Topics in this lecture

- Stationarity and Unit Roots
- Spurious Regressions
- Cointegration
- Error-Correction Models
Stationarity and Unit Roots

- We need a precise test to distinguish between stationarity and non-stationarity
  - Mean is unknown and variance explodes for non-stationary time series

- Graphical techniques were not based on any precise statistical test

- In this set of slides, we’ll discuss the "unit root", and how to identify it
  - Though a unit root has a precise definition, it basically summarizes when a autoregressive relationship is non-stationary
  - Generally, we take differences, or differences of differences, or differences of differences of...of differences of differences to purge an autoregressive relationship of non-stationary properties.
**What is a unit-root?**

- Consider the following AR(1) model
  \[ y_t = \phi_1 y_{t-1} + u_t \]

- Three possible cases for this AR(1) model:
  1. \(|\phi_1| < 1\), and the series is stationary
  2. \(\phi_1 = 1\), and the series has a unit root and is non-stationary
  3. \(\phi_1 = -1\), and the series is non-stationary without a unit root
  4. \(|\phi_1| > 1\), and the series is explosive

- To test for unit root, first subtract \(y_{t-1}\) from both sides.
  \[ y_t - y_{t-1} = (\phi_1 - 1) y_{t-1} + u_t \]
  \[ \Delta y_t = \gamma y_{t-1} + u_t \]

- \(H_0: \gamma = 0\) indicates a unit root.
  - For stationarity, we reject in favor of \(\gamma < 0\). (note this is a one-sided test)
  - In this simple form, this test is known as the "Dickey-Fuller Test".
  - Test statistics are not based on a t-distribution - Table in book, correct p-values given in R.
What is a unit-root? (graphically)

- Create 3 different AR(1) time series, at or near a unit root.

  Nobs<-100
  x<-AR1(Nobs,0.25)
  y<-AR1(Nobs,1)
  z<-AR1(Nobs,1.02)

- Plot the time series

  par(mfrow=c(1,3))
  plot(x,type='l',main="phi=0.25",xlab='t',ylab="Y")
  plot(y,type='l',main="phi=1",xlab='t',ylab="Y")
  plot(z,type='l',main="phi=1.02",xlab='t',ylab="Y")

- What are the features of these three plots?
Integrated series

- From the previous series...
  \[ \Delta y_t = \gamma y_{t-1} + u_t \]

- Again, this is stationary when \( \gamma < 0 \).
  - This type of series is called "integrated of order 0": I(0)

- If not stationary, take differences and test again. If \( \Delta y_t \) is stationary, this type of series is called "integrated of order 1"

- In general, a series is integrated order \( d \) if \( d \) differences are required to make stationary.

- In R, for our previous series, take differences and plot:
  plot (diff(x,lag=1),type='l',main="phi=0.25",xlab='t',ylab="Y")
  plot (diff(y,lag=1),type='l',main="phi=1",xlab='t',ylab="Y")
  plot (diff(z,lag=1),type='l',main="phi=1.02",xlab='t',ylab="Y")

- Do the differenced series look more stationary?
Testing for unit roots manually

- In R, we need to regress the differences of a time series on initial values and test the coefficient.

- Using our original time series x:

  ```r
  summary(lm(diff(x, lag=1) ~ x[1:(N-1)]))
  ```

- x[1:(N-1)] is the vector of matched initial time periods.

- Do the same for the other series

  ```r
  summary(lm(diff(y, lag=1) ~ y[1:(N-1)]))
  summary(lm(diff(z, lag=1) ~ z[1:(N-1)]))
  ```

- These regressions are only suggestive in significance. Must use the DF significance table from the book, or the R code that I will present in a few slides.
Testing for unit roots manually

- When series appear to be non-stationary, we need find out how many differences we need to take for it to be stationary.

- To begin, test for stationarity in the differenced data

- Formally, we are testing where $\phi$ lies relative to 1 in the following

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

- Subtracting $\Delta y_{t-1}$ from both sides

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

  $$\Delta^2 y_t = \gamma \Delta y_{t-1} + u_t$$

- In R, for some series $z$:

  ```r
d1 <- z[2:(N-1)] - z[1:(N-2)]
d2 <- z[3:N] - z[2:(N-1)]
summary(lm(I(d2-d1)~d1))
```
Stationarity tests in R

- In the package "tseries", adf.test(x,k=0) runs the standard DF test.

  ```r
  library(tseries)
  adf.test(x,k=0)
  adf.test(y,k=0)
  adf.test(z,k=0)
  ```

- The null hypothesis is that there is a unit root, and the alternative is "stationary".

- The augmented DF test runs the following regression

  \[
  \Delta y_t = \beta_0 + \alpha t + \gamma y_{t-1} + \sum_{i=1}^{k} \beta_k \Delta y_{t-k} + e_t
  \]

- \(k\) adjusts the lag length in the regression.

  ```r
  adf.test(z,k=0)
  adf.test(z,k=1)
  adf.test(z,k=2)
  ```

- Again, the null is that there is a unit root.
Why we care - Spurious Regressions

- Recall from our earlier example that when $\phi = 1$
  \[ y_t = y_{t-1} + u_{yt} \]

- If we run this process enough times what do we notice?
  - The series usually trends somewhere.

- Suppose we have an independently constructed series of the same from:
  \[ x_t = x_{t-1} + u_{xt} \]

- If we regress $y_t$ on $x_t$, what happens?
  - Since both series have a tendency to trend somewhere, there will appear to be a relationship between the two series most of the time.

- This is called a **spurious relationship**. We must therefore identify unit roots when regressing time series on one another to prevent this issue.
Why we care - Spurious Regressions

- Regressing two independently created series should not show a systematic relationship.

- But, when there is a unit root in both series, there may be a spurious relationship.
  - This is bad for macro data, since as we know aggregate variables usually trend somewhere.

- Run this code repeatedly and see how many times you get an insignificant relationship between the two series:
  ```
  x1<-AR1(N,1)
  x2<-AR1(N,1)
  summary(lm(x2~x1))
  ```

- Along with there clearly being no mechanical relationship between the series (since they are random), the standard errors are incorrect for classic OLS.
Spurious Regressions (cont.)

- Why are standard errors incorrect?

- Suppose we wish to regress $y_t$ on $x_t$ using

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

- Rearranging for $u_t$, we have:

  $$ u_t = y_t - \beta_0 - \beta_1 x_t $$

- Back-substituting for $y_t = y_{t-1} + u_{yt-1}$ and $x_t = x_{t-1} + u_{xt-1}$, we have:

  $$ u_t = (y_{t-1} + u_{yt-1}) - \beta_0 - \beta_1 (x_{t-1} + u_{xt-1}) $$

- Doing so repeatedly for back to period 1, we get:

  $$ u_t = y_1 + \sum_{i=1}^{t-1} u_{yi} + \beta_0 - \beta_1 x_1 - \beta_1 \sum_{i=1}^{t-1} u_{xi} $$

- Note that because of the $\sum_{i=1}^{t-1} u_{yi}$ and $\beta_1 \sum_{i=1}^{t-1} u_{xi}$ the variance explodes as $t$ gets large relative to the initial state.
Trending time series cause problems due to the spurious regression

- This tends to be a problem with any macro data

Differencing helps, but there are drawbacks

- Cannot speak to long-run changes, only short-run (since identifying variation is based on first differences or higher order differences)

Cointegration provides a framework for identifying and estimating time series regressions
Definition of Cointegration

- Suppose we have the following time-series model
  \[ Y_t = \beta_0 + \beta_1 X_t + u_t \]
- Estimate the model to get \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \). Constructing the residuals:
  \[ \hat{u}_t = Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t \]
- If \( \hat{u}_t \sim I(0) \), then \( Y_t \) and \( X_t \) are **cointegrated**
- This is trivial if \( Y_t \) and \( X_t \) are both \( I(0) \)
- This more interesting when \( Y_t \) and \( X_t \) are both \( I(1) \)
- How often does this occur given random time series generated by a process with a unit-root?
Run a Monte Carlo to see...

```r
for(i in 1:1000){
  x<-AR1(N,1)
  y<-AR1(N,1)
  errors<-resid(lm(y~x))
  adftest<-adf.test(errors)
  p<-adftest$p.value
  if(i==1){res<-data.frame(p)}
  if(i>1){res<-rbind(res,data.frame(p))}
}
```

Calculate how many reject a unit root in favor of stationarity

```r
mean(res$p<0.1,na.rm=TRUE)
```
Error Correction Model

- If $Y_t$ and $X_t$ are $I(1)$, but $\hat{u}_t \sim I(0)$, then we can estimate using OLS the following "Error correction model"

$$
\Delta Y_t = a_0 + b_1 \Delta X_t + \pi \hat{u}_{t-1} + e_t
$$

- This regression is not spurious because $\hat{u}_{t-1}$, $\Delta Y_t$ and $\Delta X_t$ are all $I(0)$

- Since, $Y_t$ and $X_t$ are also $I(0)$, we can obtain consistent estimates for $b_1$ using standard regression

- We can also obtain long-run equilibrium dynamics by focusing on $\pi$.

  - $\hat{u}_{t-1} \neq 0$ indicates disequilibrium between $Y$ and $X$ in $Y_{t-1} = \beta_0 + \beta_1 X_{t-1} + u_{t-1}$
  
  - $\pi = -1$ equilibrium is reached immediately
  
  - $\pi \in (-1, 0]$ equilibrium is reached gradually
  
  - $\pi < -1$ suggests an over-correction
Error Correction Model to ARDL model

- The ECM model is equivalent to a ARDL model (Autoregressive Distributed Lag), which we presented a few lectures ago when talking about VARs. To see this, note that:

\[
\Delta Y_t = a_0 + b_1 \Delta X_t + \pi \hat{u}_{t-1} + e_t
\]

- Expanding $\Delta Y_t$, $\Delta X_t$, and $\hat{u}_{t-1}$, we have:

\[
Y_t - Y_{t-1} = a_0 + b_1 (X_t - X_{t-1}) + \pi (Y_{t-1} - \beta_0 - \beta_1 X_{t-1}) + e_t
\]

- Bringing all lags to the RHS:

\[
Y_t = a_0 + Y_{t-1} + \pi Y_{t-1} + b_1 X_t - b_1 X_{t-1} - \pi \beta_0 - \pi \beta_1 X_{t-1} + e_t
\]

- Collecting terms

\[
Y_t = a_0 - \pi \beta_0 + (1 + \pi) Y_{t-1} + b_1 X_t + (-b_1 - \pi \beta_1) X_{t-1} + e_t
\]

- Thus, we have an ARDL model.
Engel and Granger have proposed a technique for evaluating data that may be spurious.

**Step 1:** Determine whether $X_t$ and $Y_t$ are cointegrated.
- If $X_t$ and $Y_t$ are $I(0)$, then use classic regression
- If only one of $X_t$ and $Y_t$ are $I(1)$, and the other $I(0)$, then need a new technique
- If $X_t$ and $Y_t$ are $I(1)$, then run
  
  $$ Y_t = \beta_0 + \beta_1 X_t + u_t $$

  and collect residuals. Go to step 2.

**Step 2:** Check whether $u_t$ is $I(0)$
- If $u_t$ is $I(0)$, move to step 3.
- If $u_t$ is $I(1)$, find a new model

**Step 3:** Estimate and interpret:

$$ \Delta Y_t = a_0 + b_1 \Delta X_t + \pi \hat{u}_{t-1} + e_t $$
Example - Co-integration of investment funds

- Download a year of daily opening prices for SPY and VOO
  ```
  getSymbols('SPY',from='2014-11-12',to='2015-11-12')
  getSymbols('VOO',from='2014-11-12',to='2015-11-12')
  prices.spy <- SPY$SPY.Open
  prices.voo <- VOO$VOO.Open
  ```

- **Step 1:** Determine whether **SPY** and **VOO** have unit root.
  ```
  adf.test(prices.spy)
  adf.test(prices.voo)
  ```
Example - Co-integration of investment funds

- **Step 2:** Regress

\[
SPY_t = a_0 + b_1 VOO_t + u_t
\]

and test whether \( u_t \) is \( I(0) \)

```r
coint <- lm(prices.spy~prices.voo)
summary(coint)
beta<-coint$coef
adf.test(resid)
```

- **Step 3:** Estimate and interpret:

\[
\Delta SPY_t = a_0 + b_1 \Delta VOO_t + \pi \hat{u}_{t-1} + e_t
\]

```r
lag.resid<-lag(resid,1)
dSPY<-prices.spy-lag(prices.spy,1)
dVVOO<-prices.voo-lag(prices.voo,1)
ecm <-lm(dSPY~dVVOO+lag.resid)
summary(ecm)
```