Economics 471: Econometrics

Department of Economics, Finance and Legal Studies
University of Alabama

Problem Set #5 – Answers

1. (a)

$$\begin{aligned} prate &= 80.29 + 5.44 mrate + 0.269 age - 0.00013 totemp \\ prate &= 97.32 + 5.02 mrate + 0.314 age - 2.26 \ln(totemp) \\ prate &= 80.62 + 5.34 mrate + 0.290 age - 0.00043 totemp + 0.0000000039 totemp^2 \end{aligned}$$

   (b) The second equation is clearly preferred, as its adjusted R-squared (0.142) is notably larger than that in the other two equations (0.098 and 0.106). The second equation contains the same number of estimated parameters as the first, and the one fewer than the third. The second equation is also easier to interpret than the third.

2. (a) The estimated equation is

$$\widehat{sat} = 997.98 + 19.81 hsize - 2.13 hsize^2$$

   The quadratic term is statistically significant, with t statistic –3.87.

   (b) We want the value of *hsize*, say $hsize^*$, where reaches its maximum. This is the turning point in the parabola, which we calculate as $hsize^* = 19.81/[2(2.13)] = 4.65$. Since *hsize* is in 100s, this means 465 students is the "optimal" class size. Of course, the very small R-squared shows that class size explains only a tiny amount of the variation in SAT score.

   (c) Only students who actually take the SAT exam appear in the sample, so it is not representative of all high school seniors. If the population of interest is all high school seniors, we need a random sample of such students who all took the same standardized exam.

   (d) With $\ln(sat)$ as the dependent variable we get

$$\widehat{\ln(sat)} = 6.896 + 0.0196 hsize - 0.00209 hsize^2$$

   The optimal class size is now estimated as about 469, which is very close to what we obtained with the level-level model.