Faculty of Economics

University of Cambridge

Lent 2019

Theory and Practice of Econometrics: II

Supervision Sheet 6

Dr. William Peterson

Question 1

A researcher has the following data for a sample of 1,498 females drawn from the United States National Longitudinal Survey of Youth: weight in kilos, height in centimetres, years of schooling, age, marital status in the form of a dummy variable *MARRIED* defined to be 1 if the respondent was married, 0 if single, and ethnicity in the form of a dummy variable *BLACK* defined to be 1 if the respondent was black, 0 otherwise.

These data were obtained for 1985 and 2000 for the same women. The respondents were aged 20–27 in 1985. Women who were divorced in either 1985 or 2000 were excluded from the sample.

The researcher fits two regressions:

(i) an ordinary least squares (OLS) regression combining the observations for 1985 and the observations for 2000 with weight as the dependent variable and years of schooling, *MARRIED*, height, age, and *BLACK* as explanatory variables

(ii) a first differences (FD) regression with the change in weight from 1985 to 2000 as the dependent variable and the change in years of schooling, the change in *MARRIED*, and the change in age (15 years for all respondents) over the same period as explanatory variables. The FD regression was fitted without a constant.

The results of these regressions are shown in Table 1. t statistics are given in parentheses.

a) Explain theoretically why OLS and FD regressions may yield different estimates of the parameters of the model.

b) The coefficients of schooling and MARRIED are negative and highly significant in the OLS regression, but near zero and not significant in the FD regression. Give an intuitive explanation of this. Explain why height and BLACK are excluded from the FD regression.

c) The change in age from 1985 to 2000 is the same for all respondents. Discuss the implications, if any, for the FD regression.

Table 1:	OLS	FD
Years of schooling	$-0.88(-7.41)$	$-0.06(-0.25)$
Married	$-3.27(-5.28)$	0.01(0.02)
Height (cm)	0.37(11.51)	–
Age	0.82(22.06)	0.72(28.26)
Black	6.12(7.43)	–
constant	-5.52	(-1.03) –
R^2	0.20	0.49
n	2,996	1,498

d) R^2 is much higher for the FD regression than for the OLS regression. Does this imply that the FD regression is a better specification?

e) Explain in principle how one might test whether individual-specific fixed effects jointly have significant explanatory power, if the number of individuals is small. (In this case, the number of individuals is large, and so the test is not practical.)

Question 2

Computer Exercise: Students should use `STATA` to answer this question.

Data is located on the course Moodle page under the name mom.dta.

Below we utilise the a dataset of 532 males for each of the 10 years from 1979 to 1988.

$$\ln hrs_{it} = \alpha_i + \beta \ln wg_{it} + \varepsilon_{it}$$

where ln hrs denotes the natural logarithm of annual hours worked;

ln wg denotes the natural logarithm of hourly wage.

α_i and β are unknown parameters

1. Estimate the equation by pooled OLS, and report the results. Are the usual standard errors reliable even if α_i is uncorrelated with all explanatory variables?

2. Estimate the wage equation by RE and FE and compare with the pooled OLS estimates

3. Briefly discuss the assumptions that underly the viability of standard errors for the OLS, FE and GLS estimators.

4. Briefly explain the distinction between robust and *panel* robust standard errors? Why might one utilise panel robust standard errors for the RE estimator?

Question 3

An econometrician interested in the effect of minimum wage on youth employment collects annual data on the panel of 48 mainland states of the US from 1976 to 2007. She estimates the following specification

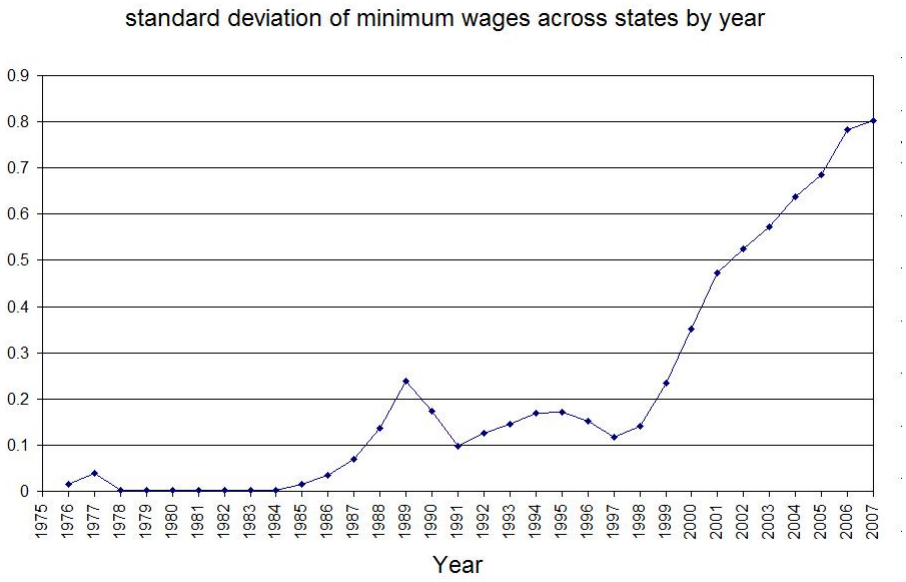
$$ln(empy_{it}) = \alpha_i + f_t + \gamma_i\, t + \beta ln(mw_{it}) + \varepsilon_{it}$$

where $empy_{it}$ is the youth employment-population ratio and mw_{it} the minimum wage in state i in year t, α_i and f_t are state- and time-specific fixed effects and t is a time trend.

She obtains the following results estimating this equation over different time periods

Sample Period	1976-1987	1976-1997	1976-2007
$\widehat{\beta}$	0.234	0.048	-0.134
s.e. $(\widehat{\beta})$	0.359	0.098	0.050
R^2	0.933	0.893	0.899

She also records the standard deviation across states of minimum wages over time as in the graph below



1. Assess the evidence for a negative effect of minimum wages on youth employment over these sample periods. To what extent does the graph help explain these findings?

2. What interpretation would you give to:

   (i) the individual fixed effects?

   (ii) the time fixed effects?

   (iii) the state specific time trend coefficients?

3. A colleague suggests that the regression is misspecified in that the explanatory variable should be real minimum wages (ie the right hand side should be $ln(mw) - ln(p)$ where p is a US-wide price index). Explain whether this criticism is valid.

4. Another colleague argues that minimum wages may be endogenous in that state authorities raise minimum wages when the economy is strong. What effect would this have on your interpretation of these regression results and how might any inferential problems be overcome?

5. A third colleague suggests that the effect of minimum wages on employment may be different in a recession. Explain how, given an indicator variable that is one if the US economy is in recession and zero otherwise, the above specification could be modified to investigate this hypothesis.

Question 4

Computer Exercise: Students should use `STATA` to answer this question.

Data is located on the course Moodle webpage under the name pums80m.dta.

The objective of this exercise is to evaluate the effects of having more than two children (`morekids`) on women's labour force participation. Utilising the data provided, answer the following questions.

1. Estimate the parameters of the following linear probability model (LPM)

   ```
   regress workedm morekids agem1 agefstm boy1st boy2nd blackm
                   hispm othracem educm, robust
   ```

   Variable definitions are given in Table 1.

   Compute the average partial effect (APE) and the partial effect at the mean (PEA) for the impact of `morekids` on labour force participation. Comment on your results

2. Using the same set of covariates, estimate a binary probit model and interpret your results. Can you compare the estimated parameters with those obtained in (i)? Compute the APE and PEA. Compare these results with those from the LPM.

Table 1: Definitions of Variables

Variable name	*Variable Definition*
`workedm`	`=1 if mother worked that year`
`morekids`	`=1 if mother had more than 2 kids`
`agem1`	`age of mother at census`
`agefstm`	`age of mother when she 1st gave birth`
`boy1st`	`=1 if 1st kid was a boy`
`boy2nd`	`=1 if 2nd kid was a boy`
`blackm`	`=1 if mother is black`
`hispm`	`=1 if mother is hispanic`
`othracem`	`=1 if mother is other race`
`educm`	`years of education of mother`
`samesex`	`=1 if 1st two kids same sex`

3. Briefly discuss why there may be one or more confounding effects which will generate a biased estimates of the impact of one more child.

   Using a linear model and the same set of covariates, extend the model in (i) to account for the potential endogeneity of `morekids`. Use the dummy variable `samesex`, which is equal to one if the 2 children are of the same sex.

   Provide a justification for this instrument and interpret your results.

4. Carefully explain what you understand by the following quote:

   > *If there are no covariates or the covariates are sparse and discrete, linear models and the associated estimation techniques like 2SLS are no less appropriate for LDVs than for other kinds of dependent variables*

   How does this quote relate to your findings above?

   See Joshua D. Angrist (2001) *Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice.* Journal of Business & Economic Statistics, Vol. 19, No. 1.

References

[1] Joshua D. Angrist and William N. Evans (1998) *Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size*. American Economic Review, Vol. 88, No. 3.

M. Weeks

Lent 2018