Econ 582
Trend-Cycle Decompositions and Unit Root Tests

Eric Zivot

April 22, 2013

Introduction

A convenient way of representing the levels of an economic time series y_t is through the so-called *trend-cycle decomposition*

$$\begin{aligned} y_t &= TD_t + Z_t \\ TD_t &= \text{deterministic trend} \\ Z_t &= \text{ random cycle/noise} \end{aligned}$$

For simplicity, assume

$$\begin{aligned} TD_t &= \kappa + \delta t \\ \phi(L)Z_t &= \theta(L)\varepsilon_t,\ \varepsilon_t \sim WN(0, \sigma^2) \end{aligned}$$

where $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ and $\theta\left(L\right) = 1 + \theta_1 L + \cdots + \theta_q L^q$. It is assumed that the polynomial $\phi(z) = 0$ has at most one root on the complex unit circle and $\theta(z) = 0$ has all roots outside the unit circle.

Trend Stationary and Difference Processes

Defn: The series y_t is called *trend stationary* if the roots of $\phi(z) = 0$ are outside the unit circle

Defn: The series y_t is called *difference stationary* if $\phi(z) = 0$ has one root on the unit circle and the others outside the unit circle.

Trend Stationary Process

If y_t is trend stationary then $\phi(L)$ is invertible and Z_t has the stationary or Wold representation

$$\begin{aligned} Z_t &= \phi(L)^{-1}\theta(L)\varepsilon_t = \psi(L)\varepsilon_t, \\ \psi(L) &= \phi(L)^{-1}\theta(L) = \sum_{k=0}^{\infty} \psi_k L^k, \psi_0 = 1 \text{ and } \psi(1) \neq 0. \end{aligned}$$

Here, y_t exhibits mean reversion around the deterministic trend $TD_t = \kappa + \delta t$

$$y_t = TD_t + \psi(L)\varepsilon_t$$

and the detrended series $y_t - TD_t = \psi(L)\varepsilon_t$ is covariance stationary.

Estimating the Trend

Suppose y_t is trend stationary and Z_t is covariance stationary ARMA(p,q)

$$y_t = TD_t + Z_t = \kappa + \delta t + \phi(L)^{-1}\theta(L)\varepsilon_t$$

- Trend parameters κ and δ can be consistently estimated by OLS ignoring the ARMA(p,q) structure in the errors

- Use Newey-West standard errors to correct for autocorrelation in Z_t

- ARMA(p,q) parameters can be consistently estimated from detrended series $\hat{Z}_t = y_t - \hat{\kappa} - \hat{\delta} t$

Example: Trend Stationary AR(2)

Let

$$\begin{aligned} y_t &= TD_t + Z_t \\ TD_t &= \kappa + \delta t \\ \phi(L)Z_t &= \varepsilon_t,\ \phi(L) = 1 - \phi_1 L - \phi_2 L^2 \end{aligned}$$

Assume that $\phi(z) = 0$ has all roots outside unit circle. Then y_t is mean-reverting about the deterministic trend $\kappa + \delta t$. That is,

$$\begin{aligned} \phi(L)y_t &= \phi(L)(\kappa + \delta t) + \phi(L)Z_t \Rightarrow \\ \phi(L)(y_t - \kappa - \delta t) &= \varepsilon_t \end{aligned}$$

or

$$\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \phi_2 \tilde{y}_{t-2} + \varepsilon_t,\ \tilde{y}_t = y_t - \kappa - \delta t$$

Difference Stationary Processes

If y_t is difference stationary then $\phi(L)$ can be factored as

$$\phi(L) = (1 - L)\phi^*(L)$$

where $\phi^*(z) = 0$ has all $p - 1$ roots outside the unit circle. In this case, ΔZ_t has the stationary ARMA($p - 1, q$) representation

$$\begin{aligned} \Delta Z_t &= \phi^*(L)^{-1}\theta(L)\varepsilon_t = \psi^*(L)\varepsilon_t \\ \psi^*(L) &= \phi^*(L)^{-1}\theta(L) = \sum_{k=0}^{\infty} \psi_k^* L^k, \psi_0^* = 1 \text{ and } \psi^*(1) \neq 0. \end{aligned}$$

and y_t does not exhibit mean reversion around the deterministic trend $TD_t = \kappa + \delta t$.

Example: Difference stationary AR(2)

Let

$$\begin{aligned} y_t &= TD_t + Z_t \\ \phi(L)Z_t &= \varepsilon_t,\ \phi(L) = 1 - \phi_1 L - \phi_2 L^2 \end{aligned}$$

Assume that $\phi(z) = 0$ has one root equal to unity and the other root real valued with absolute value less than 1. Factor $\phi(L)$ so that

$$\begin{aligned} \phi(L) &= (1 - \phi^* L)(1 - L) = \phi^*(L)(1 - L) \\ \phi^*(L) &= 1 - \phi^* L \text{ with } |\phi^*| < 1. \end{aligned}$$

Then

$$\phi(L)Z_t = (1 - \phi^* L)(1 - L)Z_t = (1 - \phi^* L)\Delta Z_t$$

so that ΔZ_t follows an AR(1) process.

Then

$$\begin{aligned}
\phi(L)y_t &= \phi(L)TD_t + \phi(L)Z_t \Rightarrow \\
(1-\phi^* L)(1-L)y_t &= (1-\phi^* L)(1-L)TD_t + (1-\phi^* L)(1-L)Z_t \Rightarrow \\
(1-\phi^* L)\Delta y_t &= (1-\phi^* L)\Delta TD_t + (1-\phi^* L)\Delta Z_t \Rightarrow \\
(1-\phi^* L)(\Delta y_t - \delta) &= \varepsilon_t
\end{aligned}$$

or

$$\Delta \tilde{y}_t = \phi^* \Delta \tilde{y}_t + \varepsilon_t,\ \Delta \tilde{y}_t = \Delta y_t - \delta$$

Note:

$$\Delta TD_t = \kappa + \delta t - (\kappa + \delta(t-1)) = \delta$$

Difference Stationary Process and Stochastic Trends

Let

$$\begin{aligned} y_t &= TD_t + Z_t \\ TD_t &= \kappa + \delta t \\ \phi^*(L)\Delta Z_t &= \varepsilon_t \Rightarrow \Delta Z_t = \psi^*(L)\varepsilon_t = u_t \end{aligned}$$

Because $Z_t = Z_{t-1} + u_t$, by recursive substitution

$$\begin{aligned} Z_t &= Z_0 + \sum_{k=1}^{t} u_k = Z_0 + TS_t \\ TS_t &= \text{ stochastic trend} \\ TS_t &= TS_{t-1} + u_t, \ TS_0 = 0 \end{aligned}$$

Then

$$y_t = Z_0 + TD_t + TS_t$$

I(1) and I(0) Processes

If the cycle series Z_t is difference stationary then we say that Z_t is *integrated of order 1* and we write $Z_t \sim I(1)$. To see why, note that

$$\begin{aligned} \Delta Z_t &= \psi^*(L)\varepsilon_t = u_t \\ & u_t \text{ is stationary} \end{aligned}$$

It follows that $Z_t = Z_{t-1} + u_t$ and by recursive substitution starting at time $t = 0$ we have

$$Z_t = Z_0 + \sum_{k=1}^{t} u_k$$

so that Z_t can be represented as the (integrated) sum of t stationary innovations $\{u_k\}_{k=1}^{t}$.

Remarks:

1. Since u_t is stationary we say that u_t is *integrated of order zero*, and write $u_t \sim I(0)$, to signify that u_t cannot be written as the sum of stationary innovations.

2. An $I(1)$ process is a random walk process when $u_t \sim iid\ (0, \sigma^2)$

3. If ΔZ_t is an ARMA(p,q) process then Z_t is called an ARIMA(p,1,q) process. The term ARIMA refers to an autoregressive *integrated* moving average process.

Trend Stationary vs. Difference Stationary: Why Do We Care?

- Different views about the nature of trends in economic data
  - TS: trends are smooth predictable deterministic functions of time and do not change quickly. Forecasts of trend are precise.
  - DS: trends are not smooth or predictable functions of time and change often. Forecasts of trend are not precise.
- Detrending a DS process can create spurious cycles
- Differencing a TS process can induce negative autocorrelation

Trend Stationary vs. Difference Stationary: Why Do We Care? (cont'd)

- Regression models with trend stationary variables can be dealt with by including trends in the regression or by detrending the variables

$$
\begin{aligned}
y_t &= \alpha + \delta t + \beta x_t + \varepsilon_t \\
\tilde{y}_t &= \beta_0 + \beta \tilde{x}_t + \varepsilon_t \\
\tilde{y}_t &= y_t - \hat{\kappa} - \hat{\delta} t, \ \tilde{x}_t = x_t - \hat{\gamma} - \hat{\pi} t
\end{aligned}
$$

- Regression models with the levels of difference stationary variables must be handled with extreme care due to the problem of spurious regression
  - Regression in levels only makes sense if variables are *cointegrated*.

Impulse Response Functions from I(1) Processes

Consider an $I(1)$ process with Wold representation $\Delta y_t = \psi^*(L)\varepsilon_t$. Since $\Delta y_t = y_t - y_{t-1}$ the level y_t may be represented as

$$y_t = y_{t-1} + \Delta y_t$$

Similarly, the level at time $t+h$ may be represented as

$$y_{t+h} = y_{t-1} + \Delta y_t + \Delta y_{t+1} + \cdots + \Delta y_{t+h}$$

The impulse response on the level of y_{t+h} of a shock to ε_t is

$$\begin{aligned}\frac{\partial y_{t+h}}{\partial \varepsilon_t} &= \frac{\partial \Delta y_t}{\partial \varepsilon_t} + \frac{\partial \Delta y_{t+1}}{\partial \varepsilon_t} + \cdots + \frac{\partial \Delta y_{t+h}}{\partial \varepsilon_t} \\ &= 1 + \psi_1^* + \cdots + \psi_h^*\end{aligned}$$

The long-run impact of a shock to the level of y_t is given by

$$\lim_{h\to\infty} \frac{\partial y_{t+h}}{\partial \varepsilon_t} = \sum_{j=1}^{\infty} \psi_j^* = \psi^*(1).$$

Hence, $\psi^*(1)$ measures the permanent effect of a shock, ε_t, to the level of y_t.

Note: If $y_t \sim I(0)$ with Wold representation $y_t = \psi(L)\varepsilon_t$ then

$$\lim_{h\to\infty} \frac{\partial y_{t+h}}{\partial \varepsilon_t} = \lim_{h\to\infty} \psi_h = 0$$

Remarks:

1. Since $\frac{\partial y_t}{\partial \varepsilon_t} = 1$ it follows that $\psi^*(1)$ can also be interpreted as the long-run effect of a shock relative to the immediate effect of a shock.

2. If $\psi^*(1) = 0$ then $Z_t \sim I(0)$. To see this suppose $Z_t \sim I(0)$ and has the Wold representation $Z_t = \psi(L)\varepsilon_t$ with $\psi(1) \neq 0$. Then

$$\begin{aligned} \Delta Z_t &= (1-L)Z_t = (1-L)\psi(L)\varepsilon_t = \psi^*(L)\varepsilon_t \\ \psi^*(L) &= (1-L)\psi(L) \end{aligned}$$

It follows that $\psi^*(1) = (1-1)\psi(1) = 0$.

Forecasting from an I(1) Process

Forecasting from an I(1) process follows directly from writing y_{t+h} as

$$y_{t+h} = y_t + \Delta y_{t+1} + \Delta y_{t+2} + \cdots + \Delta y_{t+h}$$

Then

$$\begin{aligned} y_{t+h|t} &= y_t + \Delta y_{t+1|t} + \Delta y_{t+2|t} + \cdots + \Delta y_{t+h|t} \\ &= y_t + \sum_{s=1}^{h} \Delta y_{t+s|t} \end{aligned}$$

Notice that forecasting an I(1) process proceeds from the most recent observation.

Example: Forecasting from an AR(1) model for Δy_t

Let Δy_t follow an AR(1) process

$$\Delta y_t - \mu = \phi(\Delta y_{t-1} - \mu) + \varepsilon_t,\ \ \varepsilon_t \sim WN(0, \sigma^2)$$

where $|\phi| < 1$. Using the chain-rule of forecasting, the h-step ahead forecast of Δy_{t+h}based on information at time t is

$$\Delta y_{t+h|t} = \mu + \phi^h(\Delta y_t - \mu)$$

Then, the h-step ahead forecast of y_{t+h} is

$$\begin{aligned} y_{t+h|t} &= y_t + \sum_{s=1}^{h} [\mu + \phi^s(\Delta y_t - \mu)] \\ &= y_t + h\mu + (\Delta y_t - \mu) \sum_{s=1}^{h} \phi^h \end{aligned}$$

Unit Root Tests

Consider the trend-cycle decomposition of a time series y_t

$$y_t = TD_t + Z_t$$

The basic issue in unit root testing is to determine if Z_t contains a stochastic trend, TS_t. Two classes of tests, both called unit root tests, have been developed to answer this question:

- $H_0 : TS_t \neq 0$ $(y_t \sim I(1))$ vs. $TS_t = 0$ $(y_t \sim I(0))$

- $H_0 : TS_t = 0$ $(y_t \sim I(0))$ vs. $TS_t \neq 0$ $(y_t \sim I(1))$

Autoregressive Unit Root Tests

These tests are based on the following set-up. Let

$$
\begin{aligned}
\phi(L)y_t &= u_t,\ u_t \sim I(0) \\
\phi(L) &= 1 - \phi_1 L - \cdots - \phi_p L^p,\ \phi = \phi_1 + \cdots + \phi_p
\end{aligned}
$$

The null and alternative hypothesis are

$$
\begin{aligned}
H_0 &: \phi = 1\ (\phi(z) = 0 \text{ has a unit root, } TS_t \neq 0) \\
H_1 &: |\phi| < 1\ (\phi(z) = 0 \text{ has all roots outside unit circle, } TS_t = 0)
\end{aligned}
$$

The most popular of these tests are the Dickey-Fuller (ADF) test and the Phillips-Perron (PP) test. The ADF and PP tests differ mainly in how they treat serial correlation in the test regressions.

Moving Average Unit Root Tests (Stationarity Tests)

Consider the first difference of y_t :

$$\Delta y_t = \psi^*(L)\varepsilon_t,\ \varepsilon_t \sim \text{iid}(0, \sigma^2)$$

The null and alternative hypotheses are

$$\begin{aligned} H_0 &: \psi^*(1) = 0 \ (\psi^*(z) = 0 \text{ has a unit root, } TS_t = 0) \\ H_1 &: \psi^*(1) > 0 \ (\psi^*(z) = 0 \text{ has roots outside unit circle, } TS_t \neq 0) \end{aligned}$$

The most popular stationarity tests are the Kitawoski-Phillips-Schmidt-Shin (KPSS) test and the Leyborne-McCabe test. As with the ADF and PP tests the KPSS and Leyborne-McCabe tests differ main in how they treat serial correlation in the test regressions.

Statistical Issues with Unit Root Tests

Conceptually the unit root tests are straightforward. In practice, however, there are a number of difficulties:

- Unit root tests generally have nonstandard and non-normal asymptotic distributions.

- These distributions are functions of standard Brownian motions, and do not have convenient closed form expressions. Consequently, critical values must be calculated using simulation methods.

- The distributions are affected by the inclusion of deterministic terms, e.g. constant, time trend, dummy variables, and so different sets of critical values must be used for test regressions with different deterministic terms.

Distribution Theory for Unit Root Tests

Consider the simple AR(1) model

$$y_t = \phi y_{t-1} + \varepsilon_t, \text{ where } \varepsilon_t \sim \mathsf{WN}(0, \sigma^2)$$

The hypotheses of interest are

$$\begin{aligned} H_0 &: \phi = 1 \text{ (unit root in } \phi(z) = 0) \Rightarrow y_t \sim I(1) \\ H_1 &: |\phi| < 1 \Rightarrow y_t \sim I(0) \end{aligned}$$

The test statistic is

$$t_{\phi=1} = \frac{\hat{\phi} - 1}{\mathsf{SE}(\hat{\phi})}$$
$$\hat{\phi} = \text{least squares estimate}$$

If $\{y_t\}$ is stationary (i.e., $|\phi| < 1$) then

$$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{d} N(0, (1 - \phi^2))$$

$$\hat{\phi} \overset{A}{\sim} N\left(\phi, \frac{1}{T}(1 - \phi^2)\right)$$

$$t_{\phi=\phi_0} \overset{A}{\sim} N(0, 1)$$

However, under the null hypothesis of nonstationarity $\phi = 1$ the above result gives

$$\hat{\phi} \overset{A}{\sim} N(1, 0)$$

$$t_{\phi=1} = \frac{\hat{\phi} - 1}{\mathsf{SE}(\hat{\phi})} \rightarrow \frac{0}{0} = \text{ undefined}$$

which clearly does not make any sense.

Problem: under the unit root null, $\{y_t\}$ is not stationary and ergodic, and the usual sample moments do not converge to fixed constants. Instead, Phillips (1987) showed that the sample moments of $\{y_t\}$ converge to random functions of Brownian motion:

$$T^{-3/2}\sum_{t=1}^{T} y_{t-1} \xrightarrow{d} \sigma \int_0^1 W(r)dr$$
$$T^{-2}\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 W(r)^2 dr$$
$$T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \xrightarrow{d} \sigma^2 \int_0^1 W(r)dW(r)$$

where $W(r)$ denotes a standard Brownian motion (Wiener process) defined on the unit interval.

A Wiener process $W(\cdot)$ is a continuous-time stochastic process, associating each date $r \in [0,1]$ a scalar random variable $W(r)$ that satisfies:

1. $W(0) = 0$

2. For any dates $0 \leq t_1 \leq \cdots \leq t_k \leq 1$ the changes $W(t_2)-W(t_1), W(t_3)-W(t_2), \ldots, W(t_k) - W(t_{k-1})$ are independent normal with

$$W(s) - W(t) \sim N(0, (s-t))$$

3. $W(s)$ is continuous in s.

Intuition: A Wiener process is the scaled continuous-time limit of a random walk

Using the above results Phillips showed that under the unit root null $H_0 : \phi = 1$

$$T(\hat{\phi} - 1) \xrightarrow{d} \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2 dr}$$

$$t_{\phi=1} \xrightarrow{d} \frac{\int_0^1 W(r)dW(r)}{\left(\int_0^1 W(r)^2 dr\right)^{1/2}}$$

For example,

$$
\begin{aligned}
\hat{\phi} - 1 &= \left(\sum_{t=1}^{T} y_{t-1}^2 \right)^{-1} \sum_{t=1}^{T} y_{t-1}\varepsilon_t \\
&\Rightarrow T(\hat{\phi} - 1) = \left(T^{-2} \sum_{t=1}^{T} y_{t-1}^2 \right)^{-1} T^{-1} \sum_{t=1}^{T} y_{t-1}\varepsilon_t \\
&\quad \xrightarrow{d} \left(\int_0^1 W(r)^2 dr \right)^{-1} \int_0^1 W(r) dW(r)
\end{aligned}
$$

Phillips' derivations yield some surprising results:

- $\hat{\phi}$ is *super-consistent*; that is, $\hat{\phi} \xrightarrow{p} \phi$ at rate T instead of the usual rate $T^{1/2}$.

- $\hat{\phi}$ is not asymptotically normally distributed, and $t_{\phi=1}$ is not asymptotically standard normal.

- The limiting distribution of $t_{\phi=1}$ is called the *Dickey-Fuller* (DF) distribution and does not have a closed form representation. Consequently, quantiles of the distribution must be computed by numerical approximation or by simulation.

- Since the *normalized bias* $T(\hat{\phi}-1)$ has a well defined limiting distribution that does not depend on nuisance parameters it can also be used as a test statistic for the null hypothesis $H_0 : \phi = 1$.

- $t_{\phi=1}$ is used much more often than $T(\hat{\phi} - 1)$

Distn/Quantile	**1%**	**5%**	**10%**
Normal	-2.326	-1.645	-1.282
DF	-2.565	-1.941	-1.617

Table 1: Normal and DF Left Tail Quantiles

Remarks:

- The usual one-sided 5% critical value for standard normal is -1.645
- The one-sided 5% critical value for the DF distribution is -1.941
- -1.645 is the 9.45% quantile of the DF distribution

Trend Cases

When testing for unit roots, it is crucial to specify the null and alternative hypotheses appropriately to characterize the trend properties of the data at hand.

- If the observed data does not exhibit an increasing or decreasing trend, then the appropriate null and alternative hypotheses should reflect this.

- The trend properties of the data *under the alternative hypothesis* will determine the form of the test regression used.

- The type of deterministic terms in the test regression will influence the asymptotic distributions of the unit root test statistics.

Case I: Constant Only

The test regression is

$$y_t = c + \phi y_{t-1} + \varepsilon_t$$

and includes a constant to capture the nonzero mean under the alternative. The hypotheses to be tested are

$$\begin{aligned} H_0 &: \phi = 1, c = 0 \ \Rightarrow y_t \sim I(1) \text{ without drift} \\ H_1 &: |\phi| < 1 \Rightarrow y_t \sim I(0) \text{ with nonzero mean} \end{aligned}$$

This formulation is appropriate for non-trending economic and financial series like interest rates, exchange rates, and spreads.

The test statistics $t_{\phi=1}$ and $T(\hat{\phi}-1)$ are computed from the above regression. Under $H_0 : \phi = 1, c = 0$ the asymptotic distributions of these test statistics are influenced by the presence, but not the coefficient value, of the constant in the test regression:

$$
\begin{aligned}
T(\hat{\phi} - 1) &\Rightarrow \frac{\int_0^1 W^\mu(r)dW(r)}{\int_0^1 W^\mu(r)^2 dr} \\
t_{\phi=1} &\Rightarrow \frac{\int_0^1 W^\mu(r)dW(r)}{\left(\int_0^1 W^\mu(r)^2 dr\right)^{1/2}}
\end{aligned}
$$

where

$$
W^\mu(r) = W(r) - \int_0^1 W(r)dr
$$

is a "de-meaned" Wiener process. That is,

$$
\int_0^1 W^\mu(r) = 0
$$

Remarks:

- Derivation requires special trick from Sims, Stock and Watson (1989) ECTA.
- See Hayashi Chapter 9 for details.
- See Hamilton's Time Series textbook for gory details.

Distn/Quantile	**1%**	**5%**	**10%**
Normal	-2.326	-1.645	-1.282
DF	-2.565	-1.941	-1.617
DF$^{\mu}$	-3.430	-2.861	-2.567

Table 2: Normal and DF Left Tail Quantiles

Remarks:

- Inclusion of a constant pushes the distribution of $t_{\phi=1}$ to the left.
- The 5% normal quantile, -1.645, is the 45.94% quantile of the DF$^{\mu}$ distribution!

Case II: Constant and Time Trend

The test regression is

$$y_t = c + \delta t + \phi y_{t-1} + \varepsilon_t$$

and includes a constant and deterministic time trend to capture the deterministic trend under the alternative. The hypotheses to be tested are

$$\begin{aligned} H_0 &: \phi = 1, \delta = 0 \; \Rightarrow y_t \sim I(1) \text{ with drift} \\ H_1 &: |\phi| < 1 \Rightarrow y_t \sim I(0) \text{ with deterministic time trend} \end{aligned}$$

This formulation is appropriate for trending time series like asset prices or the levels of macroeconomic aggregates like real GDP. The test statistics $t_{\phi=1}$ and $T(\hat{\phi} - 1)$ are computed from the above regression.

Under $H_0 : \phi = 1, \delta = 0$ the asymptotic distributions of these test statistics are influenced by the presence but not the coefficient values of the constant and time trend in the test regression.

$$\begin{aligned} T(\hat{\phi} - 1) &\Rightarrow \frac{\int_0^1 W^\tau(r) dW(r)}{\int_0^1 W^\tau(r)^2 dr} \\ t_{\phi=1} &\Rightarrow \frac{\int_0^1 W^\tau(r) dW(r)}{\left(\int_0^1 W^\tau(r)^2 dr\right)^{1/2}} \end{aligned}$$

where

$$W^\tau(r) = W^\mu(r) - 12(r - \frac{1}{2}) \int_0^1 (s - \frac{1}{2}) W(s) ds$$

is a “de-meaned” and “de-trended” Wiener process.

Distn/Quantile	**1%**	**5%**	**10%**
Normal	-2.326	-1.645	-1.282
DF	-2.565	-1.941	-1.617
DF$^{\mu}$	-3.430	-2.861	-2.567
DF$^{\tau}$	-3.958	-3.410	-3.127

Table 3: Normal and DF Left Tail Quantiles

Remarks:

- The inclusion of a constant and trend in the test regression further shifts the distribution of $t_{\phi=1}$ to the left.

- The 5% normal quantile, -1.645, is the 77.52% quantile of the DF$^{\tau}$ distribution!

Dickey-Fuller Unit Root Tests

- The unit root tests described above are valid if the time series y_t is well characterized by an AR(1) with white noise errors.

- Many economic and financial time series have a more complicated dynamic structure than is captured by a simple AR(1) model.

- Said and Dickey (1984) augment the basic autoregressive unit root test to accommodate general ARMA(p, q) models with unknown orders and their test is referred to as the *augmented Dickey-Fuller* (ADF) test

Basic model

$$\begin{aligned} y_t &= TD_t + Z_t = \boldsymbol{\beta}'\mathbf{D}_t + Z_t \\ \phi(L)Z_t &= \theta(L)\varepsilon_t,\ \varepsilon_t \sim \text{WN}(0, \sigma^2) \\ \mathbf{D}_t &= \text{deterministic terms} \\ & Z_t \sim ARMA(p^*, q^*) \end{aligned}$$

Question: Does $\phi(z) = 0$ have a unit root?

Said and Dickey insight: Approximate ARMA(p*,q*) process by AR(p) process for appropriately chosen p

Approximating $ARMA(p^*, q^*)$ by $AR(p)$

$$\begin{aligned}\phi(L)Z_t &= \theta(L)\varepsilon_t,\ \varepsilon_t \sim \mathrm{WN}(0,\sigma^2) \\ \phi(L) &= 1 - \phi_1 L - \cdots - \phi_{p^*}L^{p^*} \\ \theta(L) &= 1 + \theta_1 L + \cdots + \theta_{q^*}L^{q^*}\end{aligned}$$

Assume that roots of $\theta(z) = 0$ lie outside unit circle (MA poly is invertible). Then $\theta(L)^{-1}$ exists and

$$\begin{aligned}\theta(L)^{-1}\phi(L)Z_t &= \theta(L)^{-1}\theta(L)\varepsilon_t = \varepsilon_t \\ &\Rightarrow \tilde{\phi}(L)Z_t = \varepsilon_t \\ \tilde{\phi}(L) &= 1 - \tilde{\phi}_1 L - \tilde{\phi}_2 L^2 - \cdots \\ &\approx 1 - \tilde{\phi}_1 L - \cdots - \tilde{\phi}_p L^p\end{aligned}$$

The ADF test tests the null hypothesis that a time series y_t is $I(1)$ against the alternative that it is $I(0)$, assuming that the dynamics in the data have an ARMA structure. The ADF test is based on estimating the test regression

$$y_t = \boldsymbol{\beta}'\mathbf{D}_t + \phi y_{t-1} + \sum_{j=1}^{p} \psi_j \Delta y_{t-j} + \varepsilon_t$$
$$\mathbf{D}_t = \text{deterministic terms}$$
$$\Delta y_{t-j} \text{ captures serial correlation}$$

The ADF t-statistic and normalized bias statistic are

$$\begin{aligned} \mathsf{ADF}_t &= t_{\phi=1} = \frac{\hat{\phi} - 1}{\mathsf{SE}(\phi)} \\ \mathsf{ADF}_n &= \frac{T(\hat{\phi} - 1)}{1 - \hat{\psi}_1 - \cdots - \hat{\psi}_p} \end{aligned}$$

Result: ADF_t,ADF_n have same asymptotic distributions as $t_{\phi=1}$ and $T(\hat{\phi}-1)$ under white noise errors provided p is selected appropriately.

Intuition: Re-parameterize AR(2) model

$$
\begin{aligned}
y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \\
&= \phi_1 y_{t-1} + \phi_2 y_{t-1} - \phi_2 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \\
&= (\phi_1 + \phi_2) y_{t-1} - \phi_2 \Delta y_{t-1} + \varepsilon_t \\
&= \phi y_{t-1} + \psi \Delta y_{t-1} + \varepsilon_t
\end{aligned}
$$

where

$$
\begin{aligned}
\phi &= (\phi_1 + \phi_2) \\
\psi &= -\phi_2
\end{aligned}
$$

Remarks:

- $y_{t-1} \sim I(1) \Rightarrow \hat{\phi}$ has non-normal distribution

- $\Delta y_{t-1} \sim I(0) \Rightarrow \hat{\psi}$ has normal distribution!

- Derivation requires trick from Sims, Stock and Watson (1989) ECTA

Important results:

- In the AR(2) model with a unit root

  $$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

  the model may be reparameterized such that ϕ_2 is the coefficient on an $I(0)$ variable

  $$y_t = (\phi_1 + \phi_2) y_{t-1} - \phi_2 \Delta y_{t-1} + \varepsilon_t$$

  The Sims, Stock and Watson trick then shows that $\hat{\phi}_2$ has an asymptotic normal distribution.

- The model cannot be reparameterized such that $\phi = \phi_1 + \phi_2$ is the coefficient on an $I(0)$ variable. It is the coefficient on an $I(1)$ variable. Therefore, $\hat{\phi}$ has an asymptotic "unit root" distribution.

Alternative formulation of the ADF test regression:

$$\begin{aligned}\Delta y_t &= \boldsymbol{\beta}'\mathbf{D}_t + \pi y_{t-1} + \sum_{j=1}^{p} \psi_j \Delta y_{t-j} + \varepsilon_t \\ \pi &= \phi - 1\end{aligned}$$

Under the null hypothesis,

$$\Delta y_t \sim I(0) \Rightarrow \pi = 0.$$

The ADF t-statistic and normalized bias statistics are

$$\begin{aligned}\mathrm{ADF}_t &= t_{\pi=0} = \frac{\hat{\pi}}{\mathrm{SE}(\phi)} \\ \mathrm{ADF}_n &= \frac{T\hat{\pi}}{1 - \hat{\psi}_1 - \cdots - \hat{\psi}_p}\end{aligned}$$

and these are equivalent to the previous statistics.

Choosing the Lag Length for the ADF Test

An important practical issue for the implementation of the ADF test is the specification of the lag length p.

- If p is too small then the remaining serial correlation in the errors will bias the test.

- If p is too large then the power of the test will suffer.

- Monte Carlo experiments suggest it is better to error on the side of including too many lags.

Ng and Perron "Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag," JASA, 1995.

- Set an upper bound p_{max} for p.

- Estimate the ADF test regression with $p = p_{\text{max}}$.

- If the absolute value of the t-statistic for testing the significance of the last lagged difference is greater than 1.6 then set $p = p_{\text{max}}$ and perform the unit root test. Otherwise, reduce the lag length by one and repeat the process.

- A common rule of thumb for determining p_{max}, suggested by Schwert (1989), is

$$p_{\mathsf{max}} = \left[12 \cdot \left(\frac{T}{100} \right)^{1/4} \right]$$

where $[x]$ denotes the integer part of x. However, this choice is *ad hoc*!

Ng and Perron "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power," ECTA, 2001.

- Select p as $p_{mic} = \arg\min_{p \le p_{\max}} MAIC(p)$ where

$$\begin{aligned} MAIC(p) &= \ln(\hat{\sigma}_p^2) + \frac{2(\tau_T(p) + p)}{T - p_{\max}} \\ \tau_T(p) &= \frac{\hat{\pi}^2 \sum_{t=p_{\max}+1}^{T} y_{t-1}}{\hat{\sigma}_p^2} \\ \hat{\sigma}_p^2 &= \frac{1}{T - p_{\max}} \sum_{t=p_{\max}+1}^{T} \hat{\varepsilon}_t^2 \end{aligned}$$