Financial Econometrics Econ 40357
Spurious Regression Problem
Testing for Unit Roots
Brooks pp. 334-351

N.C. Mark

University of Notre Dame and NBER

September 6, 2020

Review: Random Walk with Drift

$$y_t = \mu + y_{t-1} + \sigma\epsilon_t \tag{1}$$
$$E_t\left(y_{t+1} - y_t\right) = \mu \tag{2}$$
$$E_t y_{t+k} = k\mu + y_t \tag{3}$$
$$E_t(y_{t+k} - y_t = k\mu \tag{4}$$

(1) is statement of the model
(2) is the one-step ahead forecasted *change*
(3) is the $k-$step ahead forecasted value
(4) is the $k-$step ahead forecasted change.
Nested model: Driftless random walk ($\mu = 0$).

Random Walk with Drift

Repeated backward substitution gives stochastic trend representation

$$
\begin{aligned}
y_t &= \mu + \underbrace{y_{t-1}}_{y_{t-1}=\mu+y_{t-2}+\epsilon_{t-1}} + \epsilon_t \\
&= \mu + \mu + \underbrace{y_{t-2}}_{y_{t-2}=\mu+y_{t-3}+\epsilon_{t-2}} + \epsilon_t + \epsilon_{t-1} \\
&= \mu + \mu + \mu + y_{t-3} + \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} \\
&\vdots
\end{aligned}
$$

$$
y_t = y_0 + t\mu + (\epsilon_t + \epsilon_{t-1} + \cdots + \epsilon_1)
$$

- y_0 looks like constant in regression with trend. But y_0 is not a constant. It is the realization of a random variable. Hence, the term *stochastic trend.*
- It is wrong to try to regress the random walk on a time trend to induce stationarity.
- Original Efficient Markets hypothesis based on this idea. Today's price "reflects all publically available information."

The Spurious Regression Problem

- The spurious regression problem illustrates why it is important to make sure your data are stationary.
- Two independent, driftless random walk processes, $\{y_t\}$ and $\{x_t\}$

$$
\begin{aligned}
y_t &= y_{t-1} + \sigma_y \epsilon_{t,y} \\
x_t &= x_{t-1} + \sigma_x \epsilon_{t,x}
\end{aligned}
$$

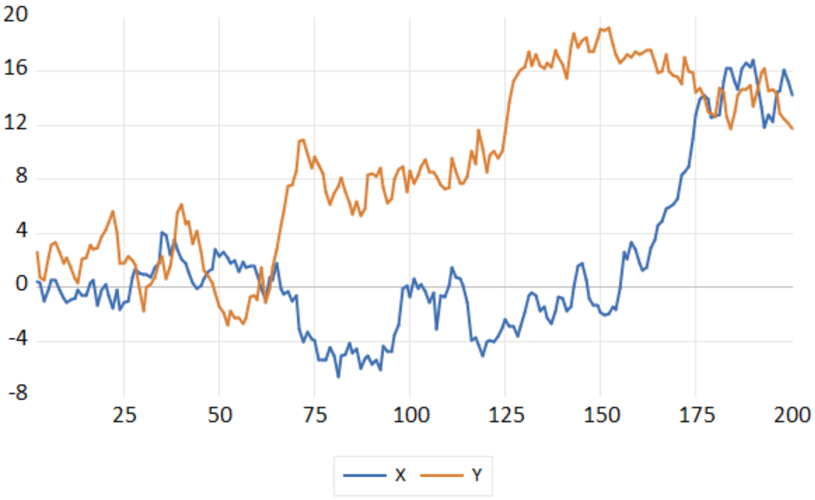
  where $\epsilon_{t,y}$ and $\epsilon_{t,x}$ each are i.i.d. Importantly, note that they are independent of each other. This means y_t is independent of x_t. (e.g., y_t is price of Tesla stock. x_t is the number of ants in a particular ant hill in the Xinjiang province of China.
- Regress y_t on x_t,

$$
y_t = \beta_0 + \beta_1 x_t + v_t
$$

  Use the usual (standard) t-ratio to test if $\beta_1 = 0$.
  The spurious regression problem is, the test will always reject the null that $\beta_1 = 0$.

Illustrate

```
' Generate two independent random walks
series e1 = nrnd
series e2 = nrnd
smpl @first @first
series sto1 = 0
series sto2 = 0
smpl @first+1 @last
series sto1 = sto1(-1) + e1
series sto2 = sto2(-1) + e2
series y = sto1
series x = sto2
equation eq01.ls y c x
```


20
16
12
8
4
0
-4
-8
25
50
75
100
125
150
175
200
X
Y

Dependent Variable: Y
Method: Least Squares
Date: 09/22/19 Time: 11:20
Sample: 2 250
Included observations: 249

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.403193	0.229460	10.47324	0.0000
X	-0.174393	0.032858	-5.307453	0.0000

R-squared	0.102370	Mean dependent var	3.026763
Adjusted R-squared	0.098736	S.D. dependent var	3.276120
S.E. of regression	3.110182	Akaike info criterion	5.115240
Sum squared resid	2389.289	Schwarz criterion	5.143492
Log likelihood	-634.8473	Hannan-Quinn criter.	5.126612
F-statistic	28.16905	Durbin-Watson stat	0.102545
Prob(F-statistic)	0.000000		

Explanation for Spurious Regression Problem

- The standard $t-$ ratio is $\hat{\beta}$ divided by its standard error. In 'deviation from mean' form, the standard error is computed as $\frac{\hat{\sigma}_v^2}{\sum_{t=1}^{T} \tilde{x}_t^2}$.
- Recall: x_t has a unit root.
  - What is happening to $\sum_{t=1}^{T} \tilde{x}_t^2$ as T gets large?
  - What happens to the standard error as T gets large?
  - What happens to $t-$ratio as T gets large?

Dickey-Fuller Tests for Unit Root

- Econometricians developed many procedures to test for unit root.
  - The DF (Dickey-Fuller) and ADF (Augmented Dickey-Fuller) tests are examples of such tests.
  - The ADF test is straightforward and easy to implement
- Why can't we just run an AR(1) on the time series and test if the autocorrelation coefficient is 1 or less than 1?
- Because! Take the AR(1) (ignore the constant)

$$y_t = \rho y_{t-1} + \sigma \epsilon_t \tag{5}$$

where $\epsilon_t \overset{iid}{\sim} N(0,1)$. Remember, in regression, we try to estimate the slope? In this case, it is

$$\rho = \frac{Cov\left(y_t, y_{t-1}\right)}{Var\left(y_{t-1}\right)}$$

but if y_t has a unit root (is not stationary), the denominator is infinity.

The DF (Dickey-Fuller) test

- Following the example above, subtract y_{t-1} from both sides of (5),

$$
\begin{aligned}
y_t &= \rho y_{t-1} + \epsilon_t \\
y_t - y_{t-1} &= \rho y_{t-1} - y_{t-1} + \epsilon_t
\end{aligned}
$$

$$\Delta y_t = (\rho - 1) y_{t-1} + \sigma \epsilon_t \tag{6}$$

- The Dickey-Fuller (DF) test: Regress Δy_t on y_{t-1}. Tests the null hypothesis that the slope coefficient is 0 (i.e., $\rho = 1$).
- The null hypothesis is there is a unit root. If you reject the null, then you conclude that the series is stationary.
- The t-ratio distribution is not (asymptotically) normal. It was worked out by Dickey and Fuller. Eviews calculates the critical values for us.

The ADF (Augmented Dickey-Fuller) test

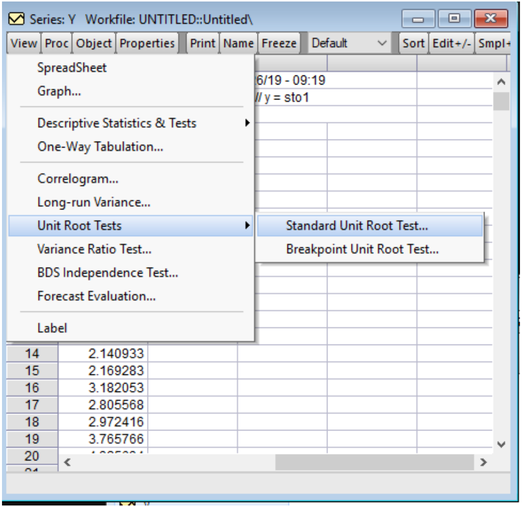
- The ADF test augments (6) to allow for additional dynamics in Δy_t.

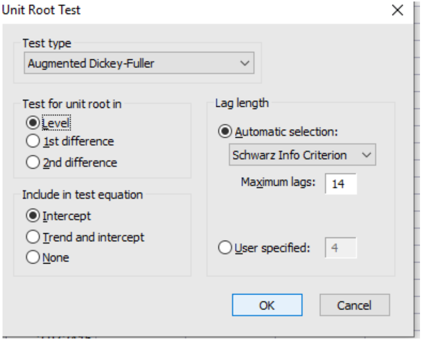
$$\Delta y_t = \underbrace{\alpha + \beta t}_{\text{time trend}} + \gamma y_{t-1} + \underbrace{\left(\delta_1 \Delta y_{t-1} + \cdots + \delta_p \Delta y_{t-p}\right)}_{**} + \sigma \epsilon_t \qquad (7)$$

  The term **, containing lagged differences of y_t is to control for possible serial correlation in Δy_t. It is like guarding against omitted variables bias.
- γ is the key parameter of interest. $\gamma = 0$, means there is a unit root. $\gamma < 0$ means the series is stationary.
- $\beta = 0$ under the null hypothesis means the series is a random walk with drift.
- $\beta \neq 0$ under the alternative hypothesis means the deviation from a deterministic time trend is a stationary AR(p) process.
- The **null hypothesis** is the series **has a unit root**.

ADF test in Eviews

'Data' is driftless random walk




Unit Root Test
Test type
Augmented Dickey-Fuller
Test for unit root in
Level
1st difference
2nd difference
Include in test equation
Intercept
Trend and intercept
None
Lag length
Automatic selection:
Schwarz Info Criterion
Maximum lags:
14
User specified:
4
OK
Cancel

Test the level

Series: Y Workfile: UNTITLED::Untitled\

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats

Augmented Dickey-Fuller Unit Root Test on Y

Null Hypothesis: Y has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=14)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.352468	0.6048
Test critical values:	1% level	-3.463405	
	5% level	-2.875972	
	10% level	-2.574541	

*MacKinnon (1996) one-sided p-values.

Test the first difference

Series: Y Workfile: UNTITLED::Untitled\

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats

Augmented Dickey-Fuller Unit Root Test on D(Y)

Null Hypothesis: D(Y) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=14)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-13.49483	0.0000
Test critical values:	1% level	-3.463576	
	5% level	-2.876047	
	10% level	-2.574581	

*MacKinnon (1996) one-sided p-values.