These notes largely concern autocorrelation

Issues Using OLS with Time Series Data

Recall main points from Chapter 10:

- Time series data <u>NOT randomly sampled</u> in same way as cross sectional—each obs not i.i.d

  Why?
  Data is a "stochastic process"—we have one realization of the process from a set of all possible realizations

<u>Leads to a Number of Common problems:</u>

1. Errors correlated over time—high errors today→ high next time (biased standard errors but not biased coefficients)

2. Effects may take a while to appear→ difficult to know how long should wait to see effects (tax cuts—is growth in Clinton years due to Clinton? Reagan?) (specification problem)

3. Feedback effects (x→ y but after seeing y, policy makers adjust x) (specification problem—can lead to biased coeffs)

4. Trending data over time→ data series can look like they are related, but really is "spurious" (biased coeffs)

Related Issue: **Prediction**—often want a prediction of future prices, GDP, etc.—Need to use properties of existing series to make that prediction

Recall Chapter 10 Models
These models dealt with problems 2 and 4 listed above

1. Static model-- Change in z has an immediate effect—in same period—on y
   $y_t = \beta_0 + \beta_1 z_t + u_t \quad t=1,2,\ldots n$

2. Finite Distributed lag Model
   $y_t = \alpha + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t \quad t=1,2,\ldots n$
   Know number of lags

3. Trending Data:
   Add a trend yt = α0 + α1t + et , t= 1,2
   Or Detrend the data

Note that if DO NOT correctly specify the model (e.g., with lagged data), can generate serial correlation. Correct specification is the first problem to address.

P&R 6.2 Serial Correlation: What is serial correlation and why is it a problem?

- Serial correlation comes when errors from one time period are carried over into future time periods (problem # 1 listed above)
- Can also occur spatially—errors in this area are correlated with errors in adjacent area
- Most authors use serial and auto-correlation interchangeably. Some use auto corr to refer to serial correlation within a series itsel and serial correlation to refer to lagged correlation between two time series. I'll use them interchangeably.

Positive serial correlation often caused by

--Inertia—some economic time series have "momentum" (?)

--Correlation in omitted variables over time

--Correlation in measurement error component of error term

--Theoretical predictions--adaptive expectations, some partial adjustment process

--Misspecification—e.g., omitted dynamic terms (lagged dependent or independent variables, trends)

--Data is already interpolated (e.g., data between Census years)

--Non-stationarity—will discuss later

Example: AR(1) Process

Very common form of serial correlation

First Order Autoregressive process: AR(1)

True model: $y_t = \beta_0+\beta_1 x_{1t} + \beta_2 x_{2t} + \ldots \beta_k X_{kt} + \varepsilon_t$

$\varepsilon_t = \rho\varepsilon_{t-1} + v_t \qquad 0\leq|\rho|\leq 1$

[If had 2 lags, would be AR(2)]

- v_t is the idiosyncratic part of the error, Indep of other errors over time, $N(0, \sigma^2_v)$
- e_t is NOT indep of other errors over time, $N(0, \sigma^2_\varepsilon)$
- error in time t is determined by the diminishing value of error in previous period (ρ) + addition of random variable v, with EV(0)

  →Implies that error in any period is reflected in all future periods

  $Var(\varepsilon_t) = E(\varepsilon_t^2) = E[(\rho\varepsilon_{t-1} + v_t)^2] = E[\rho^2\varepsilon^2_{t-1} + v^2_t + 2\rho\varepsilon_{t-1}v_t]$
  $= \rho^2 E(\varepsilon^2_{t-1}) + E(v^2_t)$ b/c ε_{t-1} and v_t are indep
  →$Var(\varepsilon_t) = \rho^2 Var(\varepsilon_t) + \sigma^2_v$ if (ε_t) is homoskedastic
  →$Var(\varepsilon_t)= Var(\varepsilon_{t-1})$

  Algebra→ $Var(\varepsilon_t)= \sigma^2_v /(1- \rho^2)$ Note that when $\rho=0$, no autocorrel.

How are the errors related over time?

$Cov(\varepsilon_{t,} \varepsilon_{t-1}) = E(\varepsilon_{t,} \varepsilon_{t-1})= E[(\rho\varepsilon_{t-1} + v_t) \varepsilon_{t-1}] = E(\rho\varepsilon^2_{t-1} + \varepsilon_{t-1}v_t) = \rho E(\varepsilon^2_{t-1}) = \rho Var(\varepsilon_t) = \rho\sigma^2_\varepsilon$

Similarly, $Cov(\varepsilon_{t,} \varepsilon_{t-2})= \rho^2\sigma^2_\varepsilon$, $Cov(\varepsilon_{t,} \varepsilon_{t-3})= \rho^3\sigma^2_\varepsilon$
Similarly, $Cov(\varepsilon_{t,} \varepsilon_{t-s})= \rho^s\sigma^2_\varepsilon$

Note that ρ is the correlation coefficient between errors at time t and t-1. Also known as coefficient of autocorrelation at lag 1

Stationarity:

Critical that $|\rho| < 1$—otherwise these variances and covariances are undefined.

If $|\rho| < 1$, we say that the series is stationary. If $\rho=1$, nonstationary.

Chapter 11 in your book discusses concept of stationarity. For now, brief definition. If mean, variance, and covariance of a series are time invariant, series is stationary.

Will discuss later tests of stationarity and what to do if data series is not stationary.

Serial correlation leads to biased standard errors

If y is positively serially correlated and x is positively serially correlated, will understate the errors

- Show figure 6.1 for why
- Note that 1st case have positive error initially, second case have negative error initially
- Both cases equally likely to occur→unbiased
- But OLS line fits the data points better than true line

With algebra:

Usual OLS Estimator $y_t = \beta_0+\beta_1 x_{1t} + \varepsilon_t$

$$var(\widehat{\beta_1}) = \frac{\sigma^2}{\sum x_i^2}$$

With AR(1)

$$var(\widehat{\beta_1}) = \frac{\sigma^2}{\sum x_t^2}\left[1 + 2\rho\frac{\sum x_t x_{t-1}}{\sum x_t^2} + 2\rho^2\frac{\sum x_t x_{t-2}}{\sum x_t^2} + \cdots + 2\rho^{n-1}\frac{\sum x_t x_n}{\sum x_t^2}\right]$$

How does this compare with standard errors in OLS case?
Depends on sign of p and type of autocorrelation in xs

If x is positively correlated over time and p is positive, OLS will understate true errors

- ➔ **T, F stats all wrong**
- ➔ **R2 wrong**

See Gujarati for a Monte Carlo experiment on how large these mistakes can be

Tests for Serial Correlation

1. Graphical method

Graph (residuals) errors in the equation---very commonly done.

Can also plot residuals against lagged residuals—see Gujarati fig 12.9

2. Durbin Watson Test Oldest test for serial correlation
P&R goes through extension when have lagged y's in model—see 6.2.3 for details

Null hypothesis: No serial correlation $\rho=0$
Alternative: $\rho\neq0$ (two tailed)
$\rho>0$ (one tailed)

Test statistic:
Step 1: Run OLS model $y_t = \beta_0+\beta_1 x_{1t} + \beta_2 x_{2t} + \ldots \beta_k X_{kt} + \varepsilon_t$
Step 2: Calculate predicted residuals
Step 3: Form test statistic

$$DW = \frac{\sum_{t=2}^{T}(\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^{T}(\hat{\varepsilon}_t)^2} \cong 2(1-\hat{\rho}) \quad \text{(See Gujarati pg 435 to derive)}$$

Assumptions:
1. Regression includes intercept term
2. Xs are fixed in repeated sampling—non-stochastic (problematic in time series context)
3. Can only be used for 1st order autoregression processes
4. Errors are normally distributed
5. No lagged dependent variables—not applicable in those models
6. No missing obs

This statistic ranges from 0 to 4

- $\hat{\varepsilon}_t$ are close to each other→ Positive serial correlation→DW will be close to zero (below 2)
- No serial correlation→ DW will be close to 2
- Negative serial correlation→ DW will be large (above 2)

Exact interpretation difficult because sequence of predicted error terms depends on x's as well→ if x's are serially correlated, correlation of predicted errors may be related to this and not serial correlation of εs

→ 2 critical values d_L and d_U

--see book for chart

```
STATA:   estat dwstat
```

3. Breusch-Godfrey test
This is yet another example of an LM test

Null hypothesis: Errors are serially independent up to order p

One X:

Step 1: Run OLS model $y_t = \beta_0 + \beta_1 x_{1t} + \varepsilon_t$
(Regression run under the null)

Step 2: Calculate predicted residuals

Step 3: Run auxiliary regression
$\hat{\varepsilon}_t = \alpha 1 + \alpha 2X + \rho\hat{\varepsilon}_{t-1} + v_t$

Step 4: T-test on $\hat{\rho}$

```
STATA: estat bgodfrey, lags(**)
```

Multiple X, multiple lags

Step 1: Run OLS model $y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots .\beta_k X_{kt} + \varepsilon_t$
(Regression run under the null)

Step 2: Calculate predicted residuals

Step 3: Run auxiliary regression
$\hat{\varepsilon}_t = \alpha 1 + \alpha 2Xs + \rho_1\hat{\varepsilon}_{t-1} + \rho_2\hat{\varepsilon}_{t-1} + ... + + \rho p\hat{\varepsilon}_{t-p} + v_t$
with higher order lags—Bruesch-Godfrey test

Step 4: $(n-p)R^2 \sim \chi^2_{(p)}$

BP test is more general than DW test—cam include laggesd Ys, moving average models

Do need to know p—order or the lag. Will talk some about this choice later.

Correcting for Serial Correlation

1. Check—is it model misspecification?
   --trend variable?
   --quadratics?
   --lagged variables?

2. Use GLS estimator—see below

3. Use Newey –West standard errors—like robust standard errors

GLS Estimators:
Correction1: Known ρ: Adjust OLS regression to get efficient parameter estimates

Want to transform the model so that errors are independent

$\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ → want to get rid of $\rho\varepsilon_{t-1}$ part

How? Linear model holds for all time periods.

$$y_{t-1} = \beta_0+\beta_1 x_{1t-1} + \beta_2 x_{2t-1} + \ldots \beta_k X_{kt-1} + \varepsilon_{t-1}$$

1. Multiply above by ρ
2. Subtract from base model:

$$y^*_t = \beta_0(1-\rho) + \beta_1 x^*_{1t} + \beta_2 x^*_{2t} + \ldots \beta_k X^*_{kt} + v_t$$

Where $y^*t = y_t - \rho y_{t-1}$, same for xs

Note that this is like a first difference, only are subtracting part and not whole of yt-1→*Generalized differences*

Now error has a mean =0 and a constant variance

→ Apply OLS to this transformed model→ efficient estimates

This is the BLUE estimator

PROBLEM: don't know ρ

Correction2: Don't Know ρ--Cochrane-Orcutt

Idea: start with a guess of ρ and iterate to make better and better guesses

Step 1: Run ols on original model

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots \beta_k X_{kt} + \varepsilon_t$$

Step 2: Obtain predicted residuals and run following regression

$$\hat{\varepsilon}_t = \rho \hat{\varepsilon}_{t-1} + v_t$$

Step 3: Obtain predicted value of ρ. Transform data using generalized differencing transformation $y_t^* = y_t - \hat{\rho} y_{t-1}$, same for X*

Step 4: Rerun regression using transformed data

$$y_t^* = \beta_0(1 - \hat{\rho}) + \beta_1 x_{1t}^* + \ldots + \beta_k x_{kt}^* + v_t$$

Obtain new estimates of betas-- $\hat{\hat{\beta}}$

Step 5: Form new estimated residuals using newly estimated betas and ORIGINAL data (not transformed data)

$$\hat{\hat{\varepsilon}}_t = yt - (\hat{\hat{\beta}}_0 + \hat{\hat{\beta}}_1 x_{1t} + \ldots + \hat{\hat{\beta}}_k x_{kt}$$

Iterate until new estimates of ρ are "close" to old estimates (differ by .01 or .005)

Correction3: Don't Know ρ--Hildreth-Lu (less popular)

Numerical minimization method

Minimize sum of squared residuals for various guesses of ρ for

$$y^*_t = \beta_0(1-\rho) + \beta_1 x^*_{1t} + \beta_2 x^*_{2t} + \ldots \beta_k X^*_{kt} + v_t$$

Choose range of potential ρ (e.g., 0, .1, .2, .3, ..., 1.0), identify best one (e.g., .3), pick other numbers close by (e.g., .25, .26, . . . , .35), iterate

Correction 4:   First difference Model

ρ lies between 0 and 1. Could run a first differenced model as the other extreme. This is the appropriate correction when series is non-stationary—talk about next time.

Recall: Correcting for Serial Correlation

1. Check—is it model misspecification?
   --trend variable?
   --quadratics?
   --lagged variables?

2. Use GLS estimator—see below

3. Use Newey –West standard errors—like robust standard errors

Newey –West standard errors

Extension of White standard errors for heteroskedasticity
Only valid in large samples

Final Notes:

Should you use OLS or FGLS or Newey-West errors?

OLS:
--unbiased
--consistent
--asymptotically normal
--t,F, r2 not appropriate

FGLS/Newey West
--efficient
--small sample properties not well documented—not unbiased
--in small samples, then, might be worse

--Griliches and Rao rule of thumb—is sample is small (<20, iffy 20-50) and $\rho<.3$, OLS better than FGLS