Panel Data Analysis with Spatial Structure

N. Artamonov

MGIMO University/MSU

March 2, 2020

Table of Content

Panel Data Models

Panel data regression with fixed individual effects

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + u_{it} \qquad \begin{array}{c} i = 1, \dots, n \\ t = t \dots, T \end{array}$$

Panel data regression with random individual effects

$$y_{it} = \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + \mu_i + u_{it} \qquad \mathrm{cov}(\mu_i, u_{it}) = 0 \qquad \begin{array}{c} i = 1, \dots, n \\ t = t \dots, T \end{array}$$

u_{it} is an idiosyncratic error.

Total number of observations $N = nT$.

Idiosyncratic Error

Spherical assumptions:

1. $\mathrm{E}(u_{it}|\mathbf{X}\ (,\boldsymbol{\mu})) = 0$
2. $\mathrm{Var}(u_{it}|\mathbf{X}\ (,\boldsymbol{\mu})) = \sigma^2$ (homoskedasticity)
3. $\mathrm{cov}(u_{it}, u_{js}|\mathbf{X}\ (,\boldsymbol{\mu})) = 0$ if $\{it\} \neq \{js\}$ (no serial and spatial correlations)

Idiosyncratic Error: Weaker Assumpltions

We allow

Heteroskedastisity

$$\mathrm{Var}(u_{it}|\mathbf{X}\ (,\boldsymbol{\mu})) \neq \sigma^2$$

Serial Correlation

$$\mathrm{cov}(u_{it}, u_{is}|\mathbf{X}\ (,\boldsymbol{\mu})) \neq 0 \qquad t \neq s$$

Spatial Correlation

$$\mathrm{cov}(u_{it}, u_{jt}|\mathbf{X}\ (,\boldsymbol{\mu})) \neq 0 \qquad i \neq j$$

Estimation and Inferences

Estimation Methods: Within (FE) or GLS (RE)

Estimation and Inferences

Estimation Methods: Within (FE) or GLS (RE)

Inferences are based on s.e. for coefficients (t-test) :

1. non-robust s.e.
2. robust s.e. by Arellano-Bond (under heteroskedasticity and serial correlation, 1987, 1991)
3. robust s.e. by Driscoll and Kraay (under both spatial and serial correlation, 1998)

Basic Tests

F-test: pooling vs FE

LM-test: pooling vs RE

Hausman test: RE vs FE. We use the robust version based on an auxiliary regression

BG-test: for serial correlation of second order

Pesaran's CD-test: for spatial correlation

Analysis with R (plm package)

Table: Basic test in R

Test	R function
F-test	pFtest
LM-test	plmtest
Hausman test	phtest(..., method='aux',..)
BG-test	pbgtest
Pesaran's CD-test	pcdtest

Table: Robust s.e. in R

s.e.	R function
Arellano-Bond	vcovHC (vcovHC.plm)
Driscoll and Kraay	vcovSCC

Variables (Russian Regional Data)

ln_grp	logged regional GRP
ln_pop	logged average annual population
ln_funds	logged value of fixed assets per unit of average annual population
distance	distance form Moscow to regional capital city
r_p_q	The proportion of the average population in age between p and q ($p = 0, 16, 25, 40, 55, 65$; $q = 15, 24, 39, 54, 64, 100$)
dummy	dummy, 1 if the proportion of extractive industries in the GVA structure of the region exceeds 15%

Regions $i = 1, \dots, 79$

Years $t = 2001, \dots, 2016$.

Overall 1264 observation.

Models

Model #1: ln_grp on ln_pop, r_16_24, r_25_39, r_40_54, r_55_64, R_65_100, ln_fund (pooling, FE, RE)

Model #2: ln_grp on ln_pop, r_16_24, r_25_39, r_40_54, r_55_64, R_65_100, ln_funds, distance, distance^2, dummy (pooling, RE)

Model #3 diff(ln_grp) on diff(ln_pop, r_16_24, r_25_39, r_40_54, r_55_64, R_65_100, ln_fund) (pooling, FE, RE)

Models

Model #1: ln_grp on ln_pop, r_16_24, r_25_39, r_40_54, r_55_64, R_65_100, ln_fund (pooling, FE, RE)

Model #2: ln_grp on ln_pop, r_16_24, r_25_39, r_40_54, r_55_64, R_65_100, ln_funds, distance, distance^2, dummy (pooling, RE)

Model #3 diff(ln_grp) on diff(ln_pop, r_16_24, r_25_39, r_40_54, r_55_64, R_65_100, ln_fund) (pooling, FE, RE)

Important

Since we deal with regional (spatial) data, it's necessary to test models for the presence of spatial correlation.

Model #1

Table: Test results for Model #1

Test	Test statistic
F-test	$F = 39.11$ (df1=78, df2=1178)
Hausmann test	$\chi^2 = 289.19$ (df=7)
BG-test (FE, $p = 2$)	$\chi^2 = 497.51$ (df=2)
BG-test (RE, $p = 2$)	$\chi^2 = 657.83$ (df=2)
Pesaran's CD-test (FE)	$z = 33.743$
Pesaran's CD-test (RE)	$z = 49.971$

Conclusion:

1. FE regression is preferable
2. We have to take into account both serial and spatial correlation, i.e. use robust SCC s.e. by Driscoll and Kraay

Model #1

Estimation results are presented in an external `html`-file

Due to these results all coefficients are significant (at 5% significant level)

Model #2

Table: Test results for Model #2

Test	Test statistic
LM-test	$\chi^2 = 2053$ (df=1)
BG-test (RE, $p = 2$)	$\chi^2 = 644.73$ (df=2)
Pesaran's CD-test (RE)	$z = 48.971$

Conclusion:

1. RE regression is preferable
2. We have to take into account both serial and spatial correlation, i.e. use robust SCC s.e. by Driscoll and Kraay

Model #2

Estimation results are presented in an external `html`-file

Due to these results at 5% significant level the following coefficients are significant:

ln_funds, distance, distance^2, r_16_24, r_p25_39, r_55_64, r_65_100,

Model #3

Table: Test results for Model #3

Test	Test statistic
F-test	$F = 0.24626$ (df1=78, df2=1099)
BG-test (FE, $p = 2$)	$\chi^2 = 50.462$ (df=2)
Pesaran's CD-test (FE)	$z = 58.71$

Conclusion:

1. Pooling regression is preferable
2. We have to take into account both serial and spatial correlation, i.e. use robust SCC s.e. by Driscoll and Kraay

Model #3

Estimation results are presented in an external `html`-file

Due to these results at 5% significant level the following coefficients are significant:

diff(r_16_24), diff(r_p25_39)

THANK YOU FOR THE ATTENTION