Pol Sci 630: Problem Set 10: Functional Form, Endogeneity, Power

Prepared by: Anh Le (anh.le@duke.edu)

Due Date: Nov 9 (Beginning of Class)

1 Functional specification

There's a famous dataset called the Anscombe quartet. You load it in R like so:

```
head(anscombe)

##   x1 x2 x3 x4   y1   y2    y3   y4
## 1 10 10 10  8 8.04 9.14  7.46 6.58
## 2  8  8  8  8 6.95 8.14  6.77 5.76
## 3 13 13 13  8 7.58 8.74 12.74 7.71
## 4  9  9  9  8 8.81 8.77  7.11 8.84
## 5 11 11 11  8 8.33 9.26  7.81 8.47
## 6 14 14 14  8 9.96 8.10  8.84 7.04
```

1.1 Explore Anscombe

There are 4 pairs of x and y. Run 4 regressions of y on x. Check out the regression result. A bit late for Halloween, but what spooky thing do you notice?

Then plot the data.

1.2 Ramsey RESET

Use Ramsey RESET on the 4 models. Which kind of functional misclassification can it catch? Can you think of why?

2 Endogeneity – Omitted Variable Bias

2.1 Sign the bias – math

Given the following data generating process (DGP)

Table 1: Signing the bias

	$\beta_2 > 0$	$\beta_2 < 0$
$\delta_1 > 0$		
$\delta_1 < 0$		

$$x_2 = \delta_0 + \delta_1 x_1 + v \tag{1}$$
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + w \tag{2}$$

The following equation (lecture 09/26) shows what happens when we regress y on x_1, omitting x_2. **Prove this equation**.

$$y = (\beta_0 + \beta_2 \delta_0) + (\beta_1 + \beta_2 \delta_1) x_1 + (\beta_2 v + w)$$

2.2 Sign the bias - simulation

In the equation you proved above, the estimated coefficient for x_1 is $\beta_1 + \beta_2\delta_1$, different from its true value β_1. The bias is $\beta_2\delta_1$. The sign of the bias thus depends on β_2 and δ_1, as discussed in the lecture and reproduced in Table 1.

Conduct 4 simulations with appropriate values of β_2 and δ_1 corresponding to the 4 cells in the table. Show that the sign of the bias is as we learned in class.

3 Power calculation

In this exercise we practice power calculation for the simplest experiment setup.

Assume that our binary treatment has an effect size of 2 on the outcome, as follows:

$$y = 1 + 2 \times \text{Treatment} + u$$
$$u \sim Normal(mean = 1, sd = 10)$$

In our experiment, we randomly assigned n experimental units into 2 groups, treated and control, i.e. treatment = 1 and treatment = 0. Calculate the power of our experiment (i.e. the probability that we can reject the null of zero treatment effect) for different values of the sample size n.

The end product I want to see is a graph with n on the x-axis and power on the y-axis. How big must your sample size be to get a power of 0.8?