Econometrics
Lab Hour – Session 2

Agustín Bénétrix
benetria@tcd.ie
Office hour:
Wednesday 4-5
Room 3021

Martin Schmitz
schmitzm@tcd.ie
Office hour:
Thursday 5-6
Room 3021

Outline

- Importing the dataset
- Regression analysis in Microfit
- Project time

Import Dataset

- Use MS Excel
- Important to keep format
  - First column: ID number/Year
  - First row: Variable name
  - Second row: Variable description
  - Save as .csv-file
- Import to Microfit (using MEAP93.csv)
  - Save as .fit file

Tasks

- Estimate the model

  Math10 = beta0 + beta1*log(*expend*) + beta2**lnchprg* + u
- Report:
  - The equation
  - Sample size, R-squared
  - Are the slope coefficients what you expected?

Tasks

```
                    Ordinary Least Squares Estimation
*******************************************************************************
 Dependent variable is MATH10
 408 observations used for estimation from    1 to  408
*******************************************************************************
 Regressor               Coefficient       Standard Error         T-Ratio[Prob]
 C                          -20.3608           25.0729           -.81206[.417]
 LEXPEND                      6.2297            2.9726            2.0957[.037]
 LNCHPRG                     -.30459           .035357           -8.6145[.000]
*******************************************************************************
 R-Squared                     .17993   R-Bar-Squared                   .17588
 S.E. of Regression            9.5262   F-stat.    F(  2, 405)   44.4293[.000]
 Mean of Dependent Variable   24.1069   S.D. of Dependent Variable     10.4936
 Residual Sum of Squares      36753.4   Equation Log-likelihood        -1497.1
 Akaike Info. Criterion       -1500.1   Schwarz Bayesian Criterion     -1506.1
 DW-statistic                  1.9028
*******************************************************************************


                           Diagnostic Tests
*******************************************************************************
*    Test Statistics  *        LM Version        *         F Version         *
*******************************************************************************
*                     *                          *                           *
* A:Serial Correlation*CHSQ(   1)=   .68241[.409]*F(   1, 404)=   .67685[.411]*
*                     *                          *                           *
* B:Functional Form   *CHSQ(   1)=   1.5241[.217]*F(   1, 404)=   1.5148[.219]*
*                     *                          *                           *
* C:Normality         *CHSQ(   2)=  83.7201[.000]*      Not applicable       *
*                     *                          *                           *
* D:Heteroscedasticity*CHSQ(   1)=   2.0906[.148]*F(   1, 406)=   2.0911[.149]*
*******************************************************************************
   A:Lagrange multiplier test of residual serial correlation
   B:Ramsey's RESET test using the square of the fitted values
   C:Based on a test of skewness and kurtosis of residuals
   D:Based on the regression of squared residuals on squared fitted values
```

Solution

$$\widehat{math10} = -20.36 + 6.23 \log(expend) - .305\, lnchprg$$

$$n = 408, \quad R^2 = .180.$$

The signs of the estimated slopes imply that more spending increases the pass rate (holding *lnchprg* fixed) and a higher poverty rate (proxied well by *lnchprg*) decreases the pass rate (holding spending fixed). These are what we expect.

Tasks

- What do you make of the intercept?
- Does it make sense to set the two variables to zero?

Solution

- As usual, the estimated intercept is the predicted value of the dependent variable when all regressors are set to zero.
- Setting *lnchprg* = 0 makes sense, as there are schools with low poverty rates. Setting log(*expend*) = 0 does not make sense, because it is the same as setting *expend* = 1, and spending is measured in dollars per student. Presumably this is well outside any sensible range. Not surprisingly, the prediction of a pass rate is nonsensical.

Tasks

- Run the simple regression of *math10* on *log(expend)*? Compare the slope coefficients with the previous result!

Tasks

```
                         Ordinary Least Squares Estimation
*******************************************************************************
 Dependent variable is MATH10
 408 observations used for estimation from    1 to  408
*******************************************************************************
 Regressor              Coefficient       Standard Error         T-Ratio[Prob]
 C                        -69.3411            26.5301            -2.6137[.009]
 LEXPEND                   11.1644             3.1690             3.5230[.000]
*******************************************************************************
 R-Squared                     .029663   R-Bar-Squared                  .027273
 S.E. of Regression            10.3495   F-stat.    F(  1, 406)   12.4115[.000]
 Mean of Dependent Variable    24.1069   S.D. of Dependent Variable     10.4936
 Residual Sum of Squares       43487.8   Equation Log-likelihood        -1531.4
 Akaike Info. Criterion        -1533.4   Schwarz Bayesian Criterion     -1537.4
 DW-statistic                   1.6146
*******************************************************************************


                                 Diagnostic Tests
*******************************************************************************
*    Test Statistics  *        LM Version        *         F Version         *
*******************************************************************************
*                     *                          *                           *
* A:Serial Correlation*CHSQ(   1)=  13.0256[.000]*F(   1, 405)=  13.3563[.000]*
*                     *                          *                           *
* B:Functional Form   *CHSQ(   1)=   4.5479[.033]*F(   1, 405)=   4.5653[.033]*
*                     *                          *                           *
* C:Normality         *CHSQ(   2)=  34.6088[.000]*       Not applicable      *
*                     *                          *                           *
* D:Heteroscedasticity*CHSQ(   1)=  17.5285[.000]*F(   1, 406)=  18.2256[.000]*
*******************************************************************************
   A:Lagrange multiplier test of residual serial correlation
   B:Ramsey's RESET test using the square of the fitted values
   C:Based on a test of skewness and kurtosis of residuals
   D:Based on the regression of squared residuals on squared fitted values
```

Solution

$$\widehat{math10} = -69.34 + 11.16 \log(expend)$$

n = 408, $R2$ = .030

- The estimated spending effect is larger than it was in part (i) – almost double.

Tasks

- Find the correlation between *lexpend* = log(*expend*) and *lnchprg.* Does the sign make sense to you? Relate to previous question!

- TYPE in command line :
  COR   LEXPEND LNCHPRG

Tasks

Sample period	:	1 to	408
Variable(s)	:	LEXPEND	LNCHPRG
Maximum	:	8.9118	79.5000
Minimum	:	8.1113	1.4000
Mean	:	8.3702	25.2015
Std. Deviation	:	.16188	13.6101
Skewness	:	1.0768	.68241
Kurtosis - 3	:	.66949	.45408
Coef of Variation	:	.019340	.54005

Estimated Correlation Matrix of Variables

****

	LEXPEND	LNCHPRG
LEXPEND	1.0000	-.19270
LNCHPRG	-.19270	1.0000

****

Solution

- The sample correlation between *lexpend* and *lnchprg* is about -0.19, which means that, on average, high schools with poorer students spent less per student. This makes sense, especially in 1993 in Michigan, where school funding was essentially determined by local property tax collections.

Solution

- Because Corr($x1$,$x2$) < 0, the simple regression estimate is larger than the multiple regression estimate.
- Intuitively, failing to account for the poverty rate leads to an overestimate of the effect of spending (positive omitted variable bias).