Pablo Estrada Cedeño

Guayaquil, Ecuador
Email: pabloestradace@gmail.com
Website: pabloestradac.github.io
GitHub: pabloestradac
Twitter: pabloestradac

# Project on Panel Data Spatial Econometrics

## Overview

*This project aims to extend the functionality of pysal.spreg to deal with panel data econometrics models. Spatial panels refer to data containing time series observations of a number of geographical units. Specifically, this project will focus on static panel models with fixed and random effects. Also, I will develop a framework to test the spatial lag, the spatial error model, and the spatial Durbin model against each other, as well as a framework to choose among fixed effects, random effects or a model without fixed/random effects. Thus, this project will handle panel data, estimate panel models and provide specification tests.*

## Introduction

Spatial econometrics deals with the estimation of spatial interaction effects among units. In general, the spatial econometrics literature uses a spatial weights matrix ***W*** to study the dependence among units across space (Elhorst 2014). Spatial panels refer to data containing time series observations of a number of geographical units.

According to Elhorst (2014), there are three generations of spatial econometrics models. The first generation consists of models based on cross-sectional data. The second is composed of static models of spatial panel data. And the third are dynamic spatial panel data models. For this

proposal, I will focus on the second generation. If time permits, tentatively I will proceed to develop functions that estimate dynamic panel data models (3rd generation).

Spatial effects for panel data can be considered in the form of spatial lag or spatial error models. As we know from the panel data econometrics literature, we need to include fixed and random effects to deal with heterogeneity. Elhorst (2014) argues that the spatial econometrics literature has focused on the random effects specifications rather than the fixed effects specification.

For this project I will focus on static panel models with fixed and random effects. Also, I will develop a framework to test the spatial lag, the spatial error model, and the spatial Durbin model against each other, as well as a framework to choose among fixed effects, random effects or a model without fixed/random effects. This framework follows Elhorst (2014):

1. Non-spatial model is estimated to test it against the spatial lag and the spatial error model (specific-to-general approach).
2. In case the nonspatial model is rejected, the spatial Durbin model is estimated to test whether it can be simplified to the spatial lag or the spatial error model (general-to-specific approach).
3. If both tests point to either the spatial lag or the spatial error model, it is safe to conclude that that model best describes the data. By contrast, if the non-spatial model is rejected in favor of the spatial lag or the spatial error model while the spatial Durbin model is not, one better adopts this more general model.

Thus, the goal of this project is to implement a module integrated in spreg from pysal.

## Deliverables

The deliverables will be divided in three parts: data handling, estimation and diagnostic functions.

- Module ***panel_utils*** for handling panel data. The function will receive 3D numpy arrays with dimension (n, 1, t) for dependent variables and dimension (n, k, t) for independent variables. Then, the arrays are converted to (nxt, 1) and (nxt, k) dimensions.

- Module ***panel_sp*** that performs the estimation of fixed effects and random effects for static panel models. Both random and fixed effects will have specifications for spatial error models and spatial lag models. In addition, they can include the option of spatial Durbin models where X = [X WX]. To estimate RE spatial error models I will follow KKP (2007). And for the rest I will follow Elhorst (2014) Sec. 3.3.
- Module ***diagnostics_panel*** that include tests for model comparison and selection such as Lagrange Multiplier test, Likelihood ratio test and R-squared.
- Documentation for the modules as well as Jupiter notebooks describing some applied examples.

## Timeline

| Dates | Tasks |
|---|---|
| Apr 27 - May 17 | Community Bonding. Discussion with Mentor and community about the best structure to adopt the new functionalities. |
| May 18 - May 31 | Write functions for handling panel data. |
| Jun 01 - Jun 21 | Write estimation function for panel models with fixed effects. |
| Jun 22 - Jul 12 | Write estimation function for panel models with random effects. |
| Jul 13 - Jul 26 | Write diagnostic tests for model comparison and selection. |
| Jul 27 - Ago 02 | Write code for unit tests |
| Ago 03 - Ago 09 | Write user guide and provide Jupyter notebooks with applications |
| Ago 10 - Ago 17 | Review and submission of the project |

## My background

My name is Pablo Estrada, and I am currently enrolled in the M.Sc. program of Economics at ESPOL University, Ecuador. As part of the Economics program, I have attended several Econometrics courses that include topics such as Panel Data Econometrics methods. In

addition to this, my formation includes two courses oriented to big data research which consist of a Data Mining and a Programming course. During my undergraduate journey, I have developed strong programming skills in general-purpose languages like Python and statistical languages like R, Stata, and Matlab to clean and prepare data, conduct experiments, and generate visualizations and tables for results.

As part of my research activities, I have experience working with Python and specifically developing models in Numpy. In one of my latest projects, I have worked with the Generalized 2SLS estimator, proposed by Lee (2003), for linear-in-means models that identifies peer effects on endogenous networks. Due to this project, I have worked with sparse matrices and Numpy linear algebra functions. Also, I have been exposed to the estimation procedures of linear-in-mean models for peer effects, that are closely related to spatial econometrics.

This project is an invaluable opportunity to work in spatial econometrics and to learn at a detailed level the estimation procedures of different spatial models in panel settings. I am confident I am an excellent candidate to participate in this project due to my coding experience and research interests.

**Link to pull request:** https://github.com/pysal/spreg/pull/38

## REFERENCES


Baltagi, BH., Song, SH., and Koh, W. (2003). Testing panel data regression models with spatial error correlation. Journal of Econometrics, 117: 123–150.

Lee LF. (2003). Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances. Econometric Reviews 22: 307–335.

Elhorst JP. (2014). Spatial Econometrics, From Cross-Sectional Data to Spatial Panels. Springer-Verlag Berlin Heidelberg.