Massachusetts Institute of Technology
Department of Economics
Time Series 14.384

Guido Kuersteiner

Lecture 7: Unit Root Asymptotics and Unit Root Tests

In this lecture we relax the stationarity assumption in the sense that we allow for processes of the form

$$x_t = x_{t-1} + u_t, \tag{7.1}$$

where $u_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$ such that $\sum |j|^{\frac{1}{2}} |c_j| < \infty$ and ε_t i.i.d.$(0,1)$. Under these conditions u_t is weakly stationary. If the polynomial $C(L) = \sum_{j=0}^{\infty} c_j L^j$ is invertible then we can also view x_t as being generated by an infinite order AR model. Thus

$$C(L)^{-1}(1-L)x_t = \varepsilon_t$$

where the AR-polynomial $C(L)^{-1}(1-L)$ now has one root on the unit circle. In other words we are considering a generalization of the model

$$x_t = \pi_1 x_{t-1} + ... + \pi_p x_{t-p} + \varepsilon_t$$

with $(1 - \pi_1 - ... - \pi_p) = 0$. In particular we are interested in estimation and testing of the unit root. It turns out, that this can be done without fully specifying the short run dynamics of the model. We can show that parameter estimates are consistent even if the model is misspecified and that they converge at a faster rate than the usual $\sqrt{T}$ asymptotics suggest. This property is often referred to as superconsistency. These facts open the way to some novel statistical procedures based on semiparametric approximations to the true underlying model.

In order to develop the necessary asymptotic theory we return to the model in equation (7.1). Expanding x_n in terms of past innovations it follows that $x_n = \sum_{t=1}^{n} u_t + x_0$. We can use the BN decomposition to analyze

$$u_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}.$$

The goal is to express u_t as the sum of an independent innovation plus the difference between two stationary processes. We first obtain a new representation of the lag polynomial $C(L)$.

$$\begin{aligned} C(L) &= \sum_{i=0}^{\infty} c_i L \\ &= \sum_{i=0}^{\infty} c_i - \sum_{i=1}^{\infty} c_i + \left(\sum_{i=1}^{\infty} c_i - \sum_{i=2}^{\infty} c_i \right) L + \left(\sum_{i=2}^{\infty} c_i - \sum_{i=3}^{\infty} c_i \right) L^2 + ... \\ &= \sum_{i=0}^{\infty} c_i + (L-1) \sum_{i=1}^{\infty} c_i + \left(L^2 - L\right) \sum_{i=2}^{\infty} c_i + ... \end{aligned}$$

We now define the coefficients $\tilde{c}_j = \sum_{k=j+1}^{\infty} c_k$ such that $C(L) = C(1) + (L-1)\tilde{C}(L)$ where $\tilde{C}(L) = \sum_{j=0}^{\infty} \tilde{c}_j L^j$. It follows immediately that $u_t = C(1)\varepsilon_t + (L-1)\,\tilde{C}(L)\varepsilon_t$ where we also commonly write $\tilde{\varepsilon}_t = \tilde{C}(L)\varepsilon_t$. The process u_t can now be written as

$$u_t = C(1)\varepsilon_t + \tilde{\varepsilon}_{t-1} - \tilde{\varepsilon}_t.$$

If we sum up the u_t terms then the differences $\tilde{\varepsilon}_{t-1} - \tilde{\varepsilon}_t$ cancel in the summation except for the first and last term. Such a sum is sometimes referred to as a telescoping sum. We have therefore

$$x_n = C(1)\sum_{t=1}^{n}\varepsilon_t + \tilde{\varepsilon}_0 - \tilde{\varepsilon}_n.$$

Note that $E\tilde{\varepsilon}_t = 0$ and that $var(\tilde{\varepsilon}_t) < \infty$. This follows from

$$\sum j^{\frac{1}{2}} |c_j| < \infty \Rightarrow \sum \tilde{c}_j^2 < \infty$$

since

$$\begin{aligned}
\sum_{j=0}^{\infty} \tilde{c}_j^2 &= \tilde{c}_o + \sum_{j=1}^{\infty}\left(\sum_{k=j+1}^{\infty}|c_k|\right)^2 \leq \tilde{c}_o + \sum_{j=1}^{\infty}\left(\sum_{k=j+1}^{\infty}k^{1/2}|c_k|\right)\left(\sum_{k=j+1}^{\infty}|c_k|/k^{1/2}\right) \\
&\leq \tilde{c}_o + \sum_{j=1}^{\infty} j^{1/2}|c_j| \sum_{j=1}^{\infty}\sum_{k=j+1}^{\infty}|c_k|/k^{1/2} \\
&= \tilde{c}_o + \sum_{j=1}^{\infty} j^{1/2}|c_j| \sum_{k=1}^{\infty}|c_k| k^{1/2} < \infty
\end{aligned}$$

where the first inequality follows from the Cauchy-Schwartz inequality and the last equality follows from counting the number of times each element $|c_k|/k^{1/2}$ appears in the double sum.

We now define a stochastic process on $[0,1]$ as

$$\begin{aligned}
X_n(r) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{[nr]} u_j + \frac{1}{\sqrt{n}}x_0 \qquad r \in [0,1] \\
&= \frac{C(1)}{\sqrt{n}}\sum_{j=1}^{[nr]}\varepsilon_j + \frac{1}{\sqrt{n}}\left[\tilde{\varepsilon}_0 - \tilde{\varepsilon}_{[nr]}\right] + \frac{1}{\sqrt{n}}x_0
\end{aligned}$$

where $[nr]$ denotes the largest integer number less than nr. The process $X_n(r)$ is right continuous and has left limits. Functions with this property are usually called CADLAG.

The problem with the function space of right continuous functions with left limits is that it is not separable under the uniform metric. Lack of separability can lead to nonmeasurability such that the standard theory of weak convergence does not apply. For the processes we consider here this is however not a serious concern. There are several ways to proceed. One can use a continuous approximation to $x_n(r)$ or one can use a different metric, the so called Skorokhod metric to make the space of CADLAG functions complete and separable..

We will not go into these technical details in this lecture but just point out a few elements of the proofs needed to establish that $X_n(r)$ converges weakly to a limit process. This result is known as Donsker's Theorem in the probability literature and the interested reader is referred to Billingsley (1968).

We have seen that $\tilde{\varepsilon}_t$ is stationary with finite variance. It can be shown that

$$\sup_r \left|\frac{1}{\sqrt{n}}(\tilde{\varepsilon}_0 - \tilde{\varepsilon}_{[nr]})\right| \underset{p}{\rightarrow} 0,$$

and it also is innocuous to assume $\frac{1}{\sqrt{n}}x_0 \xrightarrow{p} 0$. We therefore are left with

$$X_n(r) = \frac{C(1)}{\sqrt{n}}\sum_{j=1}^{[nr]}\varepsilon_j + o_p(1).$$

It is an immediate consequence of the CLT that for r fixed

$$X_n(r) = C(1)\frac{\sqrt{[nr]}}{\sqrt{n}}\frac{1}{\sqrt{[nr]}}\sum_{j}^{[nr]}\varepsilon_j \to N\left(0, C(1)^2 r\right),$$

where $\frac{\sqrt{[nr]}}{\sqrt{n}} \to \sqrt{r}$. Note that $C(1)^2 = 2\pi f_u(0)$. Moreover, for $r_1 < r_2 < r_3 ... < r_m$ fixed, we have

$$[X_n(r_1), X_n(r_2) - X_n(r_1), ..., X_n(r_m) - X(r_{m-1})] \xrightarrow{d} \left[N\left(0, C(1)^2 R\right)\right],$$

where $R = diag(r_1, r_2 - r_1, ..., r_m - r_{m-1})$. It should be pointed out that the increments $X_n(r_i) - X_n(r_{i-1})$ are independent. We have therefore established that the finite dimensional distributions of $X_n(r)$ converges to the finite dimensional distributions of Brownian Motion denoted as $B(r)$ where $B(r)$ is a process defined from standard Brownian Motion $W(r)$. Standard Brownian motion has the following properties.

1. $W(0) = 0$
2. $W(t)$ has stationary and independent increments and for all t and s such that $t > s$ we have $W(t) - W(s) \sim N(0, (t-s))$
3. $W(t) \sim N(0, t)\ \forall\ t$
4. $W(t)$ is sample path continuous

We can now define $B(r)$ as $\sigma W(r)$. We have shown that our limit process has the same finite dimensional distributions as Brownian Motion. To show convergence of $X_n(r)$ in the function space we need to establish in addition that for all $\varepsilon, \eta > 0$ there exists a $\delta > 0$ such that

$$P\left(\sup_{|r-s|<\delta} |X_n(r) - X_n(s)| > \varepsilon\right) < \eta$$

as n goes to infinity. A proof of this statement is omitted but can be found in Billingsley.

We can now use the continuous mapping theorem to analyze the behavior of certain statistics of interest. The first result is an asymptotic representation of the sample mean. Note that unlike in the stationary case the sample mean does not converge to a constant but rather to a random variable. In particular we have

$$\frac{1}{n^{3/2}}\sum^{n} x_t \Rightarrow \int_0^1 B(s)ds$$

where $\int_0^1 B(s)ds$ is a standard Riemann integral and $\Rightarrow$ denotes weak convergence in the function space. We will often write $\int_0^1 B(s)ds = \int B$ to simplify the notation. The integral is a random variable because it has a stochastic process as its argument. We now turn to the proof of this result.

We have from before

$$x_t = \sum_{j=1}^{t-1} u_j + u_t + x_0$$

such that $\sum_{t=1}^{n} x_t = \sum_{t=1}^{n}\left(S_{t-1} + u_t + x_0\right)$ with $S_{t-1} = \sum_{j=1}^{t-1} u_j$. Then

$$\begin{aligned}\frac{1}{n^{3/2}}\sum x_t &= \frac{1}{n}\sum_{t=1}^{n}\frac{S_{t-1}}{\sqrt{n}} + \frac{\sum u_t}{n^{3/2}} + \frac{x_0}{\sqrt{n}} \\ &= \sum_{t=1}^{n}\int_{\frac{t-1}{n}}^{t/n} X_n(r)dr + o_p(1)\end{aligned}$$

since $\frac{\sum u_t}{n^{3/2}} = O_p(n^{-1/2})$ and $\frac{X_0}{\sqrt{n}} = O_p(n^{-1/2})$. Also $X_n(r) = \frac{S_{[nr]}}{\sqrt{n}} + o_p(1) = \frac{S_{t-1}}{\sqrt{n}} + o_p(1)$ for $\frac{t-1}{n} \leq r < \frac{t}{n}$ and $\int_{\frac{t-1}{n}}^{t/n} dr = \frac{1}{n}$, such that

$$\frac{1}{n^{3/2}} \sum_{t=1}^{n} x_t = \int_0^1 X_n(r)dr + o_p(1).$$

Since the integral is continuous it now follows that

$$\int_0^1 X_n(r)dr \Rightarrow \int_0^1 B(r)dr$$

by the continuous mapping theorem. By linearity of the integral $\int_0^1 B(r)dr$ is Gaussian $N(0,\nu)$ where $\nu = \frac{C(1)^2}{3}$. This can be seen from noting that $E\int_0^1 B(r)dr = 0$ by linearity of the integral and $E(\int_0^1 B(r)dr)^2 = \int_0^1 \int_0^1 E[B(r)B(s)]drds$. Now $E[B(r)B(s)] = C(1)^2 \min(r,s)$ such that $\nu = 2C(1)^2 \int_0^1 \int_0^s rdrds = C(1)^2/3$.

We next turn to the analysis of the sample variance of x_t. In particular we have

$$\begin{aligned}
\frac{1}{n^2}\sum_{t=1}^{n} x_t^2 &= \frac{1}{n^2}\sum_{t=1}^{n}(S_t + x_0)^2 \\
&= \frac{1}{n^2}\sum_{t=1}^{n}\left(S_t^2 + 2x_0 S_t + x_0^2\right) \\
&= C(1)^2 \sum_{t=1}^{n} \int_{\frac{t-1}{n}}^{t/n} \frac{S_{[nr]}^2}{nC(1)^2} dr + \frac{2x_0 C(1)}{\sqrt{n}} \sum_{t=1}^{n} \int_{\frac{t-1}{n}}^{t/n} \frac{S_{[nr]}}{\sqrt{n}C(1)} + \frac{x_0^2}{n} \\
&= C(1)^2 \int_0^1 \frac{X_n^2(r)}{C(1)^2} dr + \frac{2x_0 C(1)}{\sqrt{n}} \int_0^1 \frac{X_n(r)}{C(1)} dr + \frac{x_0^2}{n} + o_p(1) \\
&\xrightarrow{d} C(1)^2 \int_0^1 W(r)^2 dr
\end{aligned}$$

which follows again by the continuous mapping theorem and the previous results, in particular the fact that $\int_0^1 \frac{X_n(r)}{C(1)} dr = O_p(1)$.

We are also interested in the behavior of score functions of the form

$$\begin{aligned}
\frac{1}{n}\sum x_{t-1}u_t &= \frac{1}{n}\sum_{t=1}^{n}(S_{t-1} + x_0)\, u_t \\
&= \frac{1}{n}\sum_{t=1}^{n} S_{t-1}u_t + op(1).
\end{aligned}$$

We use partial summation. Note that

$$\begin{aligned}
\sum_{t=1}^{n} \Delta S_t^2 &= \sum_{t=1}^{n}\left[(S_{t-1} + u_t)^2 - S_{t-1}^2\right] \\
&= \sum_{t=1}^{n}\left(u_t^2 + 2u_t S_{t-1}\right) \\
&= \sum_{t=1}^{n} u_t^2 + 2\sum_{t=1}^{n} u_t S_{t-1},
\end{aligned}$$

such that

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n} S_{t-1}u_t &= \frac{1}{2}\left(\frac{1}{n}\sum_{t=1}^{n}\Delta S_t^2 - \frac{1}{n}\sum_{t=1}^{n}u_t^2\right) \\
&= \frac{1}{2}\left(\frac{1}{n}S_n^2 - \frac{1}{n}\sum u_t^2\right) \\
&= \frac{1}{2}\left[\left(\frac{S_n}{\sqrt{n}}\right)^2 - \frac{1}{n}\sum u_t^2\right] \\
&\xrightarrow{d} \frac{1}{2}\left(B(1)^2 - \sigma^2\right)
\end{aligned}$$

where $\sigma^2 = \sum c_j^2$. Then

$$\frac{1}{2}\left(B(1)^2 - \sigma^2\right) = \frac{C(1)^2}{2}\left(W(1)^2 - 1\right) + \frac{1}{2}\underbrace{\left(C(1)^2 - \sigma^2\right)}_{\lambda}$$

It can be shown that $\frac{C(1)^2}{2}\left(W(1)^2 - 1\right)$ has a representation as a stochastic (Ito) integral $\int_0^1 BdB$. Also note that $W(1)^2 \sim \chi_1^2$. It now follows immediately that

$$n(\hat{\alpha} - 1) = \frac{n^{-1}\sum x_{t-1}u_t}{n^{-2}\sum x_t^2} \Rightarrow \frac{\int_0^1 BdB + \lambda}{\int_0^1 B^2}$$

where

$$\hat{\alpha} = \frac{\sum x_t x_{t-1}}{\sum x_t^2}$$

and $x_t = x_{t-1} + u_t$. Several things are worth noting at this point. First of all, the estimator converges at rate $O_p(n^{-1})$ to the true parameter value, regardless of whether the short run dynamics of the model are correctly specified or not. This sharply contrasts with the stationary case where misspecification of the model leads to inconsistency. However, in the case of misspecification, i.e. if the innovations u_t are serially correlated, the limit distribution has an asymptotic bias term $\lambda / \int B^2$. It is the presence of this bias term, which makes inference hard, since it is no longer possible to use tables to obtain critical values. Note that if $\lambda = 0$, i.e., $u_t \sim$ iid then

$$n(\hat{\alpha} - 1) \Rightarrow \frac{\int WdW}{\int_0^1 W^2},$$

such that the limit distribution is nuisance parameter free. In other words, there is no need to use a t-test in this case because the limit distribution of $\hat{\alpha}$ does not depend on an unknown parameter.

In this simple situation a unit root test would compare $n\left(\hat{\rho} - 1\right)$ against the critical value of a statistic with distribution

$$\frac{\frac{1}{2}\left(W(1)^2 - 1\right)}{\int W(r)^2 dr}.$$

This is possible since under $H_0 : \rho = 1$ we have shown that

$$n\left(\hat{\rho} - 1\right) \Rightarrow \frac{\frac{1}{2}\left(W(1)^2 - 1\right)}{\int W(r)^2 dr}$$

Critical values are tabulated in Hamilton, Table B5, case 1. For example, if $n = 100$, then one can reject H_0 at 5% if $n(\hat{\rho} - 1) < -7.9$ for a one-sided test. We can also look at the t-statistic, assuming again that u_t iid

$(0, \sigma^2)$

$$t = \frac{\hat{\rho}_n - 1}{\hat{\sigma}_\rho} = \frac{\hat{\rho} - 1}{\hat{\sigma}\left(\sum x_{t-1}^2\right)^{-1/2}}$$

where

$$\begin{aligned}
\hat{\sigma} &= \frac{1}{n}\sum\left(x_t - \hat{\rho}x_{t-1}\right)^2 = \frac{1}{n}\sum\left(u_t + (\rho - \hat{\rho})\,x_{t-1}\right)^2 \\
&= \frac{1}{n}\sum u_t^2 + \frac{1}{n}(\rho - \hat{\rho})\sum x_{t-1}u_t \\
&\quad + \frac{1}{n}(\rho - \hat{\rho})^2\sum x_{t-1}^2 \xrightarrow{p} \sigma_u^2
\end{aligned}$$

and

$$n(\hat{\rho} - 1)\left(\frac{1}{n^2}\sum x_{t-1}^2\right)^{\frac{1}{2}} \Rightarrow \frac{\frac{\sigma^2}{2}\left(W(1)^2 - 1\right)}{\left(\sigma^2\int_0^1 W(r)^2 dr\right)^{\frac{1}{2}}}$$

so

$$t \Rightarrow \frac{\frac{1}{2}\left(W(1)^2 - 1\right)}{\left(\int_0^1 W(r)^2 dr\right)^{\frac{1}{2}}}$$

which is tabulated in table B6, case 1. We will now see that the limit distribution of $\hat{\rho}$ depends on the fitted model, even if the true underlying model remains unchanged. In particular we consider fitting a constant while the true process is still assumed to be $x_t = x_{t-1} + u_t$ with $u_t = C(L)\varepsilon_t$. The estimator for the autoregressive parameter can now be written as

$$\begin{aligned}
\hat{\rho} &= \frac{\sum_{t=2}^n (x_t - \bar{x})(x_{t-1} - \bar{x}_{-1})}{\sum_{t=2}^n (x_{t-1} - \bar{x}_{-1})^2} = 1 + \frac{\sum_{t=2}^n (u_t - \bar{x})(x_{t-1} - \bar{x}_{-1})}{\sum_{t=2}^n (x_{t-1} - \bar{x}_{-1})^2} \\
&= 1 + \frac{\sum_{t=2}^n u_t (x_{t-1} - \bar{x}_{-1})}{\sum_{t=2}^n (x_{t-1} - \bar{x})^2} = \frac{\sum_{t=2}^n u_t x_{t-1} - \bar{x}\sum_{t=2}^n u_t}{\sum_{t=2}^n x_{t-1}^2 - n\bar{x}^2}
\end{aligned}$$

where $\bar{x} = n^{-1}\sum_{t=2}^n x_t$ and $\bar{x}_{-1} = n^{-1}\sum_{t=2}^n x_{t-1}$ so

$$n(\hat{\rho} - 1) = \frac{\frac{1}{n}\sum_{t=2}^n u_t x_{t-1} - \bar{x}\frac{1}{n}\sum_{t=2}^n u_t}{\frac{1}{n^2}\sum_{t=2}^n x_{t-1}^2 - \frac{1}{n}\bar{x}^2} + o_p(1).$$

Now, using the previous arguments we can establish the following results.

$$\frac{1}{n^2}\sum_{t=2}^n x_{t-1}^2 \Rightarrow C(1)^2\int W(r)^2 dr$$

$$\frac{1}{n}\bar{x}^2 = \left(\frac{1}{n^{3/2}}\sum_{t=2}^n x_t\right)^2 \Rightarrow \left(\int_0^1 W(r)dr\right)^2 C(1)^2$$

$$\frac{1}{n}\sum_{t=2}^n u_t x_{t-1} \Rightarrow \left(W(1)^2 - 1\right)\frac{C(1)^2}{2} + \lambda$$

$$\frac{\bar{x}}{\sqrt{n}}\frac{1}{\sqrt{n}}\sum_{t=2}^n u_t \Rightarrow \left(\int W(r)dr W(1)\right)C(1)^2.$$

This leads to the following asymptotic representation of the estimator

$$\begin{aligned} n(\hat{\rho}-1) &\Rightarrow \frac{\frac{C(1)^2}{2}\left(W(1)^2-1\right)+\lambda-W(1)\int_0^1 W(r)drC(1)^2}{C(1)^2\left[\int W(r)^2dr-\left(\int W(r)dr\right)^2\right]} \\ &= \frac{\frac{1}{2}\left(W(1)^2-1\right)-W(1)\int_0^1 W}{\left[\int W^2-\left(\int W\right)^2\right]}+\frac{\lambda}{C(1)^2\left[\int W^2-\left(\int W\right)^2\right]} \end{aligned}$$

If $\lambda = 0$ then the asymptotic distribution is free of nuisance parameters and critical values can be taken from Table B6, case 2. There are basically two possible ways to avoid the misspecification problem. One is to fit the correct parametric model to the short run dynamics. This approach is the basis of augmented Dickey-Fuller tests (ADF tests). The second approach, the Phillips Z_α test, uses a nonparametric correction to the test statistic to account for omitted serial dependence.. The advantage of this second approach is, that the unit root hypothesis can be tested without making specific assumptions about the parametric form of the dependence in u_t.

The Phillips Z_α test uses a non parametric correction to obtain a nuisance-parameter-free limit distribution. It is formulated in this way

$$n(\hat{\rho}-1)+\frac{\hat{\lambda}}{\frac{1}{n^2}\sum x_{t-1}^2}$$

with $\hat{\rho}$ being the demeaned estimator for the unit root and

$$\hat{\lambda}=\frac{1}{2}\left(2\pi\widehat{f}(0)-\hat{\sigma}_n\right)$$

where

$$\hat{\sigma}_n=\frac{1}{n}\sum \hat{u}_t^2=\frac{1}{n}\sum\left(x_t-\hat{\alpha}-\hat{\rho}x_{t-1}\right)^2$$

and

$$2\pi\widehat{f}(0)=\sum_{j=-M}^{M}\left(1-\frac{|j|}{M}\right)\hat{\gamma}(j)$$

with $\hat{\gamma}(j)=n^{-1}\sum \hat{u}_t\hat{u}_{t-j}$. Then $\hat{\lambda}-\lambda \xrightarrow{p} 0$ for $M=O(n^{1/4})$ and $\frac{1}{n^2}\sum x_{t-1}^2 \Rightarrow \int_0^1 W(r)^2drC(1)^2$ such that

$$n(\hat{\rho}-1)+\frac{\hat{\lambda}}{\frac{1}{n^2}\sum x_{t-1}^2} \Rightarrow \frac{\frac{1}{2}\left(W(1)^2-1\right)-W(1)\int_0^1 W(r)dr}{\int W(r)^2dr-\left(\int W(r)dr\right)^2}$$

We can use standard tables even though we have not fully specified the dynamics of the model. In particular the null is rejected at a certain significance level if

$$n(\hat{\rho}-1)+\frac{\hat{\lambda}}{\frac{1}{n^2}\sum x_{t-1}^2} < c_\alpha^1$$

or

$$n(\hat{\rho}-1)+\frac{\hat{\lambda}}{\frac{1}{n^2}\sum x_{t-1}^2} > c_\alpha^2$$

where c_α^1 and c_α^2 are critical values for the level alpha for a one sided test from table $B6$, case two. There is a similar test based on the t-statistic. Note that Monte Carlo studies indicate that the Z_α test has size distortions for certain models such as $\Delta x_t = \varepsilon_t - 0.8\varepsilon_{t-1}$. Here x_t is nonstationary but the Z_α test rejects H_0 too often.

As mentioned before a parametric way to remove the nuisance parameters from the limit distribution is to fully specify the short run dynamics. Assume that $\phi(L)x_t = \varepsilon_t$; $\varepsilon_t \sim iid(0,\sigma^2)$ and $\phi(L)$ such that $\phi(1)=0$.

Then we can write

$$\begin{aligned}
\phi(L) &= 1-\phi_1 L-\phi_2 L^2 \ldots-\phi_p L^p \\
&= 1-\left(\phi_1+\phi_2 \ldots+\phi_p\right) L+\left(\phi_2+\ldots+\phi_p\right) L \\
&\quad -\phi_2 L^2-\ldots \phi_p L^p \\
&= 1-\left(\phi_1+\phi_2+\ldots+\phi_\rho\right) L+\left(\phi_2+\ldots+\phi_\rho\right) L(1-L) \\
&\quad +\left(\phi_3+\ldots+\phi_\rho\right) L^2(1-L)+\ldots+\phi_p L^{p-1}(1-L)
\end{aligned}$$

so $\Delta x_t=\Pi x_{t-1}+\Pi_1 \Delta x_{t-1}+\ldots+\Pi_{\rho-1} \Delta x_{t-\rho+1}+\varepsilon_t$ where $\Pi_i=-\left(\phi_i+\ldots+\phi_\rho\right)$ and $\Pi=\left(\phi_1+\ldots+\phi_p\right)-1$.

So under the null hypothesis

$$H_0: \phi(L) \text{ has one unit root}$$

we have $\Pi=0$ and $\Delta x_t=u_t=\varepsilon_t /(1-\Pi_1 L-\ldots \Pi_{p-1} L^{p-1})$ is stationary. Stack the variables $z_t^1=(\Delta x_{t-1}, \ldots, \Delta x_{t-p}, 1, x_{t-1})$ and compute

$$\hat{\beta}=\left(\sum z_t z_t^{\prime}\right)^{-1} \sum z_t x_t$$

then

$$D_n(\hat{\beta}-\beta)=\left(D_n^{-1} \sum z_t z_t^1 D_n^{-1}\right)^{-1} D_n^{-1} \sum z_t \varepsilon_t$$

with

$$D_n=\left[\begin{array}{cccc}
n^{1/2} & & & \\
& \ddots & & \\
& & n^{1/2} & \\
& & & n
\end{array}\right].$$

Note that

$$D_n^{-1} \sum z_t z_t^1 D^{-1} \Rightarrow\left[\begin{array}{ccc}
\Gamma_\rho & 0 & 0 \\
0 & 1 & \int B \\
0 & \int B & \int B^2
\end{array}\right];$$

$$\left[\Gamma_\rho\right]_{ij}=\operatorname{cov}\left(u_i, u_j\right)$$

since $n^{-1} \sum \Delta x_t \underset{p}{\rightarrow} 0$ and $n^{-3/2} \sum \Delta x_t x_t \underset{\rho}{\rightarrow} 0$. Thus

$$\left(D_n^{-1} \sum z_t z_t^1 D_n^{-1}\right)^{-1} \Rightarrow\left[\begin{array}{cc}
\Gamma_\rho^{-1} & 0 \\
0 & \left[\begin{array}{cc}
1 & \int B \\
\int B & \int B^2
\end{array}\right]^{-1}
\end{array}\right].$$

Also,

$$D_n^{-1} \sum z_t \varepsilon_t=\left[\begin{array}{c}
\frac{1}{\sqrt{n}} \sum u_{t-1} \varepsilon_t \\
\vdots \\
\frac{1}{\sqrt{n}} \sum u_{t-p} \varepsilon_t \\
\frac{1}{\sqrt{n}} \sum \varepsilon_t \\
\frac{1}{n} \sum x_{t-1} \varepsilon_t
\end{array}\right]$$

The first $p-1$ elements converge to $N\left(0, \sigma^2 \Gamma_\rho\right)$. This is an important result in itself since it shows that the stationary part of the model can be estimated and tested for by standard inferential methods. The reason for this is that the parameters of the nonstationary part converge at a faster rate and can thus be considered as constant in the asymptotic theory for the parameters of the stationary part of the model.

The nonstationary part of the model behaves according to

$$\left(\frac{1}{\sqrt{n}}\sum\varepsilon_t, \frac{1}{n}\sum x_{t-1}\varepsilon_t\right) \Rightarrow \left(\sigma W(1), \frac{C(1)\sigma^2}{2}(W(1)-1)\right)$$

with $C(1) = (1 - \Pi_1 ... - \Pi_p)^{-1}$ from $\frac{1}{n}\sum x_{t-1}\varepsilon_t = \frac{1}{n}\sum S_{t-1}\varepsilon_t + op(1)$. Now by the BN decomposition we have

$$S_{t-1} = C(1)\sum_{j=1}^{t-1}\varepsilon_j + (\tilde{\varepsilon}_0 - \tilde{\varepsilon}_{t-1})$$

such that

$$\begin{aligned}
\frac{1}{n}\sum x_{t-1}\varepsilon_t &= C(1)\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{t-1}\varepsilon_j\varepsilon_t + \tilde{\varepsilon}_0\frac{1}{n}\sum\varepsilon_t + \frac{1}{n}\sum\varepsilon_t\tilde{\varepsilon}_{t-1} \\
&\Rightarrow \frac{C(1)\sigma^2}{2}\left(W(1)^2 - 1\right).
\end{aligned}$$

These results show that

$$n(\hat{\rho}-1) \xrightarrow{d} \frac{\frac{1}{2}\left(W(1)^2-1\right) - W(1)\int_0^1 W(r)dr}{C(1)\left(\int_0^1 W(r)dr - \left(\int_0^1 W(r)dr\right)^2\right)}$$

This can be seen by considering

$$\left[\begin{array}{cc} 1 & \int_0^1 B \\ \int_0^1 B & \int_0^1 B^2 \end{array}\right]^{-1} = \frac{1}{\int B^2 - (\int B)^2}\left[\begin{array}{cc} \int_0^1 B^2 & -\int_0^1 B \\ -\int_0^1 B & 1 \end{array}\right]$$

and

$$\int B = C(1)\sigma\int W(r)dr$$

$$\int B^2 = C(1)^2\sigma^2\int_0^1 W(r)^2 dr$$

such that it follows that

$$\begin{aligned}
n(\hat{\rho}-1) &= \frac{C(1)}{\int B^2 - \left(\int B\right)^2}\left(-\sigma^2 W(1)\int W + \frac{\sigma^2}{2}\left(W(1)^2-1\right)\right) \\
&= \frac{1}{C(1)\left[\int W^2 - \left(\int W\right)^2\right]}\left[\frac{1}{2}\left(W(1)^2-1\right) - W(1)\int W\right]
\end{aligned}$$

We see that the limit distribution is free of a bias term but still depends on nuisance parameters. The bias term vanished because we correctly modeled the short run dynamics. Unfortunately the limit distribution still depends on the unknown long run variance of the process. This problem can be overcome by considering a t-test rather than a Z test on the parameter directly. We consider a t-test of $H_0 : \rho = 1$

$$\frac{n(\hat{\rho}-1)}{s\hat{e}_\rho} = \frac{n(\hat{\rho}-1)}{\hat{\sigma}\left(D_n^{-1}\sum z_t z_t^1 D_n^{-1}\right)_{p+1}^{-1/2}}$$

where $\left(D_n^{-1}\sum z_t z_t^1 D_n^{-1}\right)_{p+1}^{-1/2}$ stands for the $p+1$ diagonal element of $\left(D_n^{-1}\sum z_t z_t^1 D_n^{-1}\right)^{-1/2}$. We have shown before that

$$\left(D_n^{-1}\sum z_t z_t^1 D_n^{-1}\right)_{p+1}^{-1/2} \Rightarrow \frac{1}{C(1)\sigma\left[\int W^2-\left(\int W\right)^2\right]^{1/2}}.$$

It now follows that

$$\frac{n\left(\hat{\rho}-1\right)}{\hat{\sigma}\left(D_n^{-1}\sum z_t z_t^1 D_n^{-1}\right)_{p+1}^{-1/2}} \Rightarrow \frac{\frac{1}{2}\left(W(1)^2-1\right)-W(1)\int_0^1 W(r)dr}{\left(\int_0^1 W^2(r)dr-\left(\int_0^1 W(r)dr\right)^2\right)^{1/2}}$$

which is free of nuisance parameters. Critical values can be obtained from Table B6, section 2. It should be pointed out that the limit distribution is independent of the number of estimated parameters for lagged Δx_{t-i}.