FINM3123 Introduction to Econometrics

Chapter 7 Exercises

Solutions

1.

(i) If $\Delta cigs$ = 10 then $\Delta\log\widehat{(bwght)}$ = –.0044(10) = –.044, which means about a 4.4% lower birth weight.

(ii) A white child is estimated to weigh about 5.5% more, other factors in the first equation fixed. Further, $t_{white} \approx 4.23$, which is well above any commonly used critical value. Thus, the difference between white and nonwhite babies is also statistically significant.

(iii) If the mother has one more year of education, the child's birth weight is estimated to be .3% less. This is not a huge effect, and the *t* statistic is only one, so it is not statistically significant.

(iv) The two regressions use different sets of observations. The second regression uses fewer observations because *motheduc* or *fatheduc* are missing for some observations. We would have to reestimate the first equation (and obtain the *R*-squared) using the same observations used to estimate the second equation.

2.

(i) Following the hint, $\widehat{colGPA} = \widehat{\beta_0} + \widehat{\delta_0}(1 - noPC) + \widehat{\beta_1}hsGPA + \widehat{\beta_2}ACT = \left(\widehat{\beta_0} + \widehat{\delta_0}\right) - \widehat{\delta_0}noPC + \widehat{\beta_1}hsGPA + \widehat{\beta_2}ACT.$ For the specific estimates in equation (7.6), $\widehat{\beta_0}$ = 1.26 and $\widehat{\delta_0}$ = .157, so the new intercept is 1.26 + .157 = 1.417. The coefficient on *noPC* is –.157.

(ii) Nothing happens to the *R*-squared. Using *noPC* in place of *PC* is simply a different way of including the same information on *PC* ownership.

(iii) It makes no sense to include both dummy variables in the regression: we cannot hold *noPC* fixed while changing *PC*. We have only two groups based on *PC* ownership so, in addition to the overall intercept, we need only to include one dummy variable. If we try to include both along with an intercept we have perfect multicollinearity (the dummy variable trap).

3.

(i) The estimated equation is

$$\widehat{points} = \underset{(1.18)}{4.76} + \underset{(.33)}{1.28}\, exper - \underset{(.024)}{.072}\, exper^2 + \underset{(1.00)}{2.31}\, guard + \underset{(1.00)}{1.54}\, forward$$

$n = 269, \quad R^2 = .091, \quad \bar{R}^2 = .077.$

(ii) Including all three position dummy variables would be redundant, and result in the dummy variable trap. Each player falls into one of the three categories, and the overall intercept is the intercept for centers.

(iii) A guard is estimated to score about 2.3 points more per game, holding experience fixed. The t statistic is 2.31, so the difference is statistically different from zero at the 5% level, against a two-sided alternative.

(iv) When *marr* is added to the regression, its coefficient is about .584 (se = .740). Therefore, a married player is estimated to score just over half a point more per game (experience and position held fixed), but the estimate is not statistically different from zero (p-value = .43). So, based on points per game, we cannot conclude married players are more productive.

(v) Adding the terms $marr \cdot exper$ and $marr \cdot exper^2$ leads to complicated signs on the three terms involving *marr*. The F test for their joint significance, with 3 and 261 *df*, gives $F = 1.44$ and p-value = .23. Therefore, there is not very strong evidence that marital status has any partial effect on points scored.

(vi) If in the regression from part (iv) we use *assists* as the dependent variable, the coefficient on *marr* becomes .322 (se = .222). Therefore, holding experience and position fixed, a married man has almost one-third more assist per game. The p-value against a two-sided alternative is about .15, which is stronger, but not overwhelming, evidence that married men are more productive when it comes to assists.

4.

(i) For women, the fraction rated as having above average looks is about .33; for men, it is .29. The proportion of women rated as having below average looks is only .135; for men, it is even lower at about .117.

(ii) The difference is about .04, that is, the percent rated as having above average looks is about four percentage points higher for women than men. A simple way to test whether the difference is statistically significant is to run a simple regression of *abvavg* on *female* and do a t test (which is asymptotically valid). The t statistic is about 1.48 with two-sided p-value = .14. Therefore, there is not strong evidence against the null that the population fractions are the same, but there is some evidence.

(iii) The regression for men is

$$\widehat{\log(wage)} = \underset{(0.024)}{1.884} - \underset{(.060)}{.199}\, belavg - \underset{(.042)}{.044}\, abvavg$$

$n = 824 \quad R^2 = .013$

and the regression for women is

$$\widehat{\log(wage)} = \underset{(0.034)}{1.309} - \underset{(.076)}{.138}\, belavg + \underset{(.055)}{.034}\, abvavg$$

$n = 436 \quad R^2 = .011.$

Using the standard approximation, a man with below average looks earns almost 20% less than a man of average looks, and a woman with below average looks earns about 13.8% less than a woman with average looks. (The more accurate estimates are about 18% and 12.9%, respectively.) The null hypothesis H_0: $\beta_1 = 0$ against H_1: $\beta_1 < 0$ means that the null is that people with below average looks earn the same, on average, as people with average looks; the alternative is that people with below average looks earn less than people with average looks (in the population). The one-sided p-value for men is .0005 and for women it is .036. We reject H_0 more strongly for men because the estimate is larger in magnitude and the estimate has less sampling variation (as measured by the standard error).

(iv) Women with above average looks are estimated to earn about 3.4% more, on average, than women with average looks. But the one-sided p-value is .272, and this provides very little evidence against H_0: $\beta_2 = 0$.

(v) Given the number of added controls, with many of them very statistically significant, the coefficients on the looks variables do not change by much. For men, the coefficient on *belavg* becomes −.143 (t = −2.80) and the coefficient on *abvavg* becomes −.001 (t = −.03). For women, the changes in magnitude are similar: the coefficient on *belavg* becomes −.115 (t = −1.75) and the coefficient on *abvavg* becomes .058 (t = 1.18). In both cases, the estimates on *belavg* move closer to zero but are still reasonably large.

(vi) The SSR for women is 83.6933, for men it is 166.0841, and so the unrestricted SSR is 249.7774 (rounded to four decimal places). The SSR obtained by adding the dummy variable *female* to the regression in part (v) is 261.1771. Therefore, the F statistic (Chow statistic) is

$$F = \frac{(261.1771 - 249.7774)}{249.7774} \cdot \frac{(1260 - 2 \cdot 14)}{13} \approx 4.33$$

With 13 and 1,232 *df* the p-value is essentially zero. So we strongly reject the common slopes formulation even if we allow for a different intercept. A look at the estimated regression functions for men and women shows that some of the slopes are quite different.