

# Lecture 6a: Unit Root and ARIMA Models

## Big Picture

- A time series is non-stationary if it contains a unit root

unit root  $\Rightarrow$  nonstationary

The reverse is not true.

- Many results of traditional statistical theory do not apply to unit root process, such as law of large number and central limit theory.
- We will learn a formal test for the unit root
- For unit root process, we need to apply ARIMA model; that is, we take difference (maybe several times) before applying the ARMA model.

## Review: Deterministic Difference Equation

- Consider the first order equation (without stochastic shock)

$$y_t = \phi_0 + \phi_1 y_{t-1}$$

- We can use the method of iteration to show that when

$$\phi_1 = 1$$

the series is

$$y_t = \phi_0 t + y_0$$

- So there is no steady state; the series will be trending if  $\phi_0 \neq 0$ ; and the initial value has permanent effect.

## Unit Root Process

- Consider the AR(1) process

$$y_t = \phi_0 + \phi_1 y_{t-1} + u_t$$

where  $u_t$  may and may not be white noise. We assume  $u_t$  is a zero-mean stationary ARMA process.

- This process has unit root if

$$\boxed{\phi_1 = 1}$$

In that case the series converges to

$$y_t = \phi_0 t + y_0 + (u_1 + u_2 + \dots + u_t) \tag{1}$$

## Remarks

- The  $\phi_0 t$  term implies that the series will be trending if  $\phi_0 \neq 0$ .
- The series is not mean-reverting. Actually, the mean changes over time (assuming  $y_0 = 0$ ):

$$E(y_t) = \phi_0 t$$

- The series has non-constant variance

$$\text{var}(y_t) = \text{var}(u_1 + u_2 + \dots + u_t),$$

which is a function of  $t$ .

- In short, the unit root process is not stationary.

## Random Walk

- Random walk (RW) is a special case of unit root process, which assumes

$$\phi_1 = 1, \phi_0 = 0$$

- RW is not trending because  $\phi_0 = 0$
- However, RW is still nonstationary because time-varying variance:  $\text{var}(y_t) = \text{var}(u_1 + u_2 + \dots + u_t)$
- If  $\phi_0 \neq 0$  we can call the series random walk with drift. The drift term  $\phi_0$  causes trending behavior.

## What causes nonstationarity?

- The RW is

$$y_t = y_{t-1} + u_t,$$

which implies

$$y_t = y_0 + (u_1 + u_2 + \dots + u_t)$$

- Basically the nonstationarity is caused by the process of summing (integrating).
- Notice that each shock  $u$  has coefficient of 1. So the impulse response never decays
- This also explains why a unit root process is highly persistent.

## High Persistence

- A unit root series is highly persistent (non-ergodic) in the sense that the autocorrelation decays to zero very slowly.
- The ACF function of a unit root series decreases to zero linearly and slowly.
- So slow-decaying ACF is signal for nonstationarity (trend is another signal).



## Why call it unit root?

We can use the lag operator to rewrite RW as

$$\begin{aligned} y_t &= y_{t-1} + u_t \\ \Rightarrow y_t &= Ly_t + u_t \\ \Rightarrow (1 - L)y_t &= u_t \end{aligned}$$

The equation

$$1 - L = 0$$

has the root

$$L = 1,$$

which is called unit root.

## Multiple Unit Roots

- A time series possibly has multiple unit roots.
- For example, consider the series with two unit roots

$$(1 - L)(1 - L)y_t = u_t$$

$$\Rightarrow (1 - 2L + L^2)y_t = u_t$$

$$\Rightarrow y_t = 2Ly_t - L^2y_t + u_t$$

$$\Rightarrow y_t = 2y_{t-1} - y_{t-2} + u_t$$

So the AR(2) process

$$y_t = 2y_{t-1} - y_{t-2} + u_t$$

is nonstationary and has two unit roots.

## Example

```
y.lag1 = c(NA, y[1:(T-1)])           # first lag
d.y = y - y.lag1                     # difference
data = data.frame(y,y.lag1,d.y)      # put data together
data                                  # display
```

|   | y           | y.lag1      | d.y           |
|---|-------------|-------------|---------------|
| 1 | 0.00000000  | NA          | NA            |
| 2 | -0.45402484 | 0.00000000  | -0.4540248368 |
| 3 | 0.89000211  | -0.45402484 | 1.3440269472  |
| 4 | -0.02792883 | 0.89000211  | -0.9179309410 |

Can we apply the ARMA model to unit root process?

The answer is No, unless we transform the series into stationary series.

## Stationary Transformation

- Define the difference operator  $\Delta$  as

$$\Delta \equiv 1 - L$$

so that

$$\Delta y_t = (1 - L)y_t = y_t - Ly_t = y_t - y_{t-1}$$

- Then the equation

$$(1 - L)(1 - L)y_t = u_t$$

suggests that we can transform the nonstationary  $y_t$  into the stationary  $u_t$  after taking difference two times

$$u_t = \Delta\Delta y_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$$

## Integrated Process and ARIMA model

- By definition, a series is integrated of order  $d$  if we need to take difference  $d$  times before it becomes stationary.
- We use letter “I” to represent “integrated”. So ARIMA( $p, d, q$ ) model means we take difference  $d$  times and then apply the ARMA( $p, q$ ) model to the differenced series.

## Why is unit root troublesome?

- For one thing, the law of large number (LLN) does not hold for a unit root process.
- For a stationary and ergodic process LLN states that as  $T \rightarrow \infty$

$$\frac{1}{T} \sum_{t=1}^T y_t \rightarrow E(y_t)$$

- Unit root may cause three troubles. First,  $E(y_t)$  may not be a constant. Second, the variance of  $y_t$  is non-constant. Third, the serial correlation between  $y_t$  and  $y_{t-j}$  decays to zero very slowly.

**Warning:** you need to be very careful when running a regression using a unit root process as regressors. Many standard results do not apply.



## Unit Root Test

- Consider the simplest case when we fit the AR(1) regression

$$y_t = \phi_1 y_{t-1} + e_t \quad (2)$$

where  $e_t$  is white noise.

- We want to test the null hypothesis that  $y_t$  has a unit root, i.e.,

$$H_0 : \phi_1 = 1$$

- We can always transform (2) into (after subtracting  $y_{t-1}$  on both sides)

$$\Delta y_t = \beta y_{t-1} + e_t \quad (3)$$

where  $\beta \equiv \phi_1 - 1$ . Now the new null hypothesis is

$$H_0 : \beta = 0$$

## Dickey-Fuller (DF) Unit Root Test

- DF test is the most popular test for unit root. It is nothing but the t test for  $H_0 : \beta = 0$  based on the transformed equation (3)
- The alternative hypothesis is

$$H_0 : \beta < 0$$

- Note this is an one-tailed test. The  $<$  sign indicates that the rejection region is on the left. That means we need to reject the null hypothesis when the t test is less than the critical value. We don't compute the absolute value of the t test.

## Why is DF test so famous

- Because under the null hypothesis the t test does NOT follow t distribution or normal distribution. Instead, the t test follows the bizarre distribution of

$$\text{DF Test} \rightarrow \frac{\int W dW}{(\int W^2)^{1/2}}$$

where  $W$  stands for Brownian motion, and the numerator is a stochastic integral!

- Intuitively, we are in this weird situation because the regressor  $y_{t-1}$  is nonstationary under the null hypothesis.

## More weird stuff

- To make it worse, the distribution of the DF test will change when we add an intercept, and change again when we add a trend
- Table A of the textbook reports the critical values for the distribution of DF test under different situations
- In practice we always include an intercept in the testing regression.

## Augmented Dickey-Fuller (ADF) Unit Root Test

- So far we assume the error is white noise in (2)
- What if the error is serially correlated? In that case, we need to add enough number of lagged differences as additional regressors until the error become white noise.
- The so called augmented Dickey Fuller test is the t test for the null hypothesis

$$H_0 : \beta = 0$$

based on the regression

$$\Delta y_t = c_0 + c_1 t + \beta y_{t-1} + \sum_{i=1}^p \gamma_i \Delta y_{t-i} + e_t \quad (4)$$

## Remarks

- The series has no unit root and is stationary if the null hypothesis is rejected.
- The trend  $c_1 t$  can be dropped if the series is not trending
- In practice we can use AIC to choose the number of lags  $p$ .
- More rigorously, you need to apply Ljung–Box test to ensure the error is serially uncorrelated.
- The ADF test follows the same distribution as the DF test.

Remember: you reject the null hypothesis of unit root (at 5% level) when the ADF (or DF) test is less than -2.86 if the regression has no trend, and when the test is less than -3.41 if the regression has a trend

## Why is a trend necessary

- A series can be trending for different reasons.
- The unit root series  $y_t = \phi_0 + y_{t-1} + u_t$  is trending because the drift term  $\phi_0$ .

- By contrast the series

$$y_t = ct + u_t$$

is trending, but it does not have unit root.

- We include a trend in the ADF test in order to distinguish those two.



## Difference Stationary vs Trend Stationary

- Also there are two ways to make a trending series stationary, depending on the cause of trend.
- For the unit root series  $y_t = \phi_0 + y_{t-1} + u_t$  we need to take difference:

$$\phi_0 + u_t = y_t - y_{t-1}$$

So it is called difference-stationary series.

- For series  $y_t = ct + u_t$  we need to subtract trend from  $y_t$  :

$$u_t = y_t - ct.$$

The process is called detrending.  $y_t$  is called trend-stationary series.

## Remarks

- Because  $u_t$  is stationary, a trend-stationary series is “trend-reverting”. That is, there is a tendency that the series will go back to its trend after deviating from it.
- By contrast, a difference-stationary series is not trend-reverting. There is tendency that once deviating from the trend, the series will move further and further away from that trend. This is caused by the  $(u_1 + u_2 + \dots + u_t)$  in (1)
- In practice, if the ADF test rejects the null for a trending series, then it is trend-stationary. Otherwise, it is difference-stationary.

## ARIMA Model

1. Suppose the series is not trending
  - (a) If the ADF test (without trend) rejects, then apply ARMA model directly
  - (b) If the ADF test (without trend) does not reject, then apply ARMA model after taking difference (maybe several times)
2. Suppose the series is trending
  - (a) If the ADF test (with trend) rejects, then apply ARMA model after detrending the series
  - (b) If the ADF test (with trend) does not reject, then apply ARMA model after taking difference (maybe several times)

## Lecture 6b: Simulating the distribution of DF test

## Why do simulation?

- You need to go to a Ph.D program to really understand that bizarre distribution of the DF test
- However, we can use simulation to visualize that distribution.

## How does simulation work

1. We need to generate a series that satisfies the null hypothesis. In this case, we need to generate a random walk that has one unit root

$$y_t = y_{t-1} + e_t, e_t \sim \text{i.i.d.n}(0, 1), y_1 = 0, t = 2, \dots, T$$

2. We then run the testing regression

$$\Delta y_t = c_0 + \beta y_{t-1} + e_t$$

and we keep the t statistics for  $H_0 : \beta = 0$

$$\text{DF test} = t_{H_0:\beta=0}$$

3. We repeat Steps 1 and 2 many many times. Then we get many many DF statistics. The distribution of them is what we want.

## R Code: Step 1

```
T = 1000  
e = rnorm(T)  
y = rep(0, T)  
for (t in 2:T) y[t] = y[t-1]+e[t]
```

1. We let the sample size be 1000,  $T = 1000$ . This is big sample.
2. We generate 1000 *i.i.d.n* random shocks  $e$ . The R function `rnorm` returns pseudo random number that seems like the normal random variable.
3. Then we use loop to generate  $y$  recursively.

## R Code: Step 2

```
y.lag1 = c(NA, y[1:(T-1)])  
d.y = y - y.lag1  
co <- coef(ms <- summary(model<-lm(d.y~y.lag1)))  
df[j] = co[2, "t value"]
```

1. We generate the first lag  $y_{t-1}$  and the difference  $\Delta y_t$ . Note the first observation of  $y_{t-1}$  is missing value NA.
2. Then we run the testing regression using command `lm`
3. The t value is saved in a vector called `df`



## R Code: Step 3

```
skewness(df)
kurtosis(df)
jarque.test(df)
dfs = sort(df)
dfs[0.05*N]
hist(df)
```

1. We obtain the skewness and kurtosis of the DF distribution.  
Those commands are in “moments” package
2. We apply Jarque–Bera test to DF distribution
3. We find the 5-th quantile (critical value) after sorting df vector
4. We draw the histogram of df

## Summary of DF Distribution

1. It is skewed because the skewness is nonzero
2. It has fatter tail than normal distribution because the kurtosis is greater than 3
3. Jarque–Bera test rejects the normality for the DF distribution
4. The histogram of df is not centered around zero
5. The 5-th quantile of the DF distribution is -2.88. By contrast, the 5-th quantile of normal distribution is -1.64. So it would be too easy to reject the null hypothesis if you incorrectly use the critical values of normal distribution.

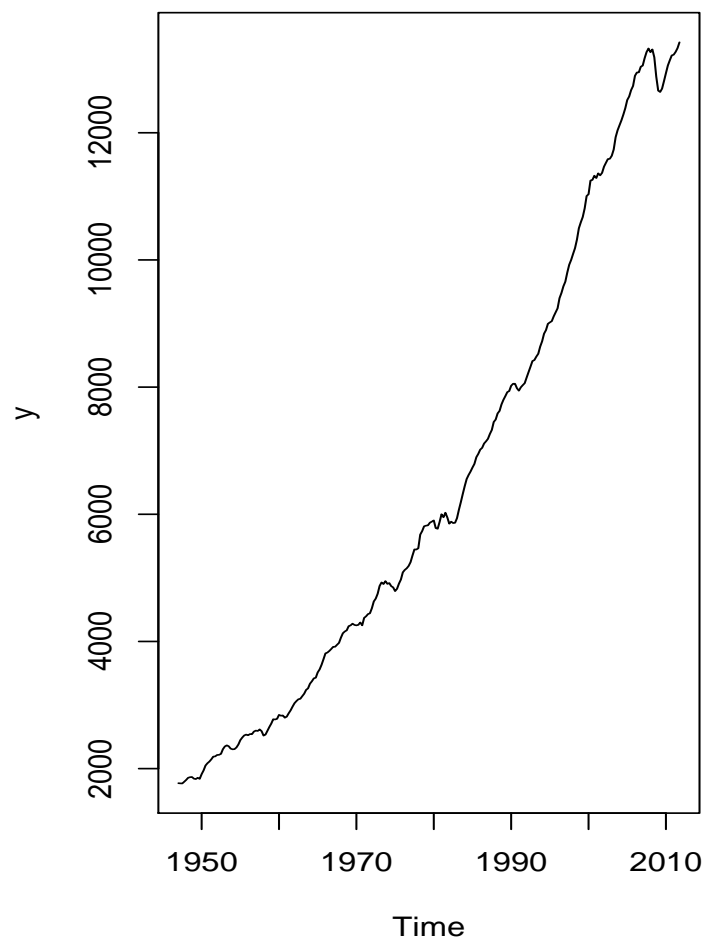
## Lecture 6c: Empirical Example

## US Quarterly Real GDP

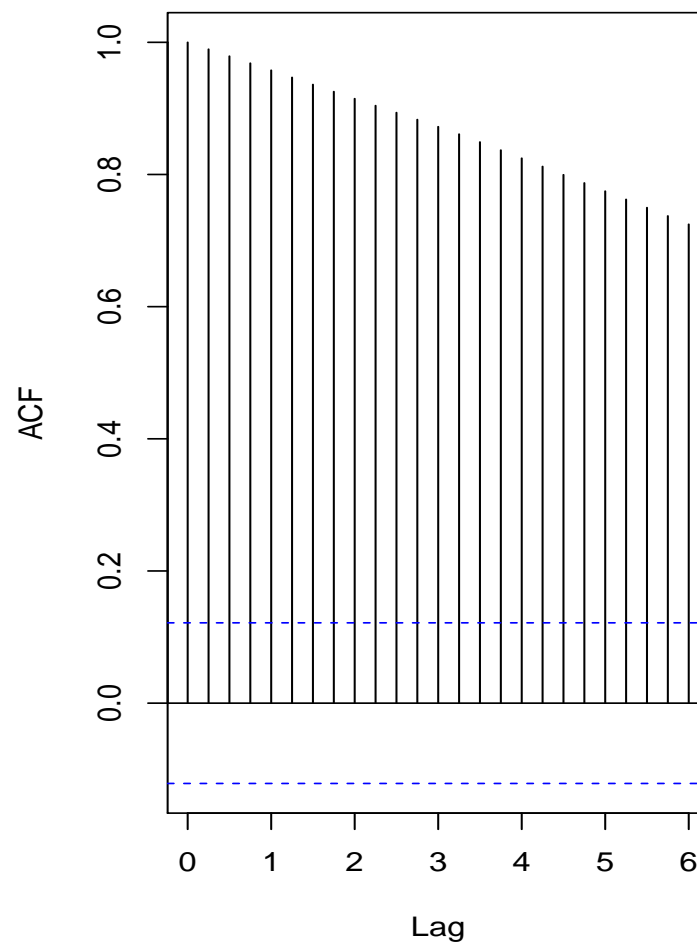
- We obtain the quarterly US real GDP from FRED data
- There are two visual signals that the GDP series is nonstationary
  1. The GDP is trending
  2. The ACF of GDP decays to zero slowly

# GDP: Time Series Plot and ACF

US Real GDP



ACF



## Unit Root Test for GDP

- The command for ADF test is `adf.test`, which is in “tseries” package.
- Because GDP is trending, the testing regression should include the trend. So we choose `alternative = c(“explosive”)` option.
- We try including one, four and eight lagged differences in the regression, given that we have quarterly data.

## Result when one lagged difference is included

```
adf.test(y, alternative = c("explosive"), k = 1)
```

```
Augmented Dickey-Fuller Test
```

```
data: y
```

```
Dickey-Fuller = -1.6428, Lag order = 1, p-value = 0.2737
```

```
alternative hypothesis: explosive
```

1. The ADF test is -1.6428, greater than the critical value -3.41. So we cannot reject the null hypothesis that there is one unit root in GDP. Put differently, the GDP is difference stationary, not trend stationary.
2.  $p\text{-value} = 0.2737 > 0.05$  leads to the same conclusion.
3. The conclusion does not change when we change lag order

## A more transparent way

Instead of using the black-box like command `adf.test`, we can run the testing regression explicitly as

```
T = length(y)
y.lag1 = c(NA, y[1:(T-1)])
d.y = y - y.lag1
dy.lag1 = c(NA, d.y[1:(T-1)])
tre = 1:T
summary(lm(d.y~y.lag1+dy.lag1+tre))
```

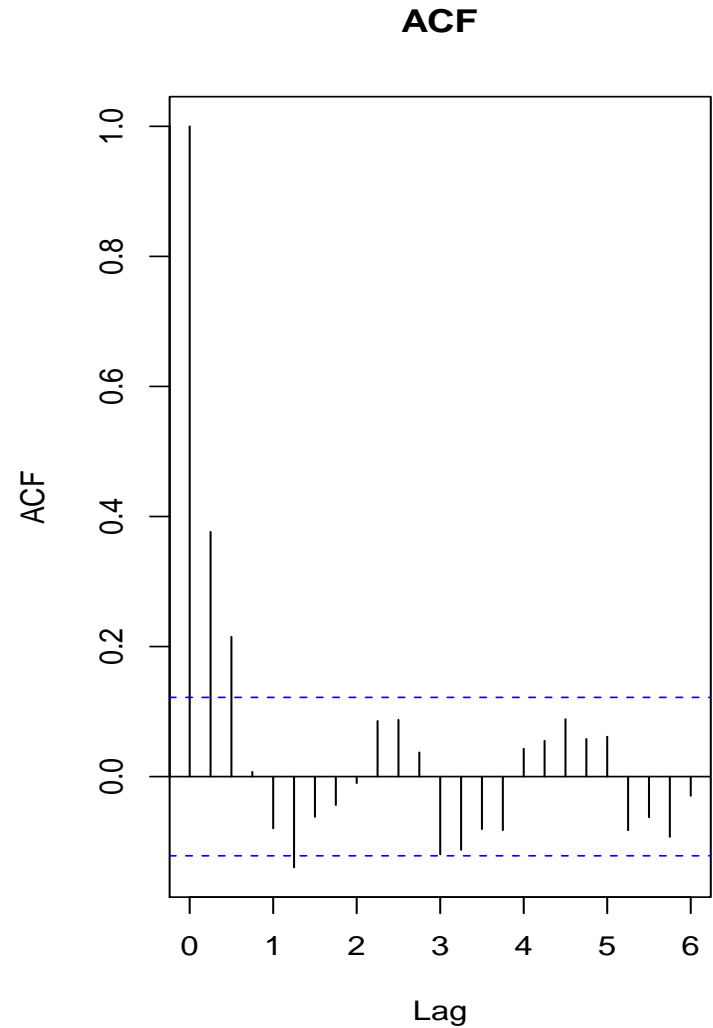
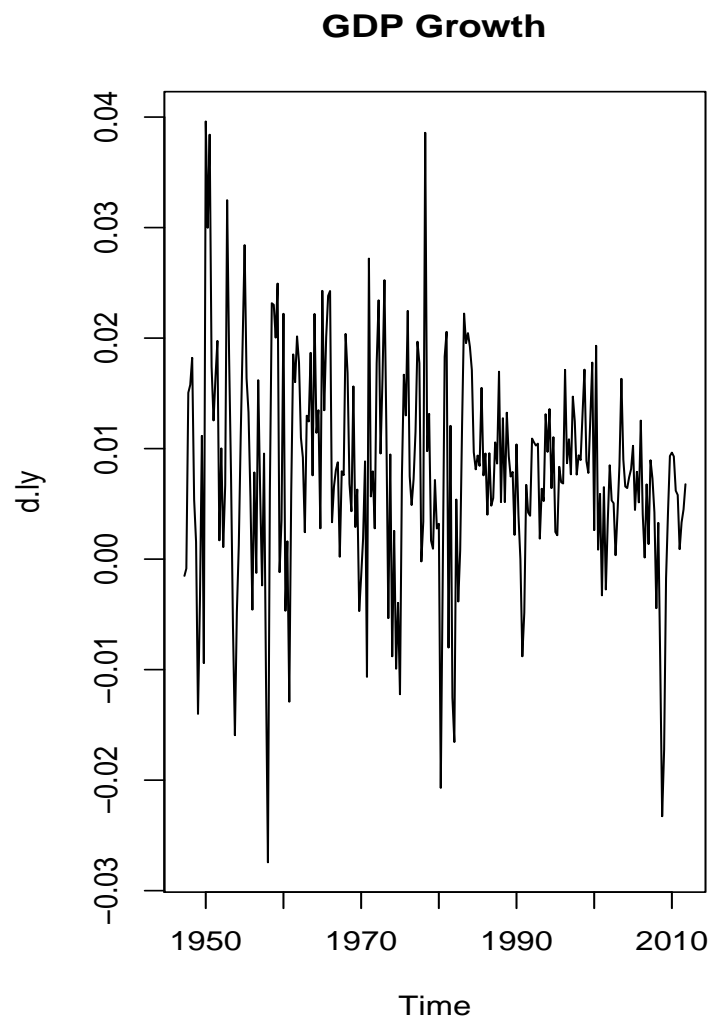
The p-value reported by `lm` command is wrong because it is based on the normal rather than DF distribution.



## GDP Growth Rate

- Then we may wonder if taking difference once can achieve stationarity.
- We get GDP quarterly growth rate after taking difference of log GDP
- Now we see two signals for stationarity
  1. The growth series is mean-reverting
  2. The ACF of growth decays to zero very quickly

# GDP Growth Rate: Time Series Plot and ACF



## GDP Growth Rate: Unit Root Test

We then apply the ADF test to the growth

```
> d.ly = diff(log(y))  
> adf.test(d.ly, alternative = c("stationary"), k = 1)  
      Augmented Dickey-Fuller Test
```

```
data:  d.ly  
Dickey-Fuller = -8.4206, Lag order = 1, p-value = 0.01  
alternative hypothesis: stationary
```

## GDP Growth Rate: Unit Root Test

1. Because the growth series is not trending we choose alternative = c(“stationary”) option
2. The ADF test is -8.4206, less than the critical value -2.86. So we can reject the null hypothesis that there is one unit root in GDP growth. In other words, GDP growth has no unit root, so is stationary
3. p-value = 0.01 < 0.05 leads to the same conclusion.
4. The conclusion does not change when we change lag order
5. Thus GDP is integrated of order one: taking difference once is enough to reach stationarity

## ARMA model applied to GDP Growth

1. We then apply ARMA model to GDP growth since it is stationary
2. We try AR(1), AR(2), AR(3), AR(4), and ARMA(1,1) models.  
The AIC picks the AR(3) model

```
arima(x = d.ly, order = c(3, 0, 0))
```

Coefficients:

|       | ar1      | ar2    | ar3     | intercept |
|-------|----------|--------|---------|-----------|
|       | 0.3545   | 0.1268 | -0.1177 | 0.0078    |
| s.e.  | 0.0616   | 0.0650 | 0.0617  | 0.0009    |
| aic = | -1689.15 |        |         |           |

## Final Model

The final model for GDP is ARIMA(3,1,0)

$$u_t = 0.3545u_{t-1} + 0.1268u_{t-2} - 0.1177u_{t-3} + e_t \quad (5)$$

$$u_t = \Delta(\log(\text{GDP}_t)) - 0.0078 \quad (6)$$

Note that

$$E(\Delta(\log(\text{GDP}_t))) = 0.0078$$

1. So on average the real GDP grows 0.78% each quarter, or  $4 * 0.78\% = 3.12\%$  each year.
2. The real GDP is a random walk with drift. The drift term is 0.0078, and is significant.
3. This drift term causes the trend in the GDP

## Forecasting

In order to use the model to forecast GDP we need to

1. first forecast  $u_{T+1}, u_{T+2}, \dots$
2. Then get  $\Delta \log(\text{GDP}_{T+1}) = u_{T+1} + 0.0078$
3. Next  $\log(\text{GDP}_{T+1}) = \log(\text{GDP}_T) + \Delta \log(\text{GDP}_{T+1})$
4. Finally,  $\text{GDP}_{T+1} = e^{\log(\text{GDP}_{T+1})}$

## Is GDP trend-stationary?

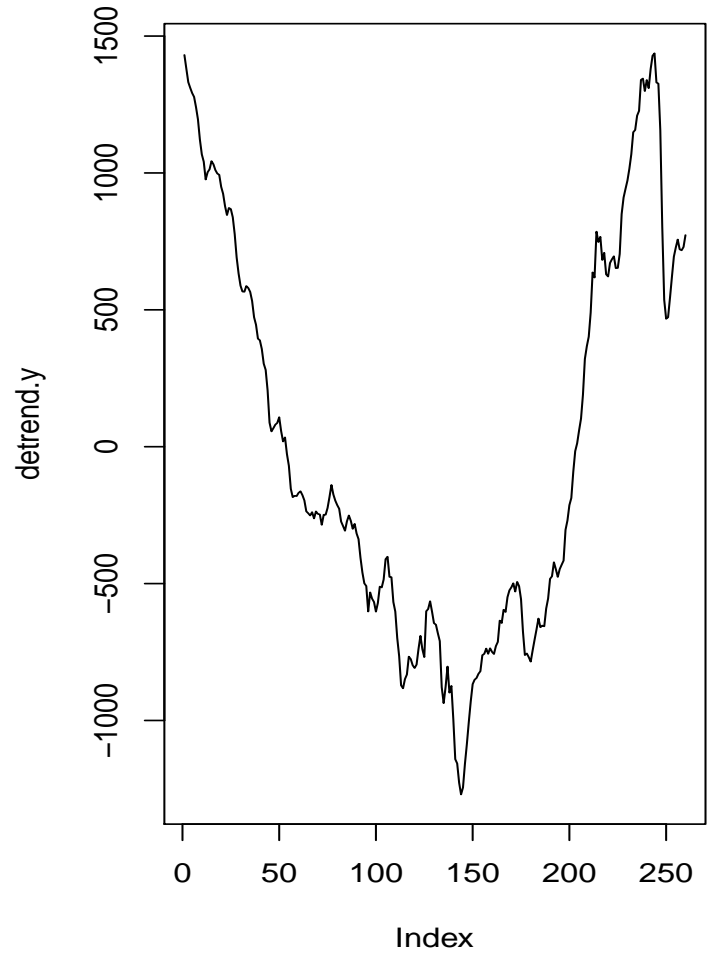
- The answer is no according to the ADF test applied to GDP
- We can get the same answer by applying the ADF test to the detrended GDP, which is the residual of regressing GDP onto a linear trend. The R code to get the detrended GDP is

```
trend = 1:length(y)
model = lm(y~trend)
detrend.y = model$res
```

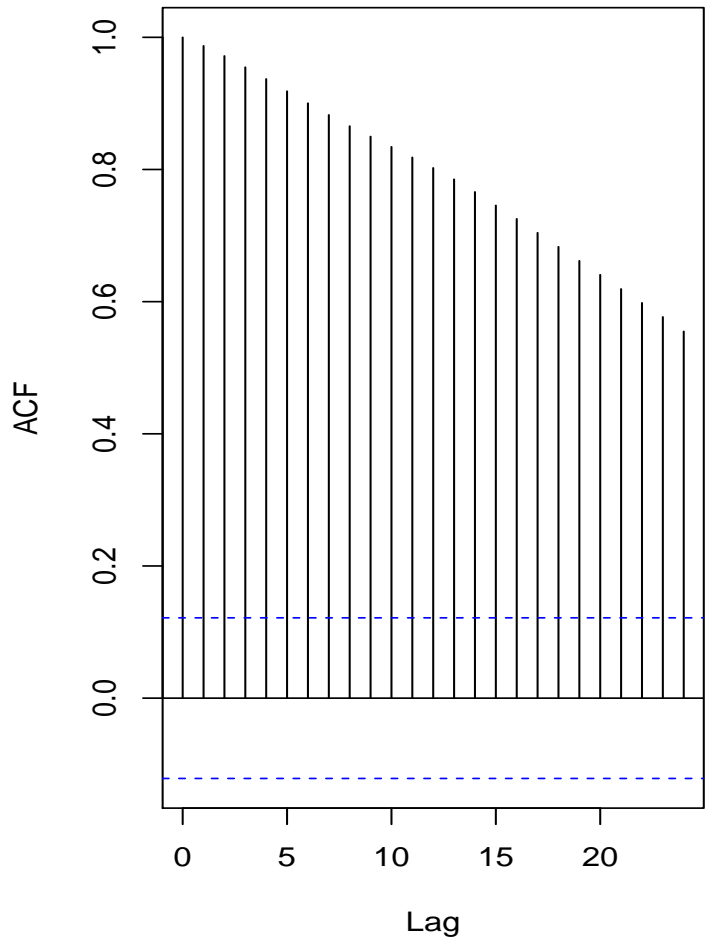


# Detrended GDP: Time Series Plot and ACF

**Detrended GDP**



**ACF**



## Remarks

We see two signals for nonstationarity for detrended GDP

1. The detrended series is very smooth (not choppy) with long swing
2. The ACF is slow-decaying

## ADF test applied to detrended GDP

```
adf.test(detrend.y, alternative = c("stationary"), k = 1)
```

Augmented Dickey-Fuller Test

```
data: detrend.y
```

```
Dickey-Fuller = -1.6428, Lag order = 1, p-value = 0.7263
```

```
alternative hypothesis: stationary
```

The p-value is greater than 0.05, not rejecting the null of unit root.

So the detrended series is not stationary. Therefore GDP is difference stationary, not trend stationary.

## Does it matter?

The issue of trend-stationary vs difference-stationary is important for GDP

1. The policy (shock) would have transitory effect (pushing the series away from its trend temporarily) if GDP was trend-stationary.
2. We just proved GDP is difference-stationary, so any policy shock will have permanent effect (pushing the series away from its trend permanently)

## Economic Theory Justification

- The fact that GDP is difference stationary can be justified by macro theory
- For example, the random walk hypothesis of consumption (Robert Hall, 1978, JPE) implies that the consumption has unit root. So GDP has unit root as well (and is difference stationary) since consumption is its biggest component.