EBC4222
Descriptive and Predictive Analytics
Tutorial 6 Exercises: Panel Data Models

Roselinde Kessels
r.kessels@maastrichtuniversity.nl

28 May 2024

Questions

We will analyze the Employment and Wages in the United Kingdom, provided in the package `plm`. These data include 1031 observations from an unbalanced panel (*panel data with missing observations*) of 140 observations from 1976 to 1984 in the UK.
The dataset defines cross-sections as 'firms' and time series as 'years'.
Our purpose is to analyze the effect of wages on employment in a panel data setting. We suspect that the relation between wages and employment can depend on the firm and sector.

1. Load the employment data `EmplUK` from the *R* package `plm`. Visually analyze the employment levels across firms and across sectors. Do you think that employment levels are different between firms and sectors?
   Hint: Two useful functions for this are `coplot` in the package *graphics*, and `plotmeans` in the package *gplots*, but you are not restricted to these two functions.

2. First consider the pooled panel data estimation. In this case, we disregard the panel setting, and the model (and OLS estimates) does not consider heterogeneity across cross-sections (groups of firms) or time (years).
   Estimate a linear regression model explaining employment with wages and an intercept. Plot the data points together with the fitted OLS line. Does employment increase with wages according to the OLS results?

3. The dataset has 9 sectors. We suspect that the average employment levels are different across sectors. For this purpose, extend the linear regression model in part (2) with 9 dummy variables for each sector. Write down this model explicitly. Estimate

this model and comment on the least squares estimates. Do the sectors have different average wages according to your results?
Hint: When you add 9 dummy variables, you will need to remove the intercept from the model to avoid perfect multicollinearity. In addition, you do not need to manually create dummy variables, instead use `factor(sector)`.

4. Compare the models in parts (2) and (3) in terms of two information criteria, AIC and BIC. Which model is more appropriate for the data according to the information criteria?

5. Plot the data and the fitted OLS line for the model with dummy variables in part (3). Add the OLS line from part (2) in this plot. According to this figure, what was wrong in the OLS estimation in part (2)?
Hint: You can use `scatterplot` in the package **car**.

6. Based on the estimation results in part (3), calculate the predicted employment in a firm with average wage 23.9, in sector 3.

7. Estimate a *fixed effects model* explaining employment with wages. Print the summary of the estimated fixed effects. Are the fixed effects significantly different from zero?

8. Based on the estimation results in part (7), calculate the predicted employment in firm 140, with average wage 23.9. Compare your result to part (6).

9. Estimate a *random effects model* explaining employment with wages. Comment on the estimates of this model.

10. The random effects model has the additional exogeneity assumption compared to the fixed effects model. We want to test whether this assumption is violated. For this, use a Hausman test (`phtest` in the **plm** package). The null and alternative hypotheses of this test are:

    - H_0 : Both fixed and random effects models can be used (the exogeneity assumption in the random effects model is not violated).
    - H_A : One of the models is inconsistent (the random effects model is inconsistent, since the fixed effects model does not have the exogeneity assumption).

    What is your conclusion about the use of the random effects model for these data?