FINM3123 Introduction to Econometrics

Chapter 7 Exercises

1. The following equation were estimated using the data in BWGHT.RData:

$$\log(\widehat{bwght}) = \underset{(.22)}{4.66} - \underset{(.0009)}{.0044}cigs + \underset{(.0059)}{.0093}\log(faminc) + \underset{(.006)}{.016}parity$$
$$+\underset{(.010)}{.027}male + \underset{(.013)}{.055}white$$
$$n = 1{,}388,\ R^2 = .0472$$

and

$$\log(\widehat{bwght}) = \underset{(.38)}{4.65} - \underset{(.0010)}{.0052}cigs + \underset{(.0085)}{.0110}\log(faminc) + \underset{(.006)}{.017}parity$$
$$+\underset{(.011)}{.034}male + \underset{(.015)}{.045}white - \underset{(.0030)}{.0030}motheduc + \underset{(.0026)}{.0032}fatheduc$$
$$n = 1{,}191,\ R^2 = .0493$$

where

bwght	=	birth weight, in pounds.
cigs	=	average number of cigarettes the mother smoked per day during pregnancy.
parity	=	the birth order of this child.
faminc	=	annual family income.
motheduc	=	years of schooling for the mother.
fatheduc	=	years of schooling for the father.
male	=	1 if the child is male and 0 otherwise.
white	=	1 if the child is classified as white and 0 otherwise.

i) In the first equation, interpret the coefficient on the variable *cigs*. In particular, what is the effect on birth weight from smoking 10 more cigarettes per day?

ii) How much more is a white child predicted to weigh than a nonwhite child, holding the other factors in the first equation fixed? If the difference statistically significant?

iii) Comment on the estimated effect and statistical significance of *motheduc*.

iv) From the given information, why are you unable to compute the F statistic for joint significance of *motheduc* and *fatheduc*? What would you have to do to compute the F statistic?

2. In order to determine the effects of computer ownership on college grade point average, we estimate the model

$$colGPA = \beta_0 + \delta_0 PC + \beta_1 hsGPA + \beta_2 ACT + u,$$

where the dummy variable PC equals one if a student owns a personal computer and zero otherwise. The variables $hsGPA$ (high school GPA) and ACT (achievement test score) are used as controls. Let *noPC* be a dummy variable equal to one if the student does not own a PC, and zero otherwise.

i) If *noPC* is used in place of *PC* in the above equation, what happens to the intercept in the estimated equation? What will be the coefficient on *noPC*? (*Hint*: Write *PC* = 1 − *noPC* and plug this into the equation $\widehat{colGPA} = \widehat{\beta_0} + \widehat{\delta_0} PC + \widehat{\beta_1} hsGPA + \widehat{\beta_2} ACT$.)

ii) What will happen to the *R*-squared if *noPC* is used in place of *PC*?

iii) Should *PC* and *noPC* both be included as independent variables in the model? Explain.

3. Use the data in NBASAL.RData for this exercise.

i) Estimate a linear regression model relating points per game to experience in the league and position (guard, forward, or center). Include experience in quadratic form and use centers as the base group. Report the results in the usual form.

ii) Why do you not include all three position dummy variables in part (i)?

iii) Holding experience fixed, does a guard score more than a center? How much more? Is the difference statistically significant?

iv) Now, add marital status to the equation. Holding position and experience fixed, are married players more productive (based on points per game)?

v) Add interactions of marital status with both experience variables. In this expanded model, is there strong evidence that marital status affects points per game?

vi) Estimate the model from part (iv) but use assists per game as the dependent variable. Are there any notable differences from part (iv)? Discuss.

4. Hamermesh and Biddle (1994) used measures of physical attractiveness in a wage equation. (The file BEAUTY.RData contains a subset of the variables reported in the article.) Each person in the sample was ranked by an interviewer for physical attractiveness, using three

categories (average, below average, and above average, where the base group is *average*. Other factors considered in the analysis include education, experience, tenure, marital status, and race, etc.

i) Find the separate fractions of men and women that are classified as having above average looks. Are more people rated as having above average or below average looks?

ii) Test the null hypothesis that the population fractions of above-average-looking women and men are the same. Report the one-sided p-value that the fraction is higher for women. (*Hint*: Estimating a simple linear probability model is easiest.)

iii) Now estimate the model

$$\log(wage) = \beta_0 + \beta_1 belavg + \beta_2 abvavg + u$$

separately for men and women, and report the results in the usual form. In both cases, interpret the coefficient on *belavg*. Explain in words what the hypothesis H_0: $\beta_1 = 0$ against H_1: $\beta_1 < 0$ means, and find the p-values for men and women.

iv) Is there convincing evidence that women with above average looks earn more than women with average looks? Explain.

v) For both men and women, add the explanatory variables *educ, exper, exper*^{2}, *union, goodhlth, black, married, south, bigcity, smllcity,* and *service*. Do the effects of the "looks" variables change in important ways?

vi) Use the SSR form of the Chow F statistic to test whether the slopes of the regression functions in part (v) differ across men and women. Be sure to allow for an intercept shift under the null.