum_{t=2}^T \hat{u}_t^2 + 2 \sum_{j=1}^{T-1} w(j/M_T) \left(T^{-1} \sum_{t=j+2}^T \hat{u}_t \hat{u}_{t-j} \right), \quad (7.12)$$

where $w(\cdot)$ is some kernel (weight) function and M_T is a bandwidth parameter (see for example, Newey and West, 1987; Andrews, 1991).⁴ The vast majority of unit root tests proposed in the literature use such kernel estimators of the long-run variance ω^2 to remove the influence of nuisance parameters in the asymptotic distributions. Notwithstanding, it has been shown by Perron and Ng (1996) that no spectral density estimator can completely eliminate size distortions and, in fact, kernel-based estimators tend to aggravate the size distortions.⁵ This finding is due to the fact that estimation of ρ and ω^2 are coupled, in the sense that the least squares estimator $\hat{\rho}$ is used in constructing \hat{u}_t and hence affects $\hat{\omega}_{KER}^2$. Because the least squares estimator $\hat{\rho}$ is well known to be seriously biased in finite (and even in large) samples when u_t exhibits strong serial correlation, the nuisance parameter estimator $\hat{\omega}_{KER}^2$ is expected to be very imprecise in critical regions of the serial correlation parameter space. A seemingly obvious way to alleviate this problem is to construct an estimator where residuals are calculated under the null hypothesis, i.e., by using $\Delta \gamma_t$ instead of \hat{u}_t in (7.12), but it has been shown by Phillips and Ouliaris (1990) that this leads to inconsistent unit root tests.

To produce an estimator that is consistent under the unit root null whilst attenuating the dependence on $\hat{\rho}$, Perron and Ng (1996, 1998), following earlier

work by Berk (1974) and Stock (1999), suggested using an autoregressive spectral density estimator based on estimation of the long autoregression (7.7):

$$\hat{\omega}_{AR}^2 = \frac{\hat{\sigma}_k^2}{\left(1 - \sum_{j=1}^{k-1} \tilde{\gamma}_j\right)^2}, \quad (7.13)$$

where k is chosen according to one of the information criteria discussed above. The motivation for using the regression (7.7) to estimate $\sum_{j=1}^{k-1} \tilde{\gamma}_j$, rather than basing estimation on an autoregressive model in first differences (i.e., the model under the null), is that this ensures a consistent unit root test (e.g., Stock, 1999). The construction (7.13) decouples estimation of ω^2 from the estimation of ρ and, therefore, helps avoid the problems caused by the bias of ρ . In particular, $\hat{\omega}_{AR}^2$ is immune to potentially severe biases in $\hat{\rho}$ that are caused by the presence of serial correlation in the errors.

In a comparison of the size properties of the Phillips–Perron tests using the estimator $\hat{\omega}_{AR}^2$ and the tests using a Bartlett kernel estimator of ω^2 , it was found that significant size improvements can be achieved in the most critical parameter space. Nevertheless, size distortions are still severe and remain so even if $\hat{\omega}_{AR}^2$ is replaced by the (unknown) true value ω^2 (Perron and Ng, 1996), a finding which indicates that biased in $\hat{\rho}$ are an important source of size distortions. The next subsection discusses developments that seek to obtain further improvements by addressing this issue.

7.2.3 Modified unit root tests with good size

Perron and Ng (1996) consider modified Phillips–Perron tests (referred to as M -tests in the following) that appear to have much improved size properties compared to any other unit root test. Moreover, the tests can be designed such that they satisfy desirable optimality criteria in terms of power; a topic which we will later return to. The tests belong to a class originally suggested by Stock (1999), which exploits the fact that a series converges at different rates under the null and the alternative hypotheses.

The first statistic reads

$$MZ_\rho = \frac{T^{-1}y_T^2 - \hat{\rho}_{AR}^2}{2T^{-2} \sum_{t=2}^T y_{t-1}^2}, \quad (7.14)$$

which can also be written in terms of Z_ρ as

$$MZ_\rho = Z_\rho + \frac{T}{2}(\hat{\rho} - 1)^2. \quad (7.15)$$

Because $\hat{\rho} - 1 = O_p(T^{-1})$ under the null (super consistency), it is seen that Z_ρ and MZ_ρ will be asymptotically equivalent (under the null), implying in particular that the limiting null distribution of MZ_ρ is the one given in (7.2).

The next statistic reads

$$MSB = \sqrt{\hat{\omega}_{AR}^{-2} T^{-2} \sum_{t=2}^T y_{t-1}^2}, \quad (7.16)$$

which is stochastically bounded under the null and is $O_p(T^{-1})$ (and thus tends to zero) under fixed alternatives. This test is related to the Sargan and Bhargava (1983) test, hence the name, and critical values are reported by Stock (1999). Finally, because $Z_t = MSB \cdot Z_\rho$, a modified Phillips–Perron t test can be defined as

$$MZ_t = Z_t + \frac{1}{2} MSB \cdot (\hat{\rho} - 1)^2. \quad (7.17)$$

Even though $\hat{\rho}$ is superconsistent (under the null), implying that the correction terms associated with Z_ρ and Z_t are asymptotically negligible, it is still the case that the correction factors can be important even in moderately large samples, the reason being that $\hat{\rho}$ is severely biased in the presence of strong serial correlation. Simulation experiments reported in Perron and Ng (1996) show that the M -tests have impressively lower size distortion compared to other unit root tests that are available in the literature. However, it is essential that the autoregressive spectral density estimator $\hat{\omega}_{AR}^2$ is used as an estimator of ω^2 to decouple estimation of the unit root and the long-run variance for the tests to have good size properties. For instance, for an MA root of, e.g., -0.8 , the actual size of the MZ_t test is about 6% at a nominal 5% level, whereas Phillips–Perron tests or modified Phillips–Perron tests using kernel estimates of ω^2 have size close to 100%.

The M -tests also appear to be robust to, e.g., measurement errors and additive outliers in the observed series. Franses and Haldrup (1994) and Haldrup, Montanés, and Sanso (2005) show that in these cases with data contamination, unit root inference using standard tests become seriously size affected. Vogelsang (1999) shows that the M -tests effectively solve these problems in terms of test size.⁶

7.2.4 Deterministics

In practical applications the underlying model will also contain deterministic components. These can be accommodated by generalizing (7.4) to a components representation of the form

$$y_t = \mu_t + z_t, \quad z_t = \rho z_{t-1} + u_t, \quad (7.18)$$

where $z_0 = 0$, u_t is as in (7.4), and μ_t is a deterministic component. Most of the specifications of μ_t used in applications are linear-in-parameters specifications of the form

$$\mu_t = d_t' \beta, \quad (7.19)$$

where d_t are known k -vectors of deterministic terms (for some $k \geq 1$), while β are k -vectors of unknown parameters. The leading special cases of linear-in-parameters

specifications are the constant mean and linear trend specifications, in which $d_t = 1$ and $d_t = (1, t)'$, respectively.⁷

Appropriate detrending of the data is needed if one tests the unit root hypothesis against a trend stationary alternative (i.e., $\rho < 1$ in (7.18) and $d_t = (1, t)'$ in (7.19)). Specifically, the Dickey–Fuller (or Said–Dickey) regressions should take the alternative form

$$y_t = \tilde{\mu}_t + \tilde{\rho}y_{t-1} + \sum_{j=1}^{k-1} \tilde{\gamma}_j \Delta y_{t-j} + \tilde{v}_{tk}. \quad (7.20)$$

Appropriate treatment of deterministic components is extremely important. For instance, failure to include a time trend regressor in the auxiliary regression when power against the trend stationary alternative is wanted will lead to a test with zero asymptotic power. Similarly, the Phillips–Perron class of tests allow inclusion of deterministic components or, alternatively, detrending of the series prior to unit root testing can be done. For all cases where the model is augmented with deterministic components the relevant distributions change accordingly, as Brownian motion processes are replaced with demeaned and detrended Brownian motions of the form

$$W^d(r) = W(r) - D(r)' \left(\int_0^1 D(s)D(s)' ds \right)^{-1} \left(\int_0^1 D(s)W(s) ds \right), \quad (7.21)$$

where $D(r) = 1$ when $d_t = 1$ and $D(r) = (1, r)$ when $d_t = (1, t)'$.

With respect to the M -tests of the previous subsection, Ng and Perron (2001) suggest an alternative way of dealing with deterministic components. The alternative detrending method, local GLS detrending, is in the spirit of Elliott, Rothenberg, Stock (1996) and has the advantage of yielding tests that are “nearly” efficient, in the sense that they nearly achieve the asymptotic power envelopes for unit root tests. (A discussion of these power envelopes will be provided in section 7.3.) The GLS detrending method can be described as follows. For any series $\{x_t\}_{t=1}^T$ of length T and any constant \bar{c} , define $x^{\bar{c}} = (x_1, \Delta x_2 - \bar{c}, T^{-1}x_1, \dots, \Delta x_T - \bar{c}T^{-1}x_{T-1})'$. The GLS detrended series $\{\tilde{y}_t\}$ is given by

$$\tilde{y}_t = y_t - d_t' \tilde{\beta}, \quad \tilde{\beta} = \arg \min_{\beta} (y^{\bar{c}} - d^{\bar{c}} \beta)' (y^{\bar{c}} - d^{\bar{c}} \beta).$$

Elliott, Rothenberg, Stock (1996) suggested $\bar{c} = -7$ and $\bar{c} = -13.5$ for $d_t = 1$ and $d_t = (1, t)'$, respectively, as these values of \bar{c} correspond to the local alternatives against which the local asymptotic power envelope for 5 percent tests equals 50 percent. The M -tests constructed using GLS detrended data (and using $\hat{\omega}_{AR}^2$ together with the modified information criteria of Section 7.2.2) are denoted, respectively, MZ_{ρ}^{GLS} , MSB^{GLS} , and MZ_t^{GLS} . These tests are shown by Ng and Perron (2001) to have excellent size and local asymptotic power.

In conclusion, unit root tests can be constructed with both excellent size and local asymptotic power properties, but to achieve these dual objectives it is necessary to use GLS detrended data.

7.3 Power envelopes for unit root tests

This section discusses power envelopes (efficiency bounds) for tests of the unit root hypothesis. Power envelopes for unit root tests have proven to be useful for two reasons. First, being attainable upper bounds on power, they give an objective standard against which the power properties of any feasible unit root test can be compared. For instance, the fact that the M -tests discussed in the previous section have local asymptotic power “close” to the appropriate power envelopes (Ng and Perron, 2001) implies that these are “nearly” efficient. Second, the derivation of power envelopes is useful because it suggests how admissible unit root tests with good overall power properties can be constructed. Indeed, the GLS detrending method, which is the key to accommodating deterministic components without sacrificing efficiency, is a natural by-product of the derivation of the power envelope in the presence of deterministic components (Elliott, Rothenberg, and Stock, 1996).

7.3.1 The leading special case

A natural starting point for the discussion of power envelopes for unit root tests is the known-variance, zero-mean Gaussian AR(1) model.⁸ In this model, the observed data $\{y_t : t = 1, \dots, T\}$ is generated as

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad (7.22)$$

where $y_0 = 0$ and $\varepsilon_t \sim i.i.d. \mathcal{N}(0, 1)$.

Any (possibly randomized) unit root test can be represented by means of a test function $\phi_T : \mathbb{R}^T \rightarrow [0, 1]$, such that H_0 , the unit root hypothesis, is rejected with probability $\phi_T(Y)$ if $Y_T = (y_1, \dots, y_T)' = Y$. The power (function) associated with $\phi_T(\cdot)$ is given by $E_\rho \phi_T(Y_T)$, where the subscript on “ E ” indicates the distribution with respect to which the expectation is taken (i.e., the argument of the power function).

When evaluating the power properties of a unit root test, a power envelope is very useful. By definition, a power envelope for a class of unit root tests gives an attainable upper bound on $E_\rho \phi_T(Y_T)$ for tests in the class. Throughout this section, the class of tests under consideration will be the class of tests of (asymptotic) size α or some subset thereof. The power envelope for size α tests is the function $\Pi_T^\alpha(\cdot)$ given by⁹

$$\Pi_T^\alpha(\rho) = \max_{\phi_T(\cdot) : E_1 \phi_T(Y_T) = \alpha} E_\rho \phi_T^T(Y_T). \quad (7.23)$$

By construction, $\Pi_T^\alpha(\rho)$ is an upper bound on $E_\rho \phi_T(Y_T)$ for test functions $\phi_T(\cdot)$ associated with tests of size α . Moreover, the power envelope is attainable (pointwise) in the sense that, for every ρ , $E_\rho \phi_T(Y_T) = \Pi_T^\alpha(\rho)$ for some $\phi_T(\cdot)$ corresponding to a test of size α .

There is a simple and constructive way to derive the power envelope. For the known-variance, zero-mean Gaussian AR (1) model, the log likelihood function, $L_T(\cdot)$, satisfies the relation

$$L_T(\rho) - L_T(1) = T(\rho - 1)S_T - \frac{1}{2}[T(\rho - 1)]^2H_T, \tag{7.24}$$

where $S_T = T^{-1} \sum_{t=2}^T y_{t-1} \Delta y_t$ and $H_T = T^{-2} \sum_{t=2}^T y_{t-1}^2$. An application of the Neyman–Pearson lemma (e.g., Lehmann, 1994, Theorem 3.1) therefore yields

$$\Pi_T^\alpha(\rho) = \Pr_\rho \left[T(\rho - 1)S_T - \frac{1}{2}[T(\rho - 1)]^2H_T > k_T^\alpha(\rho) \right], \tag{7.25}$$

where $k_T^\alpha(\rho)$ satisfies $\Pr_1 \left[T(\rho - 1)S_T - \frac{1}{2}[T(\rho - 1)]^2H_T > k_T^\alpha(\rho) \right] = \alpha$ and the subscript on “Pr” indicates the distribution with respect to which the probability is evaluated.

In addition to providing a formula for computing the power envelope, expression (7.25) delivers a characterization of the test that attains the power envelope at any given value of ρ . Specifically, it follows from (7.25) that the power envelope $\Pi_T^\alpha(\rho)$ is attained by the test which rejects for large values of $T(\rho - 1)S_T - \frac{1}{2}[T(\rho - 1)]^2H_T$. Because the functional form of the optimal test against any specific alternative $\rho < 1$ depends on ρ , the unit root testing problem does not admit a uniformly most powerful (UMP) size α test, in spite of the fact that it is a one-sided testing problem without any nuisance parameters.

Applying the preceding arguments to other simple hypotheses on ρ , it can be verified that non-existence of a UMP size α test is a property shared by all one-sided hypothesis tests on the autoregressive coefficient ρ in (7.22). In other words, the nonexistence of a UMP size α test is not specific to the unit root hypothesis. What is somewhat special about the unit root hypothesis is the fact that non-existence of a UMP size α test holds even asymptotically.¹⁰

By analogy with the finite sample situation, the asymptotic power envelope for a class of (sequences of) unit root tests gives an attainable upper bound on local asymptotic power for (sequences of) tests in the class. Assuming the limit exists, the local asymptotic power function of a sequence $\{\phi_T(\cdot)\}$ of unit root tests is the function (with argument $c \leq 0$) $\lim_{T \rightarrow \infty} E_{1+T^{-1}c} \phi_T(Y_T)$.¹¹ A sequence $\{\phi_T(\cdot)\}$ of unit root tests is said to have asymptotic size α if $\lim_{T \rightarrow \infty} E_1 \phi_T(Y_T) = \alpha$. The asymptotic power envelope for tests asymptotically of size α is the function $\Pi_\infty^\alpha(\cdot)$ given by

$$\Pi_\infty^\alpha(c) = \max_{\{\phi_T(\cdot)\}: \lim_{T \rightarrow \infty} E_1 \phi_T(Y_T) = \alpha} \overline{\lim}_{T \rightarrow \infty} E_{1+T^{-1}c} \phi_T(Y_T). \tag{7.26}$$

An explicit formula for the asymptotic power envelope is available. Because the optimal unit root test against the alternative $\rho = 1 + T^{-1}c$ rejects for large values of $cS_T - \frac{1}{2}c^2H_T$, it stands to reason that the asymptotic power envelope $\Phi_\infty^\alpha(c)$ is attained by the sequence of tests with rejection regions of the form $\{cS_T - \frac{1}{2}c^2H_T > k_\infty^\alpha(c)\}$, where $k_\infty^\alpha(c)$ is such that the sequence has asymptotic size α .

Indeed, it can be shown that

$$\begin{aligned}\Pi_{\infty}^{\alpha}(c) &= \lim_{T \rightarrow \infty} \Pr_{1+T^{-1}c} \left[cS_T - \frac{1}{2}c^2H_T > k_{\infty}^{\alpha}(c) \right] \\ &= \Pr \left[c \int_0^1 W_c(r)dW(r) + \frac{1}{2}c^2 \int_0^1 W_c(r)^2 dr > k_{\infty}^{\alpha}(c) \right],\end{aligned}\tag{7.27}$$

where $k_{\infty}^{\alpha}(c)$ satisfies $\Pr \left[c \int_0^1 W(r)dW(r) - \frac{1}{2}c^2 \int_0^1 W(r)^2 dr > k_{\infty}^{\alpha}(c) \right] = \alpha$, W is a Wiener process, and W_c is an Ornstein–Uhlenbeck process satisfying the stochastic differential equation $dW_c(r) = cW_c(r)dr + dW(r)$ with initial condition $W_c(0) = 0$.¹²

As is true of its finite sample counterpart, the asymptotic power envelope can only be attained pointwise. In other words, there does not exist a sequence of tests (asymptotically of size α) which attains $\Pi_{\infty}^{\alpha}(c)$ for all values of c . In the absence of a UMP test, it seems natural to try to derive tests enjoying weaker optimality properties in the hope that these tests will have good overall power properties. Two complementary notions of optimality, local optimality and point optimality, have been employed to derive unit root tests with demonstrable optimality properties. The notion of local optimality leads to a test which maximizes (the scaled limit of) the derivative of the power function under the null hypothesis. For the unit root testing problem, the locally most powerful test rejects for small values of S_T .¹³ Although admissible, the locally most powerful unit root test turns out to have pretty poor local asymptotic power properties (Stock, 1994). In contrast, the notion of point optimality has been found to deliver (admissible) unit root tests with excellent power properties.

By definition, a point optimal unit root test maximizes power against a specific (point) alternative (e.g., King, 1988).¹⁴ The family of point optimal unit root tests is obtained as a by-product of the power envelope. It consists of all tests with rejection regions of the form $\{cS_T - \frac{1}{2}c^2H_T > k_{\infty}^{\alpha}(c)\}$ where c indexes the local alternative against which optimal power is desired. Elliott, Rothenberg, and Stock (1996) found that point optimal unit root tests have local asymptotic power functions essentially identical to the power envelope for a wide range of values of the index c . A popular choice, advocated in the unit root context by Elliott, Rothenberg, and Stock (1996), is the value of c such that the asymptotic power envelope for 5 percent tests equals 50 percent when evaluated at c ; that is, the recommended value of c solves the equation $\Pi_{\infty}^{0.05}(c) = 0.5$.

Numerous other unit root tests have been proposed (for a review, see Stock, 1994). The most well-known examples are probably the Dickey–Fuller (1979) tests (i.e., the tests based on $T(\hat{\rho}^{-1})$ and $t_{\hat{\rho}}$) and their asymptotic equivalents (e.g., the tests based on $Z_{\hat{\rho}}$ and Z_t discussed in section 7.2.1). In the case where the error variance is known to equal unity, a Dickey–Fuller type t -test rejects for small values of

$$\frac{\hat{\rho} - 1}{1/\sqrt{\sum_{t=2}^T Y_{t-1}^2}} = \frac{S_T}{\sqrt{H_T}}.$$

This test can be interpreted as a (signed) likelihood ratio test.¹⁵ Moreover, the test has excellent local asymptotic power properties (Elliott, Rothenberg, and Stock, 1996). In spite of this, the test does not seem to enjoy any conventional optimality properties.¹⁶ On the other hand, the Dickey–Fuller estimator test, which rejects for small values of $T(\hat{\rho} - 1) = S_T/H_T$, is a member of the class of point optimal tests and is therefore admissible (Stock, 1994).

Power envelopes have been derived for a variety of extensions of the basic model (7.22). The remainder of this section discusses five such extensions. These are all of practical and theoretical interest, the practical interest being due to the empirical relevance of the extensions. To simplify the exposition, we discuss each of the extensions in isolation, in each case studying a model which departs as little as possible from the known-variance, zero-mean Gaussian AR(1) model.

7.3.2 Serial correlation

This subsection discusses the impact of serial correlation on the form of the asymptotic power envelope. As was explained in section 7.2, the presence of serial correlation introduces nontrivial complications for practitioners wanting to employ unit root tests. In contrast, the presence of serial correlation has no impact on the asymptotic properties of the unit root testing problem, because the asymptotic power envelope (7.27) remains valid under rather general assumptions on the short-run temporal dependence properties of the quasi-difference process $\{y_t - \rho y_{t-1}\}$.

Consider the Gaussian variant of (7.4) in which

$$y_t = \rho y_{t-1} + u_t, \quad u_t = \psi(L)\varepsilon_t, \quad (7.28)$$

where $y_0 = 0$, $\varepsilon_t \sim i.i.d. \mathcal{N}(0, 1)$, and $\psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$ is a lag polynomial whose (unknown) coefficients $\{\psi_j\}$ satisfy $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and $\psi(e^{ir}) = \sum_{j=0}^{\infty} \psi_j e^{irj} \neq 0$ for all $r \in \mathbb{R}$.

The construction of the asymptotic power envelope proceeds in two steps. In the first step, the Neyman–Pearson lemma is used to derive the asymptotic power envelope under the (counterfactual) assumption that $\{\psi_j\}$ is known. The second step then shows that this envelope is indeed the asymptotic power envelope for unit root tests in the model (7.28) by showing that the bound can be attained (pointwise) without knowledge of $\{\psi_j\}$.

For now, suppose the parameters $\{\psi_j\}$ are known. In that case, the derivation of an asymptotic power envelope is conceptually straightforward because ρ is the only parameter of the model. By the Neyman–Pearson lemma, the unit root test with optimal power against the local alternative $\rho = 1 + T^{-1}c$ rejects for large values of the log likelihood ratio $L_T^\psi(1 + T^{-1}c) - L_T^\psi(1)$, where $L_T^\psi(\cdot)$ is the log likelihood function. A slick derivation of the asymptotic power envelope can be based on the fact that

$$L_T^\psi(1 + T^{-1}c) - L_T^\psi(1) = cS_T^\psi - \frac{1}{2}c^2H_T^\psi + o_p(1) \quad (7.29)$$

under the null hypothesis, where $S_T^\psi = \frac{1}{2}(\omega^{-2}T^{-1}y_T^2 - 1)$, $H_T^\psi = \omega^{-2}T^{-2} \sum_{t=2}^T y_{t-1}^2$, and $\omega^2 = \psi(1)^2$ is the long-run variance of $\psi(L)\varepsilon_t$.¹⁷ Using the quadratic expansion (7.29) and the theory of limits of experiments (see, for example, Le Cam and Yang, 2000; van der Vaart, 1998), it can be shown that the asymptotic power envelope for size α tests is attained (at the point c) by the sequence of tests with rejection regions of the form $\left\{cS_T^\psi - \frac{1}{2}c^2H_T^\psi > k_\infty^{\alpha,\psi}(c)\right\}$, where $k_\infty^{\alpha,\psi}(c)$ is such that the sequence has asymptotic size α . Under the null and contiguous alternatives, the limiting distribution of $\left(S_T^\psi, H_T^\psi\right)$ does not depend on $\{\psi_j\}$. By implication, the critical value function $k_\infty^{\alpha,\psi}(c)$ does not depend on $\{\psi_j\}$. More importantly, the asymptotic power envelope is invariant with respect to $\{\psi_j\}$ and is given by the function $\Pi_\infty^\alpha(\cdot)$ defined in (7.27).

To show that the upper bound $\Pi_\infty^\alpha(\cdot)$ constitutes the asymptotic power envelope for the model (7.28), it must be shown that $\Pi_\infty^\alpha(\cdot)$ is attainable without knowledge of $\{\psi_j\}$. To do so, it suffices to exhibit a pair $\left(\hat{S}_T^\psi, \hat{H}_T^\psi\right)$, computable without knowledge of $\{\psi_j\}$, such that $\left(\hat{S}_T^\psi, \hat{H}_T^\psi\right) = \left(S_T^\psi, H_T^\psi\right) + o_p(1)$ under the unit root hypothesis (irrespective of the value of $\{\psi_j\}$). (Assuming such a pair can be found, the test which rejects for large values of $c\hat{S}_T^\psi - \frac{1}{2}c^2\hat{H}_T^\psi$ will attain the asymptotic power envelope.) The asymptotic equivalence requirement is met by $\hat{S}_T^\psi = \frac{1}{2}(\hat{\omega}^{-2}T^{-1}y_T^2 - 1)$ and $\hat{H}_T^\psi = \hat{\omega}^{-2}T^{-2} \sum_{t=2}^T y_{t-1}^2$, where $\hat{\omega}^2$ is any consistent (under the unit root hypothesis) estimator of ω^2 .¹⁸

7.3.3 Deterministics

Proceeding as in section 7.2.4, deterministic terms can be accommodated by extending the basic model (7.22) as follows:

$$y_t = \mu_t + z_t, \quad z_t = \rho z_{t-1} + \varepsilon_t, \quad (7.30)$$

where μ_t is an unknown deterministic component, $z_0 = 0$, and $\varepsilon_t \sim i.i.d. \mathcal{N}(0, 1)$.

In this model, $\{\mu_t\}$ is a nuisance feature in the unit root testing problem. When deriving an asymptotic power envelope in the presence of $\{\mu_t\}$, it is tempting to try to employ the same strategy as in the previous subsection; that is, it is tempting to first derive the asymptotic power envelope assuming $\{\mu_t\}$ is known and then attempt to find a feasible test which attains the bound obtained under the assumption that $\{\mu_t\}$ is known. That method of construction breaks down in general, however. On the one hand, because $z_t = y_t - \mu_t$ is generated by the model of section 7.3.1, it is obvious that the asymptotic power envelope is given by the function $\Pi_\infty^\alpha(\cdot)$ (defined in (7.27)) when $\{\mu_t\}$ is known. On the other hand, for many specifications of $\{\mu_t\}$ used in practice, it turns out to be impossible to find tests that attain $\Pi_\infty^\alpha(\cdot)$ without knowledge of $\{\mu_t\}$.

When $\mu_t = d_t'\beta$ (i.e., μ_t is of the linear-in-parameters form (7.19)), the principle of invariance (e.g., Lehmann, 1994, chapter 6) can be employed to eliminate the deterministic component from the unit root testing problem. Any testing problem regarding ρ is invariant under transformations of the form $g_b(y_1, \dots, y_T) = (y_1 + d_1'b, \dots, y_T + d_T'b)$ (where $b \in \mathbb{R}^k$), the induced transformation in the parameter space being $g_b(\rho, \beta) = (\rho, \beta + b)$. When β is treated as an unknown

nuisance parameter, it therefore seems natural to restrict attention to unit root tests that are invariant in the sense that their test functions $\phi_T(\cdot)$ satisfy $\phi_T(y_1 + d_1' b, \dots, y_T + d_T' b) = \phi_T(y_1, \dots, y_T)$ for every $b \in \mathcal{R}^k$. In other words, a test is invariant if the conclusion drawn by the test depends on the observed data $\{y_t\}$ only through the unobserved series $\{z_t\}$. As a consequence, an invariant test has a test function whose distribution depends only on ρ .¹⁹ Therefore, reduction by invariance eliminates the nuisance parameter β from the unit root testing problem, thereby making it possible to obtain power envelopes by means of the Neyman–Pearson lemma. Drawing on the work of King (1980) and King and Hillier (1985), Dufour and King (1991) used this insight to obtain point optimal invariant tests of simple hypotheses on ρ in the AR(1) model (7.30). Asymptotic power envelopes were obtained (in a model accommodating serial correlation) for the unit root testing problem by Elliott, Rothenberg, and Stock (1996) in the case where d_t is a polynomial trend term. The functional form of the asymptotic power envelope depends on the order of the polynomial trend. It is given by $\Pi_\infty^\alpha(\cdot)$ in the constant mean case, but not otherwise.

As was true in the model without deterministic components, the derivation of the asymptotic power envelopes in model (7.30) is constructive in the sense that tests attaining the asymptotic power envelope are obtained as a by-product. The optimal invariant test against the local alternative $\rho = 1 + T^{-1}c$ rejects for large values of the profile log likelihood ratio $\max_\beta L_T(1 + T^{-1}c, \beta) - \max_\beta L_T(1, \beta)$, where $L_T(\cdot)$ is the log likelihood function. As in the model without deterministic components, no test attains the power envelope uniformly but appropriately chosen point optimal invariant tests are “nearly efficient” in the sense that their local asymptotic power functions are “close” to the asymptotic power envelopes (Elliott, Rothenberg, and Stock, 1996). This “near-efficiency” property is not shared by the popular Dickey–Fuller (1979) tests, whose local asymptotic power functions fall well short of the asymptotic power envelope. Nevertheless, the class of “nearly efficient” invariant unit root tests contains tests other than the point optimal tests obtained in the derivation of the asymptotic power envelope. Examples include the DF-GLS test of Elliott, Rothenberg, and Stock (1996) and Ng and Perron’s (2001) M^{GLS} tests discussed in section 7.2.4. Consequently, the M^{GLS} tests have both excellent size properties and “nearly” optimal power properties.

In view of the inferiority of the Dickey–Fuller (1979) tests in the model (7.30), it would appear that practitioners ought to abandon the use of these tests. More recent research, examining the role of the initial condition y_0 , has arrived at a slightly less drastic conclusion. A brief discussion of that literature is provided in the next subsection.

7.3.4 The initial condition

In the known-variance, zero-mean Gaussian AR (1) model of section 7.3.1, the (unobserved) initial observation y_0 is assumed to be equal to zero. Analogous assumptions are made in the models of sections 7.3.2 and 7.3.3. At first sight, these would appear to be innocuous normalizations because it is easy to show that the initial observation is asymptotically negligible whenever $T^{-1/2}y_0 = o_p(1)$, a

condition that is also satisfied if y_0 is treated as a nuisance parameter (i.e., modeled as a constant) or modeled as a random variable with a fixed distribution. Now, if $\{y_t : 0 \leq t \leq T\}$ is generated by a stationary Gaussian AR(1) model with autoregressive coefficient ρ and innovation variance equal to unity, the initial observation will satisfy $y_0 \sim \mathcal{N}[0, 1/(1 - \rho^2)]$. In that case, the initial observation is not asymptotically negligible, the limiting distribution $T^{-1/2}y_0$ being $\mathcal{N}[0, 1/(-2c)]$ under local-to-unity asymptotics with $\rho = 1 + T^{-1}c$ for some $c < 0$. To the extent that stationarity is a plausible alternative to the unit root hypothesis, these considerations suggest that the role of the initial condition is worth investigating.

The role of the initial condition can be explored by means of the following stripped-down version of the model studied by Müller and Elliott (2003):

$$y_t = \mu + z_t, \quad z_t = \rho z_{t-1} + \varepsilon_t \quad (7.31)$$

where $\varepsilon_t \sim i.i.d. \mathcal{N}(0, 1)$ and $z_0 \sim \mathcal{N}(0, \kappa \sigma_0^2(\rho))$, where $\kappa \geq 0$ is a known constant, $\sigma_0^2(\rho) := 1(|\rho| < 1)/(1 - \rho^2)$, $1(\cdot)$ is the indicator function, and z_0 is independent of $\{\varepsilon_t\}$. The model (7.31) reduces to the model of section 7.3.3 when $\kappa = 0$. When $\kappa = 1$, in contrast, $\{z_t\}$ is generated by a stationary AR(1) whenever $|\rho| < 1$ and the model reduces to a special case of the model studied by Elliott (1999). Irrespective of the value of κ , the initial condition is $z_0 = 0$ under the unit root hypothesis. Due to the inclusion of the constant term μ , this assumption is simply a normalization.

The derivation of the asymptotic power envelope for the unit-root testing problem proceeds as in section 7.3.3. First, the principle of invariance can be employed to remove the nuisance parameter μ . Then, the Neyman–Pearson lemma can be used to characterize the functional form of point optimal invariant tests. The optimal invariant test against the local alternative $\rho = 1 + T^{-1}c$ rejects for large values of the profile log likelihood ratio $\max_{\beta} L_T^{\kappa}(1 + T^{-1}c, \beta) - \max_{\beta} L_T^{\kappa}(1, \beta)$, where $L_T^{\kappa}(\cdot)$ is the log likelihood function. The functional form of the point optimal tests and the shape of the asymptotic power envelope both depend on the value of the constant κ . Moreover, although the tests discussed in section 7.3.3 (“nearly efficient” when $\kappa = 0$ in the model considered here) are asymptotically similar for any value of κ , these tests have power well below the power envelopes corresponding to moderately large values of κ (Müller and Elliott, 2003).²⁰

Müller and Elliott (2003) emphasize an alternative interpretation of the power envelopes discussed in the previous paragraph. If the initial condition z_0 is treated as an unknown nuisance parameter (as opposed to a random variable with a known distribution), the unit root testing problem is complicated by the presence of an unidentified nuisance parameter under the null hypothesis, the parameters μ and z_0 appearing in the likelihood only through their sum. Müller and Elliott (2003) deal with this problem by applying a weighted average power criterion in the spirit of Andrews and Ploberger (1994) when deriving asymptotic power envelopes. The weighting functions employed by Müller and Elliott (2003) correspond to the distributional assumption on z_0 made in (7.31) and give rise to the same asymptotic power envelopes.

In addition to deriving asymptotic power envelopes for the model (7.31), Müller and Elliott (2003) explore the extent to which existing unit root tests can be “rationalized” as being point optimal tests in model (7.31) for appropriately selected values of the constant κ and the local-to-unity parameter c . They find that the tests proposed by Bhargava (1986) can be interpreted as (limiting versions of) point optimal tests in the model (7.31), as can the locally best invariant tests derived by Dufour and King (1991) and Nabeya and Tanaka (1990). Moreover, Müller and Elliott (2003) argue that, although the popular Dickey–Fuller (1979) tests cannot be “rationalized” in this way, there is a sense in which the Dickey–Fuller tests are well approximated by certain members of the class of point optimal tests in the model (7.31), albeit with rather large values of κ .

7.3.5 Non-Gaussian errors

All of the power envelopes discussed so far have been derived under the assumption that the latent errors $\{\varepsilon_t\}$ are (standard) normally distributed. Because the normality assumption is implausible in most empirical applications of unit root tests, it is of interest to develop asymptotic power envelopes for unit roots in (possibly) non-Gaussian environments.

Consider the model²¹

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad (7.32)$$

where $y_0 = 0$ and $\{\varepsilon_t\}$ are *i.i.d.* errors from an unknown (possibly non-Gaussian) distribution with mean zero and variance one. The derivation of the asymptotic power envelope in section 7.3.1 made use of two results, the Neyman–Pearson lemma and the fact that

$$(S_T - cH_T, H_T) \rightarrow d_c \left(\int_0^1 W_c(r) dW(r), \int_0^1 W_c(r)^2 dr \right), \quad c \leq 0.$$

The displayed convergence result, which was used to characterize the local asymptotic power of point optimal tests (derived by means of the Neyman–Pearson lemma), remains valid in the model (7.32). In other words, the limiting representation of (S_T, H_T) , the minimal sufficient statistic under the assumption of normality, is invariant with respect to the distribution of $\{\varepsilon_t\}$ as long as $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = 1$.²² By implication, the local asymptotic power function of the point optimal tests from section 7.3.1 does not depend on the distribution of $\{\varepsilon_t\}$ in the model (7.32). The Gaussian asymptotic power envelope of section 7.3.1 therefore gives a lower bound on maximal attainable local asymptotic power in the model (7.32).

An upper bound on the magnitude of the power gains available when the errors in the model (7.32) are non-Gaussian can be obtained by deriving the asymptotic power envelope under the (counterfactual) assumption that the underlying error distribution is known. Assuming the errors are generated by a continuous distribution with density $f(\cdot)$, it follows from the Neyman–Pearson lemma that the

point optimal unit root test against the local alternative $\rho = 1 + T^{-1}c$ rejects for large values of the log likelihood ratio $L_T^f(1 + T^{-1}c) - L_T^f(1)$, where $L_T^f(\cdot)$ is the log likelihood function. Under appropriate smoothness conditions on $f(\cdot)$, the Neyman–Pearson test admits the following quadratic expansion under the unit root hypothesis:

$$\begin{aligned} L_T^f(1 + T^{-1}c) - L_T^f(1) &= \sum_{t=2}^T \log f(\Delta y_t - cT^{-1}y_{t-1}) - \sum_{t=2}^T \log f(\Delta y_t) \\ &= cS_T^f - \frac{1}{2}c^2H_T^f + o_p(1) \end{aligned} \tag{7.33}$$

where $S_T^f = T^{-1} \sum_{t=2}^T y_{t-1} \ell_f(\Delta y_t)$, $H_T^f = \mathcal{I}_{ff} T^{-2} \sum_{t=2}^T y_{t-1}^2$, and $\ell_f(\cdot)$ is a function satisfying $E[\ell_f(\varepsilon_t)] = 0$, $E[\varepsilon_t \ell_f(\varepsilon_t)] = 1$, and $1 \leq \mathcal{I}_{ff} = E[\ell_f(\varepsilon_t)^2] < \infty$.

As the notation suggests, the function $\ell_f(\cdot)$ can be interpreted as a score function and \mathcal{I}_{ff} is the associated Fisher information for location.²³ Jeganathan (1995) gives (absolute continuity and moment) conditions on $f(\cdot)$ under which (7.33) holds with $\ell_f(\varepsilon) = \partial \log f(\varepsilon - \theta) / \partial \theta|_{\theta=0}$, while Jansson (2005) shows that differentiability in quadratic mean, an even weaker condition, is sufficient. By implication, the expansion (7.33) is valid for a wide range of error distributions.

Using (7.33) and the theory of limits of experiments (e.g., Le Cam and Yang, 2000; van der Vaart, 1998), it can be shown that an upper bound on the local asymptotic power of a unit root test (asymptotically of size α) in the model (7.32) is given by the function (of $c \leq 0$)

$$\begin{aligned} \lim_{T \rightarrow \infty} \Pr_{1+T^{-1}c} \left[\mathcal{I}_{ff}^{-1} \left(cS_T^f - \frac{1}{2}c^2H_T^f \right) > k_{\infty}^{z,f}(c) \right] \\ = \Pr \left[c\mathcal{I}_{ff}^{-1/2} \int_0^1 W_c(r) dB_f(r) + \frac{1}{2}c^2 \int_0^1 W_c(r)^2 dr > k_{\infty}^{z,f}(c) \right] \end{aligned} \tag{7.34}$$

where $k_{\infty}^{z,f}(c)$ satisfies $\Pr \left[c\mathcal{I}_{ff}^{-1/2} \int_0^1 W(r) dB_f(r) - \frac{1}{2}c^2 \int_0^1 W(r)^2 dr > k_{\infty}^{z,f}(c) \right] = \alpha$, W and B_f are correlated Wiener processes with coefficient of correlation $\mathcal{I}_{ff}^{-1/2}$, and W_c satisfies the stochastic differential equation $dW_c(r) = cW_c(r)dr + dW(r)$ with initial condition $W_c(0) = 0$.

The upper bound (7.34) depends on the density $f(\cdot)$ through \mathcal{I}_{ff} , which equals one when the error ε_t is perfectly correlated with $\ell_f(\varepsilon_t)$ and is strictly greater than one otherwise.²⁴ Rothenberg and Stock (1997) evaluated (7.34) for various values of \mathcal{I}_{ff} and found large increases in power as \mathcal{I}_{ff} increased. Although this result suggests that non-normality may be an important source of power in unit root testing, a potential problem with the result is that it is derived under the counterfactual assumption the $f(\cdot)$ is known. The upper bound (7.34) is attained by the test which rejects for large values of $cS_T^f - \frac{1}{2}c^2H_T^f$. Following Jansson (2005), adaptation is said to be possible if there exists a pair $(\hat{S}_T^f, \hat{H}_T^f)$, computable without knowledge of $f(\cdot)$, such that $(\hat{S}_T^f, \hat{H}_T^f) = (S_T^f, H_T^f) + o_p(1)$ under the unit root hypothesis. (If adaptation is possible, then the power envelope (7.34) is attained by the test which rejects for large values of $c\hat{S}_T^f - \frac{1}{2}c^2\hat{H}_T^f$.) Jansson (2005) shows that adaptation is possible when $f(\cdot)$ is known to be symmetric, but not in general.

Section 7.3.6, the final subsection of this section, discusses a source of nontrivial power gains which is available in many cases.

7.3.6 Covariates

In most applications of unit root tests, the series $\{y_t\}$ being tested for a unit root is not observed in isolation. Instead, one typically observes at least one time series, say $\{x_t\}$, in addition to the time series $\{y_t\}$ of interest. As observed by Hansen (1995), the additional time series $\{x_t\}$ contains exploitable information about $\{y_t\}$ whenever its order of integration is known.²⁵

As in section 7.3.1, suppose $\{y_t\}$ is generated by the model

$$y_t = \rho y_{t-1} + \varepsilon_t, \tag{7.35}$$

where $y_0 = 0$ and $\varepsilon_t \sim i.i.d. \mathcal{N}(0, 1)$. To accommodate a covariate with a known order of integration, suppose an additional time series $\{x_t : 1 \leq t \leq T\}$ is observed whose generating mechanism is

$$\begin{pmatrix} \varepsilon_t \\ x_t \end{pmatrix} \sim i.i.d. \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \delta \\ \delta & 1 \end{pmatrix}\right), \tag{7.36}$$

where δ is known (and satisfies $|\delta| < 1$).

The log likelihood function $L_T^\delta(\cdot)$ associated with the model (7.35)–(7.36) satisfies the relation

$$L_T^\delta(1 + T^{-1}c) - L_T^\delta(1) = cS_T^\delta - \frac{1}{2}c^2H_T^\delta, \tag{7.37}$$

where $S_T^\delta = T^{-1} \sum_{t=2}^T y_{t-1}(\Delta y_t - \delta x_t)$ and $H_T^\delta = T^{-2} \sum_{t=2}^T y_{t-1}^2$. By the Neyman–Pearson lemma, the test which rejects for large values of $cS_T^\delta - \frac{1}{2}c^2H_T^\delta$ is the point optimal unit root test against the local alternative $\rho = 1 + T^{-1}c$. Unless $\{x_t\}$ is independent of $\{y_t\}$ (in which case δ equals zero), this point optimal test makes use of the information in $\{x_t\}$. By implication, the stationary covariate $\{x_t\}$ contains exploitable information about ρ unless it is independent of $\{y_t\}$.

The magnitude of the gains in local asymptotic power achievable by exploiting the information in the covariate $\{x_t\}$ can be evaluated by deriving the asymptotic power envelope for the model (7.35)–(7.36). The asymptotic power envelope (for unit root tests asymptotically of size α) is given by the function

$$\begin{aligned} & \lim_{T \rightarrow \infty} \Pr_{1+T^{-1}c} \left[cS_T^\delta - \frac{1}{2}c^2H_T^\delta > k_\infty^{z, \delta}(c) \right] \\ &= \Pr \left[c\sqrt{1 - \delta^2} \int_0^1 W_c(r) dV(r) + \frac{1}{2}c^2 \int_0^1 W_c(r)^2 dr > k_\infty^{z, \delta}(c) \right], \end{aligned} \tag{7.38}$$

where $k_\infty^{z, \delta}(c)$ satisfies $\Pr \left[c\sqrt{1 - \delta^2} \int_0^1 W(r) dV(r) - \frac{1}{2}c^2 \int_0^1 W(r)^2 dr > k_\infty^{z, \delta}(c) \right] = \alpha$. W and V are correlated Wiener processes with coefficient of correlation $\sqrt{1 - \delta^2}$,

and W_c satisfies the stochastic differential equation $dW_c(r) = cW_c(r)dr + dW(r)$ with initial condition $W_c(0) = 0$. The functional form of the power envelope (7.38) is exactly the same as the functional form of the (infeasible) power envelope (7.34) associated with (known) non-Gaussian error distributions, the quantity $\sqrt{1 - \delta^2}$ playing the same role here as $\mathcal{I}_{\bar{\eta}}^{-1/2}$ did there. As a consequence, appreciable power gains are available whenever “good” covariates (having δ^2 moderately large) can be found.

The asymptotic power envelope (7.38) was derived by Hansen (1995). In addition to (implicitly) proposing a family of point optimal tests, Hansen (1995) proposed a regression-based unit root test, the covariate augmented Dickey–Fuller (CADF) test. The CADF test is “nearly efficient” in the model (7.35)–(7.36). If the model is extended to include deterministic components, however, the local asymptotic power of the CADF test is well below the asymptotic power envelope for invariant tests (Elliott and Jansson, 2003). In contrast, the point optimal tests of Elliott and Jansson (2003) are “nearly efficient” in the presence of deterministic components.

7.4 Conclusion

This chapter has reviewed recent advances in the literature on unit root testing, emphasizing developments aimed at reducing size distortions and/or boosting the power of unit root tests. As should be apparent from the discussion herein, significant advances in both directions have been made during the decade since the publication of Stock (1994). The literature now seems to have reached a relatively mature state and it is difficult to predict if any major advances will occur over the next decade. Nevertheless, it seems worth pointing out two potential shortcomings of the existing body of knowledge. In relation to the first of the two main themes of the present survey, size distortions, it remains an open question whether the use of refined asymptotic approximations can enhance the theoretical understanding of the properties of the bootstrap and/or guide the choice among asymptotically equivalent testing procedures. On the power front, it would appear to be useful to further investigate the extent to which non-Gaussianity and/or extraneous information (other than information about the integration properties of observed covariates, as in section 7.3.6) can be an exploitable source of power in unit root testing applications.

Notes

1. Under the stated assumptions regarding u_t , the long-run variance ω^2 equals $2\pi f_u(0)$, where $f_u(\cdot)$ is the spectral density of u_t .
2. As shown by Chang and Park (2002), the ARMA (p, q) assumption on u_t can actually be replaced by the weaker assumption that $\sum_{j=0}^{\infty} \psi_j z^j \neq 0$ for every $|z| \leq 1$ (and $\sum_{j=0}^{\infty} j|\psi_j| < \infty$).

3. A similarly modified form of the BIC is

$$MBIC(k) = \log \hat{\sigma}_k^2 + \log(T - k_{\max})(\tau_T(k) + k)/(T - k_{\max}).$$

4. When $\{y_{it}\}$ is generated by (7.4), the estimator $\hat{\omega}_{KER}^2$ is consistent if $M_T^{-1} + T^{-1/2}M_T = o(1)$ and $\omega(\cdot)$ satisfies the conditions of Jansson (2002).
5. Also, Kim and Schmidt (1990) demonstrate, by means of Monte Carlo simulations, that a range of different kernel estimators and different ways of selecting the bandwidth parameter seem to deliver unit root tests with fairly similar finite sample properties.
6. Haldrup, Montanés, and Sanso (2005) show that the presence of additive and other types of outliers (as well as measurement errors) has implications for the (moving average) serial correlation structure of the data, so Vogelsangs's (1996) results are consistent with Perron and Ng's (1996) Monte Carlo results on the behavior of the M -tests in the presence of MA errors.
7. The class of linear-in-parameters specifications also includes structural break models with a known break date. In contrast, structural break models with an unknown break date do not belong to the class of linear-in-parameters specifications. For a survey of structural break models, see Perron (2005).
8. The following discussion draws on Stock (1994).
9. More generally, the power envelope for a class Φ_T of tests functions is given by $\sup_{\phi_T \in \Phi_T} E_\rho \phi_T(Y_T)$. In (7.23), "sup" has been replaced with "max" in recognition of the fact that the sup is attained.
10. It follows from (7.24) and the properties of exponential families (e.g., Lehmann, 1994) that the model admits a two-dimensional minimal sufficient statistic whose distribution belongs to a curved exponential family (Efron, 1975). Therefore, the functional form of a most powerful test of a simple null against a simple alternative in the known-variance, zero-mean Gaussian AR(1) model depends on the alternative, implying in that one-sided testing problems do not admit UMP size α tests. The limiting experiment associated with a simple null hypothesis on ρ with $|\rho| < 1$ corresponds to a full exponential family model (the log likelihood ratios are locally asymptotically normal (Le Cam, 1960) and therefore admits a UMP size α test. In contrast, the limiting experiment associated with the unit root hypothesis corresponds to a curved exponential family model (the log likelihood ratios are locally asymptotically quadratic (Jeganathan, 1995), but not locally asymptotically (mixed) normal) and does not admit a UMP size α test.
11. For details on local asymptotic power and local-to-unity asymptotics, see Stock (1994) and the references therein [e.g., Chan Wei (1987) and Phillips (1987b)].
12. A proof of (7.27) can be based on the inequality

$$E_{1+T^{-1}c} \phi_T(Y_T) \leq \Pr \left[cS_T - \frac{1}{2}c^2H_T > k_T^{zT}(1 + T^{-1}c) \right], \quad \alpha_T = E_1 \phi_T(Y_T),$$

and the fact that $(S_T - cH_T, H_T) \rightarrow_{d_c} (\int_0^1 W_c(r)dW(r), \int_0^1 W_c(r)^2 dr)$, where \rightarrow_{d_c} signifies convergence in distribution when $\rho = 1 + T^{-1}c$.

13. Local optimality of S_T follows from a Neyman–Pearson argument and the fact that

$$\left. \frac{d}{dc} E_{1+T^{-1}c} \phi_T(Y_T) \right|_{c=0} = T^{-1} \left. \frac{d}{d\rho} E_\rho \phi_T(Y_T) \right|_{\rho=1} = E_1 [S_T \phi_T(Y_T)]$$

for any test function $\phi_T(\cdot)$.

14. Following Davies (1969), point optimal testing procedures are sometimes referred to as beta-optimal testing procedures.

15. The (signed) likelihood ratio test rejects for large values of

$$\max_{\rho \leq 1} L_T(\rho) - L_T(1) = \frac{\min(S_T, 0)^2}{2H_T},$$

a decreasing function of $S_T/\sqrt{H_T}$.

16. In fact, it appears to be unknown if the Dickey–Fuller (1979) t -test is even admissible, in the sense that there does not exist a unit root test with uniformly superior local asymptotic power. Studying a general class of models with locally asymptotically quadratic log likelihood ratios, Ploberger (2004) gives a complete class result for a two-sided testing problem and shows that the likelihood ratio test is not a member of his (essentially) complete class of tests.
17. Elliott, Rothenberg, and Stock (1992) obtained the expansion (7.29) under the assumption that $\{y_t\}$ is generated by a Gaussian AR model of finite order. Using ingenious arguments, Elliott, Rothenberg, Stock (1996) established (7.29) for model (7.28) under the additional assumption that $\sum_{j=1}^{\infty} j|\psi_j| < \infty$. The latter assumption can be removed by using a slightly modified version of the proof employed by Elliott, Rothenberg, and Stock (1996).
18. The existence of such estimators follows from Jansson (2002) who shows that standard kernel estimators of ω^2 [e.g., Newey West (1987), Andrews (1991)] are consistent under the assumptions of this subsection.
19. A formal proof of this claim can be based on Lehmann (1994, Theorem 6.3) and the fact that ρ is a maximal invariant under the group of transformations of the form $\tilde{g}_b(\rho, \beta)$, where $b \in \mathbb{R}^k$.
20. Asymptotic similarity of the tests discussed in section 7.3.3 follows from the fact that the null distribution of $\{y_t\}$ does not depend on κ . The same fact implies that the test which rejects for large values of $\max_{\beta} L_T^{\kappa}(1 + T^{-1}c, \beta) - \max_{\beta} L_T^{\kappa}(1, \beta)$ is a point optimal unit root test even if κ is treated as an unknown nuisance parameter. By implication, the power envelope for the model in which κ is treated as an unknown nuisance parameter coincides with the family (indexed by κ) of power envelopes derived under the assumption that κ is a known constant.
21. The following discussion draws on Jansson (2005).
22. This invariance result follows from Donsker’s theorem (e.g., Billingsley, 1999) and the continuous mapping theorem.
23. Indeed, $\ell_f(\cdot)$ is the score function, evaluated at $\theta = 0$, of the location model $X_i = \theta + \varepsilon_i$, where the errors $\{\varepsilon_i\}$ are *i.i.d.* with density function $f(\cdot)$.
24. The correlation between ε_t and $\ell_f(\varepsilon_t)$ is unity when the underlying distribution is Gaussian, the score function of the standard normal location model being $\ell_f(\varepsilon) = \varepsilon$.
25. An important example of a unit root testing problem in which a covariate with a known order of integration is observed is the problem of testing for absence of cointegration when the (potentially) cointegrating vector is prespecified (Elliott, Jansson, and Pesavento, 2005; Zivot, 2000).

References

- Agiakloglou, C. and P. Newbold (1992) Empirical evidence on Dickey–Fuller type tests. *Journal of Time Series Analysis* 13, 471–83.
- Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–58.
- Andrews D.W.K. and W. Ploberger (1994) Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383–1414.
- Baillie, R. (1996) Long memory processes and fractional integration in econometrics. *Journal of Econometrics* 73, 6–59.

- Berk, K.N. (1974) Consistent autoregressive spectral estimates. *Annals of Statistics* 2, 489–502.
- Bhargava, A. (1986) On the theory of testing for unit roots in observed time series. *Review of Economic Studies* 53, 369–84.
- Billingsley, P. (1999) *Convergence of Probability Measures*, 2nd edn. New York: Wiley.
- Chan, N.H. and C.Z. Wei (1987) Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics* 15, 1050–1063.
- Chang, Y. and J.Y. Park (2002) On the asymptotics of ADF tests for unit roots. *Econometric Reviews* 21, 431–47.
- Choi, I. (2005) “Nonstationary Panels,” Palgrave Handbooks of Econometrics: Vol. 1, forthcoming.
- Davidson, R. and J. MacKinnon (2005) Bootstrap methods in econometrics, Palgrave Handbooks of Econometrics: Vol. 1, forthcoming.
- Davies, R.B. (1969) Beta-optimal tests and an application to the summary evaluation of experiments. *Journal of the Royal Statistical Society, Series B* 31, 524–38.
- DeJong, D., J. Nankervis, N. Savin and C. Whiteman (1992a) Integration versus trend-stationarity in macroeconomic time series. *Econometrica* 60, 423–34.
- DeJong, D., J. Nankervis, N. Savin and C. Whiteman (1992b) The power problems of unit root tests for time series with autoregressive errors. *Journal of Econometrics* 53, 323–43.
- 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, 427–31.
- Dickey, D.A. and S.G. Pantula (1987) Determining the order of differencing in autoregressive processes. *Journal of Business and Economic Statistics* 5, 455–61.
- 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, 115–43.
- Efron, B. (1975) Defining the curvature of a statistical problem (with applications to second order efficiency). *Annals of Statistics* 3, 1189–1242.
- Elliott, G. (1999) Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution. *International Economic Review* 40, 767–83.
- Elliott, G. and M. Jansson (2003) Testing for unit roots with stationary covariates. *Journal of Econometrics* 115, 75–89.
- Elliott, G., M. Jansson and E. Pesavento (2005) Optimal power for testing potential cointegrating vectors with known parameters for nonstationarity. *Journal of Business and Economic Statistics* 23, 34–48.
- Elliott, G., T.J. Rothenberg and J.H. Stock (1992) Efficient tests for an autoregressive unit root. NBER Technical Working Paper No. 130.
- Elliott, G., T.J. Rothenberg and J.H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–36.
- Engle, R.F. and C.W.J. Granger (1987) Cointegration and error correction: representation, estimation, and testing. *Econometrica* 55, 251–76.
- Franses, P. (1996) *Periodicity and Stochastic Trends in Economic Time Series*. Oxford: Oxford University Press.
- Franses, P.H. and N. Haldrup (1994) The effects of additive outliers on tests for unit roots and cointegration. *Journal of Business and Economic Statistics* 12, 471–8.
- Ghysels, E. and D. Osborn (2001) *The Econometric Analysis of Seasonal Time Series*. Cambridge, UK: Cambridge University Press.
- Granger, C.W.J. (1981) Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* 16, 121–30.
- Granger, C.W.J. and R. Joyeux (1980) An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis* 1, 15–29.
- Haldrup, N. (1998) An econometric analysis of I(2) variables. *Journal of Economic Surveys* 12, 595–650.
- Haldrup, N., A. Montanés and A. Sanso (2005) Measurement errors and outliers in seasonal unit root testing. *Journal of Econometrics* 127, 103–28.

Please update.

- Hansen, B.E. (1995) Rethinking the univariate approach to unit root testing: using covariates to increase power. *Econometric Theory* 11, 1148–1171.
- Hylleberg, S., R.F. Engle, C.W.J. Granger and B. Yoo (1990) Seasonal integration and cointegration. *Journal of Econometrics* 44, 215–38.
- Im, K.S., M.H. Pesaran and Y. Shin (2003) Testing for unit roots in heterogeneous panels. *Journal of Econometrics* 115, 53–74.
- Jansson, M. (2002) Consistent covariance matrix estimation for linear processes. *Econometric Theory* 18, 1449–1459.
- Jansson, M. (2004) Stationarity testing with covariates. *Econometric Theory* 20, 56–94.
- Jansson, M. (2005) Semiparametric power envelopes for tests of the unit root hypothesis. Manuscript, UC Berkeley.
- Jeganathan, P. (1995) Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* 11, 818–87.
- Kim, K. and P. Schmidt (1990) Some evidence on the accuracy of Phillips-Perron tests using alternative estimates of nuisance parameters. *Economics Letters* 34, 345–50.
- 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.
- Kwiatkowski D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992) Testing the null hypothesis of stationarity against the alternative of a unit Root: how sure are we that economic time series have a unit root? *Journal of Econometrics* 54, 159–78.
- Le Cam, L. (1960) Locally asymptotically normal families of distributions. *University of California Publications in Statistics* 3, 37–98.
- Le Cam, L. and G.L. Yang (2000) *Asymptotics in Statistics: Some Basic Concepts*. 2nd edn. New York: Springer-Verlag.
- Lehmann, E.L. (1994) *Testing Statistical Hypotheses*. 2nd edn. New York: Chapman and Hall.
- Levin, A., C.-F. Lin and C.-S.J. Chu (2002) Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics* 108, 1–25.
- Maddala, G. and I.M. Kim (1998) *Unit Roots, Cointegration and Structural Change*. Cambridge: Cambridge University Press.
- Müller, U.K. and G. Elliott (2003) Tests for unit root and the initial condition. *Econometrica* 71, 1269–1286.
- Nabeya, S. and K. Tanaka (1990) Limiting power of unit-root tests in time-series regression. *Journal of Econometrics* 46, 247–71.
- Newey, W.K. and K.D. West (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–8.
- Ng, S. and P. Perron (1995) Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association* 90, 268–81.
- Ng, S. and P. Perron (2001) Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69, 1519–1554.
- Paparoditis, E. and D.N. Politis (2003) Residual-based block bootstrap for unit root testing. *Econometrica* 71, 813–55.
- Park, J.Y. (2002) An invariance principle for sieve bootstrap in time series. *Econometric Theory* 18, 469–90.
- Park, J.Y. (2003) Bootstrap unit root tests. *Econometrica* 71, 1845–1895.
- Perron, P. (1989) The great crash, the oil price shock and the unit root Hypothesis. *Econometrica* 57, 1361–1401.
- Perron, P. (2005) “Dealing with structural breaks.” *Palgrave Handbooks of Econometrics*: vol. 1, forthcoming.
- Perron P. and S. Ng (1996) Useful modifications to unit root tests with dependent errors and their local asymptotic properties. *Review of Economic Studies* 63, 435–65.

- Perron, P. and S. Ng (1998) An autoregressive spectral density estimator at frequency zero for nonstationarity tests. *Econometric Theory* 14, 560–603.
- Phillips, P.C.B. (1987a): Time series regression with a unit root. *Econometrica* 55, 277–301.
- Phillips, P.C.B. (1987b): Towards a unified asymptotic theory for autoregression. *Biometrika* 74, 535–47.
- Phillips, P.C.B. and S. Ouliaris (1990) Asymptotic properties of residual based tests for cointegration. *Econometrica* 58, 165–93.
- Phillips, P.C.B. and P. Perron (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335–46.
- Phillips, P.C.B. and Z. Xiao (1998) A primer on unit root testing. *Journal of Economic Surveys* 12, 423–69.
- Ploberger, W. (2004) A complete class of tests when the likelihood is locally asymptotically quadratic. *Journal of Econometrics* 118, 67–94.
- Rothenberg, T.J. and J.H. Stock (1997) Inference in a Nearly Integrated Autoregressive Model with Nonnormal Innovations. *Journal of Econometrics* 80, 269–86.
- Said, S.E. and D.A. Dickey (1984) Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71, 599–607.
- Saikkonen, P. and R. Luukkonen (1993a): Point optimal tests for testing the order of differencing in ARIMA models. *Econometric Theory* 9, 343–62.
- Saikkonen, P. and R. Luukkonen (1993b): Testing for a moving average unit root in autoregressive integrated moving average models. *Journal of the American Statistical Association* 88, 596–601.
- Sargan, J. and A. Bhargava (1983) Testing for residuals from least squares regression being generated by Gaussian random walk. *Econometrica* 51, 153–74.
- Schwert, G.W. (1989) Test for unit roots: a Monte Carlo investigation. *Journal of Business and Economic Statistics*, 7, 147–60.
- Stock, J.H. (1994) Unit roots, structural breaks and trends. In R.F. Engle and D.L. McFadden (eds), *Handbook of Econometrics, Volume IV*. New York: North Holland, pp. 2739–2841.
- Stock, J.H. (1999) A class of tests for integration and cointegration. In *Cointegration, Causality, and Forecasting: A Festschrift for Clive W.J. Granger*. Oxford: Oxford University Press, pp. 135–67.
- van der Vaart, A.W. (1998) *Asymptotic Statistics*. Cambridge: Cambridge University Press.
- Velasco, C. (2005) Semi-Parametric Estimation of Long Memory Models, Palgrave Handbooks of Econometrics: Vol. 1, forthcoming.
- Vogelsang, T.J. (1999) Two simple procedures for testing for a unit root when there are additive outliers. *Journal of Time Series Analysis* 20, 237–52.
- White, J.S. (1958) The limiting distribution of the serial correlation coefficient in the explosive case. *Annals of Mathematical Statistics* 29, 1188–1197.
- Zivot, E. (2000) The power of single equation tests for cointegration when the cointegrating vector is prespecified. *Econometric Theory* 16, 407–39.