Lecture 6: OLS asymptotics and further issues

Topics we’ll cover today

- Asymptotic consistency of OLS
- Lagrange multiplier test
- Data scaling
- Predicted values with logged dependent variables
- Interaction terms

Consistency

$$n \to \infty$$

$$bias \to 0$$

- Consistency is a more relaxed form of unbiasedness. An estimator may be biased, but as *n* approaches infinity, it may be consistent (or unbiased in the limit).
- Consistency of the OLS slope estimate requires a relaxed version of MLR4
  - Each x_j is uncorrelated with *u*

Inconsistency

$$n \to \infty$$

$$bias \to |c| > 0$$

- If any x_j is correlated with *u,* each slope estimate is biased, and increasing sample size does not eliminate bias, so the slope estimates are inconsistent as well.

Asymptotics of hypothesis testing

- MLR6 assumes that the error term is distributed normally, allowing us to perform t-tests and F-tests on the estimated parameters.
- In practice, the actual distribution of the error term has a lot to do with the distribution of the dependent variable. In many cases, with a highly non-normal dependent variable, the error term is nowhere near normally distributed.
- But . . .

Asymptotics of hypothesis testing

- If assumptions MLR1 through MLR5 hold,

$$n \to \infty$$

$$\hat{\beta}_j \sim N(\beta_j, se(\hat{\beta}_j))$$

$$(\hat{\beta}_j - \beta_j)/se(\hat{\beta}_j) \overset{a}{\sim} N(0,1)$$

$$se(\hat{\beta}_j) \cong c_j / \sqrt{n}$$

- This means that t and F tests are valid as sample size increases. Also, the standard error will decrease proportional the increase in the square root of the sample size.

Asymptotics of hypothesis testing

- If assumptions MLR1 through MLR5 hold,

$$n \rightarrow \infty$$

$$\hat{\beta}_j \sim N(\beta_j, se(\hat{\beta}_j))$$

$$(\hat{\beta}_j - \beta_j) / se(\hat{\beta}_j) \overset{a}{\sim} N(0,1)$$

$$se(\hat{\beta}_j) \cong c_j / \sqrt{n}$$

- We are not invoking MLR6 here. We make no assumption about the distribution of the error terms.
- This means that as n approaches infinity, our parameters are normally distributed.

Asymptotics of hypothesis testing

- But how close to infinity do we need to get before we can invoke the asymptotic properties of OLS regression?
- Some econometricians say 30. Let's say above 200, assuming you don't have too many regressors.
- Note: Reviewers in criminology are typically not sympathetic to the asymptotic properties of OLS!

Lagrange Multiplier test

- In large samples, an alternative to testing multiple restrictions using the F-test is the Lagrange multiplier test.
  1. Regress *y* on restricted set of independent variables
  2. Save residuals from this regression
  3. Regress residuals on unrestricted set of independent variables.
  4. R-squared times *n* in above regression is the Lagrange multiplier statistic, distributed chi-square with degrees of freedom equal to number of restrictions being tested.

Lagrange Multiplier test example

- Does ethnicity/race, age, delinquency frequency, school attachment, income and antisocial peers explain any variation in high school gpa?
- We will compare to a model that only includes male, middle school gpa and math knowledge.

```
. reg hsgpa male msgpa r_mk

      Source |       SS       df       MS              Number of obs =    6574
-------------+------------------------------           F(  3,  6570) = 2030.42
       Model |  1488.67547     3  496.225156           Prob > F      =  0.0000
    Residual |   1605.6756  6570  .244395069           R-squared     =  0.4811
-------------+------------------------------           Adj R-squared =  0.4809
       Total |  3094.35107  6573  .470766936           Root MSE      =  .49436

------------------------------------------------------------------------------
       hsgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |  -.1341638    .012397   -10.82   0.000     -.158466   -.1098616
       msgpa |   .4352299   .0081609    53.33   0.000     .4192319    .4512278
        r_mk |   .1728567   .0074853    23.09   0.000     .1581832    .1875303
       _cons |   1.554284   .0257374    60.39   0.000      1.50383    1.604738
------------------------------------------------------------------------------

. predict residual, r
```

← What do the residuals represent?

Lagrange Multiplier test example

```
. reg residual male hisp black other agedol dfreq1 schattach msgpa r_mk income1 antipeer

      Source |       SS       df       MS              Number of obs =    6574
-------------+------------------------------           F( 11,  6562) =   29.76
       Model |  76.3075043    11  6.93704584           Prob > F      =  0.0000
    Residual |   1529.3681  6562  .233064325           R-squared     =  0.0475
-------------+------------------------------           Adj R-squared =  0.0459
       Total |   1605.6756  6573  .244283524           Root MSE      =  .48277

------------------------------------------------------------------------------
    residual |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |  -.0232693   .0122943    -1.89   0.058    -.0473701    .0008316
        hisp |  -.0600072   .0174325    -3.44   0.001    -.0941806   -.0258337
       black |  -.1402889   .0152967    -9.17   0.000    -.1702753   -.1103024
       other |  -.0282229   .0186507    -1.51   0.130    -.0647844    .0083386
      agedol |  -.0105066   .0048056    -2.19   0.029    -.0199273    -.001086
      dfreq1 |  -.0002774   .0004785    -0.58   0.562    -.0012153    .0006606
   schattach |   .0216439   .0032003     6.76   0.000     .0153702    .0279176
       msgpa |  -.0260755   .0081747    -3.19   0.001    -.0421005   -.0100504
        r_mk |  -.0408928   .0077274    -5.29   0.000    -.0560411   -.0257445
     income1 |   1.21e-06   1.60e-07     7.55   0.000     8.96e-07    1.52e-06
    antipeer |  -.0167256   .0041675    -4.01   0.000    -.0248953   -.0085559
       _cons |   .0941165   .0740153     1.27   0.204    -.0509776    .2392106
------------------------------------------------------------------------------
```

Lagrange Multiplier test example

```
. di "This is the Lagrange multiplier statistic:",e(r2)*e(N)
This is the Lagrange multiplier statistic: 312.42022

. di chi2tail(8,312.42022)
9.336e-63
```

- Null rejected.
- The degrees of freedom in either the restricted or unrestricted model plays no part on the test statistic. This is because the test relies on large sample properties.
- The residual from the first regression represents variation in high school gpa not explained by the first three variables (sex, middle school gpa and math knowledge).
- The second regression shows us whether the excluded variables can explain any variation in the dependent variable that the included variables couldn’t.

In-class exercise

Do questions 1 through 4

Data scaling and OLS estimates

- If you multiply *y* by a constant *c*
  - the coefficients are multiplied by *c*
  - SST, SSR, SSE are multiplied by c^2
  - RMSE multiplied by *c*
  - *R-squared, F-statistic, t-statistics, p values* **unchanged**
  - If you have really small coefficients that are statistically significant, multiply your dependent variable by a constant for ease of interpretation.
- If you add a constant *c* to *y*
  - Intercept changes by same amount.
  - Nothing else changes.

Data scaling and OLS estimates

- If you multiply x_j by a constant c
  - the coefficient β_j, se(β_j), CI(β_j) are *divided* by c
  - Nothing else changes
- If you add a constant c to x_j
  - Intercept reduces by $c*\beta_j$
  - Standard error and confidence interval of intercept changes as well.
  - Nothing else changes.

Predicted values with logged dependent variables

- It is incorrect to simply exponentiate the predicted value from the regression with the logged dependent variable. The error term must be taken into account:

$$\hat{y} = \exp(\hat{\sigma}^2 / 2) \cdot \exp(\log yhat)$$

- Where σ^2 (hat) is the mean squared error of the regression.
- Even better, where alpha hat is the expected value of the exponentiated error term:

$$\hat{y} = \hat{\alpha}_0 \cdot \exp(\log yhat)$$

Predicted values with logged dependent variables

- Alpha hat can be estimated two different ways.
  - Take the average of the exponentiated residuals ("smearing estimate", I kid you not)
  - Regress y on the expected value of log(y) from the initial regression (no constant). The slope estimate is an estimate of alpha.
- Example of smearing estimate in ceosal1.dta:

Predicted values with logged dependent variables, example

```
. reg lsalary lsales

      Source |       SS       df       MS              Number of obs =     209
-------------+------------------------------           F(  1,   207) =   55.30
       Model |  14.0661688     1  14.0661688           Prob > F      =  0.0000
    Residual |  52.6559944   207  .254376785           R-squared     =  0.2108
-------------+------------------------------           Adj R-squared =  0.2070
       Total |  66.7221632   208  .320779631           Root MSE      =  .50436

------------------------------------------------------------------------------
     lsalary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lsales |   .2566717   .0345167     7.44   0.000     .1886224    .3247209
       _cons |   4.821997   .2883396    16.72   0.000     4.253538    5.390455
------------------------------------------------------------------------------
```

```
. gen wronghat=exp(4.821997+.2566717*lsales)

. summ wronghat

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    wronghat |       209     1079.71    292.9326    467.755   2370.687
```

Predicted values with logged dependent variables, example

```
. predict resid, r

. gen expresid=exp(resid)

. summ expresid
```

Variable	Obs	Mean	Std. Dev.	Min	Max
expresid	209	1.199045	1.364334	.3640822	16.63112

```
. gen righthat2=1.199045*exp(4.821997+.2566717*lsales)

. summ righthat2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
righthat2	209	1294.621	351.2394	560.8593	2842.561

- Another way to obtain an estimate of alpha-hat:

```
. reg lsalary sales

      Source |       SS       df       MS              Number of obs =     209
-------------+------------------------------           F(  1,   207) =   17.79
       Model |  5.27916955     1  5.27916955           Prob > F      =  0.0000
    Residual |  61.4429937   207  .296826056           R-squared     =  0.0791
-------------+------------------------------           Adj R-squared =  0.0747
       Total |  66.7221632   208  .320779631           Root MSE      =  .54482

------------------------------------------------------------------------------
     lsalary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       sales |    .000015   3.55e-06     4.22   0.000     7.98e-06     .000022
       _cons |    6.84665    .045003   152.14   0.000     6.757927    6.935373
------------------------------------------------------------------------------

. predict p
(option xb assumed; fitted values)

. gen ep=exp(p)

. reg salary ep, noc

      Source |       SS       df       MS              Number of obs =     209
-------------+------------------------------           F(  1,   208) =  176.55
       Model |   337331623     1   337331623           Prob > F      =  0.0000
    Residual |   397426261   208  1910703.18           R-squared     =  0.4591
-------------+------------------------------           Adj R-squared =  0.4565
       Total |   734757884   209  3515587.96           Root MSE      =  1382.3

------------------------------------------------------------------------------
      salary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ep |   1.158517   .0871908    13.29   0.000     .9866264    1.330408
------------------------------------------------------------------------------
```

In-class exercise

Do questions 5 and 6

Assumption #0: Additivity

- This assumption, usually unstated, implies that for each X_j, the effect is constant regardless of the values other independent variables.
- If we believe, on the other hand, that the effect of X_j depends on values of some other independent variable X_k, then we estimate an interactive (non-additive) model

Interactive model, non-additivity

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 \left(X_1 X_2 \right) \ldots + \beta_k X_k + u$$

- In this model, the effects of X_1 and X_2 on Y are no longer constant.
- The effect of X_1 on Y is $(\beta_1 + \beta_3 X_2)$
- The effect of X_2 on Y is $(\beta_2 + \beta_3 X_1)$

Interactive model, non-additivity

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 X_2) \ldots + \beta_k X_k + u$$

- This drastically changes the meaning of β_1 and β_2
- β_1 is now the effect of X_1 on Y ***when X_2 equals zero.***
- β_2 is now the effect of X_2 on Y ***when X_1 equals zero.***
- If X_2 never equals zero in your sample, β_1 is meaningless!
- If X_1 never equals zero in your sample, β_2 is meaningless!
- Do not interpret the magnitude of β_3 by itself. It is interpreted in combination with either β_1 or β_2.
- If β_3 is statistically significant, it means that the effect of X_1 on Y depends on X_2, or that the effect of X_2 on Y depends on X_1, or both.

Non-additivity example: Hay & Forrest 2008

Figure 1. Opportunity as an Hypothesized Moderator of the Relationship Between Low Self-Control and Crime

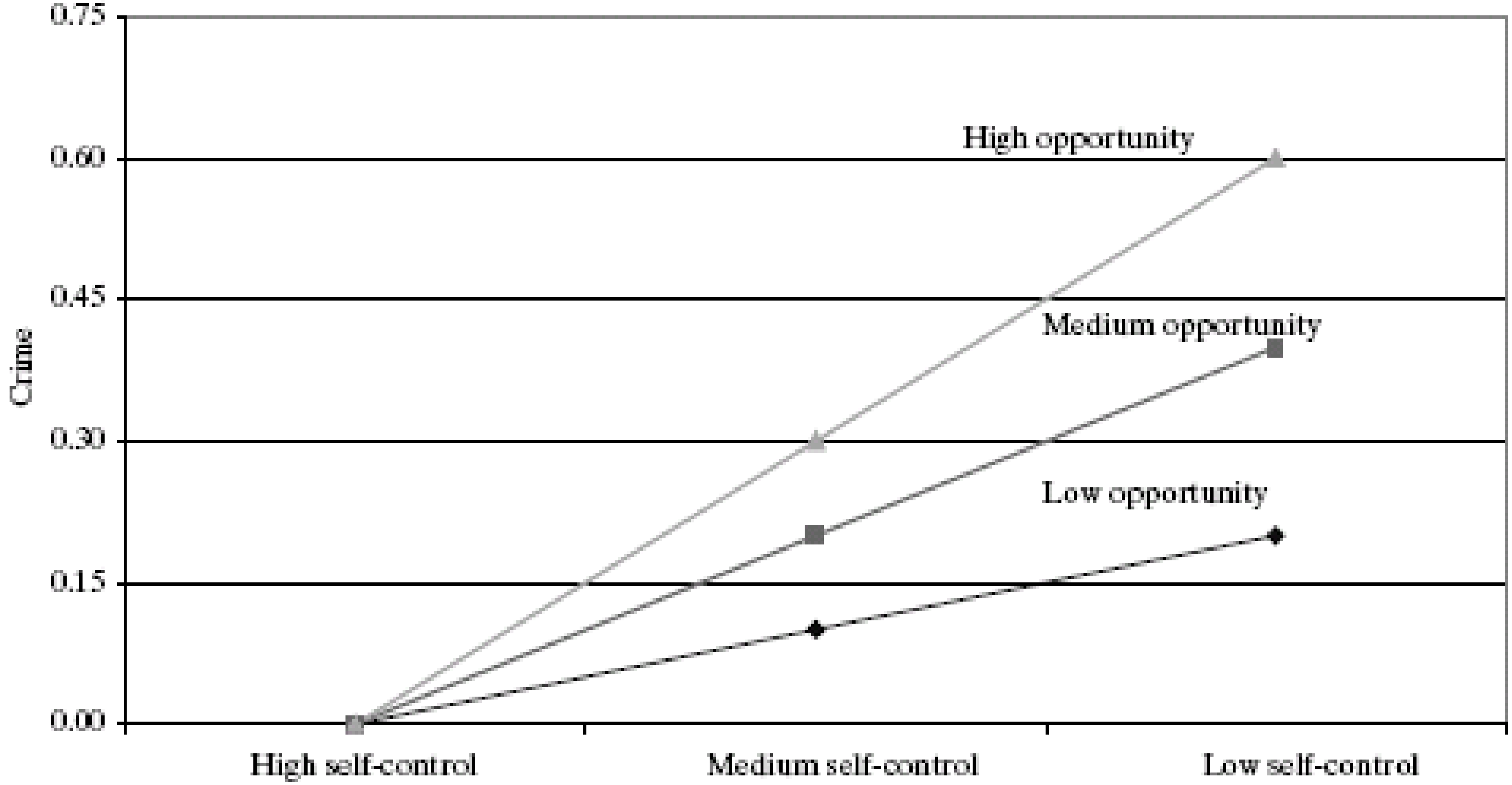

Table 2. Results for OLS Regressions of Crime on Low Self-Control, Opportunity, and the Interaction between the Two

	Unsupervised Time from Home (Mother)		Unsupervised Time from Home (Child)		Time with Peers		Adult Absence	
Intercept	−.62	(.37)	−.59	(.37)	−.68	(.38)	−.65	(.37)
	—		—		—		—	
Age	.06**	(.03)	.06**	(.03)	.07**	(.03)	.07**	(.03)
	.07		.06		.08		.07	
Male	.11**	(.05)	.10**	(.04)	.11**	(.05)	.13**	(.05)
	.08		.07		.09		.10	
Black	.10	(.06)	.12**	(.06)	.14**	(.06)	.14**	(.06)
	.07		.08		.09		.09	
Hispanic	−.04	(.06)	−.05	(.06)	−.02	(.06)	−.01	(.06)
	−.02		−.03		−.01		.00	
Parental attachment	−.13***	(.04)	−.10**	(.04)	−.15***	(.04)	−.13***	(.04)
	−.13		−.11		−.16		−.13	
Deviant peer pressure	.79***	(.16)	.78***	(.17)	.85***	(.17)	.83***	(.17)
	.26		.26		.28		.27	
Low self-control	.11***	(.03)	.13***	(.03)	.14***	(.03)	.11***	(.03)
	.16		.19		.21		.17	
Opportunity	.08***	(.02)	.11***	(.03)	.03	(.02)	.06**	(.03)
	.12		.16		.05		.10	
Low self-control × opportunity	.09***	(.03)	.06*	(.03)	.02	(.02)	.05	(.04)
	.15		.09		.03		.08	
N (Clusters)	749 (692)		743 (686)		744 (687)		740 (683)	
R^2	.240		.253		.226		.222	
F-test for R^2 differences[a]	22.79***		7.15***		.87		4.58**	

NOTES: Unstandardized coefficients and robust standard errors (in parentheses) are shown in the top row, and standardized coefficients are shown in the bottom row.

[a]The *F*-test is computed by comparing the *R*-squared in these equations to the *R*-squared that was obtained in the table 1 equations that do not contain the interactions (Bohrnstedt and Knoke, 1988: 414).

*$p < .06$; **$p < .05$; ***$p < .01$.

Non-additivity example: Hay & Forrest 2008

- The standardized coefficients in the first column can be interpreted as follows:
  - On average (when opportunity=0, the average), a 1 standard deviation decrease in self-control is associated with a .16 s.d. increase in crime
  - But for those with 1 s.d. less unsupervised time, a 1 s.d. decrease in self-control is associated with a .01 s.d. increase in crime
  - And for those with 1 s.d. more unsupervised time, a 1 s.d. decrease in self-control is associated with a .31 s.d. increase in crime.
- Or, we could focus on the standardized effect of unsupervised time, with self-control as a moderator:
  - .12 on average, for those with average self-control
  - -.03 for those with 1 s.d. higher self-control
  - .27 for those with 1 s.d. lower self-control

Non-additivity: interaction terms

- Interpretation of the main effects in non-additive models is easier if 0 has a substantive meaning for both variables in the interaction term.
- Wooldridge (p. 197) notes that if we would like the main effects to have specific meanings, we can subtract particular values from X_1 and X_2 before multiplying them.

Non-additivity: interaction terms

- To determine if interaction term adds to explanation, look at t-statistic for interaction term, or conduct F-test for restricted/unrestricted models.
- Hay and Forrest used an r-squared version of the restricted/unrestricted F-test. It's equivalent.
- In general, in order to interpret interaction effects, you have to plug in interesting values for the X_2 term and see how the effect on X_1 changes, or vice versa. Either that, or learn how to use the margins command.

In-class exercise

Do questions 7 through 9

Next time:

Homework 7 Problems C4.10, C5.2, C5.3, C6.4i, ii, iii, C6.6i, ii, iii, iv

Read: Wooldridge Chapter 7