Reportes resultados sobre modelos panel

Vicente López Díaz

April 6, 2021

Abstract

In these note there is a summary of the relations between several models that can use panel data.

The objective of these note is to give a broad overview of the possible models that can use panel data. There are several usual features to consider in a model with panel data, for example, changes on parameters for time or individual. Also, specification on error term is relevant for interpretation.

These notes are based on Hsiao (2014) It goes from the theory in the text, to the application.

1 Dummies for each level: city, brand, time

One simple regression with indicators for city, brand and time. Same effect on all the brands.

$$y_{itm} = \alpha_i^* + \gamma_m^* + \lambda * t + \beta_0^{'} janDummy_m + \beta_1^{'} taxDummy_m + u_{itm};$$

$i = 1, \ldots, N; t = 1, \ldots, T; m = 1, \ldots, M$. Results for principal brands:

VARIABLES	(1) ppu	(2) ppu	(3) ppu
m1	-0.023***	-0.073***	-0.073***
	(0.004)	(0.004)	(0.004)
m1_20	0.214***	-0.012	
	(0.010)	(0.008)	
ym	0.009***		
	(0.000)		
2.marca	-0.009***	-0.004*	-0.004
	(0.003)	(0.003)	(0.003)
3.marca	-0.590***	-0.588***	-0.587***
	(0.003)	(0.003)	(0.003)
4.marca	-0.331***	-0.311***	-0.310***
	(0.004)	(0.003)	(0.003)
5.marca	0.012***	0.012***	0.012***
	(0.003)	(0.002)	(0.002)
6.marca	-0.525***	-0.530***	-0.528***
	(0.005)	(0.004)	(0.004)
7.marca	-0.438***	-0.440***	-0.440***
	(0.003)	(0.003)	(0.003)
Observations	22,771	22,771	22,771
R-squared	0.898	0.938	0.938

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

1.1 Comparisons by segment

Results for brand type: 1 is premium, 2 is medium, 3 is low.

VARIABLES	(1) ppu	(2) ppu	(3) ppu
m1	-0.018***	-0.046***	-0.012
	(0.004)	(0.008)	(0.008)
m1_20	0.233***	0.175***	0.193***
	(0.011)	(0.019)	(0.029)
ym	0.010***	0.010***	0.007***
	(0.000)	(0.000)	(0.000)
Observations	13,598	5,256	3,917
R-squared	0.916	0.868	0.774

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

2 Parameters are different for each brand

Separate regression for each brand. The main estimation routine is a random coefficients result.

$$y_{itm} = \alpha^*_{im} + \delta'_m janDummy_m + \beta'_m taxDummy_m + \lambda'_m t_m + u_{itm};$$

$i = 1, \ldots, N; t = 1, \ldots, T; m = 1, \ldots, M.$

2.1 Comparisons by segment

Results for premium brands

VARIABLES	(1) ppu1	(2) ppu2	(3) ppu5
m1	-0.001	-0.000	-0.031***
	(0.003)	(0.003)	(0.005)
m1_20	0.282***	0.106***	0.047*
	(0.017)	(0.018)	(0.025)
ym	0.009***	0.009***	0.009***
	(0.000)	(0.000)	(0.000)
Observations	2,488	2,488	2,488
Number of cve_ciudad	45	45	45

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Results for lower segment brands

VARIABLES	(1) ppu3	(2) ppu6
m1	0.019**	-0.051***
	(0.009)	(0.012)
m1_20	0.194***	0.151***
	(0.028)	(0.038)
ym	0.008***	0.005***
	(0.000)	(0.000)
Observations	587	587
Number of cve_ciudad	43	43

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Results for mid-range segment brands

VARIABLES	(1) ppu4	(2) ppu7
m1	0.098***	-0.117***
	(0.012)	(0.010)
m1_20	-0.420	0.735
	(0.000)	(0.000)
ym	0.004***	0.004***
	(0.000)	(0.000)
Observations	1,356	1,356
Number of cve_ciudad	43	43

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

3 Parameters are constant over time

Section based on Hsiao (2014) Separate regression for each individual

$$y_{it} = \alpha_i^* + \beta_i^{'} x_{it} + u_{it}; i = 1, \dots, N; t = 1, \dots, T.$$

3.1 Parameters restricted over time

Separate regression for each individual

$$y_{it} = \alpha_i^* + \beta_i^{'} x_{it} + u_{it}; i = 1, \dots, N; t = 1, \dots, T.$$

It can be restricted in several ways. Only slope coefficients are identical, intercepts are individual.

$$y_{it} = \alpha_i^* + \beta^{'} x_{it} + u_{it}.$$

Both slope coefficients and intercepts are identical.

$$y_{it} = \alpha^* + \beta^{'} x_{it} + u_{it}.$$

Following Hsiao(2014) the first model is called unrestrincted, the second as individual-mean regression model and the last model is known as pooled model.

4 Simple Regression with Variable Intercept

This models are one way to consider unobserved heterogeneity across individuals and/or through time. The assumption is that the effects of that heterogeneity come from three types of variables: time-invariant, individual-invariant and individual time-variant. The model can be written:

$$y_{it} = \alpha_i^* + x_{it}^{'} \beta_i + u_{it}; i = 1, \dots, N; t = 1, \dots, T.$$

With the assumption that u_{it} is uncorrelated with $(x_{i1}, \ldots, x_{iT})$ and have an independent identically distributed random variable with mean 0 and constant variance.

Following Hsiao(2014), the OLS estimator is called least-squares dummy variable (LSDV), covariance (CV) estimator or within-group estimator. If the variance is constant for every individual an efficient estimator can be obtained using weighted least-squares with the initial estimator for individual variance from the individual errors.

4.1 Estimations of Variance-Components models

The individual-specific effects as random variables. The residual can be assumed to consist three components:

$$v_{it} = \alpha_i + \lambda_t + u_{it}.$$

The estimation for random-effects model assume a distribution on the conditional distribution of $f(\alpha_i, \lambda_t | x_i)$. With the assumption of constant λ_t for all t, the presence of α_i produces correlations in v_{it} over time for a given individual. Consistent estimates in finite samples can be obtained by Generalized Least-Squares (GLS). The GLS estimator is a weighted average of the between-group and the within-group estimator. In a practical situation, without knowing the constants from variance components, the estimation uses feasible GLS or two-step GLS.

4.2 Fixed or Random effects

When N is fixed and T is large LSDV and GLS are the same estimator. [The time-specific effects could be a problem?] The residual can be assumed to consist three components:

$$v_{it} = \alpha_i + \lambda_t + u_{it}.$$

The estimation for random-effects model assume a distribution on the conditional

5 Dynamic models

An alternative model is to consider dynamics in the equation, with a difference in the dependent variable. The second equation includes interactions, to consider the effect of the price change in every january and in january of 2020, when the tax was in place, different brand-types.

$$y_{it} = \gamma y_{i,t-1} + x_{it}^{'}\beta_i + \alpha_i^* + \lambda_t + u_{it}; i = 1, \ldots, N; t = 1, \ldots, T.$$

The assumption of strict exogeneity is no longer valid. The initial values become relevant. The way in which the T and N tend to infinity become relevant for asymptotic properties, like consistency.

VARIABLES	(1) ppu	(2) ppu
m1	-0.100***	-0.100***
	(0.009)	(0.009)
1.m1_20		0.547***
		(0.036)
1b.tipo#0b.m1_20		0.000
		(0.000)
1b.tipo#1o.m1_20		0.000
		(0.000)
2o.tipo#0b.m1_20		0.000
		(0.000)
2.tipo#1.m1_20		-0.163***
		(0.053)
3o.tipo#0b.m1_20		0.000
		(0.000)
3.tipo#1.m1_20		-0.125
		(0.076)
m1_20	0.476***	
	(0.028)	
Observations	21,517	21,517
R-squared	0.039	0.039
Number of gr_marca_ciudad	259	259

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

5.1 Dynamic on differences

The dependent variable is the change of price in a given city.

$$\Delta y_{it} = \gamma \Delta y_{i,t-1} + u_{it}; i = 1, \ldots, N; t = 1, \ldots, T.$$

VARIABLES	(1) D.ppu	(2) D.ppu
m1	0.025***	0.025***
	(0.001)	(0.001)
1.m1_20		0.217***
		(0.004)
1b.tipo#0b.m1_20		0.000
		(0.000)
1b.tipo#1o.m1_20		0.000
		(0.000)
2o.tipo#0b.m1_20		0.000
		(0.000)
2.tipo#1.m1_20		-0.089***
		(0.006)
3o.tipo#0b.m1_20		0.000
		(0.000)
3.tipo#1.m1_20		-0.050***
		(0.008)
m1_20	0.183***	
	(0.003)	
Observations	21,213	21,213
R-squared	0.258	0.267
Number of gr_marca_ciudad	256	256

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

With premium for the first label, it shows that the medium brands has lower impact on the tax, although, counterintuitively the lowest impact is estimated for the medium brands with a decrease of 8.9 cents while the lower brand only decreased 5 cents, both with respect to the premium brands average.

6 Dummies for trend-spline

Each dummy is one for the period that ends in the first month of the year.

$$\begin{aligned} y_{itm} = & \alpha_i^* + \lambda * t + \beta_{2012}^{'} Dummy2012_m + \ldots + \\ & \beta_{2020}^{'} Dummy2020_m + \beta_0^{'} janDummy_m + \\ & \beta_1^{'} taxDummy_m + u_{it}; \end{aligned}$$

$i = 1, \ldots, N; t = 1, \ldots, T.$

6.1 Comparisons by segment

Results for brand type 2 is medium.

The interpretation of the spline with negative signs?

	(1)	(2)
VARIABLES	ppu4	ppu7
m1	0.013	0.039***
	(0.012)	(0.007)
m1_20	0.256	0.192
	(0.000)	(0.000)
Observations	1,356	1,356
Number of cve_ciudad	43	43

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

7 Quadratic trend

Each dummy is one for the period that ends in the first month of the year.

$$y_{itm} = \alpha_i^* + \lambda_1 * t + \lambda_2 * t^2 + \beta_0^{'} janDummy_m + \beta_1^{'} taxDummy_m + u_{itm};$$

$i = 1, \ldots, N; t = 1, \ldots, T.$

7.1 Comparisons by segment

Results for medium brands.

	(1)	(2)
VARIABLES	ppu4	ppu7
m1	-0.035***	-0.039***
	(0.012)	(0.009)
m1_20	0.163	0.097
	(0.000)	(0.000)
ym	-0.035***	0.027***
	(0.007)	(0.010)
Observations	1,356	1,356
Number of cve_ciudad	43	43

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

8 Sample by brand

There is no common sample.

$$y_{itm} = \alpha_i^* + \lambda * t + \beta_0^{'} janDummy_m + \beta_1^{'} taxDummy_m + u_{itm};$$

$i = 1, \ldots, N; t = 1, \ldots, T.$

8.1 Comparisons by segment

Results for medium brands: Lucky , PallMall.

	(1)	(2)	(3)	(4)
VARIABLES	ppu4	ppu4	ppu7	ppu7
m1	-0.042***	-0.048***	-0.043***	-0.039***
	(0.012)	(0.015)	(0.009)	(0.012)
m1_20	0.153***	0.154***	0.193***	0.162***
	(0.031)	(0.037)	(0.023)	(0.031)
ym	0.010***	0.010***	0.010***	0.011***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	1,837	1,356	3,157	1,356
Number of cve_ciudad	36	29	41	29

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1