FCT Universidade de Coimbra Departamento de Engenharia Civil

Instituto Superior Técnico Departamento de Engenharia Civil e Arquitetura

Faculdade de Engenharia da Universidade do Porto

Course: Transport Demand Modelling

2020/2021

Home Assignment: (A) Panel Data Models and (B) Spatial Regression

Anabela Ribeiro

Due 6, January, 2021 (24hours)

(A) Panel Data Models

1. Objectives

The objective of this HA is to expand your ability to model real phenomena using Panel Data models over continuous measurement variables, normally estimated with simple multiple linear regression models. You should assess if your model calibrated for the Linear Regression assignment would change by introducing some of the four specifications studied, comparing the final results. You must adequate the analysis to be made with the possibilities you have through NLOGIT.

2. Software

The reference software is Excel and NLOGIT (LIMDEP).

3. Data

The data set you have to analyse in NLOGIT is the '**SwissRailways**' file.

4. Your tasks

1. To do this project, you should discuss how Total Cost (TOTCOST) in Swiss Railways (for 49 companies and during a period of 12 years) is related with number of stops ('STOPS'), length of the railway network (m) ('NETWORK'), labour expenses in 1000 CHF ('LABOREXP'), number of employees ('STAFF'), electricity expenses in 1000 CHF ('ELECEXP'), and total consumed electricity (in kWh). You should also identify if there are significant differences across state

and along the years (panel data) using one way and two way fixed models estimation. Finally discuss the use of random effects in this sample. Steps:

a. Develop the usual process as in class, namely by:

- Estimate the initial linear model that reflects the initial hypothesis;
- Reformulate the model avoiding non-significant variables;
- Test different specifications and perform statistical tests comparing the specifications, according to the exercises done in classes;
- Build a table where you can put all the values that matter for this comparison, and report the differences between the values obtained among the different specifications;
- Select the best model specification.

b. Concerning point a. results, discuss extensively the results on the tests about the use of fixed effects over normal regression and about fixed effects over random effects.

Note: Include the steps developed and the description of the test statistics that lead to your decisions. Present the outputs, tables to compare results, and comments on that.

5. Report content

Your final report should include:

1. Description of the four model specifications learned in class, fixed and random effects, one-way and two-way error estimation, plus the initial model;
2. Description and discussion of the effects introduced by a panel data formulation considering the different possibilities;
3. Presentation of the best specification after comparing the results. This presentation must be sustained by all the available test statistics in NLOGIT.
4. Besides Panel Data Models specifications, capacity of discussion of the results namely by: a. Identifying relevant variables; b. The use of random effects model in this sample.

(B) Spatial Data Models

1. Objectives

Autocorrelation studies and spatial regression analysis on the data base referring to the use of Bike Sharing Stations in Boston. The dependent variable is D_TTRIP201. The explanatory variables are the other variables presented in Table 1.

2. Software

The reference software is GeoDa.

3. Data

There are two data bases. One with the variables as they were collected/measure (BGRIdatabase), and other with the variables linearized (BGRIdatabaseLN), using log. Variables are as follow on Table 1.

Table 1 – Some variables from the Boston BS System.

Variable	Definition
D_TTRIP201	Total number of trips in 2013 – Destination at each station
OPDays2013	Number of operated days in 2013 at each station
NEmployers	Employers
BusStops	Bus stops
MBTApaxE	Number of entrances at each the MBTA station
Pop2013	Total population
Male2013	Total of male population
P15_24Y201	Total population aged between 15 to 24 years
P25_59Y201	Total population aged between 25 to 59 years
TFam2013	Total of families
TW10_14	Workers that take between 10 and 14 minutes to go to work
TW20_24	Workers that take between 20 and 24 minutes to go to work
TW25_29	Workers that take between 25 and 29 minutes to go to work
TWM60	Workers that take more than 60 minutes to go to work
P18YHSh	Population 18 years and over with high school graduate or higher
P18YBDh	Population 18 years and over with bachelor's degree or higher

4. Your tasks

With the support of GeoDa, develop the following analysis:

1. Using the both databases built LISA maps (Local Indicators of Spatial Association) with Moran Scatterplot and Significance Maps, for the variables DTTRIP201 and TW10_14 in BGRIdatabase. Are there any evidences of spatial clusters and/or outliers for the two variables? Of what type? Justify. *Note: Use the regular rook criteria to build your neighbour matrix.*

2. Using as **dependent variable D_TTRIP201, and as independent variables the rest present in the table 1**, identify if there is some evidence for Spatial Regression specifications, following the process explained in class.

     a. For data base 'BGRIdatabase':

- Estimate OLS, and present and <u>comment all the results obtained for the identification of spatial regression</u>.
- Take the non-significant variables out and develop the same process;
- Compare the results obtained.

     b. For the data base 'BGRIdatabaseLN':

- Estimate OLS, present and <u>comment all the results obtained for the identification of spatial regression</u>.
- Take the non-significant variables out and develop the same process.
- Compare the results obtained.

3. Build a table with all the final results and identify the best fitted model. Estimate that model if it is different than OLS. Answer to the following issues:

- Identify the problems that arise during estimation;
- Justify your selection of 'best model' and give also some insights on why were these variables the best ones?

5. Report content

Your final report should include:

1. You should present the graphic and numerical elements that show autocorrelation evidence (or not) for the variables suggested;
2. All the steps developed and the description of the test statistics that lead to your decisions e spatial regression analysis. Present the outputs, tables to compare results, and comments on that.

(C) Some Comments

1. The following criteria will be applied for grading:

a. Your understanding of the problem (e.g. causal relationships); b. Your understanding of the panel data/spatial data specification principles; c. Your utilization of the regression software (evidence you estimated a regression model, etc.); d. Your understanding of regression statistics and hypothesis testing (explain what the statistics mean).

2. Remember that you must always examine and comment on your results. Computer outputs without explanations are not acceptable.

3. There are no formatting rules except that you should write a concise report that isn't longer than 10 pages (approx.) without annexes (where you should include tables you might find useful to complete your report).