CERGE-EI
Summer 2014
Econometrics II
Instructors: Nikolas Mittag, Dragana Stanišič
TA: Jelena Plazonja, Gega Todua
Date: 05/05/2014

Exercise Session 2

Concepts

- *Partialling Out*

  Suppose we have a model:

$$y = \mathbf{X_1}\hat{\mathbf{b_1}} + \mathbf{X_2}\hat{\mathbf{b_2}} + \epsilon$$

  We can get $\hat{\mathbf{b_2}}$ by the following way:

  1) reg Y on $\mathbf{X_1}$ $\rightarrow$ predict residuals $\hat{\mathbf{U}}_\mathbf{i}$

  2) reg $\mathbf{X_2}$ on $\mathbf{X_1}$ $\rightarrow$ predict residuals $\hat{\eta}_\mathbf{i}$

  3) reg $\hat{\mathbf{U}}_\mathbf{i}$ on $\hat{\eta}_\mathbf{i}$: $\hat{\mathbf{U}}_\mathbf{i} = \hat{\eta}_\mathbf{i}\hat{b_2} + \alpha_i$

  Alternatively, Suppose we have a model:

$$Y_i = \beta_0 + \beta_1 X_1 + ...\beta_k X_k + \epsilon$$

  We can get a coefficient β_j by the following way:

  1) reg X_j on $X_1, X_2....X_k$ $\rightarrow$ predict residuals $\hat{v}_i$

  2) reg y on $\hat{v}_i$.

- *Omitted Variable Bias (OVB)*

  Suppose we have a model

$$Y_i = \beta_0 + \beta_1 X_1 + ...\beta_{k-1} X_{k-1} + \beta_k X_k + \epsilon$$

  and we omit X_k from the regression. The estimated coefficient will be:

$$\tilde{\beta_j} = \hat{\beta_j} + \hat{\beta_k}\tilde{\delta}_{kj}$$

  Where $\tilde{\beta_j}$ is a coefficient from $Y_i = \beta_0 + \beta_1 X_1 + ...\beta_{k-1} X_{k-1} + \epsilon$, $\hat{\beta_j}$ and $\hat{\beta_k}$ are coefficients from a true model $Y_i = \beta_0 + \beta_1 X_1 + ...\beta_{k-1} X_{k-1} + \beta_k X_k + \epsilon$, $\tilde{\delta}_{kj}$ is a slope coefficient from simple regression of X_j on $X_1, X_2....X_k$.

  Therefore, the OVB will be equal to $\hat{\beta_k}\tilde{\delta}_{kj}$ and the sign of it depend on the signs of both coefficients.

Exercises

1. What is the interpretation of $b > 0$ in the following regression models? Let's say x is monthly disposable income and y is monthly consumption measured in dollars. Derive.

   $y_i = a + bx_i + u_i$

   $y_i = a + bln(x_i) + u_i$

   $ln(y_i) = a + bx_i + u_i$

   $ln(y_i) = a + bln(x_i) + u_i$

2. **Linear transformation of the data**

   Consider the least squares regression of $\mathbf{Y}$ on K variables (with a constant) $\mathbf{X}$. Consider an alternative set of regressors $\mathbf{Z} = \mathbf{XP}$, where $\mathbf{P}$ is a nonsingular KxK matrix. Thus, each column of $\mathbf{Z}$ is a mixture of some of the columns $\mathbf{X}$. Prove, that the residual vectors in the regressions of $\mathbf{Y}$ on $\mathbf{X}$ and $\mathbf{Y}$ on $\mathbf{Z}$ are identical. What relevance does this have to the question of changing the fit of a regression by changing the units of measurement of the independent variables?

3. **Demeaning data** In the OLS regresion of $\mathbf{Y}$ on a constant and $\mathbf{X}$, will we get the same coefficients if we demean both $\mathbf{Y}$ and $\mathbf{X}$ and run regression of $\mathbf{Y}$ on $\mathbf{X}$ without constant? What if we only demean $\mathbf{X}$? If only $\mathbf{Y}$?

4. Suppose we have a model

   $$\mathbf{Y} = \beta_{\mathbf{1}}\mathbf{X_1} + \beta_{\mathbf{2}}\mathbf{X_2} + \epsilon$$

   and suppose, $\mathbf{b_1}$ is a coefficient vector after regressing of Y on only X_1. Show, that $E[\mathbf{b_1}|\mathbf{X}] = \beta_{\mathbf{1}} + \mathbf{P_{1,2}}\beta_{\mathbf{2}}$, where $\mathbf{P_{1,2}}$ is the column of slopes in the regression of the corresponding column of $\mathbf{X_2}$ on the columns of $\mathbf{X_1}$

5. Suppose that your data set consists of three observations of $(y, x) : (1, 1), (4, 2), (2, 3)$. Define a dummy variable D which is equal to 1 for $x > 3/2$ and zero otherwise. We would like to estimate the following regression equation, $y = A_0 + A_1D + e$

   (a) Calculate A_0 and A_1 using OLS.

   (b) Plot the three data points and your regression line.

   (c) Explain, in one or two sentences, what the coefficient of the dummy variable measures.

6. Suppose we have a model:

   $$Y_i = a + bD_{1i} + cD_{2i} + \epsilon$$

   where $D_{2i} = 1 - D_{1i}$

   Show, that the model suffers from multicolinearity.

7. Suppose we observe the following model:

$$log(wage_i) = \beta_0 + \beta_1 train_i + \beta_2 educ_i + \beta_3 exper_i + u_i$$

Where *train* is a binary variable equal to unity if a worker participated in the program.

(a) What is the interpretation of coefficients β_1 and β_2.

(b) Think of the error term u as containing unobserved worker ability. if less able workers have a greater chance of being selected from the program, and you use an OLS analysis, what can you say about the likely bias in the OLS estimator of β_1?