Job-education mismatch and spatial mismatch evidence for a developing country

David Baena
Department of Economics
Universidad EAFIT
dbaenac@eafit.edu.co

Gustavo A. García
Research in Spatial Economics (RiSE-group)
Department of Economics
Universidad EAFIT
ggarci24@eafit.edu.co

Outline

Motivation

- The spatial mismatch hypothesis argues that workers residing far away from jobs may experience poor labor market outputs because they are disconnected from job opportunities.
- Mechanisms that explain how distance to job opportunities could be harmful (Gobillon, Selod, & Zenou, 2007):
  - Workers may refuse a job that involves commutes that are too long.
  - Workers' job search efficiency may decrease with distance to jobs.
  - Workers residing far away from jobs may not search intensively.
  - Workers may incur high search costs that cause them to restrict their spatial search horizon at the vicinity of their neighborhood.
- In particular, there is a substance account of individuals employed in jobs that do not correspond to their qualification level and this so-called job-education mismatch can be in part explain by the lack of accessibility.
- The idea is that the lack of transport (public or private) can limit both the quantity and quality of jobs that individual can be accessed with which the job search will not be efficient, and they could accept jobs that are below or above their qualifications.

Motivation

- This relationship between job-education mismatch and job accessibility could be more relevant in developing countries than in developed countries, where the cities are less accessible in terms of having poor public transport system and being socio-spatial segmented.
- Our research seeks to examine the link between job accessibility (both public and private) and job-education mismatch at the urban level in the case of a city in a developing country: Medellín, Colombia.
- Medellín is interesting because its urban transport development especially in lower-income zones has attracted international attention like a successful case of urban management that efficiently has increased accessibility (Bocarejo et al., 2014):
  - Sistema Integrado Metro
  - Proyecto Urbano Integral
  - Housing related policies
  - Policies to promote public transit

Contribution

- This study contributes to the understanding of the relationship between job-education mismatch and job accessibility in an urban labor market in developing economies, studying the case of Medellín (Colombia).

- We provide new insights about this relationship proposing alternatives measures of both public and private accessibility that take into account the transit connectivity by using *Google Matrix Distance API* and *Bings Maps Distance Matrix API*. Estimating, then, an econometric model that simultaneously control by employment selectivity and endogeinity over our measures of job accessibility.

- To the best of our knowledge, this study is the first work to examine the link between spatial connectivity to potential employment and job-education mismatch at the urban level in the context of a developing country.

- Therefore, the research aims at providing new empirical evidence studying the case of Medellín, offering conceptual and methodological insights that could be relevant for both developing and developed economies.

Main findings

Job-education mismatch $\Longrightarrow \downarrow$ Public transport accessibility
$\uparrow$ Private transport accessibility

Accessibility

As surveyed in Páez, Scott, and Morency (2012):

- accessibility typically comprised two elements; the cost of travelling and the quality of opportunities. It could be interpreted as a measure that indicates how simple it is to someone, from one place, to get some kind of good, service or remuneration offered in any other place of the city.
- it is usually classified as a positive or normative indicator, according to the type of data the analyst use to build its representation.
- All measures of accessibility could be classified as utility-based, gravity-based or cumulative-opportunities-based measures.
- could be generalized as,

$$A_{jk}^{p} = \sum_{i} g(P_{ik}^{p}) f(c_{ij}^{p}), \tag{1}$$

where accessibility A of the k opportunity, over some subset p of population P (in zone j); is explained by the sum of opportunities in other zones, as function of its populations $g(P_{ik}^{p})$, weighted by the costs $f(c_{ij}^{p})$, of getting from zone j to any other zone i

Spatial mismatch

- Kain (1968) was the first one to argue that residential discrimination affected negatively labour market outcomes of ethnic minorities.
- Kain (1992), Ihlanfeldt and Sjoquist (1998) and Holzer (1998) surveyed empirical evidence for Europe and USA.
- Mechanisms, through which this friction could appear, are concisely explained by Gobillon and Selod (2014) and Gobillon et al. (2007).
- Some theoretical models have appeared to illustrate spatial mismatch (Brueckner & Martin, 1997; Zenou, 2009). All of them are usually based on Fujita and Ogawa (1982) and Lucas and Rossi-Hansberg (2002).
- We follow two models. One with skill and residential space, where there are two sources of heterogeneity (Brueckner, Thisse, & Zenou, 2002). And another with two sets of households, one for employed and other unemployed workers (Patacchini & Zenou, 2005).

Job-education mismatch

- Job-education mismatch was first proposed on literature with Frank (1978). He tried to explain why women earn less. They were supposed to find it harder to be employed, as local labor demand was insufficient; hence, they would be forced to accept bad job offers.
- There have been empirical approaches testing his theory in Germany (Büchel & Battu, 2003; Büchel & Van Ham, 2003), Italy (Devillanova, 2013), Finland (Jauhiainen, 2011) and Spain (Ramos & Sanromá, 2013).
- Even when many of this papers are aware of commuting costs effects on overeducation, none of them is actually concerned of the accessibility concept in the broader sense we explained above.

Data

- We use data from Origin-Destiny Survey 2012, for the urban area of Medellín.
- We took, as unit of analysis, individuals without handicaps who were between 18 and 60 years old.
- This database is georrefered by 456 SIT zones, which are homogeneous zones that respect neighborhood limits (de Aburrá, 2012). As we had polygons of Medellín for an older zonification system; we estimated 409 centroids of SIT zones, equivalent to those more actual 456 zones.
- The spatial coordinates of these centroids were used as inputs to calculate a distance matrix, using *Google Distance Matrix API* for transit commuting time and *Bings Maps Distance Matrix API* for private automobile commuting time. Its elements were the seconds it took on average to travel, in year 2018, from one centroid to another. Whenever commuting time was not possible to compute, we assumed it to be the maximum of the distance matrix (this happened with 28.25% of the SIT zones with public transport).

Accessibility measures

- We define the public cost of commuting to jobs, from the j SIT zone to the other SIT zones i as

$$Cost\ \ of\ \ Commuting_j = \sum_{\substack{i=1\\ i\neq j}} \frac{Potential\ \ Employment\ \ Rate_i}{Public\ \ Time_{ji}} \tag{2}$$

- We define public accessibility, for the h household, as

$$\hat{\pi}_h = \sum_{\substack{i=1\\ i\neq j}} \frac{1}{Other\ \ means_h} * \frac{Potential\ \ Employment\ \ Rate_i}{Public\ \ Time_{ji}} \tag{3}$$

- We define the private cost of commuting to jobs, from the j SIT zone to the other SIT zones i as

$$Cost\ \ of\ \ Commuting_j = \sum_{\substack{i=1 \\ i \neq j}} \frac{Potential\ \ Employment\ \ Rate_i}{Private\ \ Time_{ji}} \tag{4}$$

- Thus we could define private accessibility, for the h household, as

$$\hat{\rho}_h = \sum_{\substack{i=1 \\ i \neq j}} I_h[Cars > 0] * \frac{Potential\ \ Employment\ \ Rate_i}{Private\ \ Time_{ji}} \tag{5}$$

Descriptive evidence

Table 1. Descriptive Statistics at Individual Level

	Mean	Standard Deviation	Min.	25%	50%	75%	Max.	Missing Values	Obs.
Job-Education Mismatch	0.0000	4.5049	-16.3060	-3.7843	1.1568	2.8373	10.8373	56540	886123
Public Job-commuting cost	0.0498	0.0128	0.0173	0.0421	0.0503	0.0596	0.0819	0	942663
Private Job-commuting cost	0.2047	0.0403	0.1061	0.1763	0.2088	0.2340	0.2841	0	942663
Public Accessibility	0.0468	0.0153	0.0045	0.0387	0.0482	0.0587	0.0819	0	942663
Private Accessibility	0.0909	0.1060	0.0000	0.0000	0.0000	0.2057	0.2841	0	942663
Employed	0.9399	0.2378	0	1	1	1	1	0	942663

Source: Encuesta de Origen & Destino, 2012

- In regression, job-education mismatch, public accessibility and private accessibility are scaled so that β's were easier to interpret.
- Employed people seemed to be over represented: measurement error.

Descriptive evidence

Table 2. Descriptive Statistics at SIT Level

	Mean	Standard Deviation	Min.	25%	50%	75%	Max.
Job-Education Mismatch	0.3843	2.4897	-6.2372	-1.4903	0.5387	2.4180	6.6725
Public Job-commuting cost	0.0528	0.0132	0.0173	0.0447	0.0534	0.0624	0.0819
Private Job-commuting cost	0.2114	0.0394	0.1061	0.1872	0.2139	0.2408	0.2841
Public Accessibility	0.0494	0.0144	0.0116	0.0408	0.0496	0.0606	0.0819
Private Accessibility	0.0995	0.0589	0.0000	0.0550	0.0917	0.1415	0.2699
Employed	3120	2651	0	1327	2461	4364	22879

Source: Encuesta de Origen & Destino, 2012

- All variables but employed people were aggregated by its mean at SIT level. Employed, for this table, is the amount of employed people per SIT zone.

SMR

- To better visualize mismatch we computed *standardized "mortality" ratios*.
- As Banerjee, Carlin, and Gelfand (2004), we assume with these ratios that

$$
\begin{aligned}
Y_i \mid \eta_i &\overset{ind}{\sim} Po(E_i \eta_i), \\
\eta_i &\overset{iid}{\sim} G(0.0001,\ \ 0.0001),
\end{aligned}
$$

such that,

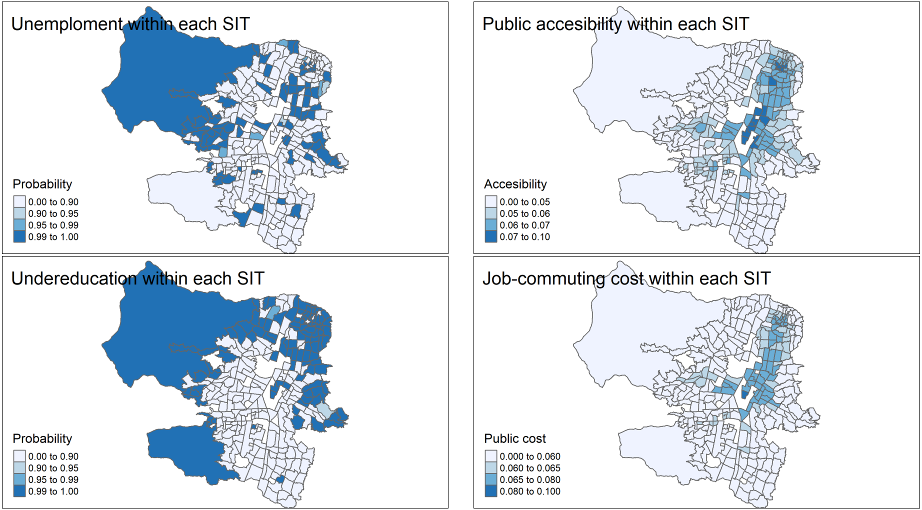
$$
\eta_i \mid Y_i \overset{ind}{\sim} G(Y_i + 0.0001,\ \ E_i + 0.0001).
$$

For SIT i with $E_i = \frac{Population_i}{\sum_i Population_i} \cdot Y_i$, given some attribute Y_i.

- This would show us how credible it is that Y_i concentrates more within SIT zone i than any other.

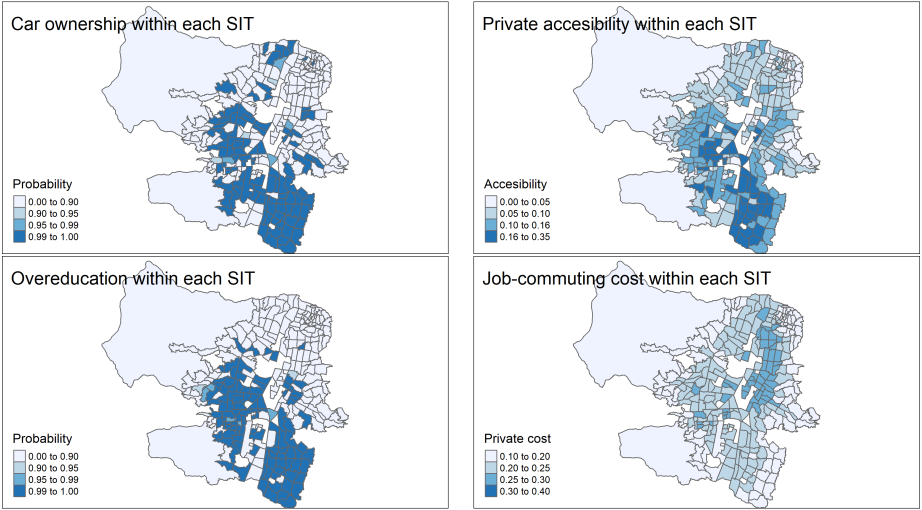
Maps

Figure 1. Standardized ratios and accessibility measures



Source: Encuesta de Origen & Destino, 2012

Figure 2. Standardized ratios and accessibility measures



Source: Encuesta de Origen & Destino, 2012

Econometric model

- Our econometric framework let us control for self-selection, since we can only observe job-education mismatch on employed people. We also consider endogeneity over our two variables of interest: private and public accessibility. We use instrumental variables.

- This might allow us to control for truncation bias, attenuation bias and omitted variable bias.

- Our model is

$$mismatch_h = \beta_0 + \beta_1 \hat{\rho}_h + \beta_2 \hat{\pi}_h + X_h' \alpha + \mu_h, \tag{6}$$

$$\hat{\rho}_h = \lambda_0 + W_h' \alpha_h + \epsilon_h, \tag{7}$$

$$\hat{\pi}_h = \gamma_0 + W_h' \gamma_h + \nu_h, \tag{8}$$

$$Pr(Employed_h = 1) = \Phi(\delta_0 + Z_h' \delta), \tag{9}$$

Scope

- Our econometric framework let us control for self-selection, since we can only observe job-education mismatch on employed people. We also consider endogeneity over our two variables of interest: private and public accessibility. We use instrumental variables.

- This might allow us to control for truncation bias, attenuation bias and omitted variable bias.

- Since unemployed population is underestimated in our O-D survey, we used a pseudo-unemployment rate as exclusion restriction in the selection equation. This variable might served us as proxy of unemployment within each SIT.

Control variables

We controlled for:

1. the number of inhabitants between 0-4 years old, 5-9 years old, 10-15 years old and 16-19 years old within each household, for married couples,
2. if the persona was spouse, son or daughter, relative, housekeeper, visitant, tenant, grandfather or grandmother and other within its household (our reference category is the head of household),
3. age, age squared,
4. the number of people within each household (different to the person himself) who had none, primary, secondary, non-formal, technological, technical and graduate education.

Entire sample

Table 3. Conditional Mixed Process

	(1) Job-education mismatch	(2) Employed	(3) Private accessibility	(4) Public accessibility
Public accessibility	0.0223	0.0211	-	-
	(0.0481)	(0.0555)	-	-
Private accessibility	0.6671***	-0.2177	-	-
	(0.1289)	(0.1501)	-	-
Pseudo-unemployment rate	-	-3.4902***	-	-
	-	(0.3791)	-	-
Minimum distance to nearest bycicle parking lot	-	-	-0.0001***	-0.0002***
	-	-	(0.0000)	(0.0000)
Minimum distance to nearest taxi stop	-	-	-0.0001***	-0.0001***
	-	-	(0.0000)	(0.0000)
Minimum distance to nearest bus stop	-	-	-0.0001***	-0.0007***
	-	-	(0.0000)	(0.0001)
Most common & newest car model per SIT	-	-	0.0003***	-0.0003***
	-	-	(0.0000)	(0.0001)
ρ_{21}	-1.4537***			
	(0.137)			
ρ_{31}	-0.4816***			
	(0.1335)			
ρ_{41}	0.0015			
	(0.0744)			
ρ_{34}	-0.3115***			
	(0.0201)			
ρ_{32}	0.2663*			
	(0.1616)			
ρ_{42}	-0.1029			
	(0.0878)			

*Notes: *** significant at 1%, ** significant at 5%, * significant at 10%, standard errors clustered at SIT zone level*
Source: Encuesta de Origen & Destino, 2012

Overeducated sub-sample

Table 4. Conditional Mixed Process

	(1) Job-education mismatch	(2) Private accessibility	(3) Public accessibility
Public accessibility	-0.0395**	-	-
	(0.0171)	-	-
Private accessibility	0.0777*	-	-
	(0.0409)	-	-
Minimum distance to nearest public parking lot	-	0.0000	-0.0006***
	-	(0.0001)	(0.0002)
Minimum distance to nearest taxi stop	-	-0.0002***	0.0001**
	-	(0.0000)	(0.0001)
Minimum distance to nearest bus stop	-	-0.0004***	-0.0006***
	-	(0.0001)	(0.0001)
Minimum distance to nearest Metro station	-	0.0001	-0.0006***
	-	(0.0000)	(0.0001)
Most common newest car model per SIT	-	0.0007***	-0.0007***
	-	(0.0001)	(0.0001)
ρ_{23}	-0.2509***		
	(0.0248)		
ρ_{21}	-0.2118*		
	(0.1168)		
ρ_{31}	0.1615**		
	(0.0623)		

*Notes: *** significant at* 1%*, ** significant at* 5%*, * significant at* 10%*, standard errors clustered at SIT zone level*
Source: Encuesta de Origen & Destino, 2012

Undereducated sub-sample

Table 4. Conditional Mixed Process

	(1) Job-education mismatch	(2) Private accessibility	(3) Public accessibility
Public accessibility	-0.0038	-	-
	(0.0238)	-	-
Private accessibility	0.2871***	-	-
	(0.1086)	-	-
Minimum distance to nearest bycicle parking lot	-	-0.0001***	-0.0002***
	-	(0.0000)	(0.0000)
Minimum distance to nearest bus stop	-	0.0000	-0.0007***
	-	(0.0000)	(0.0001)
Most common newest car model per SIT	-	0.0002***	-0.0003**
	-	(0.0000)	(0.0001)
ρ_{23}	-0.1525***		
	(0.031)		
ρ_{21}	-0.4636**		
	(0.182)		
ρ_{31}	0.041		
	(0.0578)		

*Notes: *** significant at 1%, ** significant at 5%, * significant at 10%, standard errors clustered at SIT zone level*
Source: Encuesta de Origen & Destino, 2012

Entire Sample

Table 1A. Control Function Approach

	Job-education mismatch
Public accessibility	-0.0028
	(0.0323)
Private accessibility	0.6742***
	(0.0731)
Inverse Mills ratio	1.2817***
	(0.3931)
Kleibergen-Paap LM (*p-value*)	0.0000
Sargan Test (*p-value*)	0.5195
Craig-Donald F-Test	43.223
Yogo-Stock critical values (5% *maximal IV relative bias*)	11.04
Yogo-Stock critical values (10% *maximal IV relative size*)	16.87

*Notes: *** significant at* 1%, *** significant at* 5%, ** significant at* 10%, *bootstrap standard errors*
Source: Encuesta de Origen & Destino, 2012

Overeducated sub-sample

Table 2A. GMM

	Job-education mismatch
Public accessibility	-0.0337**
	(0.0154)
Private accessibility	0.0690*
	(0.0382)
Kleibergen-Paap LM (*p-value*)	0.0000
Hansen J Test (*p-value*)	0.5249
Kleibergen-Paap F-Test	28.313
Yogo-Stock critical values (5% *maximal IV relative bias*)	13.97
Yogo-Stock critical values (10% *maximal IV relative size*)	19.45

*Notes: *** significant at* 1%, *** significant at* 5%, ** significant at* 10%, *standard errors clustered at SIT zone level*
Source: Encuesta de Origen & Destino, 2012

Undereducated sub-sample

Table 3A. GMM

	Job-education mismatch
Public accessibility	-0.0069 (0.0217)
Private accessibility	0.3037*** (0.0955)
Kleibergen-Paap LM (*p-value*)	0.0009
Hansen J Test (*p-value*)	0.7265
Kleibergen-Paap F-Test	26.076
Yogo-Stock critical values (10% *maximal IV relative size*)	13.43

*Notes: *** significant at* 1%, *** significant at* 5%, ** significant at* 10%, *standard errors clustered at SIT zone level*
Source: Encuesta de Origen & Destino, 2012

Conclusions

- Private accessibility in developing economies seems not to diminish job-education mismatch, but to proxy household income; thus, the effect is exactly the contrary of what would be established by theory, and what Di Paolo, Matas, and Raymond (2017) found for Barcelona, Spain.
- Public accessibility would diminish job-education mismatch on overeducated people, as expected: less human capital would be wasted (Combes & Gobillon, 2015; Duranton, 2016), and commuting costs would become cheaper.
- We found evidence in favour of job-education mismatch in Medellin, Colombia.

References I

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2004). *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC.

Bocarejo, J. P., Portilla, I. J., Velásquez, J. M., Cruz, M. N., Peña, A., & Oviedo, D. R. (2014). An innovative transit system and its impact on low income users: the case of the metrocable in medellín. *Journal of Transport Geography*, *39*, 49–61.

Brueckner, J. K., & Martin, R. W. (1997). Spatial mismatch: an equilibrium analysis. *Regional Science and Urban Economics*, *27*(6), 693–714.

Brueckner, J. K., Thisse, J.-F., & Zenou, Y. (2002). Local labor markets, job matching, and urban location. *International Economic Review*, *43*(1), 155–171.

Büchel, F., & Battu, H. (2003). The theory of differential overqualification: Does it work? *Scottish Journal of Political Economy*, *50*(1), 1–16.

Büchel, F., & Van Ham, M. (2003). Overeducation, regional labor markets, and spatial flexibility. *Journal of Urban Economics*, *53*(3), 482–493.

Combes, P.-P., & Gobillon, L. (2015). The empirics of agglomeration economies. In *Handbook of regional and urban economics* (Vol. 5, pp. 247–348). Elsevier.

References II

Devillanova, C. (2013). Over-education and spatial flexibility: New evidence from italian survey data. *Papers in Regional Science*, *92*(3), 445–464.

Di Paolo, A., Matas, A., & Raymond, J. L. (2017). Job accessibility and job-education mismatch in the metropolitan area of barcelona. *Papers in regional science*, *96*(S1).

Duranton, G. (2016). Agglomeration effects in colombia. *Journal of Regional Science*, *56*(2), 210–238.

Frank, R. H. (1978). Why women earn less: the theory and estimation of differential overqualification. *The American Economic Review*, *68*(3), 360–373.

Fujita, M., & Ogawa, H. (1982). Multiple equilibria and structural transition of non-monocentric urban configurations. *Regional science and urban economics*, *12*(2), 161–196.

Gobillon, L., & Selod, H. (2014). Spatial mismatch, poverty, and vulnerable populations. In *Handbook of regional science* (pp. 93–107). Springer.

Gobillon, L., Selod, H., & Zenou, Y. (2007). The mechanisms of spatial mismatch. *Urban studies*, *44*(12), 2401–2427.

References III

Holzer, H. J. (1991). The spatial mismatch hypothesis: What has the evidence shown? *Urban Studies*, *28*(1), 105–122.

Ihlanfeldt, K. R., & Sjoquist, D. L. (1998). The spatial mismatch hypothesis: a review of recent studies and their implications for welfare reform. *Housing policy debate*, *9*(4), 849–892.

Jauhiainen, S. (2011). Overeducation in the finnish regional labour markets. *Papers in Regional Science*, *90*(3), 573–588.

Kain, J. F. (1968). Housing segregation, negro employment, and metropolitan decentralization. *The Quarterly Journal of Economics*, *82*(2), 175–197.

Kain, J. F. (1992). The spatial mismatch hypothesis: three decades later. *Housing policy debate*, *3*(2), 371–460.

Lucas, R. E., & Rossi-Hansberg, E. (2002). On the internal structure of cities. *Econometrica*, *70*(4), 1445–1476.

Páez, A., Scott, D. M., & Morency, C. (2012). Measuring accessibility: positive and normative implementations of various accessibility indicators. *Journal of Transport Geography*, *25*, 141–153.

Patacchini, E., & Zenou, Y. (2005). Spatial mismatch, transport mode and search decisions in england. *Journal of Urban Economics*, *58*(1), 62–90.

References IV

Ramos, R., & Sanromá, E. (2013). Overeducation and local labour markets in spain. *Tijdschrift voor economische en sociale geografie*, *104*(3), 278–291.

Zenou, Y. (2009). *Urban labor economics*. Cambridge University Press.