MPhil Econometrics
Exercises[1]

1) *Interpretation of OLS coefficient estimates in multiple regression*

Using the Stata dataset MRW.DTA:

i) Regress the log of real GDP per worker in 1985 on a constant and Mankiw, Romer and Weil's `https://www.jstor.org/stable/2118477` adjusted population growth rate variable, using the sample of non-oil countries. Save the residuals from this regression.

ii) Regress the log of investment as a share of GDP on a constant and Mankiw, Romer and Weil's adjusted population growth rate variable, using the sample of non-oil countries. Save the residuals from this regression.

iii) Now regress the residuals obtained in part (i) on the residuals obtained in part (ii). Compare the estimated coefficient and its standard error to the results obtained from the multiple regression of the log of real GDP per worker in 1985 on a constant, the log of investment as a share of GDP, and the adjusted population growth rate variable, using the sample of non-oil countries. Does it matter whether or not a constant is included in the residual regression?

2) *OLS regression with proxy variables*

The Stata dataset NLS80.DTA contains data on 935 working men aged between 28 and 38 who were part of the 1980 US National Longitudinal Survey of Young Men. This data was used to estimate earnings equations by M. Blackburn and D. Neumark, *Quarterly Journal of Economics*, 1992. The dataset includes a measure of the log of monthly earnings (lwage), and

[1]Originator: Steve Bond, Modified: Stefan Hubner

measures of years of education (educ), years of labour market experience (exper), years of tenure with the current employer (tenure), a dummy variable equal to one if the individual is married (married), a dummy variable equal to one if the individual is black (black), a dummy variable equal to one if the individual lives in the south (south), and a dummy variable equal to one if the individual lives in a standard metropolitan statistical area (urban).

i) Obtain the OLS estimates of a linear model that relates the dependent variable lwage to a constant and the explanatory variables exper, tenure, married, south, urban, black, and educ. This should replicate the OLS results reported on p65 of J. Wooldridge, *Econometric Analysis of Cross Section and Panel Data* (2002 edition).

ii) Assuming that the error term in this model satisfies conditional homoskedasticity (so that $\widehat{avar}(\widehat{\beta}_{OLS}) = \widehat{\sigma}^2(X'X)^{-1}$), use your results to test the null hypothesis that the coefficient on exper is equal to zero.

iii) Construct a new explanatory variable that is the square of years of labour market experience (i.e. exper^2). Add this to the model as an additional explanatory variable. Would you reject the null hypothesis that the coefficient on exper^2 is equal to zero? What about the coefficient on exper in this specification?

iv) Test the hypothesis that the coefficients on exper and exper^2 are both equal to zero. How can this be reconciled with the results of the individual hypothesis tests considered in part (iii)?

The dataset also includes two proxies for the individual's ability, an IQ test score (iq) and a knowledge of the world of work test score (kww).

v) Add the variable iq to the initial specification considered in part (i). This should replicate the OLS results reported on p65 of J. Wooldridge, *Econometric Analysis of Cross Section and Panel Data* (2002 edition). How does the inclusion of the iq variable affect the estimated coefficient on years of education?

vi) Add both the variables iq and kww to the initial specification. How does the further inclusion of the kww variable affect the estimated coefficient on years of education?

vii) Repeat this last specification using heteroskedasticity-consistent standard errors. Use these to test the null hypothesis that the coefficient on kww is equal to zero, against a two-sided alternative, at the 5% significance level. What is the marginal significance level of the test statistic used here?