ECON3133 Introductory Econometrics

Chapter 6 Exercises

Solutions

1.

(i) The answer is not entirely obvious, but one must properly interpret the coefficient on *alcohol* in either case. If we include *attend*, then we are measuring the effect of alcohol consumption on college GPA, holding attendance fixed. Because attendance is likely to be an important mechanism through which drinking affects performance, we probably do not want to hold it fixed in the analysis. If we do include *attend*, then we interpret the estimate of $\beta_{alcohol}$ as measuring those effects on *colGPA* that are not due to attending class. (For example, we could be measuring the effects that drinking alcohol has on study time.) To get a total effect of alcohol consumption, we would leave *attend* out.

(ii) We would want to include *SAT* and *hsGPA* as controls, as these measure student abilities and motivation. Drinking behavior in college could be correlated with one's performance in high school and on standardized tests. Other factors, such as family background, would also be good controls.

2.

(a) \$23,400 (recall that *Price* is measured in \$1000s).
(b) In this case $\Delta BDR = 1$ and $\Delta Hsize = 100$. The resulting expected change in price is 23.4 + 0.156 × 100 = 39.0 thousand dollars or \$39,000.
(c) The loss is \$48,800.
(d) From the text $\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1-R^2)$, so $R^2 = 1 - \frac{n-k-1}{n-1}(1-\bar{R}^2)$, thus, $R^2 = 0.727$.

3.

(a) The estimated equation is

$$\underset{}{\widehat{points}} = \underset{(6.99)}{35.22} + \underset{(.405)}{2.364}\, exper - \underset{(.0235)}{.0770}\, exper^2 - \underset{(.295)}{1.074}\, age - \underset{(.451)}{1.286}\, coll$$

$n = 269$, $R^2 = .141$, $\bar{R}^2 = .128$.

(b) The turnaround point is 2.364/[2(.0770)] ≈ 15.35. So, the increase from 15 to 16 years of experience would actually reduce salary. This is a very high level of experience, and we can essentially ignore this prediction: only two players in the sample of 269 have more than 15 years of experience.

(c) Many of the most promising players leave college early, or, in some cases, forego college altogether, to play in the NBA. These top players command the highest salaries. It is not more college that hurts salary, but less college is indicative of super-star potential.

(d) When age^2 is added to the regression from part (i), its coefficient is .0536 (se = .0492). Its t statistic is barely above one, so we are justified in dropping it. The coefficient on *age* in the same regression is –3.984 (se = 2.689). Together, these estimates imply a negative, increasing, return to *age*. The turning point is roughly at 37 years old. In any case, the linear function of *age* seems sufficient.

(e) The OLS results are

$$\widehat{\log(wage)} = \underset{(.85)}{6.78} + \underset{(.007)}{.078}\, points + \underset{(.050)}{.218}\, exper - \underset{(.0028)}{.0071}\, exper^2 - \underset{(.035)}{.048}\, age - \underset{(.053)}{.040}\, coll$$

$n = 269$, $R^2 = .488$, $\bar{R}^2 = .478$

(f) The joint F statistic produced by Stata is about 1.19. With 2 and 263 df, this gives a p-value of roughly .31. Therefore, once scoring and years played are controlled for, there is no evidence for wage differentials depending on age or years played in college.

4.

(a) The youngest age is 25, and there are 99 people of this age in the sample with $fsize = 1$.

(b) One literal interpretation is that β_2 is the increase in *nettfa* when *age* increases by one year, holding fixed *inc* and age^2. Of course, it makes no sense to change *age* while keeping age^2 fixed. Alternatively, because $\partial nettfa / \partial age = \beta_2 + 2\beta_3 age$, β_2 is the approximate increase in *nettfa* when *age* increases from zero to one. But in this application, the partial effect starting at $age = 0$ is not interesting; the sample represents single people at least 25 years old.

(c) The OLS estimates are

$$\widehat{nettfa} = \underset{(15.28)}{-1.20} + \underset{(.060)}{.825}\, inc - \underset{(0.767)}{1.322}\, age + \underset{(.0090)}{.0256}\, age^2$$

$n = 2{,}017$, $R^2 = .1229$, $\bar{R}^2 = .1216$

Initially, the negative coefficient on *age* may seem counterintuitive. The estimated relationship is a U-shape, but, to make sense of it, we need to find the turning point in the quadratic. From equation (6.13), the estimated turning point is $1.322/[2(.0256)] \approx 25.8$. Interestingly, this is very close to the youngest age in the sample. In other words, starting at roughly *age* = 25, the relationship between *nettfa* and *age* is positive – as we might expect. So, in this case, the negative coefficient on *age* makes sense when we compute the partial effect.

(d) I follow the hint, form the new regressor $(age - 25)^2$, and run the regression *nettfa* on *inc*, *age*, and $(age - 25)^2$. This changes the intercept (which we are not concerned with, anyway) and the coefficient on *age*, which is simply $\beta_2 + 2\beta_3(25)$ – the partial effect evaluated at *age* = 25. The results are

$$\widehat{nettfa} = \underset{(9.97)}{-17.20} + \underset{(.060)}{.825}\, inc - \underset{(.325)}{.0437}\, age + \underset{(.0090)}{.0256}\, (age - 25)^2$$

$n = 2{,}017$, $R^2 = .1229$, $\bar{R}^2 = .1216$

Therefore, the estimated partial effect starting at *age* = 25 is only –.044 and very statistically insignificant ($t = -.13$). The two-sided *p*-value is about .89.

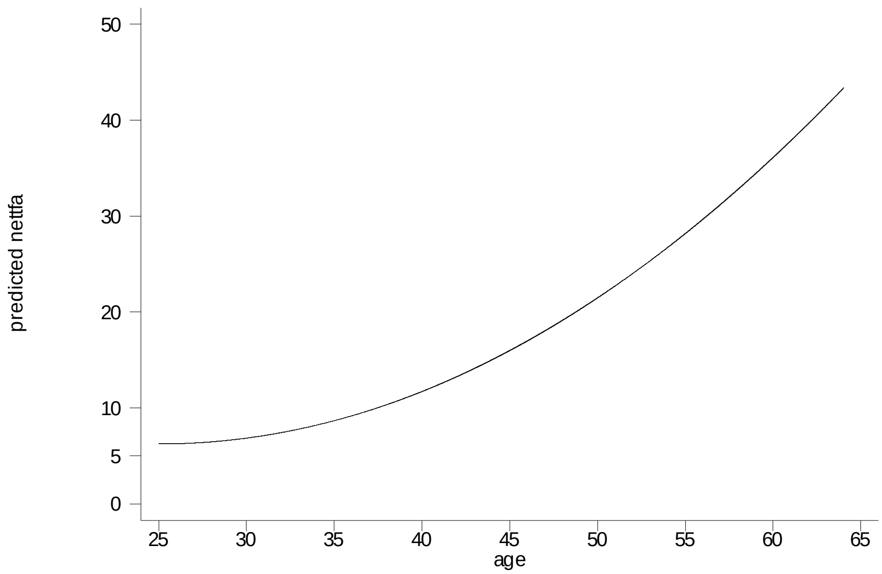
(e) If we drop *age* from the regression in part (iv) we get

$$\widehat{nettfa} = \underset{(2.18)}{-18.49} + \underset{(.060)}{.824}\, inc + \underset{(.0025)}{.0244}\, (age - 25)^2$$

$n = 2{,}017$, $R^2 = .1229$, $\bar{R}^2 = .1220$

The adjusted *R*-squared is slightly higher when we drop *age*. But the real reason for dropping *age* is that its *t* statistic is quite small, and the model without it has a straightforward interpretation.

(f) The graph of the relationship estimated in (v), with *inc* = 30, is



The slope of the relationship between $\widehat{nettfa}$ and *age* is clearly increasing. That is, there is an increasing marginal effect. The model is constructed so that the slope is zero at *age* = 25; from there, the slope increases.

(g) When inc^2 is added to the regression in part (v) its coefficient is only −.00054 with $t = -0.27$. Thus, the linear relationship between *nettfa* and *inc* is not rejected, and we would exclude the squared income term.