Income disparities in Colombia: Market potential and neighborhood effects

Jesus Lopez-Rodriguez[a,†], Julian S.Vasquez-Roldan [a,b]

[a] Jean Monnet Group on Competition and Development (GCD), Departament of Economics, School of Economics and Business, Universidade da Coruña, 15071 A Coruña, Spain.
e-mail: jesus.lopez.rodriguez@udc.es
[b] Instituto para el Desarrollo de Antioquia (IDEA), Calle 42 N° 52 - 259 Medellín, Colombia
e-mail: julian.vasquez@udc.es

Corresponding author: Jesus Lopez Rodriguez (jesus.lopez.rodriguez@udc.es)
words: 2000

* Corresponding author: Jesus Lopez Rodriguez (jesus.lopez.rodriguez@udc.es)

Income disparities in Colombia: Market potential and neighborhood effects

Abstract

Market potential has been shown to have an important impact on the explanation of observed income disparities both across countries and across regions within countries. However, the importance of neighbourhood for "income levels-market potential" regressions is largely neglected in the empirical studies on the subject. This paper tries to fill up this gap by estimating fixed effects spatial panel data models using Colombian regional data over the period 1990-2015. Our results reveal that around half of the impact of market potential on regional income disparities can be attributed to neighbouring regions.

Key words: Spatial panel data; Market potential; Income disparities; Colombia

JEL Classification: R11; R12; R13; R14

1. Introduction

Market potential, a variable which has its roots in the Newtonian physics, has been successfully used in the applied regional economics literature to characterize the spatial income and human capital gradients across different areas (López-Rodríguez and Faíña 2006, López-Rodríguez, 2007). New Economic Geography (Krugman, 1991) meant a revival for the concept of market potential through a sound theoretical micro foundation. One of the workhorses of applied-New Economic Geography studies is the estimation of "income levels-market potential" regressions (Hanson 2005, Head and Mayer, 2011, Karahasan and Bilgel 2020, Karahasan et al. 2016, Lopez-Rodriguez et al., 2007). However, the importance of neighbourhood in these types of analysis has been largely neglected. This paper tries to fill up this gap by estimating augmented "income levels-market potential" regressions using spatial panel data models. Our results using Colombian regional data over the period 1990-2015 show that half of the impact of market potential on income disparities is accrued to neighbourhood effects.

2. Data and Methodology

Colombia represents an interesting case study for our analysis (core-periphery pattern, sizeable income disparities-see Table 1- and a strong regional income-gradient relative to region´s distance from Bogota-see Figure 1). Core-periphery type of economic structures are structures where the relative regions´ market access play an important role and are best suited to be explained by "income levels-market potential" regressions.
To account for both the role of region´s market access and the patterns of spatial dependence in our data, we propose to estimate the following spatial econometrics models: a) A spatial autoregressive model (SAR) [1]; b) A spatial error model (SEM) [2]; and finally c) a spatial Durbin model (SDM) [3]. The specifications are as follows:

$$\ln Ypc_{it} = \rho W lnYpc_{it} + \beta_1 lnMP_{it} + u_{it} \quad [1]$$

$$\ln Ypc_{it} = \beta_1 lnMP_{it} + \lambda W u_{it} + u_{it} \quad [2]$$

$$\ln Ypc_{it} = \rho W lnYpc_{it} + \beta_1 lnMP_{it} + \theta_1 W lnMP_{it} + u_{it} \quad [3]$$

Where Ypc represents regional gross value added per head (measured at constant 2005 pesos colombianos[1]), MP represents regional market potential computed à la Harris (1954) using the following formula:

$$MP(K)_{it} = \sum_{j=1}^{n} \frac{M_{jt}}{d_{ij}} = \frac{M_{it}}{d_{ii}} + \sum_{j \neq i}^{n-1} \frac{M_{jt}}{d_{ij}} \qquad [4]$$

Where M represents regional gross value added; d measures the distance in kilometers between pairs of regions ($i \neq j$) and the internal distance within each region (i= j). In making the calculations of the internal distance (d_{ii}) the standard methodology assumes that locations are circular and the internal distance is approximated by a function that is proportional to the radius of the location. The radius of a circular-shaped location "i" of size equal to "$area_i$" is $r_i = \sqrt{area_i/\pi}$. In this paper and following the work of Keeble et al. (1982), we will use $d_{ii} = 1/3 \cdot r_i = 0.188\sqrt{area_i}$ as our first option. On the other hand, following other authors such as Crozet (2004), Head and Mayer (2000), and Nitsch (2000) we will use $d_{ii} = 2/3 \cdot r_i = 0.376\sqrt{area_i}$ as our second option. Both formulas have been frequently used in the literature and give the average distance in a circular locations under the assumption that production takes place in the center and consumers are evenly spread across locations; n is the number of regions in which Colombia is divided (26). K take the value 1 when $d_{ii} = 1/3 \cdot r_i = 0.188\sqrt{area_i}$ and 2 when $d_{ii} = 2/3 \cdot r_i = 0.376\sqrt{area_i}$; "i", "j" and "t" are sub-indexes for region and time; W represents the spatial weights matrix of the models which in our case will be based in the 5-nearest neighbors. Finally, "u" is the disturbance term. All data have been taking from the Colombian National Statistical Institute (DANE[2]).

"Please insert Table 1 around here"

"Table 1. Per capita income across Colombian regions"

"Please insert Figure 1 around here"

"Figure 1. Income in the Colombian regions and distance to Bogota (km.)"

3. Findings

A potential concern of simple "income levels-market potential" regressions is related to the presence of patterns of spatial dependence in our data. This is confirmed by the results of the Moran's I and the battery of Lagrange Multiplier (LM) tests of spatial dependence shown in columns 1 and 2 of Table 2. The tests conclude in favor of significant residual spatial dependence. Failure of controlling for this spatial dependence leads to results based on the OLS estimator that are inefficient and biased.

[1] "Peso colombiano" is the name of the Colombian currency

[2] DANE: Departamento Administrativo Nacional de Estadística, www.dane.gov.co

"Please insert Table 2 around here"

Augmented "income levels-market potential" regressions corresponding to the three specifications are reported in columns 3 to 5 (6 to 8) when the market potential is defined as MP1 (MP2). The results of the Hausman tests of the null hypothesis of no systematic difference in coefficients between the random and fixed effects estimator is rejected in all specifications and therefore the models have been estimated by fixed effects. The estimation results show that the spatial parameters (rho and lambda) are strongly significant in all cases and large in magnitude. As for the effect of market potential, the results continue to support the hypotheses the standard "income levels-market potential" regressions, i.e, the positive impact of market potential in shaping the spatial income structure observed across the Colombian regions. However, controlling for spatial autocorrelation by either SAR, SEM or SDM specifications reveals a sizeable drop in the magnitude of the coefficient estimates for market potential.

Spatial Durbin models (SDM) have a more complex structure than spatial lag models (SAR). However, SDM can be reduced to a spatial error model (SEM). In order to test this hypothesis, we apply the test of common factors (COMFAC test) which under the null means that the model can be merged into a SEM.

The value of the COMFAC statistic when we run our "income levels-market potential" regression where market potential is defined as MP1 is 8.50 (p-value 0.0035) which means that we reject the null and therefore the correct spatial specification would be a SDM. However, the results of the estimation of the SDM show that the spatial lag of market potential does not show up as statistically significant and therefore we conclude that the best spatial specification corresponds to a SAR. The same conclusions are obtained when our market potential variable is defined as MP2. The value of the COMFAC statistic is 13.03 (p-value 0.000) and therefore the test points to SDM. Again, the spatial lag of market potential is not statistically significant leading us to choose a SAR-type of spatial specification.

The econometric results of Table 2 are complemented by the computation of the marginal effects (Table 3) for the spatial lag (SAR) which are broken down into direct, indirect and total effects.

The direct effects of the market potential on per capita income levels can be attributed to the self-effect that a shock in terms of market potential of region "j" generates in the expected per capita income of region "j". In contrast, the indirect effects of market potential which in this case are spatial effects of global-type that are dynamically dispersed throughout the system represent the impact that a shock in the market potential of region "j" generates in the expected per capita income of region "j" but in an indirect way; in other words, understood as the effects on "j" throughout the impacts onto the expected per capita income of region "i≠j".

The results in Table 3 show that both the direct and indirect effects of market potential MP1 and MP2 on the per capita income levels of the Colombian regions are statistically significant at the standard significance levels and economically important. Moreover, the results show that the spillover effects are of the same importance than the direct effects. Ceteris Paribus, doubling the market potential of a region increases the expected per capita income of the region by between 17%-21%. Half of this increment in expected per capita income can be attributed to the spillover effects.

"Please insert Table 3 around here"

4. Conclusions

“income levels-market potential” type of regressions are usually carried out without considering the presence of potential patterns of spatial dependence in the data with the econometric limitations that this imposed in the estimated coefficients. Our paper shows that for the case of Colombia this pattern of spatial dependence is present in the data. We control for this problem by estimating initially three different types of spatial models (SAR, SDM, SEM). The results of the COMFAC statistic were in favour of a SDM. However, the spatial lag of market potential in this specification was not statistically significant leading us to choose a SAR model. Breaking down the computations of the marginal effects for the spatial lag model (SAR) showed that the estimated direct impacts of market potential (MP(1) and MP(2)) were approximatively of the same size as the indirect impacts suggesting that neighbourhood or spillover effects are very important in explaining the impact of market potential on the spatial structure of income across Colombian regions. A fruitful research avenue along these lines will be consider other spatial contexts.

Acknowledgements:
This wok was supported by the Spanish Ministry of Economics and Competitiveness (grant number ECO2015-68597-P (MINECO/FEDER))

5. References

Crozet, M. (2004), “Do Migrants Follow Market accesss? An Estimation of a New Economic Geography Model”, *Journal of Economic Geography* 4: 439-458

Hanson, G.H. (2005), “Market potential, increasing returns and geographic concentration”, *Journal of International Economics*, 67(1):1–24

Harris, C.D. (1954), “The Market as a Factor in the Localization of Industry in the United States”, *Annals of the Association of American Geographers* 44(4): 315–348

Head, K. and Mayer, T. (2000), “Non-Europe: The Magnitude and Causes of Market Fragmentation in the EU”, *Review of World Economics* 136: 284-314

Head, K. and Mayer, T. (2011), “Gravity, market potential and economic development”, *Journal of Economic Geography* 11(2): 281–294

Karahasan, B. C. and Bilgel, F. (2020), “Market access and regional dispersion of human capital accumulation in Turkey”, *Review of Development Economics*, 24(3):1073-1101

Karahasan, B.C.; Dogruel F. and Dogruel, S. (2016), “Can Market Potential Explain Regional Disparities in Developing Countries? Evidence from Turkey”, *The Developing Economies*, 54(2):162-197

Keeble, D. Owens, PL. Thompson, C. (1982), “Regional accessibility and economic potential in the European Community”, *Regional Studie*s 16(6):419–432

Krugman, P. (1991), “Increasing returns and economic geography”, *Journal of Political Economy* 99(3): 483–499

Lopez-Rodriguez, J. (2007), "Is market potential robust in explaining regional educational levels in Europe?," *Applied Economics Letters*, 14(7): 489-492.

Lopez-Rodriguez, J. and Faiña, A. (2006), "Does distance matter for determining regional income in the European Union? An approach through the market potential concept," *Applied Economics Letters*, 13(6): 385-390.

Lopez-Rodriguez, J.; Faiña, A. and Lopez Rodriguez, J. (2007), "Human capital accumulation and geography: Empirical evidence from the European Union", *Regional Studies* 41(2): 217–234

Nitsch, V. (2000) "National borders and international trade: evidence from the European Union", *Canadian Journal of Economics* 22:091–1105

Regions	1995-2004	2005-2015
Antioquia	4.961.629	12.391.907
Atlántico	3.860.427	9.556.760
Bogotá D. C.	7.520.156	19.359.223
Bolívar	3.652.430	11.535.620
Boyacá	4.003.562	12.623.776
Caldas	3.665.616	8.919.007
Caquetá	2.244.126	5.487.713
Cauca	2.199.115	6.461.040
Cesar	3.336.917	10.992.903
Córdoba	2.785.394	6.540.630
Cundinamarca	4.695.636	11.524.085
Chocó	1.511.731	4.950.413
Huila	3.412.450	9.358.566
La Guajira	3.069.034	7.806.901
Magdalena	2.372.906	6.116.497
Meta	5.254.702	26.658.220
Nariño	2.039.700	5.241.788
Norte Santander	2.705.913	7.215.165
Quindío	3.254.584	7.834.535
Risaralda	3.559.837	9.029.497
Santander	5.805.887	20.917.072
Sucre	2.028.162	5.448.391
Tolima	3.439.160	9.057.595
Valle del Cauca	5.316.166	12.439.703
Amazonas	2.427.492	5.357.587
Arauca	7.949.647	19.718.389
Casanare	17.131.330	33.084.502
Guanía	2.083.511	4.876.989
Guaviare	2.683.196	4.845.939
Putumayo	1.814.622	8.040.120
San Andrés y Providencia	5.320.951	11.397.477
Vaupés	1.850.472	3.557.343
Vichada	3.007.640	5.229.207
Average 1	**3.968.609**	**10.411.350**
Max/Average1	**4,32**	**3,18**
Min/Average1	**0,38**	**0,34**
Max	**17.131.330**	**33.084.502**
Mín	**1.511.731**	**3.557.343**
Average 2*	**3.269.743**	**8.386.082**

*Note: Without oil-producing regions (Arauca, Casanare, Meta and Santander; Figures are expressed in Colombian pesos
Source: Own elaboration based on DANE

	Levels (lYrpc)							
Dept. variable	(OLS)		(SAR)	(SEM)	(SDM)	(SAR)	(SEM)	(SDM)
Reg.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Const.	11.46***	11.93***						
	(0.18)	(0.19)						
lnMP(1)	0.30**		0.11***	0.18***	0.11***			
	(0.013)		(0.01)	(0.01)	(0.03)			
Wln MP(1)					-0.00			
					(0.03)			
lnMP(2)		0.27***				0.08***	0.14***	0.05**
		(0.01)				(0.01)	(0.01)	(0.02)
Wln MP(2)								0.03
								(..03)
rho			0.49***		0.49***	0.53***		0.52***
			(0.04)		(0.04)	(0.04)		(0.04)
lambda				0.50***			0.53***	
				(0.05)			(0.05)	
Resid Moran's I	13.431***	13.828***						
	[0.000]	[0.000]						
LM-ERR	174.988***	186.144***						
	[0.000]	[0.000]						
LM-LAG	149.236***	215.130***						
	[0.000]	[0.000]						
Robust LM-ERR	10.132***	11.324***						
	[0.000]	[0.001]						
Robust LM-LAG	14.685***	40.310***						
	[0.000]	[0.000]						
Hausman test			17.73***	16.50***	11***	20.73***	9.76***	16.83***
			[0.000]	[0.000]	[0.01]	[0.000]	[0.000]	[0.000]
COMFAC test					8.50***			13.03***
					[0.003]			[0.000]
FERegion /year	No/No	No/No	Yes/No	Yes/No	Yes/No	Yes/No	Yes/No	Yes/No
Obs.	624	624	624	624	624	624	624	624
	(24/26)	(24/26)	(24/26)	(24/26)	(24/26)	(24/26)	(24/26)	(24/26)
	0.45	0.37	0.46	0.52	0.46	0.38	0.49	0.26
Est.	OLS	OLS	MLE					

Note: *, **, *** represent significance at 10 percent, 5 percent and 1 percent respectively. *p-values* for the statistics are in brackets. LM-ERR== Lagrange Multiplier test for spatial error dependence; LM-LAG==Lagrange Multiplier test for spatial lag dependence. Spatial weights matrix based on 5-nearest neighbors. Standard errors for coefficient estimates are in parentheses

	SAR		
VARIABLES	Direct	Indirect	Total
lnMP(1)	0.11***	0.10***	0.21***
	(0.01)	(0.01)	(0.01)
lnMP(2)	0.087***	0.087***	0.17***
	(0.01)	(0.01)	(0.01)

Standard errors between brackets ***p<0.01,**p<0.05,*p<0.1.
Source: Own Elaboration