Chapter 13

UNIT ROOT NONSTATIONARY PROCESSES

This chapter concerns univariate stochastic processes. Since the seminal work of Nelson and Plosser (1982), much theoretical and empirical research has been done in the area of unit root nonstationarity. They found that the null hypothesis of unit root nonstationarity was not rejected for many macroeconomic series. When one or more variables of interest are unit root nonstationary, standard asymptotic distribution theory does not apply to the econometric system involving these variables. The spurious regression results discussed in Section 14.2 are concrete examples of this type of problem.

When a variable is unit root nonstationary, it has a stochastic trend. If linear combinations of two or more unit root nonstationary variables do not contain stochastic trends, then these variables are said to be cointegrated. Then the cointegrating vector, which eliminates the stochastic trends, can be estimated consistently by regressions without the use of instrumental variables, even when no variables are exogenous. If the cointegrating vector includes structural parameters, then the econometrician can estimate these structural parameters without making exogeneity assumptions.[1]

The rest of this chapter is organized as follows. In Section 13.1, univariate unit root econometrics is discussed. It begins with definitions of basic concepts such as difference stationarity and trend stationarity. Then a decomposition of a difference stationary variable into a deterministic trend, a stochastic trend, and a stationary component is discussed. Spurious regression results, tests for the null of difference stationarity, and tests for the null of stationarity are reviewed.

13.1 Definitions

Consider a univariate stochastic process, $\{x_t : t = \cdots, -2, -1, 0, 1, 2, \cdots\}$, which is a sequence of random variables. Many macroeconomic variables tend to grow over time, so that their distributions shift upward over time. Hence they are not stationary. However, there are many possible forms of nonstationarity, and it is not clear which form of nonstationarity is appropriate in representing macroeconomic variables. It may be reasonable to assume that the growth rate or the first difference of the (natural) log of a variable is stationary for many macroeconomic variables. Let us now assume that the first difference of x_t ($\Delta x_t = x_t - x_{t-1}$) is stationary. Then x_t is either difference stationary or trend stationary. If x_t is stationary after removing a deterministic time trend, then x_t is said to be trend stationary. Since Δx_t is assumed to be stationary, x_t has a linear time trend when x_t is trend stationary:

$$x_t = \theta + \mu t + \epsilon_t, \tag{13.1}$$

[1] Stock and Watson (1988), Diebold and Nerlove (1990), Campbell and Perron (1991), and Watson (1994) are examples of surveys for unit root econometrics.

where ϵ_t is stationary with mean zero.[2] If Δx_t is stationary but x_t is not trend stationary, then x_t is said to be difference stationary. Alternatively, it is called unit root nonstationary or integrated of order one. The trend stationary and difference stationary processes have different properties on their long-run variances. The long-run variance of a stationary variable y_t is defined by

$$\omega^2 = \sum_{\tau=-\infty}^{\infty} E\{[y_t - E(y_t)][y_{t-\tau} - E(y_t)]\}. \tag{13.2}$$

After taking the first difference, a difference stationary process has a positive long-run variance, while trend stationary process has a long-run variance of zero.

A special case of a difference stationary process is a random walk. If $E(x_{t+1}|\ x_t, x_{t-1}, x_{t-2}, \cdots) = x_t$ and if $E((\Delta x_{t+1})^2|x_t, x_{t-1}, x_{t-2}, \cdots)$ is constant over time, then x_t is a random walk. In general, if x_t is difference stationary, then Δx_t has nonzero serial correlation; however, if x_t is a random walk, then Δx_t does not have serial correlation.

13.2 Decompositions

It is often convenient to decompose a difference stationary process into components representing a deterministic trend, a stochastic trend, and a stationary component.

Let x_t be a difference stationary process:

$$x_t - x_{t-1} = \mu + \epsilon_t \tag{13.3}$$

for $t \geq 1$ where ϵ_t is stationary with mean zero. Here μ is called a drift, which is the

[2]Note that ϵ_t is not assumed to be *iid* because serial correlation is allowed in a stationary process.

mean of Δx_t. Then

$$x_t = \mu + x_{t-1} + \epsilon_t \tag{13.4}$$
$$= 2\mu + x_{t-2} + \epsilon_{t-1} + \epsilon_t$$
$$= 3\mu + x_{t-3} + \epsilon_{t-2} + \epsilon_{t-1} + \epsilon_t$$
$$= \cdots$$
$$= \mu t + x_0 + \sum_{\tau=1}^{t} \epsilon_\tau .$$

Hence

$$x_t = \mu t + x_t^0, \tag{13.5}$$

where x_t^0 is

$$x_t^0 = x_0 + \sum_{\tau=1}^{t} \epsilon_\tau , \tag{13.6}$$

where x_0 is an initial value. Relation (13.5) decomposes the difference stationary process x_t into a deterministic trend arising from drift μ, and the difference stationary process without drift, x_t^0.

Let us now consider Beveridge and Nelson (1981) decomposition, which further decompose x_t^0 into a random walk component and a stationary component. Since Δx_t^0 is covariance stationary, it has the Wold representation:

$$(1 - L)x_t^0 = A(L)\nu_t, \tag{13.7}$$

where L is the lag operator, $A(L) = \sum_{\tau=0}^{\infty} A_\tau L^\tau$, $A_0 = 1$, $\nu_t = x_t^0 - \hat{E}(x_t^0 | x_{t-1}^0, x_{t-2}^0, \cdots)$, and $\hat{E}(\cdot | x_{t-1}^0, x_{t-2}^0, \cdots)$ is the linear projection operator. Then

$$x_t^0 = z_t + c_t, \tag{13.8}$$

where

$$z_t = z_{t-1} + A(1)\nu_t, \tag{13.9}$$

is the random walk component or a stochastic trend, and

$$c_t = -\{(\sum_{\tau=1}^{\infty} A_\tau)\nu_t + (\sum_{\tau=2}^{\infty} A_\tau)\nu_{t-1} + (\sum_{\tau=3}^{\infty} A_\tau)\nu_{t-2} + \cdots\} \tag{13.10}$$

is the stationary component of x_t. Thus a difference stationary process x_t is decomposed into a deterministic trend, a stochastic trend, and a stationary component.

The variance of the random walk component, $\text{Var}(\Delta z_t)$, is equal to $A(1)^2\text{Var}(\nu_t)$, which in turn is equal to the long-run variance of Δx_t and 2π times the spectral density of Δx_t at frequency zero. If the long-run variance is zero, then $x_t = \mu t + c_t$, and x_t is trend stationary.

Cochrane (1988), among others, uses $\frac{\text{Var}(\Delta z_t)}{\text{Var}(\Delta x_t)}$ as a measure of the persistence of x_t. This measure is zero for trend stationary x_t and is one for a random walk. He estimates $\text{Var}(\Delta z_t)$ by $\frac{1}{k}$ times the variance of k-differences of x_t, $\frac{1}{k}Var(\Delta^k x_t)$, for a large enough k. His estimator is essentially the same as the Bartlett estimator, which was advocated by Newey and West (1987) in a different context. Any estimator of the long-run variance or the spectral density at frequency zero can be used for the purpose of estimating Cochrane's measure of persistence.

13.3 Tests for the Null of Difference Stationarity

This section explains Dickey-Fuller (1979), Said-Dickey (1984), Phillips-Perron (1988), and Park's (1990) tests for the null of difference stationarity. More recent work to improve small sample properties of tests includes Kahn and Ogaki (1990), Elliott, Rothenberg, and Stock (1996), and Hansen (1993).

13.3.1 Dickey-Fuller Tests

Dickey and Fuller (1979) propose to test for the null of a unit root in an AR(1) model:[3]

$$x_t = \theta + \mu t + \alpha x_{t-1} + \epsilon_t. \tag{13.11}$$

where ϵ_t is NID. One of their tests is based on $T(\hat{\alpha}-1)$, where T is the sample size and $\hat{\alpha}$ is the OLS estimator for α in (13.11). Another test is based on the t-ratio for the hypothesis $\alpha = 1$. These test statistics do not have standard distributions. Depending on whether or not a constant and a linear time trend are included, distributions of these tests under the null are different.[4] Fuller (1976, Tables 8.5.1 and 8.5.2) tabulates critical values for Dickey-Fuller tests.

Whether or not a constant and a linear time trend should be included in the regression depends on what type of alternative is appropriate. If the alternative hypothesis is that x_t is stationary with mean zero, then no deterministic terms should be included. This alternative is not appropriate for most macroeconomic time series. If the alternative hypothesis is that x_t is stationary with unknown mean, then a constant should be included. This alternative is appropriate for the time series that exhibit no consistent tendency to grow (or shrink) over time. On the other hand, if the alternative is that x_t is trend stationary, then a constant and a linear time trend should be included. This alternative is appropriate for the time series that exhibit a consistent tendency to grow (or shrink) over time. When these test statistics are

[3]It should be noted that Dickey and Fuller's (1981) joint tests with deterministic terms can have significantly lower power than Dickey and Fuller's (1979) one-tailed single unit root tests as explained by Park (1989).

[4]If the data are demeaned prior to the regression, then the test statistics have the same distributions as those from the regression with a constant in (13.11). If the data are detrended prior to the regression, then the test statistics have the same distributions as those from the regression with a constant and a linear time trend.

negative and greater than the appropriate critical value in absolute value, then the null of a unit root is rejected in favor of one of these alternatives.

Dickey-Fuller tests assume that the econometrician knows the order of autoregression. The following tests treat the case of unknown order of autoregression.

13.3.2 Said-Dickey Test

Said and Dickey (1984) extend the Dickey-Fuller's t-ratio test to the case where the order of autoregression is unknown. Consider an AR process of order p:

$$x_t = \theta + \mu t + a_1 x_{t-1} + a_2 x_{t-2} + \cdots + a_p x_{t-p} + \nu_t. \tag{13.12}$$

We assume that this process' autoregressive roots are less than or equal to one in absolute value, and that there is at most one root whose absolute value is equal to one. If there is a root with absolute value equal to one, then the root is assumed to be one, so that the process is unit root nonstationary. It should be noted that the null hypothesis that $a_1 = 1$ in (13.12) does not have anything to do with the unit root hypothesis if $p > 1$. The unit root hypothesis is concerned with the autoregressive roots, and not with autoregressive coefficients. The first order autoregressive coefficient is equal to the autoregressive root only for an AR(1) process. For the purpose of testing for a unit root, it is convenient to reparameterize (13.12) as follows:[5]

$$\Delta x_t = \theta + \mu t + \rho x_{t-1} + \beta_1 \Delta x_{t-1} + \cdots + \beta_{p-1} \Delta x_{t-p+1} + \nu_t, \tag{13.13}$$

where

$$\rho = -(1 - a_1 - a_2 - \cdots - a_p), \tag{13.14}$$

[5] For example, consider an AR(2) process. Rearranging (13.12) yields $x_t - x_{t-1} = \theta + \mu t - (1 - a_1 - a_2)x_{t-1} - a_2(x_{t-1} - x_{t-2}) + \nu_t$. Therefore, we obtain $\Delta x_t = \theta + \mu t + \rho x_{t-1} + \beta_1 \Delta x_{t-1} + \nu_t$, where $\rho = -(1 - a_1 - a_2)$ and $\beta_1 = -a_2$.

and

$$\beta_i = -[a_{i+1} + a_{i+2} + \cdots + a_p] \quad \text{for} \quad i = 1, 2, \cdots, p-1. \tag{13.15}$$

With this reparameterization, Δx_t has an invertible autoregressive representation when $\rho = 0$. Hence x_t is unit root nonstationary if and only if $\rho = 0$, and one can test the null hypothesis of unit root nonstationarity by testing $\rho = 0$. Said and Dickey show that the t-ratio for the hypothesis $\rho = 0$ has the same asymptotic distribution as the Dickey-Fuller t-ratio test. Some authors call this test the augmented Dickey-Fuller (ADF) test while others reserve the word ADF for the corresponding cointegration test. A constant and a linear time trend are included or excluded according to the appropriate alternative hypothesis as before.

In many applications, the Said-Dickey test results are very sensitive to the choice of the order of autoregression, p. Ng and Perron (1995) analyze the choice of truncation lag, and categorize the existing methods into two rules: rules of thumb and data dependent rules. The former includes fixing p regardless of the sample size, T, or choosing p as a fixed function of T according to

$$p = int\{c(\frac{T}{100})^{\frac{1}{d}}\}, \tag{13.16}$$

where $c = 4, 12$ and $d = 4$ are used in Schwert (1989). The latter includes information-based rules such as Akaike information criterion (AIC) and Schwartz information criterion (SIC) according to

$$I_p = log\hat{\sigma}_p^2 + p\frac{C_T}{T}, \tag{13.17}$$

where $\hat{\sigma}_p^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\nu}_t^2$, and $C_T = 2$ for AIC and $C_T = logT$ for SIC. Sequential tests for the significance of the coefficients on lags also fall into this category. Based on

Hall's (1994) work, Campbell and Perron (1991) recommend starting with a reasonably large value of p that is chosen a priori and decrease p until the coefficient on the last included lag is significant.[6] Ng and Perron (1995) show that rules of thumb are dominated by data dependent rules. They also show that general-to-specific sequential tests are better than information-based rules since the latter has severe size distortion.

13.3.3 Phillips-Perron Tests

Phillips (1987) and Phillips and Perron (1988) use a nonparametric method to correct for serial correlation of ϵ_t. Their modification of the Dickey-Fuller $T(\hat{\alpha} - 1)$ test is called $Z(\alpha)$ test, while their modification of the Dickey-Fuller t-ratio test is called $Z(t)$ test. These corrections are based on a nonparametric estimate of the long run variance of ϵ_t. See Chapter 6 for a discussion of nonparametric estimation methods. Phillips-Perron tests are constructed so that they have the same asymptotic distributions as corresponding Dickey-Fuller tests.

An advantage of the Phillips-Perron tests over the Said-Dickey test is that they tend to be more powerful as shown in the Monte Carlo experiments of Phillips and Perron. A drawback of the Phillips-Perron tests is that they are subject to more severe size distortions than the Said-Dickey test (see Monte Carlo results of Phillips and Perron, 1988; Schwert, 1989). Size distortion exists when the actual size of a test in small samples is very different from the size of the test indicated by asymptotic theory. Such differences are due to approximations involved in the asymptotic theory.

[6] According to Hall (1994), compared to general-to-specific rules, specific-to-general rules are not generally asymptotically valid.

Table 13.1: Critical Values of Park's $J(p,q)$ Tests for the Null of Difference Stationarity

Size	J(0,3)	J(1,5)	J(2,6)	J(3,8)	J(4,10)	J(5,11)
.010	.1118	.1228	.0886	.1093	.1348	.1157
.025	.2072	.1977	.1409	.1684	.1974	.1652
.050	.3385	.2950	.2050	.2394	.2660	.2210
.100	.5773	.4520	.3101	.3425	.3642	.3076
.150	.8042	.5959	.4034	.4299	.4516	.3800
.200	.9243	.7326	.4968	.5177	.5335	.4470

Source: Park and Choi's (1988) Table 1-B.

13.3.4 Park's J Tests

Park's (1990) J tests based on a variable addition method were originally proposed by Park and Choi (1988). These tests are based on spurious regression results. Consider a regression

$$x_t = \sum_{\tau=0}^{p} \mu_\tau t^\tau + \sum_{\tau=p+1}^{q} \mu_\tau t^\tau + \eta_t. \tag{13.18}$$

Here the maintained hypothesis is that x_t possesses the deterministic time polynomials up to the order of p (typically, p is zero or one). The additional time polynomials are spurious time trends. Let $F(p,q)$ be the standard Wald test statistic (without any correction for serial correlation of η_t) for the null hypothesis $\mu_{p+1} = \cdots = \mu_q = 0$. Under the null hypothesis that η_t is unit root nonstationary, spurious regression results imply that $F(p,q)$ explodes, but $\frac{1}{T}F(p,q)$ has an asymptotic distribution. The $J(p,q)$ test is defined as $\frac{1}{T}F(p,q)$. The null hypothesis of difference stationarity is rejected against the alternative of trend stationarity when $J(p,q)$ is *small* because $J(p,q)$ converges to zero under the alternative hypothesis of trend stationarity. Part of Park and Choi's table of critical values for J tests are reproduced in Table 13.1 for convenience.

The $J(p,q)$ tests do not require the estimation of the long-run variance of η_t, and thus have an advantage over the Said-Dickey and Phillips-Perron tests in that neither the order of autoregression nor the lag truncation number need to be chosen. Park and Choi's Monte Carlo experiments show that J tests have relatively stable sizes and are not dominated by Said-Dickey and Phillips-Perron tests in terms of size-adjusted power.

13.4 Testing the Null of Stationarity

In some cases, it is useful to test the null of stationarity (or trend stationarity) rather than the null of difference stationarity. For example, if an econometrician plans to apply econometric theory that assumes stationarity, a natural procedure is to test the null of stationarity rather than test the null of difference stationarity. Tests for the null of stationarity will also lead to tests for the null of cointegration as will be discussed in Chapter 14. However, most of the tests in the unit root literature take the null of a unit root rather than the null of stationarity. Only recently, Fukushige, Hatanaka, and Koto (1994), Kahn and Ogaki (1992), Kwiatkowski, Phillips, Schmidt, and Shin (1992), Bierens and Guo (1993), and Choi and Ahn (1999) among others have developed tests for the null of stationarity.

Park's (1990) G tests for the null of stationarity were first developed by Park and Choi's (1988). These tests, which have been used in empirical work by several researchers, are based on the same spurious regression results as Park's J tests. With the notations in Section 13.3.4, $G(p,q) = F(p,q)\frac{\hat{\sigma}^2}{\hat{\omega}^2}$, where $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\eta_t}^2$, $\hat{\omega}^2$ is an estimate of the long-run variance of η_t, and $\hat{\eta}_t$ is the estimated residual in regression (13.18). Under the null that x_t is stationary after removing the maintained deterministic time terms of time polynomial of order p, the $G(p,q)$ test statistic has asymptotic chi-square distribution with the $q - p$ degrees of freedom. Under the alternative hypothesis that x_t is difference stationary (after removing the maintained deterministic terms), the $G(p, q)$ statistic diverges to infinity. This result is due to the spurious regression result that time polynomials tend to mimic a stochastic trend.

Unlike Park's J tests, Park's G tests require estimation of the long-run variance. Kahn and Ogaki's (1992) Monte Carlo experiments on Park's G tests suggest that it is advisable to use relatively small q when the sample size is small and not to use the prewhitening method discussed in Section 6.2.

13.5 Near Observational Equivalence

Most of the tests described in sections 13.3 and 13.4 seek to discriminate between difference stationary and trend stationary processes. In the finite samples that we observe, there is a conceptual difficulty with this task. In finite samples, any difference stationary process can be approximated arbitrarily well by a series of trend stationary processes. This evaluation can be done by driving the dominant autoregressive root of trend stationary processes to one from below. After all, it is very difficult to discriminate between the dominant autoregressive root of 0.999 and that of one. This type of problem exists for virtually any hypothesis testing. Hypothesis testing for unit root nonstationarity is special because the opposite is also true: any trend stationary process can be approximated arbitrarily well by a series of difference stationary processes. This approximation can be done by driving the long-run variance of the first difference of difference stationary processes to zero. Some authors call this problem the near observational equivalence problem (see, e.g., Cochrane, 1988; Campbell and Perron, 1991; Christiano and Eichenbaum, 1991; Blough, 1992; Faust, 1996).

13.6 Asymptotics for Unit Root Processes

This Appendix explains asymptotic theory for unit root proceeses. Many of the results depend on the Functional Central Limit Theorem (FCLT) explained in Appendix 5.B.

13.6.1 Continuous Mapping Theorem

Theorem 13.1 *Let $h : \mathcal{R} \to \mathcal{R}$ be a measurable function with discontinuity points confined to a set D where $P(D) = 0$. If $X_n \Rightarrow X$, then $h(X_n) \Rightarrow h(X)$.*

It is instructive to illustrate how the CMT can be used in the AR(1) model when $\beta = 1$:

$$y_t = \beta y_{t-1} + \varepsilon_t.$$

Consider the sampling error of the OLS estimator,

$$n(\widehat{\beta} - 1) = \frac{\frac{1}{n}\sum_{t=2}^{n} y_{t-1}\varepsilon_t}{\frac{1}{n^2}\sum_{t=2}^{n} y_{t-1}^2}.$$

Asymptotic properties of the denominator can be established by the FCLT and the CMT. Let $W_n(r) = \frac{y_{[nr]}}{\sqrt{n}}$. Note that the denominator can be written

$$\frac{1}{n^2}\sum_{t=2}^{n} y_{t-1}^2 = \frac{1}{n}\sum_{t=2}^{n}\left(\frac{y_{t-1}}{\sqrt{n}}\right)^2 = \int_0^1 [W_n(r)]^2\, dr.$$

Since $W_n(r) \Rightarrow W(r)$ and the integral is the continuous function of $W_n(r)$, by the above theorem,

$$\int_0^1 [W_n(r)]^2\, dr \Rightarrow \int_0^1 [W(r)]^2\, dr.$$

13.6.2 Dickey-Fuller test with serially uncorrelated disturbances

We consider two cases for DF tests with the null: when the true process is a random walk with or without a drift, and when the equation is estimated with or without a trend. See Hamilton (1994) for details.

The regression equation includes a constant term but no time trend when the true process is a random walk

Suppose that the data are generated by a random walk without drift

$$y_t = y_{t-1} + \epsilon_t,$$

where ϵ_t follows an i.i.d. sequence with mean zero, and variance σ^2. Consider a regression equation

$$\begin{aligned}\Delta y_t &= \alpha + \rho y_{t-1} + \epsilon_t \\ &= \mathbf{x}_t' \boldsymbol{\beta} + \epsilon_t,\end{aligned}$$

where $\mathbf{x}_t = (1, y_{t-1})'$, and $\boldsymbol{\beta} = (\alpha, \rho)'$. Define a scaling matrix

$$\mathbf{S}_T = \begin{bmatrix} \sqrt{T} & \mathbf{0} \\ \mathbf{0} & T \end{bmatrix}$$

and write the deviation of the OLS estimates using the scaling matrix

$$\begin{aligned}\mathbf{S}_T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \left\{ \mathbf{S}_T^{-1} \left[\sum_{i=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right] \mathbf{S}_T^{-1} \right\}^{-1} \left\{ \mathbf{S}_T^{-1} \left[\sum_{i=1}^{T} \mathbf{x}_t \epsilon_t \right] \right\} \\ &= \begin{bmatrix} 1 & T^{-\frac{3}{2}} \sum_{i=1}^{T} y_{t-1} \\ \cdot & T^{-2} \sum_{i=1}^{T} y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-\frac{1}{2}} \sum_{i=1}^{T} \epsilon_t \\ T^{-1} \sum_{i=1}^{T} y_{t-1} \epsilon_t \end{bmatrix},\end{aligned}$$

where under the null of $\Delta y_t = \epsilon_t$

$$\begin{bmatrix} 1 & T^{-\frac{3}{2}} \sum_{i=1}^{T} y_{t-1} \\ \cdot & T^{-2} \sum_{i=1}^{T} y_{t-1}^2 \end{bmatrix} \xrightarrow{L} \begin{bmatrix} 1 & \sigma \int W(r) dr \\ \cdot & \sigma^2 \int W(r)^2 dr \end{bmatrix} \quad \text{and}$$

$$\begin{bmatrix} T^{-\frac{1}{2}} \sum_{i=1}^{T} \epsilon_t \\ T^{-1} \sum_{i=1}^{T} y_{t-1} \epsilon_t \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \sigma W(1) \\ \sigma^2 \int W dW \end{bmatrix}.$$

Thus, we get

$$\begin{aligned}\begin{bmatrix} T^{\frac{1}{2}}\hat{\alpha} \\ T\hat{\rho} \end{bmatrix} &\xrightarrow{L} \begin{bmatrix} 1 & \sigma \int W(r)dr \\ \cdot & \sigma^2 \int W(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} \sigma W(1) \\ \sigma^2 \int W dW \end{bmatrix} \\ &\xrightarrow{L} \begin{bmatrix} \sigma & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \int W(r)dr \\ \cdot & \int W(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ \int W dW \end{bmatrix}.\end{aligned}$$

In particular,

$$\begin{aligned}T\hat{\rho} &\xrightarrow{L} \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \int W(r)dr \\ \cdot & \int W(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ \int W dW \end{bmatrix} \\ &= \frac{\int W dW - W(1)\int W(r)dr}{\int W(r)^2 dr - (\int W(r)dr)^2},\end{aligned}$$

which is the DF ρ test. Note that the coefficients on Δy_{t-i} follow a normal distribution asymptotically so that the usual test can be applied for restrictions on these variables.

Similarly, the variance of $\hat{\boldsymbol{\beta}}$ follows

$$\begin{aligned}\mathbf{S}_T \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}} \mathbf{S}_T &= \hat{\sigma}^2 \left\{ \mathbf{S}_T^{-1} \left[\sum_{i=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right] \mathbf{S}_T^{-1} \right\}^{-1} \\ &= \hat{\sigma}^2 \begin{bmatrix} 1 & T^{-\frac{3}{2}} \sum_{i=1}^{T} y_{t-1} \\ \cdot & T^{-2} \sum_{i=1}^{T} y_{t-1}^2 \end{bmatrix}^{-1} \\ &\xrightarrow{L} \sigma^2 \begin{bmatrix} 1 & \sigma \int W(r)dr \\ \cdot & \sigma^2 \int W(r)^2 dr \end{bmatrix}^{-1}.\end{aligned}$$

In particular, the standard error of $\hat{\rho}$ follows

$$T\mathbf{s}_{\hat{\rho}} \xrightarrow{L} \frac{1}{\left[\int W(r)^2 dr - (\int W(r)dr)^2\right]^{\frac{1}{2}}}.$$

Therefore, we get the DF t-test

$$\begin{aligned}t_{\hat{\rho}} = \frac{T\hat{\rho}}{Ts_{\hat{\rho}}} &\xrightarrow{L} \frac{\left[\int W dW - W(1)\int W(r)dr\right] / \left[\int W(r)^2 dr - (\int W(r)dr)^2\right]}{\left\{1/\left[\int W(r)^2 dr - (\int W(r)dr)^2\right]\right\}^{\frac{1}{2}}} \\ &\xrightarrow{L} \frac{\int W dW - W(1)\int W(r)dr}{\left[\int W(r)^2 dr - (\int W(r)dr)^2\right]^{\frac{1}{2}}}.\end{aligned}$$

The regression equation includes a constant term and a time trend when the true process is a unit root process with or without a drift

Now, suppose that the data are generated by a random walk with or without a drift

$$y_t = \mu + y_{t-1} + \epsilon_t.$$

Consider a regression equation

$$\Delta y_t = \mu + \delta t + \rho y_{t-1} + \epsilon_t.$$

Note that the regression is subject to collinearity because y_{t-1} contains a deterministic time trend component if $\mu \neq 0$. To avoid the possible collinearity, rewrite the equation using a detrended series $\xi_t = y_t - \mu t$

$$\begin{aligned}
\Delta y_t &= \mu + \delta t + \rho(\xi_{t-1} + \mu(t-1)) + \epsilon_t \\
&= (1-\rho)\mu + (\delta + \rho\mu)t + \rho\xi_{t-1} + \epsilon_t \\
&= \alpha + \tau t + \rho\xi_{t-1} + \epsilon_t \\
&= \mathbf{x}_t'\boldsymbol{\beta} + \epsilon_t,
\end{aligned}$$

where $\mathbf{x}_t = (1, t, \xi_{t-1})'$, and $\boldsymbol{\beta} = (\alpha, \tau, \rho)'$. Define a scaling matrix

$$\mathbf{S}_T = \begin{bmatrix} \sqrt{T} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sqrt[3]{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & T \end{bmatrix}$$

and write the deviation of the OLS estimates using the scaling matrix

$$\begin{aligned}
\mathbf{S}_T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \left\{ \mathbf{S}_T^{-1} \left[\sum_{i=1}^{T} \mathbf{x}_t\mathbf{x}_t' \right] \mathbf{S}_T^{-1} \right\}^{-1} \left\{ \mathbf{S}_T^{-1} \left[\sum_{i=1}^{T} \mathbf{x}_t\epsilon_t \right] \right\} \\
&= \begin{bmatrix} 1 & T^{-2}\sum_{i=1}^{T} t & T^{-\frac{3}{2}}\sum_{i=1}^{T} \xi_{t-1} \\ \cdot & T^{-3}\sum_{i=1}^{T} t^2 & T^{-\frac{5}{2}}\sum_{i=1}^{T} t\xi_{t-1} \\ \cdot & \cdot & T^{-2}\sum_{i=1}^{T} \xi_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-\frac{1}{2}}\sum_{i=1}^{T} \epsilon_t \\ T^{-\frac{3}{2}}\sum_{i=1}^{T} t\epsilon_t \\ T^{-1}\sum_{i=1}^{T} \xi_{t-1}\epsilon_t \end{bmatrix},
\end{aligned}$$

where under the null of $\Delta\xi_t = \epsilon_t$

$$\begin{bmatrix} 1 & T^{-2}\sum_{i=1}^T t & T^{-\frac{3}{2}}\sum_{i=1}^T \xi_{t-1} \\ \cdot & T^{-3}\sum_{i=1}^T t^2 & T^{-\frac{5}{2}}\sum_{i=1}^T t\xi_{t-1} \\ \cdot & \cdot & T^{-2}\sum_{i=1}^T \xi_{t-1}^2 \end{bmatrix} \xrightarrow{L} \begin{bmatrix} 1 & \frac{1}{2} & \sigma\int W(r)dr \\ \cdot & \frac{1}{3} & \sigma\int rW(r)dr \\ \cdot & \cdot & \sigma^2\int W(r)^2dr \end{bmatrix} \quad \text{and}$$

$$\begin{bmatrix} T^{-\frac{1}{2}}\sum_{i=1}^T \epsilon_t \\ T^{-\frac{3}{2}}\sum_{i=1}^T t\epsilon_t \\ T^{-1}\sum_{i=1}^T \xi_{t-1}\epsilon_t \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \sigma W(1) \\ \sigma\int rdW \\ \sigma^2\int WdW \end{bmatrix}.$$

Due to the block diagonal property, we can write

$$\begin{aligned} \begin{bmatrix} T^{\frac{1}{2}}\hat{\alpha} \\ T^{\frac{3}{2}}\hat{\tau} \\ T\hat{\rho} \end{bmatrix} &\xrightarrow{L} \begin{bmatrix} 1 & \frac{1}{2} & \sigma\int W(r)dr \\ \cdot & \frac{1}{3} & \sigma\int rW(r)dr \\ \cdot & \cdot & \sigma^2\int W(r)^2dr \end{bmatrix}^{-1} \begin{bmatrix} \sigma W(1) \\ \sigma\int rdW \\ \sigma^2\int WdW \end{bmatrix} \\ &\xrightarrow{L} \begin{bmatrix} \sigma & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2dr \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ \int rdW \\ \int WdW \end{bmatrix}. \end{aligned}$$

In particular,

$$T\hat{\rho} \xrightarrow{L} \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2dr \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ \int rdW \\ \int WdW \end{bmatrix},$$

which is the DF ρ test.

Similarly, the variance of $\hat{\boldsymbol{\beta}}$ follows

$$\begin{aligned} \mathbf{S}_T\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}\mathbf{S}_T &= \hat{\sigma}^2\left\{\mathbf{S}_T^{-1}\left[\sum_{i=1}^T \mathbf{x}_t\mathbf{x}_t'\right]\mathbf{S}_T^{-1}\right\}^{-1} \\ &= \hat{\sigma}^2 \begin{bmatrix} 1 & T^{-2}\sum_{i=1}^T t & T^{-\frac{3}{2}}\sum_{i=1}^T \xi_{t-1} \\ \cdot & T^{-3}\sum_{i=1}^T t^2 & T^{-\frac{5}{2}}\sum_{i=1}^T t\xi_{t-1} \\ \cdot & \cdot & T^{-2}\sum_{i=1}^T \xi_{t-1}^2 \end{bmatrix}^{-1} \\ &\xrightarrow{L} \sigma^2 \begin{bmatrix} 1 & \frac{1}{2} & \sigma\int W(r)dr \\ \cdot & \frac{1}{3} & \sigma\int rW(r)dr \\ \cdot & \cdot & \sigma^2\int W(r)^2dr \end{bmatrix}^{-1}. \end{aligned}$$

In particular, the standard error of $\hat{\rho}$ follows

$$T\mathbf{s}_{\hat{\rho}} \xrightarrow{L} \left\{\begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2dr \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right\}^{\frac{1}{2}}.$$

Therefore, we get the DF t-test

$$t_{\hat{\rho}} = \frac{T\hat{\rho}}{Ts_{\hat{\rho}}} \xrightarrow{L} \frac{\begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ \int rdW \\ \int WdW \end{bmatrix}}{\left\{ \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}^{\frac{1}{2}}}.$$

13.6.3 Said-Dickey test with serially correlated disturbances

We consider two cases for Said-Dickey tests with the null: when the true process is a random walk with or without a drift, and when the equation is estimated with or without a trend. See Hamilton (1994) for details.

The regression equation includes a constant term but no time trend when the true process is a unit root process without a drift

Consider a DGP:

$$a(L)y_t = \epsilon_t,$$

where ϵ_t follows an i.i.d. sequence with mean zero, and variance σ^2. Let

$$a(L) = a(1)L + b(L)(1 - L),$$

where $b(L) = 1 - \sum_{i=1}^{p-1} b_i L^i$ and $b_i = -\sum_{j=i+1}^{p} a_i$, and rearrange the equation

$$\begin{aligned} b(L)\Delta y_t &= -a(1)y_{t-1} + \epsilon_t \quad \text{or} \\ \Delta y_t &= \rho y_{t-1} + \sum_{i=1}^{p-1} b_i \Delta y_{t-i} + \epsilon_t, \end{aligned}$$

where $\rho = -a(1) = -1 + \sum_{i=1}^{p} a_i$. Note that the assumption of a single unit root in the DGP implies $\rho = 0$. Under the null, we get an MA representation

$$\begin{aligned} \Delta y_t &= c(L)\epsilon_t \\ &= u_t \end{aligned}$$

where $c(L) = b(L)^{-1} = 1 + \sum_{i=1}^{\infty} c_i L^i$.

Consider a regression equation

$$\begin{aligned}
\Delta y_t &= \alpha + \rho y_{t-1} + \sum_{i=1}^{p-1} b_i \Delta y_{t-i} + \epsilon_t \\
&= \alpha + \rho y_{t-1} + \mathbf{z}_t'\mathbf{b} + \epsilon_t \\
&= \mathbf{x}_t'\boldsymbol{\beta} + \epsilon_t,
\end{aligned}$$

where $\mathbf{z}_t = (\Delta y_{t-1}, \cdots, \Delta y_{t-p+1})'$, $\mathbf{b} = (b_1, \cdots, b_{p-1})'$, $\mathbf{x}_t = (1, y_{t-1}, \mathbf{z}_t')'$, and $\boldsymbol{\beta} = (\alpha, \rho, \mathbf{b}')'$. Define a scaling matrix

$$\mathbf{S}_T = \begin{bmatrix} \sqrt{T} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & T & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sqrt{T}\,\mathbf{I}_{p-1} \end{bmatrix}$$

and write the deviation of the OLS estimates using the scaling matrix

$$\begin{aligned}
\mathbf{S}_T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \left\{ \mathbf{S}_T^{-1} \left[\sum_{i=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right] \mathbf{S}_T^{-1} \right\}^{-1} \left\{ \mathbf{S}_T^{-1} \left[\sum_{i=1}^{T} \mathbf{x}_t \epsilon_t \right] \right\} \\
&= \begin{bmatrix} 1 & T^{-\frac{3}{2}} \sum_{i=1}^{T} y_{t-1} & T^{-1} \sum_{i=1}^{T} \mathbf{z}_t' \\ \cdot & T^{-2} \sum_{i=1}^{T} y_{t-1}^2 & T^{-\frac{3}{2}} \sum_{i=1}^{T} y_{t-1}\mathbf{z}_t' \\ \cdot & \cdot & T^{-1} \sum_{i=1}^{T} \mathbf{z}_t \mathbf{z}_t' \end{bmatrix}^{-1} \begin{bmatrix} T^{-\frac{1}{2}} \sum_{i=1}^{T} \epsilon_t \\ T^{-1} \sum_{i=1}^{T} y_{t-1}\epsilon_t \\ T^{-\frac{1}{2}} \sum_{i=1}^{T} \mathbf{z}_t \epsilon_t \end{bmatrix},
\end{aligned}$$

where under the null of $\Delta y_t = u_t$

$$\begin{bmatrix} 1 & T^{-\frac{3}{2}} \sum_{i=1}^{T} y_{t-1} & T^{-1} \sum_{i=1}^{T} \mathbf{z}_t' \\ \cdot & T^{-2} \sum_{i=1}^{T} y_{t-1}^2 & T^{-\frac{3}{2}} \sum_{i=1}^{T} y_{t-1}\mathbf{z}_t' \\ \cdot & \cdot & T^{-1} \sum_{i=1}^{T} \mathbf{z}_t \mathbf{z}_t' \end{bmatrix} \xrightarrow{L} \begin{bmatrix} 1 & \lambda \int W(r)dr & \mathbf{0} \\ \cdot & \lambda^2 \int W(r)^2 dr & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{V} \end{bmatrix} \text{ and}$$

$$\begin{bmatrix} T^{-\frac{1}{2}} \sum_{i=1}^{T} \epsilon_t \\ T^{-1} \sum_{i=1}^{T} y_{t-1}\epsilon_t \\ T^{-\frac{1}{2}} \sum_{i=1}^{T} \mathbf{z}_t \epsilon_t \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \sigma W(1) \\ \sigma\lambda \int W dW \\ \mathbf{h} \end{bmatrix}.$$

Due to the block diagonal property, we can write

$$\begin{aligned}
\begin{bmatrix} T^{\frac{1}{2}}\hat{\alpha} \\ T\hat{\rho} \end{bmatrix} &\xrightarrow{L} \begin{bmatrix} 1 & \lambda \int W(r)dr \\ \cdot & \lambda^2 \int W(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} \sigma W(1) \\ \sigma\lambda \int W dW \end{bmatrix} \\
&\xrightarrow{L} \begin{bmatrix} \sigma & 0 \\ 0 & \frac{\sigma}{\lambda} \end{bmatrix} \begin{bmatrix} 1 & \int W(r)dr \\ \cdot & \int W(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ \int W dW \end{bmatrix} \quad \text{and} \\
T^{-\frac{1}{2}}(\hat{\mathbf{b}} - \mathbf{b}) &\xrightarrow{L} \mathbf{V}^{-1}\mathbf{h} \sim N(\mathbf{0}, \sigma^2 \mathbf{V}^{-1}).
\end{aligned}$$

In particular,

$$T\hat{\rho} \xrightarrow{L} \frac{\sigma}{\lambda}\begin{bmatrix} 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & \int W(r)dr \\ \cdot & \int W(r)^2dr \end{bmatrix}^{-1}\begin{bmatrix} W(1) \\ \int WdW \end{bmatrix}.$$

From $\frac{\lambda}{\sigma} = c(1) = b(1)^{-1}$, we get the Said-Dickey ρ test

$$\frac{T\hat{\rho}}{1-\sum_{i=1}^{p-1}\hat{b}_i} \xrightarrow{L} \frac{\int WdW - W(1)\int W(r)dr}{\int W(r)^2dr - (\int W(r)dr)^2},$$

which follows the same asymptotic distribution as the DF ρ test. Note that the coefficients on Δy_{t-i} follow a normal distribution asymptotically so that the usual test can be applied for restrictions on these variables.

Similarly, the variance of $\hat{\boldsymbol{\beta}}$ follows

$$\begin{aligned} \mathbf{S}_T\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}\mathbf{S}_T &= \hat{\sigma}^2\left\{\mathbf{S}_T^{-1}\left[\sum_{i=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right]\mathbf{S}_T^{-1}\right\}^{-1} \\ &= \hat{\sigma}^2\begin{bmatrix} 1 & T^{-\frac{3}{2}}\sum_{i=1}^{T}y_{t-1} & T^{-1}\sum_{i=1}^{T}\mathbf{z}_t' \\ \cdot & T^{-2}\sum_{i=1}^{T}y_{t-1}^2 & T^{-\frac{3}{2}}\sum_{i=1}^{T}y_{t-1}\mathbf{z}_t' \\ \cdot & \cdot & T^{-1}\sum_{i=1}^{T}\mathbf{z}_t\mathbf{z}_t' \end{bmatrix}^{-1} \\ &\xrightarrow{L} \sigma^2\begin{bmatrix} 1 & \lambda\int W(r)dr & \mathbf{0} \\ \cdot & \lambda^2\int W(r)^2dr & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{V} \end{bmatrix}^{-1}. \end{aligned}$$

In particular, the standard error of $\hat{\rho}$ follows

$$T\mathbf{s}_{\hat{\rho}} \xrightarrow{L} \frac{\sigma/\lambda}{\left[\int W(r)^2dr - (\int W(r)dr)^2\right]^{\frac{1}{2}}}.$$

Therefore, we get the Said-Dickey t-test

$$\begin{aligned} t_{\hat{\rho}} = \frac{T\hat{\rho}}{Ts_{\hat{\rho}}} &\xrightarrow{L} \frac{(\sigma/\lambda)\left[\int WdW - W(1)\int W(r)dr\right]/\left[\int W(r)^2dr - (\int W(r)dr)^2\right]}{(\sigma/\lambda)\left\{1/\left[\int W(r)^2dr - (\int W(r)dr)^2\right]\right\}^{\frac{1}{2}}} \\ &\xrightarrow{L} \frac{\int WdW - W(1)\int W(r)dr}{\left[\int W(r)^2dr - (\int W(r)dr)^2\right]^{\frac{1}{2}}}, \end{aligned}$$

which follows the same asymptotic distribution as the DF t test.

The regression equation includes a constant term and a time trend when the true process is a unit root process with or without a drift

Now, consider a DGP:

$$a(L)y_t = \mu + \epsilon_t$$

and rearrange the equation

$$\begin{aligned} b(L)\Delta y_t &= \mu - a(1)y_{t-1} + \epsilon_t \quad \text{or} \\ \Delta y_t &= \mu + \rho y_{t-1} + \sum_{i=1}^{p-1} b_i \Delta y_{t-i} + \epsilon_t. \end{aligned}$$

Under the null, we get an MA representation

$$\begin{aligned} \Delta y_t &= \theta + c(L)\epsilon_t \\ &= \theta + u_t \end{aligned}$$

where $\theta = c(1)\mu$.

Consider a regression equation

$$\Delta y_t = \mu + \delta t + \rho y_{t-1} + \sum_{i=1}^{p-1} b_i \Delta y_{t-i} + \epsilon_t.$$

Note that the regression is subject to collinearity because y_{t-1} contains a deterministic time trend component if $\mu \neq 0$. To avoid the possible collinearity, rewrite the equation using a detrended series $\xi_t = y_t - \mu t$

$$\begin{aligned} \Delta y_t &= \mu + \delta t + \rho(\xi_{t-1} + \mu(t-1)) + \sum_{i=1}^{p-1} b_i(\Delta \xi_{t-i} + \mu) + \epsilon_t \\ &= (1 - \rho + \sum_{i=1}^{p-1} b_i)\mu + (\delta + \rho\mu)t + \rho\xi_{t-1} + \mathbf{z}_t'\mathbf{b} + \epsilon_t \\ &= \alpha + \tau t + \rho\xi_{t-1} + \mathbf{z}_t'\mathbf{b} + \epsilon_t \\ &= \mathbf{x}_t'\boldsymbol{\beta} + \epsilon_t, \end{aligned}$$

where $\mathbf{z}_t = (\Delta\xi_{t-1}, \cdots, \Delta\xi_{t-p+1})'$, $\mathbf{b} = (b_1, \cdots, b_{p-1})'$, $\mathbf{x}_t = (1, t, \xi_{t-1}, \mathbf{z}_t')'$, and $\boldsymbol{\beta} = (\alpha, \tau, \rho, \mathbf{b}')'$. Define a scaling matrix

$$\mathbf{S}_T = \begin{bmatrix} \sqrt{T} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sqrt[3]{T} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & T & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sqrt{T}\,\mathbf{I}_{p-1} \end{bmatrix}$$

and write the deviation of the OLS estimates using the scaling matrix

$$\begin{aligned} \mathbf{S}_T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \left\{ \mathbf{S}_T^{-1} \left[\sum_{i=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right] \mathbf{S}_T^{-1} \right\}^{-1} \left\{ \mathbf{S}_T^{-1} \left[\sum_{i=1}^{T} \mathbf{x}_t \epsilon_t \right] \right\} \\ &= \begin{bmatrix} 1 & T^{-2}\sum_{i=1}^T t & T^{-\frac{3}{2}}\sum_{i=1}^T \xi_{t-1} & T^{-1}\sum_{i=1}^T \mathbf{z}_t' \\ \cdot & T^{-3}\sum_{i=1}^T t^2 & T^{-\frac{5}{2}}\sum_{i=1}^T t\xi_{t-1} & T^{-2}\sum_{i=1}^T t\mathbf{z}_t' \\ \cdot & \cdot & T^{-2}\sum_{i=1}^T \xi_{t-1}^2 & T^{-\frac{3}{2}}\sum_{i=1}^T \xi_{t-1}\mathbf{z}_t' \\ \cdot & \cdot & \cdot & T^{-1}\sum_{i=1}^T \mathbf{z}_t\mathbf{z}_t' \end{bmatrix}^{-1} \begin{bmatrix} T^{-\frac{1}{2}}\sum_{i=1}^T \epsilon_t \\ T^{-\frac{3}{2}}\sum_{i=1}^T t\epsilon_t \\ T^{-1}\sum_{i=1}^T \xi_{t-1}\epsilon_t \\ T^{-\frac{1}{2}}\sum_{i=1}^T \mathbf{z}_t\epsilon_t \end{bmatrix}, \end{aligned}$$

where under the null of $\Delta\xi_t = u_t$

$$\begin{bmatrix} 1 & T^{-2}\sum_{i=1}^T t & T^{-\frac{3}{2}}\sum_{i=1}^T \xi_{t-1} & T^{-1}\sum_{i=1}^T \mathbf{z}_t' \\ \cdot & T^{-3}\sum_{i=1}^T t^2 & T^{-\frac{5}{2}}\sum_{i=1}^T t\xi_{t-1} & T^{-2}\sum_{i=1}^T t\mathbf{z}_t' \\ \cdot & \cdot & T^{-2}\sum_{i=1}^T \xi_{t-1}^2 & T^{-\frac{3}{2}}\sum_{i=1}^T \xi_{t-1}\mathbf{z}_t' \\ \cdot & \cdot & \cdot & T^{-1}\sum_{i=1}^T \mathbf{z}_t\mathbf{z}_t' \end{bmatrix} \xrightarrow{L} \begin{bmatrix} 1 & \frac{1}{2} & \lambda\int W(r)dr & \mathbf{0} \\ \cdot & \frac{1}{3} & \lambda\int rW(r)dr & \mathbf{0} \\ \cdot & \cdot & \lambda^2\int W(r)^2dr & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V} \end{bmatrix} \text{ and}$$

$$\begin{bmatrix} T^{-\frac{1}{2}}\sum_{i=1}^T \epsilon_t \\ T^{-\frac{3}{2}}\sum_{i=1}^T t\epsilon_t \\ T^{-1}\sum_{i=1}^T \xi_{t-1}\epsilon_t \\ T^{-\frac{1}{2}}\sum_{i=1}^T \mathbf{z}_t\epsilon_t \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \sigma W(1) \\ \sigma\int rdW \\ \sigma\lambda\int WdW \\ \mathbf{h} \end{bmatrix}.$$

Due to the block diagonal property, we can write

$$\begin{aligned} \begin{bmatrix} T^{\frac{1}{2}}\hat{\alpha} \\ T^{\frac{3}{2}}\hat{\tau} \\ T\hat{\rho} \end{bmatrix} &\xrightarrow{L} \begin{bmatrix} 1 & \frac{1}{2} & \lambda\int W(r)dr \\ \cdot & \frac{1}{3} & \lambda\int rW(r)dr \\ \cdot & \cdot & \lambda^2\int W(r)^2dr \end{bmatrix}^{-1} \begin{bmatrix} \sigma W(1) \\ \sigma\int rdW \\ \sigma\lambda\int WdW \end{bmatrix} \\ &\xrightarrow{L} \begin{bmatrix} \sigma & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & \frac{\sigma}{\lambda} \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2dr \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ \int rdW \\ \int WdW \end{bmatrix} \quad \text{and} \\ T^{-\frac{1}{2}}(\hat{\mathbf{b}} - \mathbf{b}) &\xrightarrow{L} \mathbf{V}^{-1}\mathbf{h} \sim N(\mathbf{0}, \sigma^2\mathbf{V}^{-1}). \end{aligned}$$

In particular,

$$T\hat{\rho} \xrightarrow{L} \frac{\sigma}{\lambda}\begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2dr \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ \int rdW \\ \int WdW \end{bmatrix}.$$

From $\frac{\lambda}{\sigma} = c(1) = b(1)^{-1}$, we get the Said-Dickey ρ test

$$\frac{T\hat{\rho}}{1-\sum_{i=1}^{p-1}\hat{b}_i} \xrightarrow{L} \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ \int rdW \\ \int WdW \end{bmatrix} \quad \text{and}$$

which follows the same asymptotic distribution as the DF ρ test. Note that the coefficients on Δy_{t-i} follow a normal distribution asymptotically so that the usual test can be applied for restrictions on these variables.

Similarly, the variance of $\hat{\boldsymbol{\beta}}$ follows

$$\begin{aligned} \mathbf{S}_T \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}} \mathbf{S}_T &= \hat{\sigma}^2 \left\{ \mathbf{S}_T^{-1} \left[\sum_{i=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right] \mathbf{S}_T^{-1} \right\}^{-1} \\ &= \hat{\sigma}^2 \begin{bmatrix} 1 & T^{-2}\sum_{i=1}^T t & T^{-\frac{3}{2}}\sum_{i=1}^T \xi_{t-1} & T^{-1}\sum_{i=1}^T \mathbf{z}_t' \\ \cdot & T^{-3}\sum_{i=1}^T t^2 & T^{-\frac{5}{2}}\sum_{i=1}^T t\xi_{t-1} & T^{-2}\sum_{i=1}^T t\mathbf{z}_t' \\ \cdot & \cdot & T^{-2}\sum_{i=1}^T \xi_{t-1}^2 & T^{-\frac{3}{2}}\sum_{i=1}^T \xi_{t-1}\mathbf{z}_t' \\ \cdot & \cdot & \cdot & T^{-1}\sum_{i=1}^T \mathbf{z}_t\mathbf{z}_t' \end{bmatrix}^{-1} \\ &\xrightarrow{L} \sigma^2 \begin{bmatrix} 1 & \frac{1}{2} & \lambda\int W(r)dr & \mathbf{0} \\ \cdot & \frac{1}{3} & \lambda\int rW(r)dr & \mathbf{0} \\ \cdot & \cdot & \lambda^2\int W(r)^2 dr & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V} \end{bmatrix}^{-1}. \end{aligned}$$

In particular, the standard error of $\hat{\rho}$ follows

$$T\mathbf{s}_{\hat{\rho}} \xrightarrow{L} \frac{\sigma}{\lambda} \left\{ \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}^{\frac{1}{2}}.$$

Therefore, we get the Said-Dickey t-test

$$t_{\hat{\rho}} = \frac{T\hat{\rho}}{Ts_{\hat{\rho}}} \xrightarrow{L} \frac{\begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ \int rdW \\ \int WdW \end{bmatrix}}{\left\{ \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}^{\frac{1}{2}}},$$

which follows the same asymptotic distribution as the DF t test.

13.6.4 Phillips-Perron test

We consider two cases for PP tests with the null: when the true process is a random walk with or without a drift, and when the equation is estimated with or without a trend. See Hamilton (1994) for details.

The regression equation includes a constant term but no time trend when the true process is a unit root process without a drift

Consider a DGP:

$$a(L)y_t = \epsilon_t,$$

where ϵ_t follows an i.i.d. sequence with mean zero, and variance σ^2. Let

$$a(L) = a(1)L + b(L)(1 - L),$$

where $b(L) = 1 - \sum_{i=1}^{p-1} b_i L^i$ and $b_i = -\sum_{j=i+1}^{p} a_i$, and rearrange the equation

$$\begin{aligned} b(L)\Delta y_t &= -a(1)y_{t-1} + \epsilon_t \quad \text{or} \\ \Delta y_t &= \rho y_{t-1} + \sum_{i=1}^{p-1} b_i \Delta y_{t-i} + \epsilon_t, \end{aligned}$$

where $\rho = -a(1) = -1 + \sum_{i=1}^{p} a_i$. Note that the assumption of a single unit root in the DGP implies $\rho = 0$. Under the null, we get an MA representation

$$\begin{aligned} \Delta y_t &= c(L)\epsilon_t \\ &= u_t \end{aligned}$$

where $c(L) = b(L)^{-1} = 1 + \sum_{i=1}^{\infty} c_i L^i$.

Consider a regression equation

$$\begin{aligned} \Delta y_t &= \alpha + \rho y_{t-1} + u_t \\ &= \mathbf{x}_t' \boldsymbol{\beta} + u_t, \end{aligned}$$

where $\mathbf{x}_t = (1, y_{t-1})'$, $\boldsymbol{\beta} = (\alpha, \rho)'$, and u_t is a regression error with mean zero and variance σ_u^2. Define a scaling matrix

$$\mathbf{S}_T = \begin{bmatrix} \sqrt{T} & \mathbf{0} \\ \mathbf{0} & T \end{bmatrix}$$

and write the deviation of the OLS estimates using the scaling matrix

$$\begin{aligned}\mathbf{S}_T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \left\{\mathbf{S}_T^{-1}\left[\sum_{i=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right]\mathbf{S}_T^{-1}\right\}^{-1}\left\{\mathbf{S}_T^{-1}\left[\sum_{i=1}^{T}\mathbf{x}_t u_t\right]\right\} \\ &= \begin{bmatrix} 1 & T^{-\frac{3}{2}}\sum_{i=1}^{T} y_{t-1} \\ \cdot & T^{-2}\sum_{i=1}^{T} y_{t-1}^2 \end{bmatrix}^{-1}\begin{bmatrix} T^{-\frac{1}{2}}\sum_{i=1}^{T} u_t \\ T^{-1}\sum_{i=1}^{T} y_{t-1}u_t \end{bmatrix},\end{aligned}$$

where under the null of $\Delta y_t = u_t$

$$\begin{bmatrix} 1 & T^{-\frac{3}{2}}\sum_{i=1}^{T} y_{t-1} \\ \cdot & T^{-2}\sum_{i=1}^{T} y_{t-1}^2 \end{bmatrix} \xrightarrow{L} \begin{bmatrix} 1 & \lambda\int W(r)dr \\ \cdot & \lambda^2\int W(r)^2 dr \end{bmatrix} \quad \text{and}$$

$$\begin{bmatrix} T^{-\frac{1}{2}}\sum_{i=1}^{T} u_t \\ T^{-1}\sum_{i=1}^{T} y_{t-1}u_t \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \lambda W(1) \\ \lambda^2\int W dW \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{2}(\lambda^2 - \gamma_0) \end{bmatrix}.$$

Thus,

$$\begin{bmatrix} T^{\frac{1}{2}}\hat{\alpha} \\ T\hat{\rho} \end{bmatrix} \xrightarrow{L} \begin{bmatrix} 1 & \lambda\int W(r)dr \\ \cdot & \lambda^2\int W(r)^2 dr \end{bmatrix}^{-1}\left\{\begin{bmatrix} \lambda W(1) \\ \lambda^2\int W dW \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{2}(\lambda^2 - \gamma_0) \end{bmatrix}\right\}.$$

In particular,

$$\begin{aligned}T\hat{\rho} &\xrightarrow{L} \begin{bmatrix} 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & \int W(r)dr \\ \cdot & \int W(r)^2 dr \end{bmatrix}^{-1}\left\{\begin{bmatrix} \lambda W(1) \\ \lambda^2\int W dW \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{\lambda^2-\gamma_0}{2\lambda^2} \end{bmatrix}\right\} \\ &= \frac{\int W dW - W(1)\int W(r)dr}{\int W(r)^2 dr - (\int W(r)dr)^2} + \frac{(\lambda^2 - \gamma_0)/2\lambda^2}{\int W(r)^2 dr - (\int W(r)dr)^2}\end{aligned}$$

Note that the second component can be consistently estimated by

$$\frac{T^2 s_{\hat{\rho}}^2}{\hat{\sigma}_u^2}\frac{\hat{\lambda}^2 - \hat{\gamma}_0}{2}$$

because

$$T^2\mathbf{s}_{\hat{\rho}}^2 \xrightarrow{L} \frac{\sigma_u^2/\lambda^2}{\int W(r)^2 dr - (\int W(r)dr)^2}.$$

Accordingly, we get the PP ρ test

$$T\hat{\rho} - \frac{T^2 \mathbf{s}_{\hat{\rho}}^2}{\hat{\sigma}_u^2} \frac{\hat{\lambda}^2 - \hat{\gamma}_0}{2} \xrightarrow{L} \frac{\int W dW - W(1) \int W(r) dr}{\int W(r)^2 dr - (\int W(r) dr)^2},$$

which follows the same asymptotic distribution as the DF ρ test.

Similarly, the variance of $\hat{\boldsymbol{\beta}}$ follows

$$\begin{aligned}
\mathbf{S}_T \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}} \mathbf{S}_T &= \hat{\sigma}_u^2 \left\{ \mathbf{S}_T^{-1} \left[\sum_{i=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right] \mathbf{S}_T^{-1} \right\}^{-1} \\
&= \hat{\sigma}_u^2 \begin{bmatrix} 1 & T^{-\frac{3}{2}} \sum_{i=1}^{T} y_{t-1} \\ \cdot & T^{-2} \sum_{i=1}^{T} y_{t-1}^2 \end{bmatrix}^{-1} \\
&\xrightarrow{L} \sigma_u^2 \begin{bmatrix} 1 & \lambda \int W(r) dr \\ \cdot & \lambda^2 \int W(r)^2 dr \end{bmatrix}^{-1}.
\end{aligned}$$

In particular, the standard error of $\hat{\rho}$ follows

$$T\mathbf{s}_{\hat{\rho}} \xrightarrow{L} \frac{\sigma_u / \lambda}{\left[\int W(r)^2 dr - (\int W(r) dr)^2\right]^{\frac{1}{2}}}$$

and the t-test follows

$$\begin{aligned}
t_{\hat{\rho}} = \frac{T\hat{\rho}}{T s_{\hat{\rho}}} &\xrightarrow{L} \frac{\left\{ \left[\int W dW - W(1) \int W(r) dr\right] + \frac{\lambda^2 - \gamma_0}{2\lambda^2} \right\} / \left[\int W(r)^2 dr - (\int W(r) dr)^2\right]}{(\sigma_u / \lambda) \left\{ 1 / \left[\int W(r)^2 dr - (\int W(r) dr)^2\right] \right\}^{\frac{1}{2}}} \\
&\xrightarrow{L} \frac{\lambda}{\sigma_u} \left\{ \frac{\int W dW - W(1) \int W(r) dr}{\left[\int W(r)^2 dr - (\int W(r) dr)^2\right]^{\frac{1}{2}}} + \frac{\frac{\lambda^2 - \gamma_0}{2\lambda^2}}{\left[\int W(r)^2 dr - (\int W(r) dr)^2\right]^{\frac{1}{2}}} \right\}.
\end{aligned}$$

Note that the second component can be consistently estimated by

$$\frac{T s_{\hat{\rho}}}{\hat{\sigma}_u} \frac{\hat{\lambda}^2 - \hat{\gamma}_0}{2\hat{\lambda}}$$

because

$$T^2 \mathbf{s}_{\hat{\rho}}^2 \xrightarrow{L} \frac{\sigma_u^2 / \lambda^2}{\int W(r)^2 dr - (\int W(r) dr)^2}.$$

Accordingly, we get the PP t test

$$\frac{\hat{\sigma}_u}{\hat{\lambda}} t_{\hat{\rho}} - \frac{T s_{\hat{\rho}}}{\hat{\sigma}_u} \frac{\hat{\lambda}^2 - \hat{\gamma}_0}{2\hat{\lambda}} \xrightarrow{L} \frac{\int W dW - W(1) \int W(r) dr}{\left[\int W(r)^2 dr - (\int W(r) dr)^2\right]^{\frac{1}{2}}},$$

which follows the same asymptotic distribution as the DF t test. Note that $\gamma_0(= E(u_t^2))$ can be consistently estimated by $\hat{\sigma}_u^2(= \frac{1}{T-2}\sum_{t=1}^{T}\hat{u}_t^2)$ and that λ can be consistently estimated by the Newey-West estimator

$$\hat{\lambda}^2 = \hat{\gamma}_0 + 2\sum_{j=1}^{q}(1 - \frac{j}{q+1})\hat{\gamma}_j,$$

where $\hat{\gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T}\hat{u}_t\hat{u}_{t-j}$.

The regression equation includes a constant term and a time trend when the true process is a unit root process with or without a drift

Now, consider a DGP:

$$a(L)y_t = \mu + \epsilon_t$$

and rearrange the equation

$$\begin{aligned} b(L)\Delta y_t &= \mu - a(1)y_{t-1} + \epsilon_t \quad \text{or} \\ \Delta y_t &= \mu + \rho y_{t-1} + \sum_{i=1}^{p-1} b_i \Delta y_{t-i} + \epsilon_t. \end{aligned}$$

Under the null, we get an MA representation

$$\begin{aligned} \Delta y_t &= \theta + c(L)\epsilon_t \\ &= \theta + u_t \end{aligned}$$

where $\theta = c(1)\mu$.

Consider a regression equation

$$\Delta y_t = \mu + \delta t + \rho y_{t-1} + u_t.$$

Note that the regression is subject to collinearity because y_{t-1} contains a deterministic time trend component if $\mu \neq 0$. To avoid the possible collinearity, rewrite the equation using a detrended series $\xi_t = y_t - \mu t$

$$\begin{aligned}\Delta y_t &= \mu + \delta t + \rho(\xi_{t-1} + \mu(t-1))u_t \\ &= (1-\rho)\mu + (\delta + \rho\mu)t + \rho\xi_{t-1}u_t \\ &= \alpha + \tau t + \rho\xi_{t-1} + u_t \\ &= \mathbf{x}_t'\boldsymbol{\beta} + \epsilon_t,\end{aligned}$$

where $\mathbf{x}_t = (1, t, \xi_{t-1})'$, and $\boldsymbol{\beta} = (\alpha, \tau, \rho)'$. Define a scaling matrix

$$\mathbf{S}_T = \begin{bmatrix} \sqrt{T} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sqrt[3]{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & T \end{bmatrix}$$

and write the deviation of the OLS estimates using the scaling matrix

$$\begin{aligned}\mathbf{S}_T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \left\{\mathbf{S}_T^{-1}\left[\sum_{i=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right]\mathbf{S}_T^{-1}\right\}^{-1}\left\{\mathbf{S}_T^{-1}\left[\sum_{i=1}^{T}\mathbf{x}_t u_t\right]\right\} \\ &= \begin{bmatrix} 1 & T^{-2}\sum_{i=1}^T t & T^{-\frac{3}{2}}\sum_{i=1}^T \xi_{t-1} \\ \cdot & T^{-3}\sum_{i=1}^T t^2 & T^{-\frac{5}{2}}\sum_{i=1}^T t\xi_{t-1} \\ \cdot & \cdot & T^{-2}\sum_{i=1}^T \xi_{t-1}^2 \end{bmatrix}^{-1}\begin{bmatrix} T^{-\frac{1}{2}}\sum_{i=1}^T u_t \\ T^{-\frac{3}{2}}\sum_{i=1}^T tu_t \\ T^{-1}\sum_{i=1}^T \xi_{t-1}u_t \end{bmatrix},\end{aligned}$$

where under the null of $\Delta\xi_t = u_t$

$$\begin{bmatrix} 1 & T^{-2}\sum_{i=1}^T t & T^{-\frac{3}{2}}\sum_{i=1}^T \xi_{t-1} \\ \cdot & T^{-3}\sum_{i=1}^T t^2 & T^{-\frac{5}{2}}\sum_{i=1}^T t\xi_{t-1} \\ \cdot & \cdot & T^{-2}\sum_{i=1}^T \xi_{t-1}^2 \end{bmatrix} \xrightarrow{L} \begin{bmatrix} 1 & \frac{1}{2} & \lambda\int W(r)dr \\ \cdot & \frac{1}{3} & \lambda\int rW(r)dr \\ \cdot & \cdot & \lambda^2\int W(r)^2 dr \end{bmatrix} \text{ and}$$

$$\begin{bmatrix} T^{-\frac{1}{2}}\sum_{i=1}^T u_t \\ T^{-\frac{3}{2}}\sum_{i=1}^T tu_t \\ T^{-1}\sum_{i=1}^T \xi_{t-1}u_t \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \lambda W(1) \\ \lambda\int rdW \\ \lambda^2\int WdW \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \frac{1}{2}(\lambda^2 - \gamma_0) \end{bmatrix}.$$

Thus, we get

$$\begin{aligned}\begin{bmatrix} T^{\frac{1}{2}}\hat{\alpha} \\ T^{\frac{3}{2}}\hat{\tau} \\ T\hat{\rho} \end{bmatrix} &\xrightarrow{L} \begin{bmatrix} 1 & \frac{1}{2} & \lambda\int W(r)dr \\ \cdot & \frac{1}{3} & \lambda\int rW(r)dr \\ \cdot & \cdot & \lambda^2\int W(r)^2 dr \end{bmatrix}^{-1}\left\{\begin{bmatrix} \lambda W(1) \\ \lambda\int rdW \\ \lambda^2\int WdW \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \frac{1}{2}(\lambda^2 - \gamma_0) \end{bmatrix}\right\} \\ &\xrightarrow{L} \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2 dr \end{bmatrix}^{-1}\left\{\begin{bmatrix} W(1) \\ \int rdW \\ \int WdW \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \frac{1}{2}(\lambda^2 - \gamma_0)/\lambda^2 \end{bmatrix}\right\}.\end{aligned}$$

In particular,

$$T\hat{\rho} \xrightarrow{L} \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ \int rdW \\ \int WdW \end{bmatrix}$$
$$+ \frac{\lambda^2 - \gamma_0}{2\lambda^2} \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Note that the second component can be consistently estimated by

$$\frac{T^2 s_{\hat{\rho}}^2}{\hat{\sigma}_u^2} \frac{\hat{\lambda}^2 - \hat{\gamma}_0}{2}$$

because

$$T^2 \mathbf{s}_{\hat{\rho}}^2 \xrightarrow{L} \frac{\sigma_u^2}{\lambda^2} \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Accordingly, we get the PP ρ test

$$T\hat{\rho} - \frac{T^2 \mathbf{s}_{\hat{\rho}}^2}{\hat{\sigma}_u^2} \frac{\hat{\lambda}^2 - \hat{\gamma}_0}{2} \xrightarrow{L} \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ \int rdW \\ \int WdW \end{bmatrix}$$

which follows the same asymptotic distribution as the DF ρ test.

Similarly, the variance of $\hat{\boldsymbol{\beta}}$ follows

$$\begin{aligned} \mathbf{S}_T \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}} \mathbf{S}_T &= \hat{\sigma}_u^2 \left\{ \mathbf{S}_T^{-1} \left[\sum_{i=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right] \mathbf{S}_T^{-1} \right\}^{-1} \\ &= \hat{\sigma}_u^2 \begin{bmatrix} 1 & T^{-2} \sum_{i=1}^{T} t & T^{-\frac{3}{2}} \sum_{i=1}^{T} \xi_{t-1} \\ \cdot & T^{-3} \sum_{i=1}^{T} t^2 & T^{-\frac{5}{2}} \sum_{i=1}^{T} t\xi_{t-1} \\ \cdot & \cdot & T^{-2} \sum_{i=1}^{T} \xi_{t-1}^2 \end{bmatrix}^{-1} \\ &\xrightarrow{L} \sigma_u^2 \begin{bmatrix} 1 & \frac{1}{2} & \lambda \int W(r)dr \\ \cdot & \frac{1}{3} & \lambda \int rW(r)dr \\ \cdot & \cdot & \lambda^2 \int W(r)^2 dr \end{bmatrix}^{-1}. \end{aligned}$$

In particular, the standard error of $\hat{\rho}$ follows

$$T\mathbf{s}_{\hat{\rho}} \xrightarrow{L} \frac{\sigma_u}{\lambda} \left\{ \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}^{\frac{1}{2}}$$

and the t-test follows

$$t_{\hat{\rho}} = \frac{T\hat{\rho}}{Ts_{\hat{\rho}}} \xrightarrow{L} (\frac{\lambda}{\sigma_u})\frac{\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2dr \end{bmatrix}^{-1}\begin{bmatrix} W(1) \\ \int rdW \\ \int WdW \end{bmatrix}}{\left\{\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2dr \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right\}^{\frac{1}{2}}}$$

$$+ \quad (\frac{\lambda}{\sigma_u})\frac{\lambda^2-\gamma_0}{2\lambda^2}\left\{\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2dr \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right\}^{\frac{1}{2}}.$$

Thus, we get

$$(\frac{\sigma_u}{\lambda})t_{\hat{\rho}} \xrightarrow{L} \frac{\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2dr \end{bmatrix}^{-1}\begin{bmatrix} W(1) \\ \int rdW \\ \int WdW \end{bmatrix}}{\left\{\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2dr \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right\}^{\frac{1}{2}}}$$

$$+ \quad \frac{\lambda^2-\gamma_0}{2\lambda^2}\left\{\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2dr \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right\}^{\frac{1}{2}}.$$

Note that the second component can be consistently estimated by

$$\frac{Ts_{\hat{\rho}}}{\hat{\sigma}_u}\frac{\hat{\lambda}^2-\hat{\gamma}_0}{2\hat{\lambda}}$$

because

$$T^2\mathbf{s}_{\hat{\rho}}^2 \xrightarrow{L} \frac{\sigma_u^2}{\lambda^2}\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2dr \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Accordingly, we get the PP t test

$$\frac{\hat{\sigma}_u}{\hat{\lambda}}t_{\hat{\rho}} - \frac{Ts_{\hat{\rho}}}{\hat{\sigma}_u}\frac{\hat{\lambda}^2-\hat{\gamma}_0}{2\hat{\lambda}} \xrightarrow{L} \frac{\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2dr \end{bmatrix}^{-1}\begin{bmatrix} W(1) \\ \int rdW \\ \int WdW \end{bmatrix}}{\left\{\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & \frac{1}{2} & \int W(r)dr \\ \cdot & \frac{1}{3} & \int rW(r)dr \\ \cdot & \cdot & \int W(r)^2dr \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right\}^{\frac{1}{2}}},$$

which follows the same asymptotic distribution as the DF t test.

Appendix

13.A Asymptotic Theory

13.A.1 Functional Central Limit Theorem

For the purpose of deriving asymptotic distributions for unit root tests, it is convenient to generalize the concept of convergence in distribution. Instead of considering a sequence of random variables or random vectors, we will consider a sequence of random functions. This consideration leads to a generalized version of the central limit theorem.

Let $(\mathrm{S}, \mathcal{F}, Pr)$ be a probability space and S be a metric space with a metric d. The class $\mathcal{B}$ of *Borel sets* in M is the σ-field generated by the open sets of M. If a function x which maps S into M is measurable $\mathcal{F}/\mathcal{B}$, then x is a *random element.* A random element x induces a probability measure Pr^* on $(\mathrm{M}, \mathcal{B})$ when we define $Pr^*(B) = Pr(x \in B)$ for any B in $\mathcal{B}$. A sequence $\{x_j : j \geq 1\}$ of random elements is said to *converge in distribution* to a random element x_0 if

(13.A.1) ???? Masao needs to check this!

13.B Procedures for Unit Root Tests

13.B.1 Said-Dickey Test (ADF.EXP)

Said-Dickey test with the general-to-specific rules proceeds as follows:

(i) Choose whether or not a constant and a time trend should be included in the regression by selecting an appropriate alternative hypothesis. If the variable of interest does not exhibit any secular trend, an appropriate alternative hypothesis should be that the variable is stationary with non-zero mean and without a time trend. In this case, the regression should include a constant but no time trend. On the other hand, if the variable of interest exhibits a secular trend, an appropriate alternative hypothesis is that the variable is trend stationary. Therefore, the regression should include both a constant and a linear time trend.

(ii) Select the maximum order of lagged polynomials (the corresponding variable to be determined is `P`).

(iii) Determine the order of autoregressive process by following Campbell and Perron (1991)'s recommendation.

(iv) If the t ratio consistent with the specification of the regression form is negative and greater than the appropriate critical value in absolute value, then reject the null of a unit root.

13.B.2 Park's J Test (JPQ.EXP)

Park's $J(p, q)$ test proceeds as follows:

(i) Choose the order of the maintained trend in the regression (the corresponding variable in the program is `P`). If the variable of interest does not exhibit a secular time trend, the maintained hypothesis is that it includes only a constant (set `P=0`). However, if it shows a secular time trend, the maintained hypothesis is that it possesses a linear time trend (set `P=1`).

(ii) Select the largest order of additional time polynomials (the corresponding variable in the program is `Q`) and its range (the corresponding variable in the program is `DQ`) in the regression. If the variable of interest does not exhibit a secular time trend, the maintained hypothesis is that it includes only a constant (set `Q=1`). However, if it shows a secular time trend, the maintained hypothesis is that it possesses a linear time trend (set `Q=2`). Choose an appropriate `DQ` depending on how many test results you want. We recommend either `DQ=2` or `DQ=3`.

(iii) If `J(p,q)` is smaller than the appropriate critical value, then reject the null of difference stationarity.

13.B.3 Park's G Test (GPQ.EXP)

Park's $G(p,q)$ test proceeds as follows:

(i) Choose the order of the maintained trend in the regression (the corresponding variable in the program is `P`). If the variable of interest does not exhibit a secular time trend, the maintained hypothesis is that it includes only a constant (set `P=0`). However, if it shows a secular time trend, the maintained hypothesis is that it possesses a linear time trend (set `P=1`).

(ii) Select the largest order of additional time polynomials (the corresponding variable in the program is `Q`) and its range (the corresponding variable in the program is `DQ`) in the regression. If the variable of interest does not exhibit a secular time trend, the maintained hypothesis is that it includes only a constant (set `Q=1`). However, if it shows a secular time trend, the maintained hypothesis is that it possesses a linear time trend (set `Q=2`). Choose an appropriate `DQ` depending on how many test results you want. We recommend either `DQ=2` or `DQ=3`.

(iii) Specify an appropriate method to estimate the long-run covariance matrix, Ω_T.

See chapter 6 for more details (the corresponding variables to be specified are `MAXD, ST, BST,` and `MSERHO`).

(iv) If `G(p,q)` is greater than the appropriate critical value, then reject the null of stationarity.

Exercises

13.1 Imagine that you are applying the Said-Dickey (augmented Dickey-Fuller) test to the log real GDP for the United States. Explain the Said-Dickey test (the definition, the null and alternative hypotheses that are appropriate in this context, and the small sample properties compared with the Phillips and Perron test). If the test statistic takes the value of -3.33, do you reject the null hypothesis at the 5 percent level? What if the value is -1.47? What if the value is +3.99? The critical values for the Said-Dickey test are given in Table 13.2, in which p is the order of time polynomial included in the regression.

Table 13.2: Probability of smaller values

0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
$p = 0$ (a constant)							
-3.43	-3.12	-2.86	-2.57	-0.44	-0.07	0.23	0.60
$p = 1$ (a constant and a time trend)							
-3.96	-3.66	-3.41	-3.12	-1.25	-0.94	-0.66	-0.33

13.2 Imagine that you are applying the Said-Dickey (augmented Dickey-Fuller) test to the log real exchange rate for the United States and United Kingdom for the purpose of testing Purchasing Power Parity. Explain the Said-Dickey test (the definition, the null and alternative hypotheses which are appropriate in this context, and the small sample properties compared with the Phillips and Perron test.) If the test statistic takes the value of -2.93, do you reject the null hypothesis at the 5 percent level? What if the value is -2.67? What if the value is +3.99. The critical values for the Said-Dickey test are given in 13.2, in which p is the order of time polynomial included in the regression.

References

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.

Bierens, H. J., and S. Guo (1993): "Testing Stationarity and Trend Stationarity Against the Unit Root Hypothesis," *Econometric Reviews*, 12, 1–32.

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.

Campbell, J. Y., and P. Perron (1991): "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots," in *NBER Macroeconomics Annual*, ed. by O. J. Blanchard, and S. Fischer, pp. 141–201. MIT Press, Cambridge, MA.

Choi, I., and B. C. Ahn (1999): "Testing the Null of Stationarity for Multiple Time Series," *Journal of Econometrics*, 88(1), 41–77.

Christiano, L. J., and M. Eichenbaum (1991): "Unit Roots in Real GNP: Do We Know, and Do We Care?," *Carnegie-Rochester Conference Series on Public Policy*, 32, 7–61.

Cochrane, J. H. (1988): "How Big is the Random Walk in GNP?," *Journal of Political Economy*, 96(5), 893–920.

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.

——— (1981): "Likelihood Ratio Statistics for Autoregressive Time Series With a Unit Root," *Econometrica*, 49(4), 1057–1072.

Diebold, F. X., and M. Nerlove (1990): "Unit Roots in Economic Time Series: A Selective Survey," in *Advances in Econometrics: Cointegration, Spurious Regressions, and Unit Roots*, ed. by T. B. Fomby, and G. F. Rhodes, pp. 3–69. JAI Press, Greenwich, Connecticut.

Elliott, G., T. J. Rothenberg, and J. H. Stock (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64(4), 813–836.

Faust, J. (1996): "Near Observational Equivalence and Theoretical Size Problems with Unit Root Tests," *Econometric Theory*, 12, 724–731.

Fukushige, M., M. Hatanaka, and Y. Koto (1994): "Testing for the Stationarity and the Stability of Equilibrium," in *Advances in Econometrics, Sixth World Congress*, ed. by C. A. Sims, vol. I, pp. 3–45. Cambridge University Press.

Fuller, W. A. (1976): *Introduction to Statistical Time Series.* John Wiley and Sons, New York.

Hall, A. R. (1994): "Testing for a Unit Root in Time Series with Pretest Data-Based Model Selection," *Journal of Business and Economic Statistics*, 12, 461–470.

Hamilton, J. D. (1994): *Time Series Analysis.* Princeton University Press, Princeton.

Hansen, L. P. (1993): "Semiparametric Efficiency Bound for Linear Time-Series Models," in *Models, Methods, and Applications of Econometrics: Essays in Honor of A.R. Bergstrom*, ed. by P. C. B. Phillips. Blackwell, Oxford.

Kahn, J. A., and M. Ogaki (1990): "A Chi-Square Test for a Unit Root," *Economics Letters*, 34, 37–42.

——— (1992): "A Consistent Test for the Null of Stationarity Against the Alternative of a Unit Root," *Economics Letters*, 39, 7–11.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. C. 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(1–3), 159–178.

Nelson, C. R., and C. I. Plosser (1982): "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics*, 10, 139–162.

Newey, W. K., and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

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(429), 268–281.

Park, J. Y. (1989): "On the Joint Test of A Unit Root and Time Trend," Manuscript, Cornell University.

——— (1990): "Testing for Unit Roots and Cointegration by Variable Addition," in *Advances in Econometrics: Cointegration, Spurious Regressions and Unit Roots*, ed. by T. Fomby, and G. Rhodes, vol. 8, pp. 107–133. JAI Press, Greenwich, CT.

Park, J. Y., and B. Choi (1988): "A New Approach to Testing for a Unit Root," CAE Working Paper No. 88-23, Cornell University.

Phillips, P. C. B. (1987): "Time Series Regression with a Unit Root," *Econometrica*, 55(2), 277–301.

Phillips, P. C. B., and P. Perron (1988): "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75(2), 335–346.

Said, E. S., and D. A. Dickey (1984): "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika*, 71(3), 599–607.

Schwert, G. W. (1989): "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business and Economic Statistics*, 7, 147–159.

Stock, J. H., and M. W. Watson (1988): "Variable Trends in Economic Time Series," *Journal of Economic Perspectives*, 2(3), 147–74.

Watson, M. W. (1994): "Vector Autoregressions and Cointegration," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. IV, chap. 47, pp. 2843–2915. Elsevier Science Publishers.