Panel Data:
A Case Study Analysis

5

Vera Costa
 https://orcid.org/0000-0003-0801-2113
FEUP, University of Porto, Portugal

Rui Portocarrero Sarmento
 https://orcid.org/0000-0003-1891-5295
Statsframe, University of Porto, Portugal

INTRODUCTION

In Statistical Data Analysis, when analyzing a dataset containing variables observed through time, Panel Data regression analysis methods are commonly used. Panel data research, also associated with longitudinal or cross-sectional time-series data ($t=1,\ldots,T$). can be used in the study of varied types of entities, from companies or countries to individuals ($i=1,\ldots,N$). From the perspective of data structure, spatial panel data models are the combination of conventional cross-sectional and time series data models, as represented by (Zhou & Yamaguchi, 2018):

Figure 1. Structure of panel data models

(a) Spatial panel data models

i	t	y	x
1	1	$y_{1,1}$	$x_{1,1}$
1	2	$y_{1,2}$	$x_{1,2}$
1	3	$y_{1,3}$	$x_{1,3}$
2	1	$y_{2,1}$	$x_{2,1}$
2	2	$y_{2,2}$	$x_{2,2}$
2	3	$y_{2,3}$	$x_{2,3}$
3	1	$y_{3,1}$	$x_{3,1}$
3	2	$y_{3,2}$	$x_{3,2}$
3	3	$y_{3,3}$	$x_{3,3}$

(b) Cross-sectional data models

i	t	y	x
1	1	$y_{1,1}$	$x_{1,1}$
2	1	$y_{2,1}$	$x_{2,1}$
3	1	$y_{3,1}$	$x_{3,1}$

(c) Time series data models

i	t	y	x
1	1	$y_{1,1}$	$x_{1,1}$
1	2	$y_{1,2}$	$x_{1,2}$
1	3	$y_{1,3}$	$x_{1,3}$

Panel data research can provide means to control subjacent variables not observed or measured. Thus, it accounts for individual characteristics, for example, when studying the evolution of several communities or groups in social media through time, differences in behaviors across communities or variables that change over time but not across communities (i.e., global rules, agreements between communities or social media platforms rules)

This document is focused in two techniques used to analyze panel data:

- Fixed effects
- Random effects

Thus, the authors initiate the research by introducing the reader to the background state of the art regarding Panel Data. Then, the primary focus of the chapter considers the introduction to the case study data and obtained results. Results start with the model calibration for the Linear Regression. Additionally, the authors present and explain the case study results for four specifications, one-way or two-way, fixed or random effects, and compare the final results.

BACKGROUND

Historically, econometric and statistical models have been developed by using cross-sectional or time-series data (Washington, Karlaftis, & Mannering, 2003). However, in several cases, there is an availability of data based on cross-sections of individuals observed over time (or other observational units such as firms, geographic entities, and so on). Data which combines cross-sectional and time-series characteristics, can be called **panel data, pooled data** or **longitudinal data** (Dougherty, 2006).

Panel data can provide predictions on the evolution of a certain dependent variable according to other variables that are measured among distinct entities (cross-sectional) and time intervals (time series). Thus, it allows researchers to construct and test realistic behavioral models that cannot be identified using only cross-sectional or time-series data. Formally, a panel data has the following form (Kunst, 2011)

$X_{it}, i = 1, \ldots, N, \ t = 1, \ldots, T.$

Panel data can be represented in a rectangular form, like a board. Dimension i is called the "individual dimension," and t is the time dimension. X can be a scalar (real) variable or also a vector-valued variable. Additionally, a general panel data regression model is written as (Hauser, 2013)

$$y_{it} = \beta_0 + x_{it}'\beta + \epsilon_{it} .$$

where:

x_{it}' is a K dimensional vector of explanatory variables, without a constant term,

β_0 the intercept, is independent of i and T.

β is a $(K \times 1)$ vector, the slopes, is independent of i and T.

ϵ_{it} the error, varies over i and T.

Individual characteristics (which do not vary over time), z_i may be included. In this case, the panel data regression model is written as

$$y_{it} = \beta_0 + x_{it}'\beta_1 + z_i'\beta_2 + \epsilon_{it} .$$

where z_i' is a K dimensional vector of individual characteristics (time-invariant).

Related Content

Intelligent Logistics Vehicle Path Planning Using Fused Optimization Ant Colony Algorithm With Grid
Liyang Chu, Haifeng Guoand Qingshi Meng (2024). *International Journal of Information Technologies and Systems Approach (pp. 1-20).*
www.irma-international.org/article/intelligent-logistics-vehicle-path-planning-using-fused-optimization-ant-colony-algorithm-with-grid/342613

Multimodal Literacy
Maryann Tatum Tobin (2018). *Encyclopedia of Information Science and Technology, Fourth Edition (pp. 1508-1516).*
www.irma-international.org/chapter/multimodal-literacy/183865

Gendering Information and Communication Technologies in Climate Change
Sam Wong (2021). *Encyclopedia of Information Science and Technology, Fifth Edition (pp. 1408-1422).*
www.irma-international.org/chapter/gendering-information-and-communication-technologies-in-climate-change/260275

Stock Price Trend Prediction and Recommendation using Cognitive Process
Vipul Bagand U. V. Kulkarni (2017). *International Journal of Rough Sets and Data Analysis (pp. 36-48).*
www.irma-international.org/article/stock-price-trend-prediction-and-recommendation-using-cognitive-process/178161

Mobile App Stores
Michael Curran, Nigel McKelvey, Kevin Curranand Nadarajah Subaginy (2015). *Encyclopedia of Information Science and Technology, Third Edition (pp. 5679-5685).*
www.irma-international.org/chapter/mobile-app-stores/113023