

Econ 327 - Spring 2010
Introduction to Econometrics
Dr. R. Pete Parcells
Assignment Two
Due Tuesday March 2, 2010

1) Estimate the effects of smoking during pregnancy on infant health

(Econ folder contains bwght.txt and bwght.xls)

A problem of interest to health officials (and others) is to determine the effects of smoking during pregnancy on infant health. One measure of infant health is birth weight; a birth rate that is too low can put an infant at risk for contracting various illnesses. Since factors other than cigarette smoking that affect birth weight are likely to be correlated with smoking, we should take those factors into account. For example, higher income generally results in access to better prenatal care, as well as better nutrition for the mother. An equation that recognizes this is:

$$\text{bwght}_i = \beta_0 + \beta_1 \text{cigs}_i + \beta_2 \text{faminc}_i + e_i$$

with:

bwght_i = birth weight, ounces

cigs_i = cigs smoked per day while pregnant

faminc_i = 1988 family income in \$1000s

- (a) What do you think are the most likely signs for β_1 and β_2 ? Explain.
- (b) Do you think cigs and faminc are likely to be correlated? Explain why the correlation might be positive or negative.
- (c) Estimate the equation with and without faminc . Report the results in equation form, (write out the equation with the values for the estimated coefficients and their t-stats) including the sample size and R^2 . Discuss your results, focusing on whether adding faminc substantially changes the estimated effect of cigs on bwght .
- (d) Generate a dummy variable indicating whether the mother has been smoking while pregnant ($\text{dummy} = 1$) or not ($\text{dummy} = 0$). Then estimate the following equation:

$$\text{bwght}_i = \beta_0 + \beta_1 \text{dummy}_i + \beta_2 \text{faminc}_i + e_i$$

- (e) Compare the estimated β_1 with your results from (c). Why do the estimated coefficients differ from each other? Interpret them?

2) Mincerian Earnings Equation

(Econ folder contains **germansoep.xls**)

The dataset contains labor-market-related information on 974 full-time employees from the year 1996. The data is taken from the German Socio-Economic Panel Study (SOEP). It contains the following:

ID = identification number

GESCHL = gender (0=male, 1=female)

ALTER = age

BILDUNG = number of years of education

EINK = yearly gross wage

STUDEN = hours worked

SEKTOR = (1=primary, 2=secondary, 3=tertiary)

OEFFD= (0=private, 1=public)

BGROESSE= size of firm

SAMPLE = employee's origin (1=West German, 2= East German, 3=guest worker).

Analyze the factors influencing wages in Germany by estimating a Mincerian earnings function of the following form:

$$\ln(w_i) = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 E_i + \gamma Z_i + e_i$$

with:

w_i = wage rate

X_i = work experience

E_i = number of years of education

Z_i = other variables

- (a) Generate the logarithmic wage rate. Generate the potential work experience of the employees according to the following rule: potential work experience = age - number of years of education - 6. Run a regression of the logarithmic wage rate on work experience and squared work experience (no other variables). Display the predicted wage profile graphically. After how many years of work experience do you earn the highest wage rate?

- (b) Extend the model from (a) by adding a gender dummy and years of education. Estimate the new model. Do women, on average, earn less than men?
- (c) Calculate the average wage rates of men and women separately. Do they differ? Do women with similar educational background and similar potential work experience earn less than men? To answer this question, run estimations for women and men separately. Is gender discrimination the cause of the results? Can you think of other possible explanations?
- (d) Calculate the rate of return to education for West Germans and guest workers. Is the discrepancy statistically different from zero? (Hint: Look at the interaction effect.) Based on your results, would you say there is discrimination against guest workers on the German labor market?
- (e) Generate a dummy variable for the sector of the company. In which sector are the highest wage rates being paid? In which sector do full-time employees earn the lowest wage rates? Test the hypothesis that the dummy variables jointly do not have any influence on the wage rate.

**2) Discrimination against certain customer groups by fast food restaurants
(Econ folder contains discrim.txt and discrim.xls)**

Data are ZIP code-level data on prices for various items at fast-food restaurants, along with characteristics of the zip code population, in New Jersey and Pennsylvania. The idea is to see whether fast-food restaurants charge higher prices in areas with larger concentration of African Americans. The data set contains the following variables:

psoda = price of medium soda

prpbck = proportion African American, zip code

prppov = proportion in poverty, zip code

income = median family income, zip code

- (a) Find the average values of prpbck and income in the sample, along with their standard deviations. What are the units of measurement of prpbck and income?
- (b) Consider a model to explain the price of soda, psoda, in terms of the proportion of the population that is African American and median income:

$$psoda_i = \beta_0 + \beta_1 prpbck_i + \beta_2 income_i + e_i$$

Estimate this model and report the results in equation form, including the sample size and R^2 . Interpret the coefficient on prpbck. Do you think it is economically large?

- (c) Compare the estimate from (b) with the simple regression from psoda on prpbck. Is the

discrimination effect larger or smaller when you control for income?

- (d) A model with constant price elasticity with respect to income may be more appropriate. Estimate the model:

$$\ln(\text{psoda})_i = \beta_0 + \beta_1 \text{prpbck}_i + \beta_2 \ln(\text{income})_i + e_i$$

If prpbck increases by 0.20 (20 percentage points), what is the estimated change in psoda ?

- (e) Now add the variable prppov to the regression in (d). What happens to the coefficient on prpbck ?
- (f) Find the correlation between $\ln(\text{income})$ and prppov . Is it roughly what you expected?
- (g) Evaluate the following statement: "Because $\ln(\text{income})$ and prppov are so highly correlated, they have no business being in the same regression."