Econ C142 - Section 3

Ingrid Haegele and Pablo Muñoz

February 6, 2019

1 Frisch Waugh theorem[1]

Consider a population regression model in which an outcome y_i is related to two covariates x_{1i},x_{2i} as follows:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$$

1. State the Frisch Waugh theorem relating the population regression coefficient β_2 to a univariate regression model for y_i which does not include x_{1i}. *NOTE: do not prove FW. Just state it as carefully and as clearly as you can in this case.*

2. In the case where $x_{i1} = 1$ (i.e. the first regressor is a constant), *prove* that your answer in part 1. implies that:
$$\beta_2 = \frac{E\left[y_i(x_{2i} - \mu_2)\right]}{E\left[(x_{2i} - \mu_2)^2\right]}$$
where $\mu_2 = E[x_{2i}]$.

3. Assume (as above) that $x_{1i} = 1$ and suppose that $x_{2i} > 0$ (i.e. that x_{2i} is a random variable that only takes on positive values), and that $y_i = x_{2i}^{\rho}$. Find the values of β_2 for $\rho = 1$, $\rho = 0$, and $\rho = -1$. *Extra points*: prove that when $\rho = -1$, $\beta_2 < 0$. (Hint: Jensen's inequality).[2]

2 Ommited Variable Bias

The omitted variable bias formula is very important. We are going to work through part of the example from that lecture here. Suppose we want to run the following population regression.

$$y_i = \beta_1 + \beta_2 Im_i + \beta_3 Ed_i + u_i$$

Where y_i is log wage, Im_i is a dummy for whether a person is an immigrant or not and Ed_i is the years of education a person has attained. Suppose our research assistant has only collected data on log wages and immigration status for the population. We *do not* observe Education. Suppose we go ahead and run the population regression with the data we have.

$$y_i = \gamma_1 + \gamma_2 Im_i + \nu_i$$

Where the γ's are the coefficients from this *short* regression, and ν_i is the error.

1. Write an expression that relates the coefficient on immigrant status in the *short* regression of log wages on Immigrant status to the coefficient from the *long* regression of log wages on Immigrant status and education, and the coefficient from an auxiliary regression of education on immigrant status (call the coefficient from this regression π_1). Hint: write out the auxiliary regression in a way that lets you substitute in back into the long regression.

2. Here is a copy of the regression table from class with results from short, long, and auxiliary regressions. I am omitting standard errors from this table (we'll get to those next class). Confirm that your answer to the previous part is correct.

[1] This exercise comes from the 2015 Midterm.
[2] This hint was not included on the midterm exam.

Table 1: Ln Wages, Immigrant status, and Education

	$Wage$	$Wage$	Im_i	Ed_i	$Wage$
Im_i	-0.18			-1.49	-0.01
Ed_i		0.11	-0.03		0.114

3. Examine the expression you derived in 1. Under what two possible conditions is the coefficient γ_2 equal to the the coefficient on immigrant status in the long regression, β_2?

3 Population and Sample OLS

So far, we have studied the population regression:

$$y_i = x_i'\beta^* + u_i$$

Where y_i and u_i are scalars; x_i and β^* are a $K \times 1$ vectors.

1. Write out $x_i'\beta^*$, does this expression make sense?
2. Write out the FOC in the population regression.
3. Write out the FOC in a sample regression of size N.
4. Finally, write the OLS sample regression (size N) in matrix form and show that:

$$X'(Y - X\hat{\beta}) = \frac{1}{N}\sum_i^N x_i(y_i - x_i'\hat{\beta}) = 0$$

Where Y is a $N \times 1$ vector; X is a $N \times K$ matrix and $\hat{\beta}$ is a $K \times 1$ vector.