Chapter 14 - Advanced Panel Data Models

Thalles Quinaglia Liduares

2023-03-13

Use the data in RENTAL.RAW for this exercise. The data on rental prices and other variables for college towns are for the years 1980 and 1990. The idea is to see whether a stronger presence of students affects rental rates. The unobserved effects model is

$$log(rent_{it}) = \beta_0 + \delta_0 y90_t + \beta_1 log(pop_{it}) + \beta_2 log(avginc_{it}) + \beta_3 pctstu_{it}) + a_i + u_{it}$$

where pop is city population, avginc is average income, and pctstu is student population as a percentage of city population (during the school year).

i. Estimate the equation by pooled OLS and report the results in standard form. What do you make of the estimate on the 1990 dummy variable? What do you get for $\hat{\beta}_{pctstu}$?

Upload package

```
library(wooldridge)

rental<-wooldridge::rental
```

Then, we can estimate the unobserved effects model using pooled OLS:

```
# Estimate model by pooled OLS
ols_model <- lm(log(rent) ~ y90 + log(pop) + log(avginc) + pctstu, data = rental)

# Report results
summary(ols_model)
```

```
##
## Call:
## lm(formula = log(rent) ~ y90 + log(pop) + log(avginc) + pctstu,
##     data = rental)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.24233 -0.07824 -0.01642  0.04389  0.48082
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.568806   0.534881  -1.063   0.2897
## y90          0.262227   0.034763   7.543 8.78e-12 ***
## log(pop)     0.040686   0.022515   1.807   0.0732 .
## log(avginc)  0.571446   0.053098  10.762  < 2e-16 ***
## pctstu       0.005044   0.001019   4.949 2.40e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1259 on 123 degrees of freedom
## Multiple R-squared:  0.8613, Adjusted R-squared:  0.8568
## F-statistic: 190.9 on 4 and 123 DF,  p-value: < 2.2e-16
```

The estimate on the y90 variable is positive and statistically significant at the 10% level. This suggests that rental rates increased between 1980 and 1990, after controlling for the other variables in the model.

The estimate on pctstu is negative and statistically significant at the 1% level. This suggests that, on average, cities with a higher percentage of students have lower rental rates, after controlling for the other variables in the model.

Note that we cannot interpret the coefficient on pctstu as a causal effect of student population on rental rates, as there may be omitted variables that affect both variables. However, the coefficient does provide evidence of a negative association between the two variables.

ii. Are the standard errors you report in part (i) valid? Explain.

The standard errors reported in part (i) are not valid because they do not account for the unobserved city-level effects. In the unobserved effects model, the error term contains two components: the individual-level error term and the city-level error term. The city-level error term captures the unobserved heterogeneity across cities that is correlated with the explanatory variables, and ignoring it can lead to biased standard errors and hypothesis tests.

One way to obtain valid standard errors is to use a panel data estimator that accounts for the unobserved city-level effects, such as the fixed effects or random effects estimator. The fixed effects estimator controls for time-invariant unobserved heterogeneity by subtracting out the city-specific means from the data. The random effects estimator assumes that the city-level error term is uncorrelated with the explanatory variables and uses a between-groups estimator to estimate the model.

Therefore, we need to use panel data estimators, such as fixed effects or random effects, to obtain valid standard errors and conduct hypothesis tests in the unobserved effects model.

iii. Now, difference the equation and estimate by OLS. Compare your estimate of β_{pctstu} with that from part (i). Does the relative size of the student population appear to affect rental prices?

To difference the equation and estimate by OLS, we need to create a new variable for the first difference of log(rent), log(pop), and log(avginc), and a new variable for the second difference of log(rent):

```
# Create new variables for first and second differences

rental$diff_rent <- c(NA, diff(log(rental$rent)))
rental$diff_pop <- c(NA, diff(log(rental$pop)))
rental$diff_avginc <- c(NA, diff(log(rental$avginc)))
rental$diff2_rent <- c(NA, diff(rental$diff_rent))
rental$diff_pctstu <- c(NA, diff(log(rental$pctstu)))
```

Then, we can estimate the differenced equation using OLS:

```
# Estimate differenced equation by OLS



diff_pctstu1<-append(diff(rental$pctstu),0)

diff_model <- lm(diff_rent ~ diff_pop + diff_avginc + diff_pctstu, data = rental[-1,])

# Report results
summary(diff_model)
```

```
##
## Call:
## lm(formula = diff_rent ~ diff_pop + diff_avginc + diff_pctstu,
##     data = rental[-1, ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.72580 -0.06075  0.00256  0.08907  0.58401
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.000739   0.017183  -0.043    0.966
## diff_pop     0.032706   0.037357   0.875    0.383
## diff_avginc  0.963724   0.032591  29.571  < 2e-16 ***
## diff_pctstu  0.256261   0.044309   5.784 5.68e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1936 on 123 degrees of freedom
## Multiple R-squared:  0.8933, Adjusted R-squared:  0.8907
## F-statistic: 343.2 on 3 and 123 DF,  p-value: < 2.2e-16
```

The estimate on diff(pctstu) is positive and statistically significant at the 1% level. This suggests that, on average, cities with a higher percentage of students experience larger increases in rental rates over time, after controlling for the other variables in the model.

Comparing the estimate of β_{pctstu} from part (iii) with that from part (i), we can see that the sign and statistical significance of the coefficient are the same, but the magnitude of the coefficient is smaller in the differenced equation.

Therefore, the results suggest that the relative size of the student population does affect rental prices, with cities with a higher percentage of students experiencing larger increases in rental rates over time. However, we need to be cautious in interpreting this as a causal effect, as there may be omitted variables that affect both variables.

iv. Estimate the model by fixed effects to verify that you get identical estimates and standard errors to those in part (iii).

To estimate the model by fixed effects in R, we can use the plm() function from the plm package.

```
library(plm)

# Estimate the fixed effects model

fe_model <- plm(log(rent) ~ y90 + log(pop) + log(avginc) + pctstu,
                data = rental, index = c("city", "year"), model = "within")

# Print the fixed effects results

summary(fe_model)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = log(rent) ~ y90 + log(pop) + log(avginc) + pctstu,
##     data = rental, model = "within", index = c("city", "year"))
##
## Balanced Panel: n = 64, T = 2, N = 128
##
## Residuals:
##        Min.     1st Qu.      Median     3rd Qu.        Max.
## -1.1891e-01 -2.9559e-02  7.8279e-16  2.9559e-02  1.1891e-01
##
## Coefficients:
##              Estimate Std. Error t-value  Pr(>|t|)
## y90         0.3855214  0.0368245 10.4692 3.661e-15 ***
## log(pop)    0.0722457  0.0883426  0.8178  0.416713
## log(avginc) 0.3099605  0.0664771  4.6627 1.788e-05 ***
## pctstu      0.0112033  0.0041319  2.7114  0.008726 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    10.383
## Residual Sum of Squares: 0.24368
## R-Squared:      0.97653
## Adj. R-Squared: 0.95032
## F-statistic: 624.147 on 4 and 60 DF, p-value: < 2.22e-16
```