Unit Root Tests

Hüseyin Taştan[1]

[1]Yıldız Technical University
Department of Economics

Econometrics II

Lecture Plan

- Revisiting highly persistent time series
- Random Walk and Random Walk with Drift
- Dickey-Fuller (DF) approach to unit root testing
- DF distributions
- Augmented Dickey-Fuller (ADF) test

Highly Persistent Time Series

- So far we learned that we need **weakly dependent** variables for OLS to be consistent and the standard inference procedures to be valid.
- On the other hand, it is highly likely that many economic time series **cannot** be characterized by weak dependence, but **strong dependence**.
- **Highly persistent or strongly dependent** time series display high correlation with its past values.
- What happens if we run a regression involving highly persistent variables?

Highly Persistent Time Series

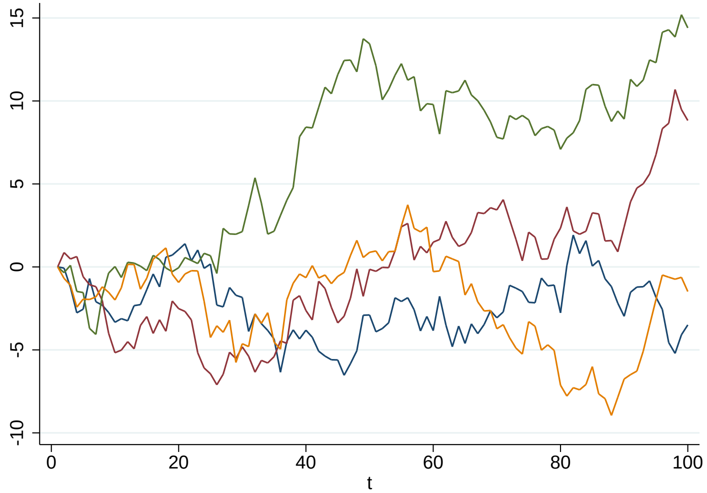
- We also learned that that many economic and financial time series are better characterized by the AR(1) model with $\rho = 1$.

$$y_t = \alpha + \rho y_{t-1} + e_t, \quad t = 1, 2, \ldots$$

  As ρ approaches 1, the more the time series gets persistent.
- When $\rho = 1$, AR(1) process is called **Random Walk**.
- A random walk is a special case of what is known as a unit root process. The name comes from the fact that $\rho = 1$ in the AR(1) model.
- If a time series follows a random walk process then the only way to make it stationary is to take the first difference.

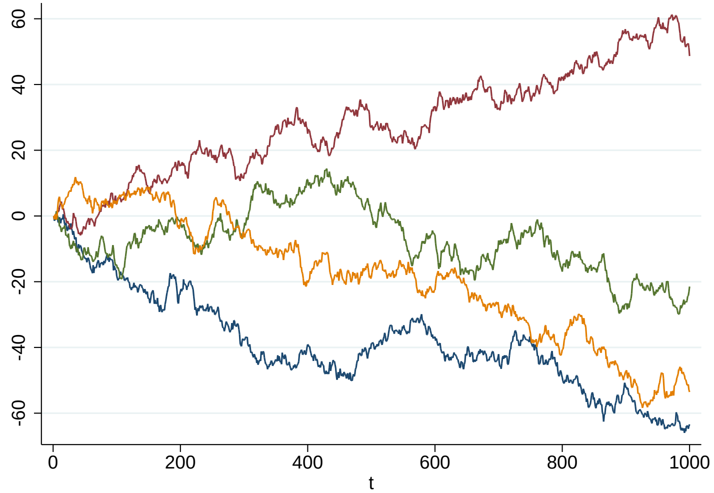
Random Walk

Four realizations of random walk process with $y_0 = 0$, $T = 100$



Random Walk

Four realizations of random walk process with $y_0 = 0$, $T = 1000$



Random Walk

- From a policy perspective, it is often important to know whether an economic time series is highly persistent or not. Consider the case of gross domestic product in Turkey. If GDP is asymptotically uncorrelated, then the level of GDP in the coming year is at best weakly related to what GDP was, say, thirty years ago. This means a policy that affected GDP long ago has very little lasting impact.
- GDP is strongly dependent, then next year's GDP can be highly correlated with the GDP from many years ago. Then, we should recognize that a policy which causes a discrete change in GDP can have persisting and long-lasting effects.
- Effect of shocks are very persistent and lasting in random walk process.

Random Walk with Drift

- It is often the case that a highly persistent series also contains a clear trend.
- One model that leads to this behavior is called the **random walk with drift**.

$$y_t = \alpha + y_{t-1} + e_t, t = 1, 2, ...$$

- where $\{e_t : t = 1, 2, ...\}$ and y_0 satisfy the same properties as in the random walk model. What is new is the parameter α, which is called the **drift** term. Essentially, to generate y_t, the constant α is added along with the random noise et to the previous value y_{t-1}
- We can show that the expected value of y_t follows a linear time trend by using repeated substitution

$$y_t = \alpha t + e_t + e_{t-1} + ... + e_1 + y_0$$

Random Walk with Drift

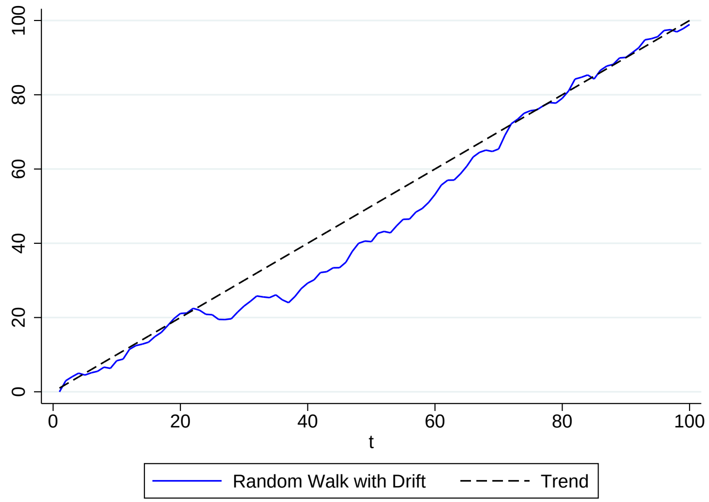
- Therefore, if $y_0 = 0$, $E(y_t) = \alpha t$: the expected value of y_t is growing over time if $\alpha > 0$ and shrinking over time if $\alpha < 0$.
- By reasoning as we did in the pure random walk case, we can show that

$$E(y_{t+h}|y_t) = \alpha h + y_t$$

- So the best prediction of y_{t+h} at time t is y_t plus the drift αh. The variance of y_t is the same as it was in the pure random walk case.

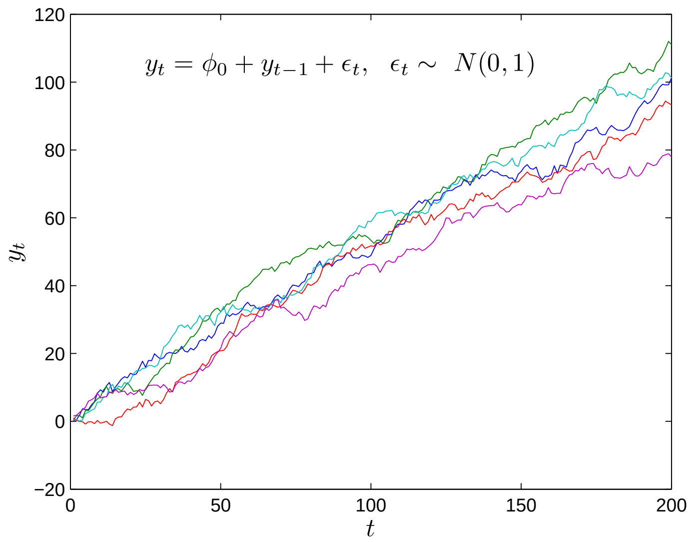
Random Walk with Drift

A single realization of random walk with drift process with $y_0 = 0$, $T = 100$



Random Walk with Drift

Simulated Realizations of Random Walk with Drift, $\alpha = 0.5$



Difference-stationary Process (DSP)

- If a nonstationary process can be made stationary after taking differences then it is called a **difference-stationary process**.
- A nonstationary process with a stationary first difference is called **integrated of order one** and denoted $I(1)$
- Similarly, if the second difference is stationary then it is called integrated of order two, $I(2)$.
- Clearly, RW and RWD process are $I(1)$. Their first differences are $I(0)$, stationary.

$$\Delta y_t = \alpha + e_t \sim \; I(0)$$

$e_t \sim iid(0, \sigma^2)$

Trend-stationary Process (TSP)

- A TSP fluctuates around a deterministic trend and it has a tendency to revert back to that trend over time.
- A simple example is

$$x_t = \beta_0 + \beta_1 t + \epsilon_t, \quad \epsilon_t \sim \ wn(0, \sigma^2)$$

- With expected value and variance

$$\mu_t = \mathsf{E}(x_t) = \beta_0 + \beta_1 t$$

$$\gamma_{0t} = \mathsf{Var}(\epsilon_t) = \sigma^2$$

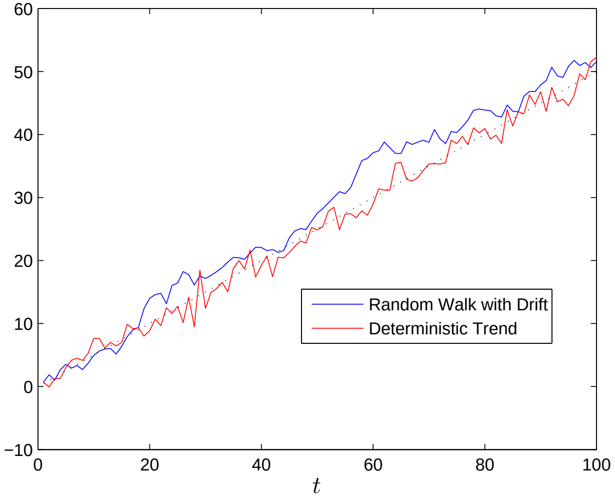
Trend-stationary Process (TSP)

- Its population mean is time-dependent, thus, nonstationary. But when we remove (or detrend) it we obtain a stationary process:

$$x_t - \beta_1 t = \beta_0 + \epsilon_t \sim \ I(0)$$

- The following graph displays 100 realizations of a DSP and TSP
- Notice that both series display positive trend. But the trend in DSP (which is random walk with drift) is **stochastic**
- Both DSP and TSP have similar sample autocorrelations. Hence, inspecting correlogram (ACF) alone may not be sufficient to decide whether a series is DSP or TSP. We need to use specialized tests for that purpose.

DSP vs. TSP



Trend-stationary Process

- Correct transformation of a nonstationary time series is essential
- Detrending a DSP or differencing a TSP may lead to wrong inference
- For instance detrending a random walk with drift we obtain:

$$y_t - \alpha t = \sum_{j=1}^{t} e_j \sim \quad I(1)$$

which is not stationary.
- Similarly, taking the first difference of a deterministic linear trend process results in a non-invertible MA(1) process:

$$\Delta x_t = x_t - x_{t-1} = \beta_1 + \epsilon_t - \epsilon_{t-1}$$

Regression Analysis with Nonstationary Data

- Classical regression analysis using time series data requires all variables to be I(0) and weakly dependent for the asymptotic properties of OLS estimators to be valid (consistency and asymptotic normality)
- Using I(1) variables in a regression model may result in spurious results. Adding a linear trend to the model, or detrending variables may not work if they are I(1)
- If all variables are I(1) then one may check whether they share a long-run trend, in other words, if they are cointegrated.
- If they are all I(1) but there exists a unique linear combination is I(0) then the appropriate model for the short-run behavior is an error correction model (ECM).
- If they are not cointegrated then the regression model needs to be re-specified using the first differences

Unit root tests

- In practice it may be important to decide whether a time series is DSP or TSP
- Unit root tests may be used for that purpose
- Some of widely used tests: ADF test, PP test, DF-GLS test, ERS Point-Optimal test, NP test, KPSS test
- Existence of large structural breaks may result in high persistence. There are many unit root tests that can handle structural breaks: Perron test, Zivot-Andrews test, Flexible Fourier ADF tests
- In this lecture, we will only consider Dickey-Fuller (DF) tests and its extension (ADF).

Deciding Whether a Time Series Is I(1) or I(0)

- How do we know if a time series contains a unit root, i.e., I(1) process or integrated of order 1?
- The natural starting point is to consider the following AR(1) model

$$y_t = \alpha + \rho y_{t-1} + e_t, \quad t = 1, 2, \dots$$

  given an observed initial value y_0.
- The error term is assumed to follow

$$\mathrm{E}[e_t | y_{t-1}, y_{t-2}, \dots, y_0] = 0$$

  The error term e_t is assumed to be iid with zero mean and independent of the past information.

Unit Roots

$$y_t = \alpha + \rho y_{t-1} + e_t, \quad t = 1, 2, \dots$$

- The process $\{y_t\}$ will have a **unit root** if and only if $\rho = 1$
- If $\alpha = 0$ and $\rho = 1$ we have a random walk process
- If $\alpha \neq 0$ and $\rho = 1$ we have a random walk with drift process
- In both cases e_t must follow the process mentioned in the previous slide

21

Unit Root Null Hypothesis

$$y_t = \alpha + \rho y_{t-1} + e_t, \quad t = 1, 2, \dots$$

- It is common to leave α unspecified and focus on the AR(1) parameter. The null hypothesis of unit root can be stated as

$$H_0 : \rho = 1$$

- Against the one-sided alternative

$$H_1 : \rho < 1$$

- Note that the other alternative $H_1 : \rho > 1$ is not considered in practice because in that case the variable will be explosive - very unrealistic for economic time series.
- Also it is almost always makes sense to expect a positive AR(1) parameter, $0 < \rho < 1$. So a negative unit root is not considered.

22

Unit Root Null Hypothesis

$$y_t = \alpha + \rho y_{t-1} + e_t, \quad t = 1, 2, \dots$$

- When $\rho < 1$ then we have a stable AR(1) process for which

$$Corr(y_t, y_{t-h}) = \rho^h \to 0$$

as $h \to 0$

23

Dickey-Fuller Test

- Dickey-Fuller unit root tests (Dickey-Fuller (1979) and Dickey-Fuller (1981)) use the standard regression approach and compute the t statistic.
- Subtracting y_{t-1} from both sides of $y_t = \alpha + \rho y_{t-1} + e_t$ and rearranging we get

$$\Delta y_t = \alpha + \theta y_{t-1} + e_t, \tag{1}$$

where $\Delta y_t = y_t - y_{t-1}$ and $\theta = \rho - 1$. The null and alternative hypotheses are

$$H_0 : \theta = 0 \quad (\rho = 1),$$

$$H_1 : \theta < 0 \quad (\rho < 1).$$

- Notice that the alternative hypothesis is still one-sided, we only consider the left tail.

24

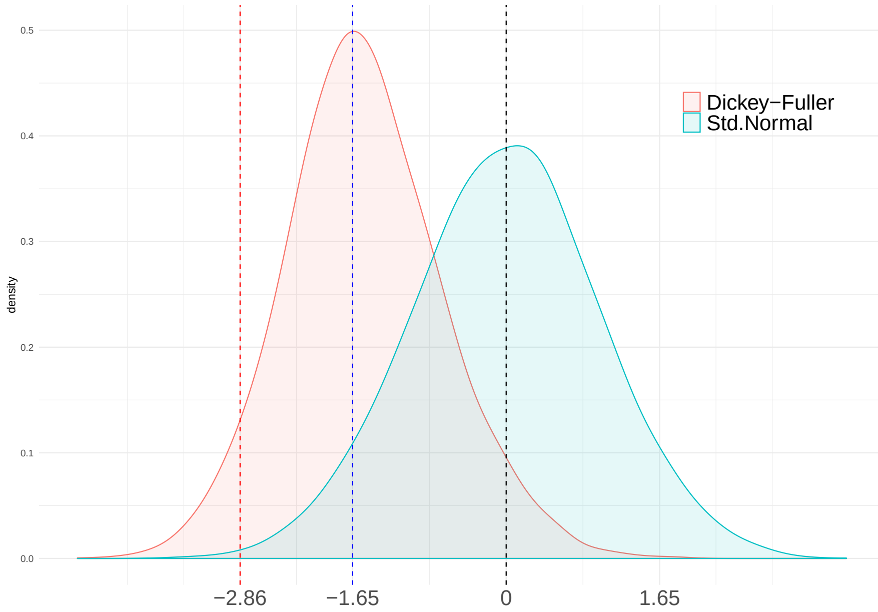
Dickey-Fuller Test

- The DF test statistic is just the t-ratio:

$$DF = \frac{\hat{\theta}}{s.e.(\hat{\theta})}.$$

- DF does not follow the usual t distribution under the null.
- The reason is that, under the null hypothesis the usual Central Limit Theorem (CLT) does not apply.
- So DF does not follow approximate normal distribution even in large sample sizes. Its distribution is **nonstandard**.
- Dickey and Fuller (1979) has computed the asymptotic distribution and critical values
- The asymptotic distribution of DF test statistic does not have a closed-form expression. Critical values need to be approximated using simulations.

Dickey-Fuller vs. Standard Normal Distribution



Asymptotic distribution of DF test statistic

- Previous graph shows an approximation of DF distribution
- DF distribution is skewed to left as compared to standard normal distribution. This implies that its critical values are smaller than the standard normal critical values.
- Critical values for the case of constant but no trend:

Critical Values of DF Test: Constant only, No Trend

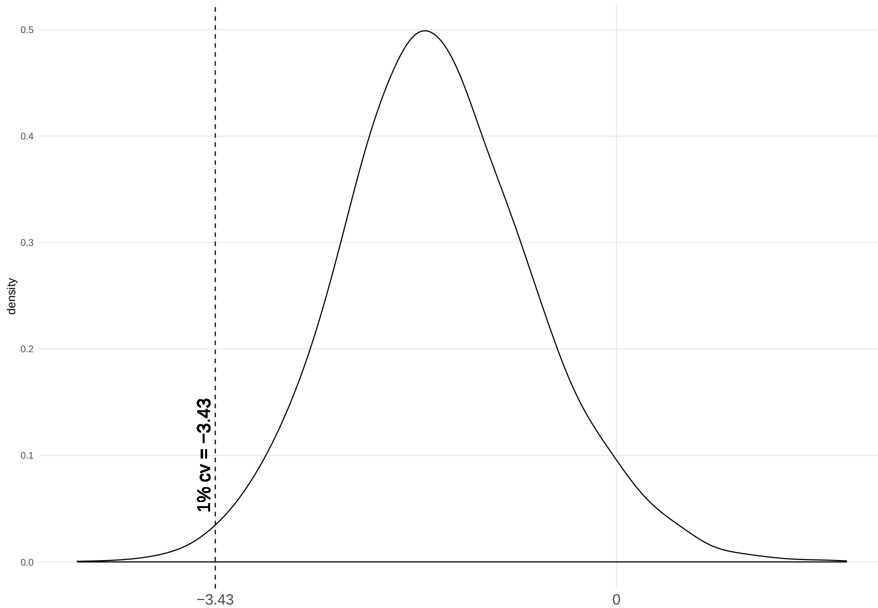
%1	%5	%10
-3.43	-2.86	-2.57

Critical Values of Standard Normal Distribution

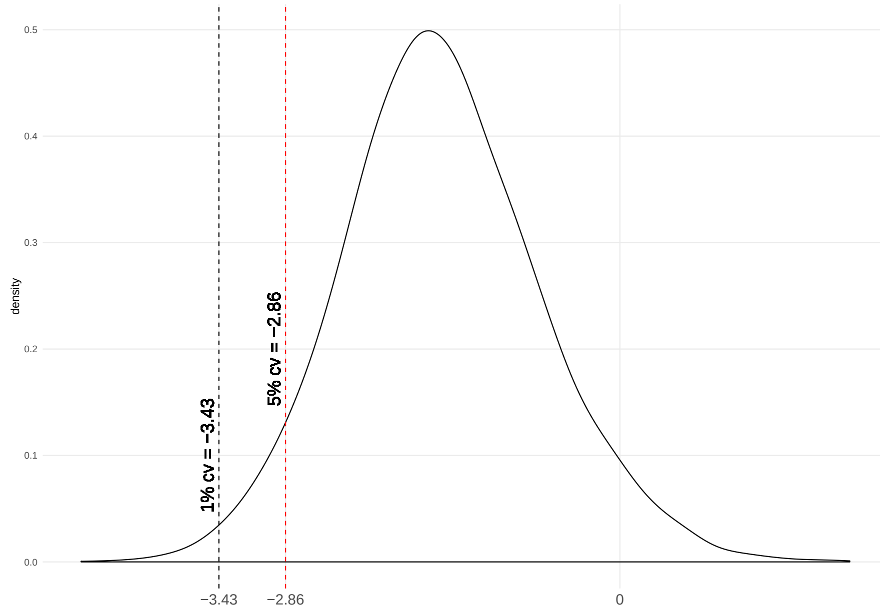
%1	%5	%10
-2.33	-1.645	-1.28

Note that the critical values of DF statistic is larger in absolute value.

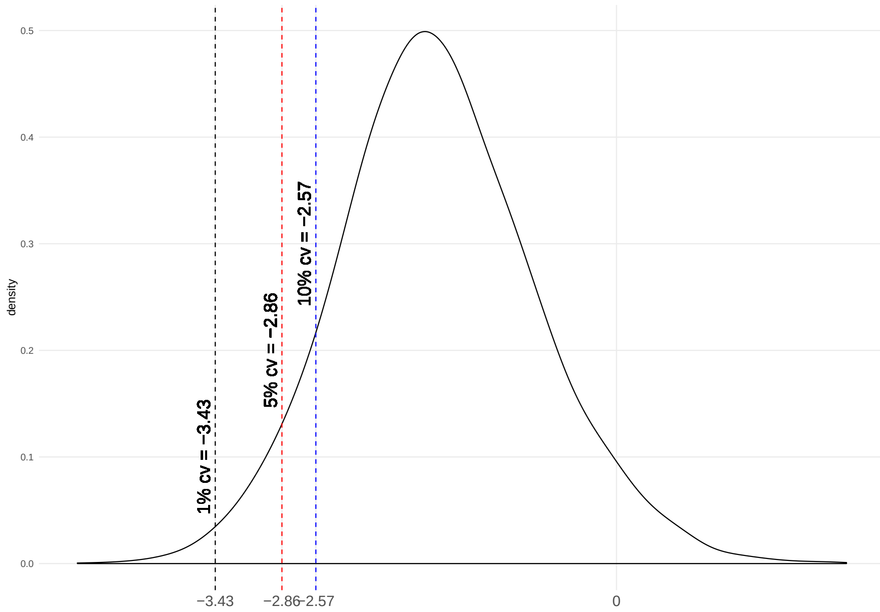
Dickey-Fuller 1% Critical Value (No Trend)



Dickey-Fuller 5% Critical Value (No Trend)



Dickey-Fuller Critical Values (No Trend)



Dickey-Fuller Test

- DF distribution depends on the deterministic components, constant term, trend, etc.
- In practice the following specifications may be used:

$$\text{no constant + no trend}: \Delta y_t = \theta y_{t-1} + e_t$$

$$\text{constant + no trend}: \Delta y_t = \alpha + \theta y_{t-1} + e_t$$

$$\text{constant + trend}: \Delta y_t = \alpha + \beta t + \theta y_{t-1} + e_t$$

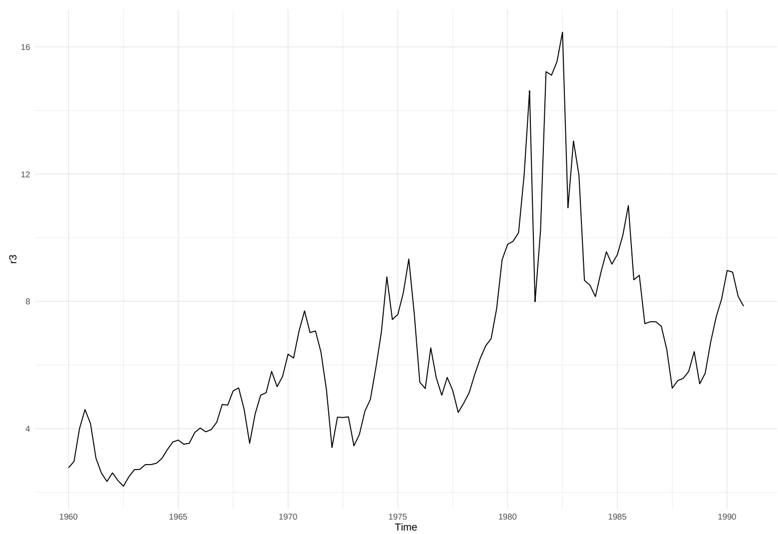
  where $\theta = \rho - 1$ in all cases.
- Also, notice that we assume $e_t \sim \; iid(0, \sigma^2)$. If the residuals are serially correlated then the tests are invalid.

Dickey-Fuller Test

- Although the theory behind the DF unit root test is highly complicated, its practical usage is very easy.
- Because the test is conducted on the left tail, the critical values are negative.
- If the computed DF statistic is less than the appropriate critical value at predetermined significance level, then the null of unit root is rejected.
- Asymptotic distribution, hence the critical values, depend on the deterministic terms (constant, trend, etc.) included in the AR(1) regression.
- In practice, we generally consider two cases: **constant only** (no time trend), and **constant + trend**.

Example: US interest rates (data=intqrt)

3-month treasury bill rates



Example: interest rates

Run the DF test regression without the trend:

$$\widehat{\Delta r3_t} = 0.625 - 0.091\ r3_{t-1}$$
$$(0.261)\quad(.037)$$
$$n = 123, \quad R^2 = 0.048$$

From $\hat{\theta} = \hat{\rho} - 1$ we see that $\hat{\rho} = 0.909$ which is less than 1. But is it statistically different from 1? Test statistic is simply the t-ratio:

$$DF = \frac{\hat{\theta}}{se(\hat{\theta})} = \frac{-0.091}{0.037} = -2.46$$

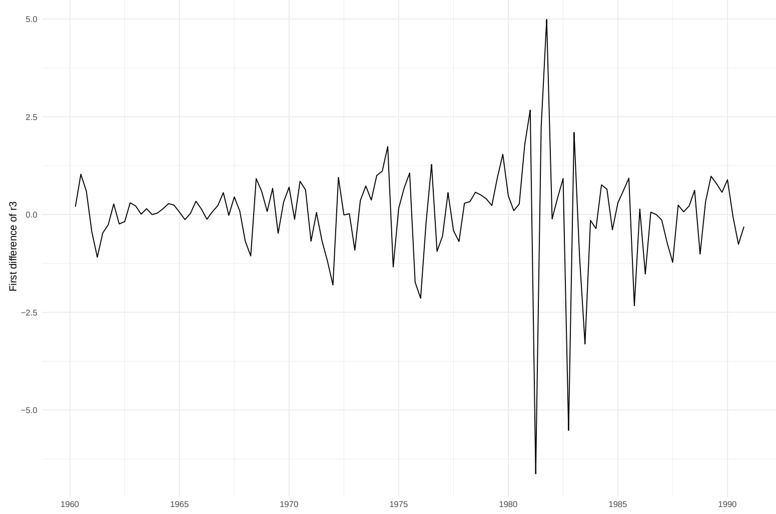
From the critical value table below, we see that $c = -2.57$ at 10% level. Because $DF > c$, we **fail to reject** H_0 at 10% level. Thus, $r3$ may be characterized as a nonsta-
tionary process. If its first difference is stationary then $r3$ will be I(1).

TABLE 18.2 Asymptotic Critical Values for Unit Root *t* Test: No Time Trend

Significance level	1%	2.5%	5%	10%
Critical value	−3.43	−3.12	−2.86	−2.57

First difference of interest rates

Is the first difference, $dr3_t = r3_t - r3_{t-1}$, stationary?



Is the first difference stationary?

Let $dr3_t$ be the first difference of $r3_t$, i.e., $dr3_t = r3_t - r3_{t-1}$. Running the DF test regression:

$$\Delta dr3_t = \theta dr3_{t-1} + e_t,$$

we obtain

$$\widehat{\Delta dr3_t} = 0.045 - 1.116\ dr3_{t-1}$$
$$(0.114)\quad(.091)$$
$$n = 122, \quad R^2 = 0.56$$

Unit root test statistic:

$$DF = \frac{\hat{\theta}}{se(\hat{\theta})} = \frac{-1.116}{0.091} = -12.3$$

Because $DF < -3.43$ we reject the null hypothesis at 1% level. The first difference of interest rate, $dr3_t$, is stationary. Thus, 3-month T-bill rate, $r3_t$, is an I(1) process.

Constant + Trend

- If a time series has a clear trend, the deterministic components in the DF test regression need to be modified
- In this case the appropriate specification is the one that includes a trend

  $$y_t = \alpha + \beta t + \rho y_{t-1} + e_t, \quad t = 1, 2, \ldots$$

  or, equivalently,

  $$\Delta y_t = \alpha + \beta t + \theta y_{t-1} + e_t,$$

  where $\theta = \rho - 1$.
- The null and alternative hypotheses are

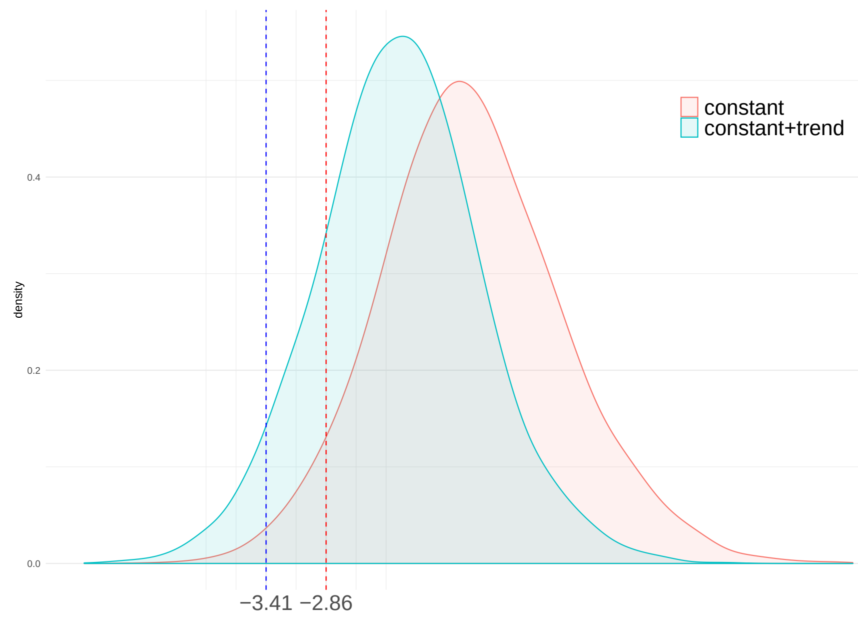
  $$H_0 : \theta = 0 \quad (\rho = 1), \quad \text{Nonstationary}$$
  $$H_1 : \theta < 0 \quad (\rho < 1), \quad \text{Trend-stationary}$$
- Note that under the alternative y_t follows a trend-stationary process.

Constant + Trend

$$\Delta y_t = \alpha + \beta t + \theta y_{t-1} + e_t,$$

- The DF test statistic is the same as before. It's just the t-statistic on the lagged dependent variable.
- The distribution of DF statistic depends on the deterministic specification as we mentioned before.
- The critical values are different. Because the distribution shifts to the left, they are smaller than the case without the trend (see the next graph).
- Decision rule is the same as before. If the DF test statistic is less than the critical value, we reject H_0 and conclude that the process may be characterized as a trend-stationary process.
- If the DF is larger than the critical value then we do not reject H_0 suggesting that the process is nonstationary. In that case, the proper transformation to achieve stationarity will be differencing.

DF 5% CV (Constant + Trend)



Constant + Trend

- When we include a trend in the DF test regression, it becomes harder to reject the null hypothesis. What's the intuition behind this behavior?
- Remember that when we include a time trend in a regression, it's like detrending the series. So when we detrend a unit root process, it tends to look more like an I(0) process
- Thus, a larger critical value (in absolute value) is required to reject the null of nonstationarity.
- As an example, to reject H_0, we need a DF test statistic (which is just a t-ratio on $\hat{\theta}$) smaller than -3.41 as compared with -2.86 without a time trend.

Example: interest rates revisited

Re-run the DF test regression with both constant and trend:

$$\begin{aligned} \widehat{\Delta r3}_t = \ & 0.53 + \ 0.03\ t - \ 0.15\ r3_{t-1} \\ & (0.264) \ \ (.0165) \ \ (0.048) \\ n = & 123, \quad R^2 = 0.058 \end{aligned}$$

Test statistic is simply the t-ratio:

$$DF = \frac{\hat{\theta}}{se(\hat{\theta})} = \frac{-0.148561}{0.0484} = -3.07$$

From the critical value table below, we see that $c = -3.41$ at 5% level. Because $DF > c$, we **fail to reject** H_0 at 5% level. Thus, $r3$ may be characterized as a nonstationary process.

TABLE 18.3 Asymptotic Critical Values for Unit Root *t* Test: Linear Time Trend

Significance level	1%	2.5%	5%	10%
Critical value	−3.96	−3.66	−3.41	−3.12

Serial Correlation in DF test regression

- In both constant and constant+trend cases, one of the important assumptions underlying the DF test was the error term e_t must be serially uncorrelated. The error term is assumed to follow a white noise process with mean 0 and a constant, nonzero, finite variance.
- For many economic and financial time series, the assumption is very likely to fail in practice.
- Note that if there is serial correlation then the DF test results will be invalid. Thus the inference will be misleading.
- The solution is to correct for the serial correlation by adding appropriate number of lags of the dependent variable (say, p lags).
- That is, we just need to augment the test regression by adding $\Delta y_{t-1}, \Delta y_{t-2}, \dots, \Delta y_{t-p}$ as explanatory variables.

Augmented Dickey-Fuller (ADF) Test

- Adding p lags of the dependent variable and then computing the DF test statistic as usual is known as the Augmented Dickey-Fuller (ADF) test.
- For example, suppose that when we add the first lag the residuals are serially uncorrelated in the following test regression:

$$\Delta y_t = \alpha + \beta t + \theta y_{t-1} + \gamma_1 \Delta y_{t-1} + e_t,$$

- ADF test statistic is defined as the same as DF:

$$ADF = \frac{\hat{\theta}}{se(\hat{\theta})}$$

- What's the distribution of the ADF test statistic?
- It is exactly the same as the DF test; adding lagged terms does not alter the critical values. So the decision rule is the same as before (but pay attention to the deterministic part!).

Augmented Dickey-Fuller (ADF) Test

Consider the following test regression:

$$\Delta y_t = \alpha + \theta y_{t-1} + \gamma_1 \Delta y_{t-1} + e_t,$$

- When the null hypothesis is true $H_0 : \theta = 0$ and $|\gamma_1| < 1$ then the first difference follows a stable AR(1) process, i.e., $\Delta y_t \sim AR(1)$
- It can be shown that under the alternative hypothesis $H_1 : \theta < 0$, y_t follows a stable AR(2) process.
- More generally, to compute the ADF test statistic we first run the regression of

$$\Delta y_t \text{ on } y_{t-1},\ \Delta y_{t-1},\ \Delta y_{t-2},\ \dots,\ \Delta y_{t-p}$$

- And then compute the t-statistic on the coefficient of y_{t-1}

Augmented Dickey-Fuller (ADF) Test

$$\Delta y_t = \alpha + \theta y_{t-1} + \sum_{j=1}^{p} \gamma_j \Delta y_{t-j} + e_t,$$

- Adding p lags of the dependent variable Δy_t is intended to obtain residuals that are serially uncorrelated.
- In practice, adding appropriate number of lags is important.
- If too few lags are included the test results will be misleading because the the critical values will not be valid. In this case, we may incorrectly decide that the series is stationary.
- If too many lags are included the variance of the test regression will increase and the power of the test suffers.

Augmented Dickey-Fuller (ADF) Test

$$\Delta y_t = \alpha + \theta y_{t-1} + \sum_{j=1}^{p} \gamma_j \Delta y_{t-j} + e_t,$$

- There are several methods to choose p. For example we may use the Akaike Information Criterion (AIC) to select p. Over a pre-selected number of maximum lags, the one with the minimum AIC is preferred.
- Another method is to use sequential t tests on the lagged terms. Starting with the maximum number of lags, we sequentially apply t-test and include the highest lag that is significant at, say, 10% level.
- The frequency of the time series is also important in selecting the number of lags.
- For annual data, usually 1 or 2 lags suffice.
- For monthly data we may try 12 lags.
- Similarly, for quarterly data considering 4 lags is a good idea.

Example: Is inflation nonstationary?

According to AIC, $p = 1$. The ADF test regression is

$$\widehat{\Delta inf}_t = \underset{(0.517)}{1.36} - \underset{(.103)}{0.31}\, inf_{t-1} + \underset{(0.126)}{0.138} \Delta inf_{t-1}$$

$$n = 47, \quad R^2 = 0.172$$

Test statistic is simply the t-ratio on inf_{t-1}:

$$ADF = \frac{\hat{\theta}}{se(\hat{\theta})} = \frac{-0.31}{0.103} - 3.01$$

From the critical value table (see previous slides), we see that $c = -2.86$ at 5% level. Because $ADF < c$, we **reject** H_0 at 5% level. Thus, there is a strong evidence against a unit root process for inflation.
Question: What is the estimate of ρ?

Is Turkish Real GDP Nonstationary?

Is Turkish Real GDP Nonstationary?

To answer this question we compute the ADF test. We take natural log of the real GDP. We use AIC to select the lag order. Optimal lag order is chosen as 1. Also, because there is a clear positive trend in the series, we add a trend term in the test regression:

$$\Delta \widehat{\log(GDP)}_t = \underset{(1.279)}{4.18} - \underset{(0.05)}{0.162}\, \log(GDP)_{t-1} + \underset{(0.0012)}{0.008t} + \underset{(0.104)}{0.20} \Delta \log(GDP)_{t-1}$$

$$n = 85, \quad R^2 = 0.14$$

The ADF test statistic is $ADF = -3.24$. Critical value at 5% is -3.41. Because ADF is above the critical value, we fail to reject the null hypothesis.
Thus, Real GDP series may be characterized as a nonstationary process. This implies that shocks to GDP are permanent.

Is the Growth Rate of Turkish Real GDP Nonstationary?

$ADF = -5.95$. What's your decision?