

Homework 2

Jacob Puthipiroj

February 12, 2020

1 Heterogeneity in Returns to Schooling

Koop and Tobias [1] studied schooling in the German education system in order to (provide summary)

Let X_1 be a vector of education, experience, and ability (the individual's own characteristics). Let X_2 contain the mother's education, the father's education, and the number of siblings (the household characteristics). Let y be the log wage.

- a. Compute the least squares regression coefficients in the regression of y on X_1 . Report and interpret the coefficients.
- b. Compute the least squares regression coefficients in the regression of y on X_1 and X_2 . Report and interpret the coefficients.
- c. Compute the R^2 for the regression of y on X_1 and X_2 manually using the *SSE* and *SST* from the output. Repeat the computation for the case in which the constant term is omitted. You need use the *noconstant* option, which suppresses the constant in a regression model. What happens to R^2 ?
- d. Compute the adjusted R^2 for the full regression with and without the constant term. Interpret your results. Do we need the constant term? (Hint: Make sure to refer to the economic theory to discuss whether one should have the constant term regardless of statistical significance)
- e. Are any of the classical assumptions violated in part a or part b? Refer to the assumptions MR1, MR2, MR5, and MR6.

2 The U.S. Gasoline Market

- a. Compute the multiple regression of per capita consumption of gasoline on per capita income, the price of gasoline, all the other prices and a time trend. Report all results. Do the signs of the estimates agree with your expectations?
- b. Test the hypothesis that at least in regard to demand for gasoline, consumers do not differentiate between changes in the prices of new and used cars.
- c. Estimate the own price elasticity of demand, the income elasticity, and the cross-price elasticity with respect to changes in the price of public transportation. Do the computations at the 2004 point in the data, which means that the covariates should take the values corresponding to 2004 (i.e., use the "if" command instead of "at").
- d. Reestimate the regression in logarithms so that the coefficients are direct estimates of the elasticities. (Do not use the log of the time trend). How do your estimates compare with the results in the previous question? Which specification do you prefer?
- e. Compute the simple correlations of the price variables. Would you conclude that multicollinearity is a "problem" for the regression in part a or part d?
- f. Notice that the price index for gasoline is normalized to 100 in 2000, whereas the other price indices are anchored at 1983 (roughly). If you were to renormalize the indices so that they were all 100.00 in 2004, then how would the results of the regression in part a change? How would the results of the regression in part d change?

Word Count (excluding questions, tables, charts, graphs and appendix):

3 Appendix

References

- [1] Gary Koop and Justin L Tobias. Learning about heterogeneity in returns to schooling. *Journal of Applied Econometrics*, 19(7):827–849, 2004.