Wp

WORKING PAPERS

CITY SIZE, DISTANCE AND FORMAL EMPLOYMENT CREATION
06/10/2016
N° 2016/09

CITY SIZE, DISTANCE AND FORMAL EMPLOYMENT CREATION

O´Clery, Neave
Lora, Eduardo

DEVELOPMENT BANK OF LATIN AMERICA

WORKING PAPERS

CITY SIZE, DISTANCE AND FORMAL EMPLOYMENT CREATION

O´Clery, Neave
Lora, Eduardo

CAF – Working paper N° 2016/09
06/10/2016

ABSTRACT

Cities thrive through the diversity of their occupants because the availability of complementary skills enables firms in the formal sector to grow, delivering increasingly sophisticated products and services. The appearance of new industries is path dependent in that new economic activities build on existing strengths, leading cities to both diversify and specialize in distinct areas. Hence, the location of necessary capabilities, and in particular the distance between firms and people with the skills they need, is key to the success of urban agglomerations.

Using data for Colombia, this paper assesses the extent to which cities benefit from skills and capabilities available in their surrounding catchment areas. Without assuming a priori a definition for cities, we sequentially agglomerate the 96 urban municipalities larger than 50,000 people based on commuting time. We show that a level of agglomeration equivalent to between 45 and 75 minutes of commuting time, corresponding to between 62 and 43 cities, maximizes the impact that the availability of skills has on the ability of agglomerations to generate formal employment. Smaller urban municipalities stand to gain more in the process of agglomeration. A range of policy implications are discussed.

WORKING PAPERS

Wp

EL TAMAÑO DE LAS CIUDADES, LAS DISTANCIAS Y LA GENERACIÓN DE EMPLEO FORMAL

O´Clery, Neave
Lora, Eduardo

CAF - Documento de trabajo N° 2016/09
06/10/2016

RESUMEN

La prosperidad de cualquier ciudad depende de la diversidad de sus habitantes porque las complementariedades entre las habilidades permiten a las empresas crear empleo formal con el fin de producir bienes y servicios cada vez más sofisticados. La aparición de nuevas industrias depende de las que ya existen, pues los nuevos sectores se construyen a partir de los conocimientos productivos que ya se tienen, en un proceso de diversificación y especialización en distintas actividades. Esto implica que la distancia entre las empresas y las personas con las habilidades necesarias debe influir en las posibilidades de generación de empleo formal de las aglomeraciones urbanas.

Utilizando datos de Colombia, este estudio evalúa el grado en que las ciudades se benefician de las habilidades y capacidades disponibles en sus zonas de influencia circundantes. Sin asumir a priori una definición de ciudad, secuencialmente se agregan los 96 municipios urbanos de más de 50.000 personas basándose en los tiempos de trayecto entre ellos. Se demuestra que el máximo aprovechamiento de las habilidades para generar empleo formal se consigue en las aglomeraciones urbanas de varios municipios localizados en un radio hasta entre 45 y 75 minutos de tiempo de viaje en automóvil (lo que corresponde a entre 62 y 43 ciudades. Los municipios urbanos pequeños pueden ganar más que los grandes en el proceso de aglomeración. Como conclusión del estudio se discuten diversas implicaciones para las políticas urbanas.

City size, distance and formal employment creation

Neave O'Clery & Eduardo Lora
Final version: October 6, 2016

ABSTRACT

Cities thrive through the diversity of their occupants because the availability of complementary skills enables firms in the formal sector to grow, delivering increasingly sophisticated products and services. The appearance of new industries is path dependent in that new economic activities build on existing strengths, leading cities to both diversify and specialize in distinct areas. Hence, the location of necessary capabilities, and in particular the distance between firms and people with the skills they need, is key to the success of urban agglomerations.

Using data for Colombia, this paper assesses the extent to which cities benefit from skills and capabilities available in their surrounding catchment areas. Without assuming *a priori* a definition for cities, we sequentially agglomerate the 96 urban municipalities larger than 50,000 people based on commuting time. We show that a level of agglomeration equivalent to between 45 and 75 minutes of commuting time, corresponding to between 62 and 43 cities, maximizes the impact that the availability of skills has on the ability of agglomerations to generate formal employment. Smaller urban municipalities stand to gain more in the process of agglomeration. A range of policy implications are discussed.

Key words: labor formality, complexity, city size, commuting, diversification, networks.

1 Introduction

Although the relationship between urbanization and informality has been central to the economic development literature since the seminal work by Fields ([1], see also Ghani and Kanbur [2]), the role that city size and commuting distances play in the creation of formal employment (and the reduction of informality) in developing countries has been largely ignored. The objective of this paper is to assess the extent to which the ability of cities to create formal employment depends on their population size due to the diversity of skills available, and the extent to which commuting distances limit that ability.

Cities exist because of the tendency of firms to cluster together. Many arguments have been advanced to explain why it is convenient for firms to locate near other firms. The traditional view has been that access to valuable resources (water, energy sources, etc.) encourages firms to cluster. This in turn attracts people, creating an additional and no less important reason to cluster: access to a pool of workers. Well-educated workers further stimulate productivity, and make cities less constrained by traditional determinants [3]. Under this view, the dynamism of a city comes from making good use of the resources available to firms and individuals through more efficient production, distribution and consumption processes. The expansion and prosperity of a city can be halted or reversed by forces that operate in the opposite direction, such as overuse of natural resources, environmental degradation or congestion. Local government interventions, such as public infrastructure investments, urban planning and housing controls, can prevent or postpone these events. Education and training services, as well as providing the urban amenities most valued by firms and workers, may help cities remain competitive. On the other hand, excessive or misguided regulation and planning may raise housing and transportation costs, prompting firms and workers to migrate to other cities [4].

Although useful to explain the rise and decay of some cities, these views of why firms cluster fail to explain why large cities keep growing, in spite of their higher costs with respect to smaller cities. Since the invention of the automobile, many have predicted the demise of the city as cars and trucks free workers and firms to locate in more convenient and less expensive places. As Robert Lucas noted: "If we postulate only the usual list of economic forces, cities should fly apart. The theory of production contains nothing to hold a city together... What can people be paying Manhattan or downtown Chicago rents for, if not for being near other people?" [5] Since Lucas' assertion in the late eighties, scholars have recognized that firms and workers concentrate in cities because of "externalities" associated with human capital: larger pools of educated people facilitates critical interaction between workers. Such interactions accelerate the diffusion of ideas within and across industries, and lead to faster innovation and productivity increases.

Gradually, and more recently, this view has led many to rediscover the importance of diversity in urban life, as stressed by Jane Jacobs in the late fifties and sixties to oppose the urban planning ideas then in vogue [6]. While urban planners recommended displacing large neighborhood populations to facilitate massive-scale rebuilding, and clear demarcation of zones by activity, Jacobs "extolled density, complexity and diversity and pointed out the advantages of narrow streets, short blocks, mixtures of old and new buildings, and mixtures of commercial, cultural, and residential uses" [7]. In Jacobs' view, the economic and social success of a city depended on its ability to facilitate human interactions and a diversity of economic activities.

In this paper, we focus on skill diversity and its relation to city size and commuting times within cities, although other dimensions of diversity may also be relevant. Skill diversity is a source of externality because of the complementarities across skills. Larger cities are more productive to the extent that they have a wider variety of skills that can be combined in ways smaller cities cannot facilitate. Larger cities are more productive not because they have increased numbers of similar individuals, but because there tends to be more variety in larger social settings.

Following this line of reasoning, Bettencourt, Samaniego and Youn [8] recently analyzed the diversity of US metropolitan areas in terms of professional diversity, and proposed a theoretical framework that explains the higher productivity of larger cities via the appearance of new occupations as the result of specialization and coordination of labor. In other words, as workers become more specialized and diverse, their interactions become more productive. The larger the city, the stronger these forces become. In this approach, cities are seen as living organisms that tend to change due to the interactions of individuals and the resulting discovery of new occupations and new ways of producing an always changing basket of goods and services. Alien to this perspective is the concept of static equilibrium that is the backbone of many economic models applied to urban issues.

Although cities can be thought as large cauldrons where capabilities are constantly combined and recombined, the actual scale at which this takes place is the firm. Firms are productive to the extent to which they are able to coordinate a range of skills in order to produce existing goods more efficiently, or develop new more sophisticated goods that consumers value more than current goods. Since the number and size of firms is not fixed, as firms become more diversified and productive they can absorb more workers. As a result, something short of miraculous takes place: larger cities generate proportionally more employment (at firms) than smaller cities. To understand this it is convenient to see it from the opposite angle: larger cities have lower self-employment and informality rates than smaller cities. In larger cities, occupation differentiation makes it possible for firms to combine a diversity of skills to produce what in smaller cities is produced by family-run or individual businesses. The patterns just described are more pronounced in the developing world, but hold true also in the US [9]. Figure 1 shows the strong relationship between city size and the percentage of working age population in formal jobs in Colombia and Mexico.

The fact that formal occupation rates are higher in larger cities is due to the expansion of existing sectors and the appearance of new ones. But new sectors do not pop up randomly as cities grow. If they did, cities would tend to become similar to each other. But the opposite happens: New York is a financial center and San Francisco is a high-tech hub. Houston is strong in the petrochemical industry while Los Angeles thrives in the entertainment sector. New capabilities of individuals and firms build on existing strengths, meaning occupation and industry differentiation processes are strongly path-dependent [10, 11], leading cities to both diversify and specialize in distinct areas.

Beyond the city-wide availability of skills, the location of capabilities, and in particular the distance between firms and workers with the skills they need, is key to the success of the complex economic activities that drive the growth of the formal sector. For example, cities may benefit from skills and capabilities available in their surrounding catchment areas, and neighboring cities. Boston, for example, profits from a high density of world-class universities in neighboring Cambridge, and has become a leading hub for technology and education start-ups in recent years. The importance of distance, location and accessibility in the growth of cities has

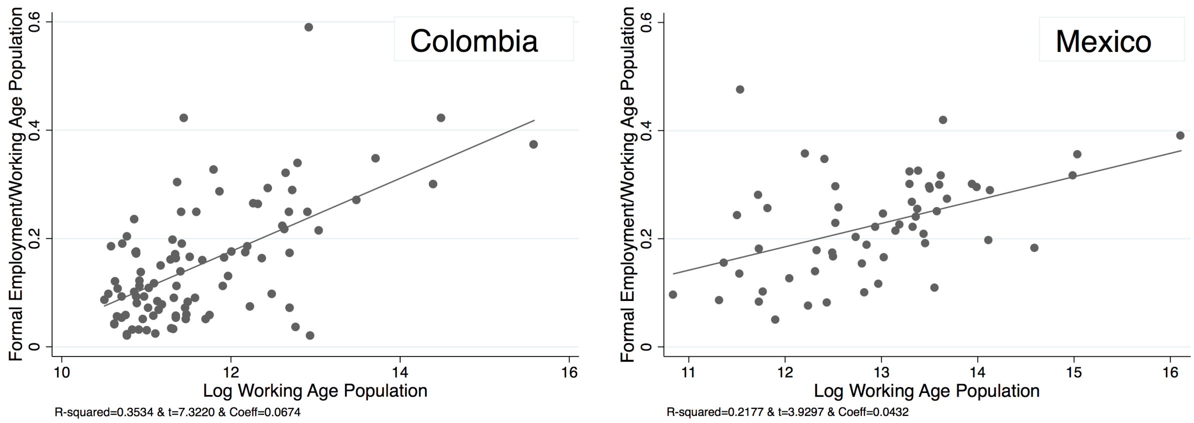

Figure 1: The ratio of formal employment to working age population versus the working age population for Colombia (2013) and Mexico (2013). We observe that the proportion of working age population in formal employment is higher in larger cities.

been well-studied by those interested in commuting zones [12, 13], transportation links [14, 15] and industry clusters [16]. Here we examine the growth of formality in relation to skill availability both in the city, and in nearby municipalities and cities based on commuting times.

Importantly, we do not assume *a priori* a definition for urban agglomerations in terms of their constituent municipalities. We develop a new approach, based on the economic impact of skill availability for the growth of formal employment, to identify an optimal level of aggregation of municipalities based on commuting time. This approach differs from existing definitions of cities or metropolitan areas (e.g., [13]) as it does not rely on an arbitrary threshold or criterion which might be inefficient for policy purposes and manipulated for political gain.

We focus our analysis on Colombia, partly due to data availability, but also because Colombia is a particularly interesting case due to profound regional differences rooted in a distinct geographical and historical context [17]. As a result, cities are very heterogenous in their industry composition and in their degree of connection to other cities and to the rest of the world. By some indicators, Colombia is one of the most geographically fragmented countries in the world [18]. Urban formal occupation rates increased considerably in Colombia between 2008 and 2013, with an average increase of 4 percent points and a standard deviation of 3.8 points across cities. As roughly two thirds of Colombian exports are mining products, the commodity boom fuelled economic growth, which reached 4.2% per annum between 2008 and 2013. A system of redistributing oil rents (and other tax revenues) to all municipalities spurred local government expenditures. The elimination in May of 2013 of payroll taxes representing 5% of the wage bill may have also encouraged formal employment, as intended by the government.

1.1 Objectives

Within this context, this paper attempts to shed light on several issues about the relationship between formal employment creation, city size and commuting times across urban municipalities and cities:

1. Is there a strong (contemporaneous) relationship, as suggested by the theory, between city size and formal occupation rate (defined as formal employment/working age population), and between city size and skill diversity and sophistication (which is not observed directly, see below)?
2. Although we will not be able to assess causality in the strict sense, is there strong evidence that the availability of diverse skills is associated with subsequent growth of formal employment at the city level?
3. Does traveling distances affect the ability of cities to create formal employment?

A range of policy implications can be derived from our answers to this list of research questions. One important issue is whether huge urban agglomerations should be discouraged or not, and how to take advantage of the tendency of big cities to absorb neighboring municipalities. Our results provide arguments for coordinating transportation, planning and industry re-localization policies across urban municipalities of different

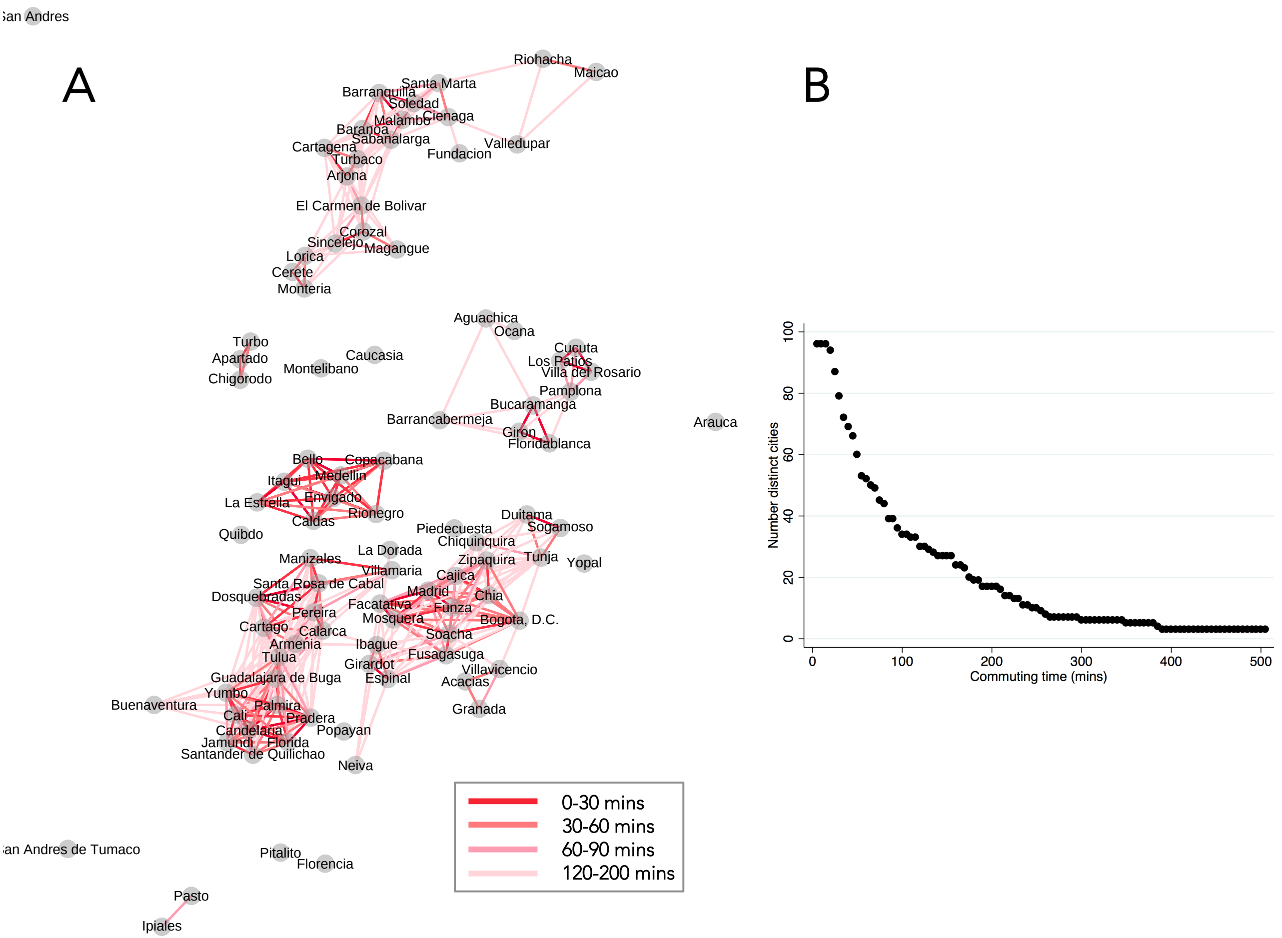

Figure 2: In subfigure A we show a network with nodes which represent urban municipalities in Colombian, roughly geographically located in space. Municipalities are connected by a colored edge if the commuting time is below the threshold shown in the legend. At each time-step we agglomerate connected municipalities in the network to produce a reduced set of cities (this is equivalent to finding all connected components in the network). In subfigure B we show the number of distinct cities (i.e. the size of this reduced set) as time increases. Eventually all municipalities will be connected (provided there is a road link) as shown.

sizes, especially if they are part of the same urban agglomeration. Our analysis sheds light on the importance of improving transportation infrastructure to connect small municipalities to urban agglomerations.

2 Data, Methodology and Key Variables

Although our units of analysis are cities, we do not predefine the number of cities. We use population data for the 96 municipalities with urban population larger than 50,000 according to DANE (the National Statistics Office) and construct cities as combinations of one or more of these 96 'urban municipalities' using commuting time radius (computed from Google Maps) around their centroids. If the time radius is zero, the number of cities is, by definition, the same number of urban municipalities. As the commuting time radius increases, the number of cities falls, as some urban municipalities become part of multi-municipality agglomerations (further details of the computing algorithm are provided in Section 5.1 below). As Figure 2 B shows, with a 60-minute radius, the number of cities falls to 52 and with a 120-minute radius it falls to 30. We will let the econometric results tell us what is the commuting time radius (and therefore the number of cities) beyond which urban agglomerations do not longer operate effectively as cauldrons of skill diversity. Reassuringly, as we will see, the radius at which this occurs is roughly consistent with the concept of (multi-municipality) metropolitan areas defined on the basis of labor commuting flows across municipalities, a concept previously applied to Colombia by [13].

The dependent variable in our regressions will be formal occupation rates (in levels or changes) by city (further details below). The main explanatory variable will be 'complexity potential' by city, a measure of the possibilities open to the existing labor pool (given their skills) to move to more complex industries that require similar skills but are not yet present in the city. Its computation involves measures of 'industry presence', 'industry complexity', and 'skill similarity between industries', which are explained below.

The main database to compute all these variables at the city level will be the PILA (the Integrated Report of Social Security Contributions), managed by the Ministry of Health. It contains information on formal employment, wages and number of firms by municipality and industry for 2008-2013. All types of sectors, including goods and services, are included (note that we use the terms 'sector' and 'industry' interchangeably). This is the source for constructing the following variables:

- *Formal employment of industry i in a city c* ($femp_{c,i}$) is defined as employment covered by the health social security system and/or the pension system (the self-employed are not included). Formal employment by industry and city in 2008 and 2013 is computed as the number of formal employees in an average month [1].

- *Formal occupation rate of any industry i in a city c* ($FOR_{c,i}$) is defined as formal employment in the industry divided by population older than 15 in city ($wpop_c$):

$$FOR_{c,i} = \frac{femp_{c,i}}{wpop_c}.$$

  Population data come from estimates by DANE.

- *Presence of industry i in city c* is measured by a location quotient, LQ (also known as revealed comparative advantage), which reflects the relative importance of an industry in a city given its overall distribution. An industry is present in a city when the corresponding location quotient is larger than one.

$$LQ_{c,i}(t) = \frac{femp_{c,i}(t)/\sum_i femp_{c,i}(t)}{\sum_c femp_{c,i}(t)/\sum_c \sum_i femp_{c,i}(t)}$$

- *Industry complexity C_i* is a measure of the range of capabilities needed by industry i. Capabilities or skills are multidimensional variables that are not directly observable, which should not be confused with years or type of education. What is observable is the outcome of productive capabilities, namely the diversity and uniqueness of the goods and services produced by teams of workers deploying their collective know-how. Originally proposed by Hausmann and Hidalgo [10, 19, 20] for export products and adapted by us [21] for industries in Colombia, industry complexity is computed on the basis of 'diversity', which is the number of industries present in a city, and 'ubiquity', the number of cities where an industry is present. The computation proceeds as follows.

  Let M be a matrix with entry $M_{c,i} = 1$ if city i has LQ>1 in industry i. The diversity of city c and ubiquity of industry i are defined as

$$k_{c,0} = \sum_i M_{c,i} \text{ and } k_{i,0} = \sum_c M_{c,i}$$

  Then the average diversity of city c and the average ubiquity of industry i may be expressed as:

$$k_{c,1} = \frac{1}{k_{c,0}} \sum_i \hat{M}_{c,i} k_{i,0} \text{ and } k_{i,1} = \frac{1}{k_{i,0}} \sum_c \hat{M}_{c,i} k_{c,0}.$$

  In this way, the diversity of a city is weighed by the ubiquity of the industries present in the city and, similarly, the ubiquity of an industry is weighed by the diversity of the cities where the industry is present. If this calculation is done iteratively to step n, the two previous expressions become:

$$k_{c,n} = \frac{1}{k_{c,0}} \sum_i \hat{M}_{c,i} k_{i,n-1} \text{ and } k_{i,n} = \frac{1}{k_{i,0}} \sum_c \hat{M}_{c,i} k_{c,n-1}$$

  which can be expressed in closed form for industry i:

$$k_{i,n} = \sum_{i'} \tilde{M}_{ii'} k_{i',n-2}$$

[1] Notice that data on informal employment at the municipality level is not available. There is data for the 13 official 'metropolitan areas', which do not correspond to our aggregations of municipalities or to Duranton's definition of metropolitan areas. Labor participation rates are positively correlated with city size for the 13 official mets and informality rates (as defined by DANE) are negatively correlated with city size.

with entries of matrix $\tilde{M}$:

$$\tilde{M}_{ii'} = \sum_c \frac{M_{c,i} M_{c,i'}}{k_{c,0} k_{i,0}}$$

Hence, if k_n is a vector whose ith element is $k_{i,n}$ then:

$$\mathsf{k}_n = \tilde{M} \mathsf{k}_{n-2}$$

The long-run solution of this system is found by the 'Method of Reflections' applying eigenvalue methods. The second largest eigenvector of $\tilde{M}$ is taken as the industry complexity (C_i). We use the calculations made by us for the Colombian Atlas of Economic Complexity [21], where further details can be found. Industry complexity is used to calculate complexity potential (see below), the key explanatory variable of our model.

- *City complexity E_c of a city c* is a measure of the range of skills or capabilities available in a city. It can be computed jointly with industry complexity. For that purpose, city complexity can be taken as the second largest eigenvector of $\tilde{M}$ computed by city, instead of by industry. Alternatively, city complexity can be computed as the average of the industry complexity of the industries present in the city (which is mathematically equivalent to the second largest eigenvalue when the computation is done jointly). This is the way it is computed for this paper. We use city complexity only to assess whether the contemporaneous correlation between formal occupation and city size may be due to the availability of skills (see Table 1).
- *Skill similarity between industries.* Skills are not observed directly, but rather the availability of skills specific to a particular industry is captured via the presence of other industries requiring similar capabilities. We develop a measure of *skill relatedness* between industry pairs by counting job switches as the most direct proxy for 'cognitive distance' or skill transferability between industries [22]. In turn, skill similarity is used to compute a measure of *skill availability* for each industry at the city level. The degree to which an industry is related to any other industry is computed as a matrix with entries:

  $$S_{i,j} = \frac{\phi_{i,j} / \sum_j \phi_{i,j}}{\sum_i \phi_{i,j} / \sum_{i,j} \phi_{i,j}}$$

  where $\phi_{i,j}$ is the number of job switches between industry i and industry j (between year t and year $t+1$). Since this matrix is asymmetric, in order to make it symmetric it is averaged with its transpose. Then, it is re-scaled in a range from -1 to 1, and the re-scaled values are used to compute:

  $$A_{i,j} = \frac{S_{i,j} + S_{j,i} - 2}{S_{i,j} + S_{j,i} + 2}. \tag{1}$$

  Finally, only positive values of this matrix are taken into account, since those represent pair-wise industries with more job switches than expected[2].
- *Complexity potential* of a city (CP_c) is, as mentioned, a measure of the possibilities facing the city to move to more complex industries that are not yet present in the city, taking into account the existing skills of the labor force. To compute *complexity potential* in the initial period (2008), the following expression is used

  $$CP_c = \frac{1}{|M_c|} \sum_{i \in M_c} d_{c,i} C_i, \tag{2}$$

  where M_c denotes the set of 'missing' industries for city c (LQ<1), and the $C_i \in [0,1]$ is the normalized complexity of industry i. The distance weighting factor $d_{c,i}$ (also known as 'density' in the literature [10, 23, 24]) is defined as

  $$d_{c,i} = \frac{\sum_{j \in N_c} A_{i,j}}{\sum_j A_{i,j}}.$$

  where N_c is the set of industries that is present in city c. Essentially, for each missing industry in a city, we compute industry complexity times density, the latter of which can be thought of as the likelihood of an industry appearance based on presences in industries with similar skills. The complexity potential for a city is the mean value of this product over all missing industries.

[2]More precisely, if $(S_{i,j} + S_{j,i})/2 > 1$, which occurs when the average of the bi-directional switches between industries i and j is more frequent than expected, then $A_{i,j} > 0$ in Equation 1.

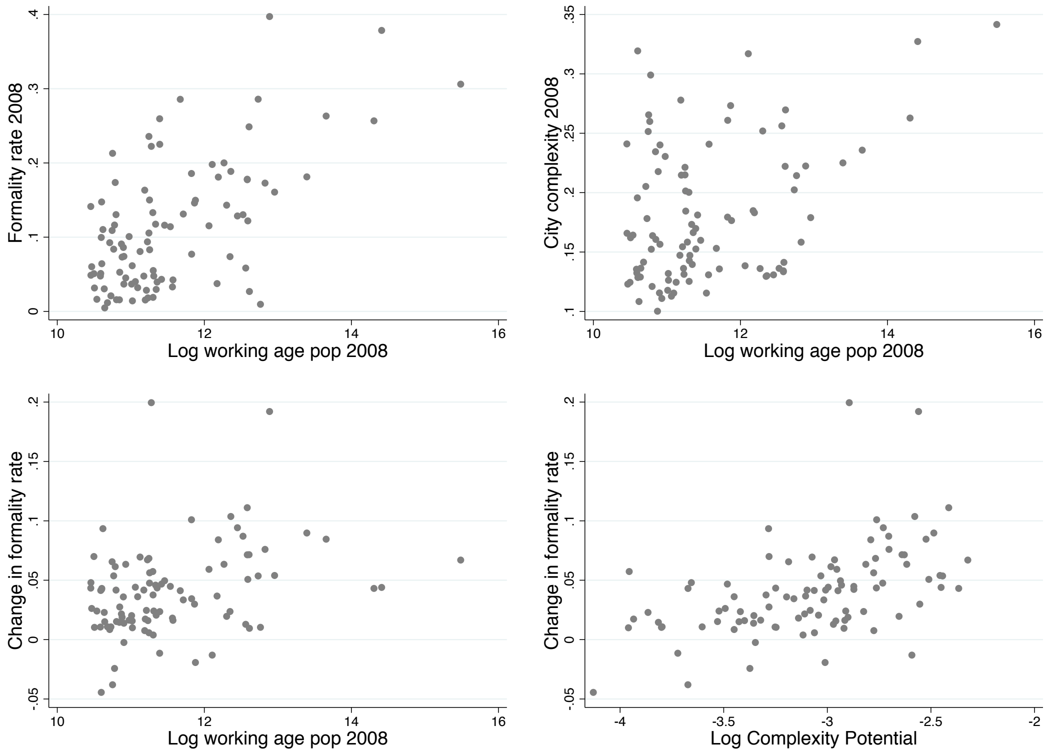

Figure 3: The two top panels show that both the formality rate and the city complexity level increase with working age population for our cross-section of 96 cities (2008). The two bottom panels show that the change in formality rate (2008-2013) is larger in cities that have more population and higher complexity potential.

Other data sources/explanatory variables include:

- *Working age population* ($wpop$), as defined above.
- *GDP per capita* ($GDPpc$), which is available from 2011 onwards, is calculated from GDP by municipality estimated by DANE.
- *Oil producing city*, a binary variable, which takes the value of one if the city has more than one oil well in production per thousand inhabitants. Oil well data refers to 2014, as reported by Ecopetrol (the Colombian hydrocarbon company) for their own internal records.
- *Government spending shock* is the change between 2008 and 2013 in total government spending (in 2008 prices) per working-age person. It is computed from municipality-level government spending data compiled by CEDE [25].
- *Sectoral demand shocks*, sds_c, is a Bartik-style measure [26] that summarizes for each city the mix of nationwide sectoral demand shocks facing the city. It is computed as

$$sds_c = FOR_c \sum_i \frac{femp_{c,i}(2008)}{femp_c(2008)} g_{i,c} \tag{3}$$

  where $g_{i,c} = log[femp_i(2013)] - log[femp_i(2008)]$ is growth of employment of industry i *excluding employment in industry i in city c*. In other words, here $femp_i = \sum_{j \in J} femp_{i,j}$ with set J containing all cities except city c. It can be interpreted as the expected change in the formal occupation rate of the city given the nationwide sectoral demand shocks (exogenous to the city).
- The *traveling distance* between each pair of cities/urban municipalities is measured using commuting time. Given the latitude and longitude for the centroid of each municipality, we use the Google Maps API to obtain typical commuting times by private car between each pair.

Figure 3 shows how some of the variables just described are related for the 96 urban municipalities. The two top panels show that larger cities have higher formal occupation rates and more complex productive systems. The left bottom panel suggests that larger cities tend to (subsequently) have even higher formal occupation rates, while the right bottom panel suggests the same for cities with higher complexity potential.

VARIABLES	(1)	(2)	(3)
Log working age pop 2008	0.055***	0.048***	0.046***
	(0.007)	(0.007)	(0.007)
Log GDP per capita 2011		0.066***	0.063***
		(0.012)	(0.013)
City complexity 2008			0.101
			(0.114)
Constant	-0.517***	-1.050***	-1.019***
	(0.082)	(0.113)	(0.120)
Observations	96	96	96
R-squared	0.371	0.551	0.555

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

VARIABLES	(1)	(2)	(3)
Log working age pop 2013	0.067***	0.059***	0.062***
	(0.009)	(0.008)	(0.008)
Log GDP per capita 2013		0.092***	0.098***
		(0.017)	(0.020)
City complexity 2013			-0.189
			(0.147)
Constant	-0.633***	-1.391***	-1.448***
	(0.104)	(0.168)	(0.186)
Observations	96	96	96
R-squared	0.353	0.558	0.565

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1: Here we show the cross-sectional relationship between the formal occupation rate of cities and several explanatory variables for 2008 and 2013. The size of the working age population and income level of cities have the strongest correlations with the formal occupation rate, both in 2008 and 2013. However, the coefficients of these two variables are substantially larger in 2013, suggesting that the contemporaneous correlations are not stable and do not capture the processes underlying formal employment creation *changes*.

VARIABLES	(1)	(2)	(3)	(4)	(5)
Log Complexity Potential	0.036***	0.039***	0.033***	0.028**	0.033***
	(0.012)	(0.012)	(0.011)	(0.012)	(0.011)
Log working age pop 2008	-0.004	-0.004	-0.002	-0.005	-0.003
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Log GDP per capita 2011	0.008	-0.014*	-0.000	0.001	-0.014*
	(0.009)	(0.007)	(0.009)	(0.008)	(0.007)
Binary oil: one well/1000		0.077***			0.054**
		(0.026)			(0.022)
Govt spending pc			0.029**		0.015
			(0.013)		(0.010)
Sectoral demand				1.340***	0.484
				(0.467)	(0.515)
Formality rate 2008	0.097	0.131	0.108	-0.217*	0.013
	(0.084)	(0.083)	(0.084)	(0.123)	(0.150)
Constant	0.110	0.310***	0.131	0.153	0.277***
	(0.104)	(0.107)	(0.109)	(0.107)	(0.102)
Observations	96	96	95	96	95
R-squared	0.309	0.435	0.382	0.390	0.463

Table 2: This set of regressions explores the determinants of the change of the formal occupation rate between 2008-2013 for a set of 96 single-municipality cities. The results show a consistent positive relationship between complexity potential - the proximity of a city to potential complex industries - and the change of formality, controlling for relevant factors. Several exogenous shocks also influence formal employment.

3 Econometric Results Ignoring Distance

We start our empirical analysis running regressions for the 96 urban municipalities, ignoring the degree of connectedness between them. The role of commuting distance is studied further below.

We assess first the contemporaneous correlation, in 2008 and 2013, between formal occupation rate, working age population, GDP per capita and city complexity. Regressions are of the form:

$$FOR_c(t) = \alpha_0 + \alpha_1 \log wpop_c(t) + \alpha_2 logGDPpc_c(t) + \alpha_3 E_c(t). \quad (4)$$

Table 1 confirms that formal occupation rate across cities (the dependent variable) are contemporaneously associated with city size, measured by working age population, and with aggregate productivity, measured by GDP per capita. Interestingly, city complexity does not appear to be significantly correlated with formal occupation, after controlling for the variables mentioned. While the results hold both for 2008 and 2013 (left and right panels, respectively), the coefficients of the significant variables are higher in the latter (the point estimates fall beyond the initial 95% confidence interval), suggesting that cross-sectional relationships between the variables are not stable and cannot be used to predict changes in the dependent variable. In other words, had the initial coefficients remained constant, formal employment growth between 2008 and 2013 would have been much lower than observed. How can we explain *changes* in formal employment?

As argued above, our main hypothesis is that the ability of larger cities to create (proportionally) more formal employment is a result of the larger availability of skills, which may generate productive possibilities that do not exist in smaller cities. If this is so, our measure of complexity potential must help to predict future changes in formal occupation. Formally, we test this hypothesis with regressions of the form:

$$\Delta FOR_c(t+1) = \beta_0 + \beta_1 \log CP_c(t) + \alpha_2 FOR_c(t) + \text{controls} \tag{5}$$

where the controls include the working age population in 2008, and GDP per capita in 2011.

Table 2 presents the results. The ability of cities to create formal employment strongly depends on complexity potential, a measure of the availability of skills to develop more complex industries. The variable is strongly significant in all the regressions. The coefficient of complexity potential is very stable, suggesting that a 10% increase in complexity potential is followed in a five-year period by an increase of about 0.28-0.39 percent points of the formal occupation rate.

Working age population, our measure of city size, often gets the wrong sign and shows no robust relation with the dependent variable. This suggests that, reliant on population size alone, a larger city cannot be expected to create (proportionally) more formal employment than a smaller city. Neither is initial per capita income level a good predictor of formal employment growth. This is relevant because it suggests that the initial availability of capital (physical or human) per capita, or its productivity, does not facilitate further formal employment creation.

Formal employment creation may be influenced by a variety of exogenous shocks coming from the presence of the oil sector in the city, from changes in government spending or from nationwide sectoral demand fluctuations. Each of the shocks taken separately does have a significant influence on formal employment creation, but barely affects the coefficient of complexity potential. This sheds light on an important issue: if cities are seen as economic systems that move back towards equilibrium when subject to shocks, complexity potential could not play an independent role in the ability of cities to create formal employment. However, if the concept of economic equilibrium is abandoned (as suggested in the introduction), complexity potential may operate as an endogenous or organic source of formal employment creation.

Therefore, our initial results, which ignore distance, suggest that the availability of a diverse and sophisticated pool of skills (as measured by complexity potential), rather than sheer size, is what allows larger cities to generate more formal employment. Exogenous shocks may also have an influence in formal employment creation but are by no means a necessary condition for a city to make better use of its existing skills or for developing new ones (as it would be the case if cities gravitated towards equilibrium).

So far we have equated cities with urban municipalities (with populations larger than 50,000), in spite of the fact that urban agglomerations may consist of several municipalities. Is the influence of complexity potential on (subsequent) formal employment growth enhanced or weakened as we expand our definition of a city to include more than one urban municipality? And, if so, at what traveling distance do agglomerates optimally combine skills resulting in the growth of formal employment?

4 Distance

Here, we wish to investigate whether collections of neighboring or proximate urban municipalities can be aggregated, based on commuting distance, to form economically relevant urban agglomerations. There are many reasons why we might wish to consider a collection of urban municipalities as a single group. In most cases, such a grouping constitutes an integrated labor market, where workers from outer zones commute to work in an urban core. There may also be clusters of industries located in more peripheral areas, but benefiting from proximity to a rich supply of skills and inputs located in the urban center. In such cases, previous municipal boundaries may no longer be practically relevant, and funding, policy and development priorities may be segmented, ill-targeted and inefficient if defined and implemented at the municipality level. Hence, the identification of distinct city boundaries, using quantitive and evidence-based means that are not prone to political manipulation, is critical to the effective allocation and organization of resources for cities.

A previous effort to identify such agglomerations in Colombia was made by Duranton [13], who developed an algorithm to sequentially connect municipalities to their urban core (municipality) using a threshold for the magnitude of commuting flows as key criteria. While conceptually similar to our approach, which uses commuting times as opposed to actual labor flows, this work does not assess whether the agglomerations identified are economically meaningful. Furthermore, a somewhat arbitrary threshold (although checked for robustness) is used to produce a final list of cities.

Here, we employ commuting times as a measure of distance between the 96 urban municipalities identified in the previous section. We use commuting times instead of flows as we wish to measure the potential for one municipality to access the skills and diverse labor force of a proximate municipality. We then measure the extent to which complexity potential explains the growth of formal employment via our model introduced above. We test this model for all commuting time thresholds, and identify at what level of aggregation the model performs better. At this level, the growth of formality in the agglomerations identified is well-described by their proximity to potential complex industries. Importantly, this algorithm does not supply any input parameters or thresholds, it simply allows us to assess the validity of the agglomeration in question via the ability of complexity potential to describe the growth of its formal occupation rate.

In the following section, we describe the algorithm in more detail.

4.1 Algorithm

The algorithm proceeds as follows.

Each matrix $G^{\tau} \in \mathbb{R}^{n \times n}$ where $n = 96$ represents the network connectivity at time τ. Initially, G^0 contains only zeros. The matrix $D \in \mathbb{R}^{n \times n}$ contains the commuting times between all 96 initial urban municipalities (in minutes), obtained from Google Maps.

For each commuting time $\tau = 1, .., 200$,

1. We set $G^{\tau}_{i,j} = 1$ if the commuting time between city i and j is less than time τ, i.e., $D_{i,j} < \tau$ (here τ is in minutes).
2. We use Stata package *nwcommands* to detect connected components in the network G^{τ}. Each component maps to a set of municipalities (or a single municipality) from the initial set of 96. In the following steps, we will refer to each of these sets as a 'city' (which may be either single- or multiple-municipality).
3. We aggregate employment by municipality and industry - combining municipalities within cities identified in step 1. We then compute the Location Quotient. We compute the set of present (LQ>1) and missing industries (LQ<1) for each city, and then discard information about all non-urban municipalities (i.e., those smaller than 50,000 inhabitants or larger but rural).
4. We then compute the complexity potential for each city as outlined in (2) above, as well as the resulting formal employment rates in 2008 and 2013 and their changes between the two periods.
5. For each new city we also compute the control and shock variables, namely working age population, GDP per capita, a binary variable for at least one oil well per 1000 people, the government spending shocks, and the Bartik-style sectoral demand shocks.
6. And finally we run the specification in (5), and store the coefficient and adjusted R2 (shown in Figure 4).

Note: the industry complexity C_i and the matrix of industry proximities $A_{i,j}$ are fixed, and only the set of present and missing industries (N_c and M_c) vary for each τ/set of cities.

4.2 Econometric Results Including Distance

The first subfigure of Figure 4 plots the coefficient of complexity potential obtained from regressions that include the same regressors of column 5 of Table 2), allowing the set of cities to change as a function of commuting-time as described above. The coefficient is shown in red, with the confidence intervals shown in grey. We observe that the coefficient is significant up to about 90 minutes, which is equivalent to 35 individual cities. After this point the confidence interval strays into negative territory. We can interpret this as suggesting that cities benefit from skilled labor located up to about a 90 minute radius. The diversity and sophistication of this labor allows a city to move into more complex industries and increase its formal occupation rate. Within this significantly non-negative range, the coefficient is largest between 45 and 75 minutes, equivalent to between 62 and 43 cities, shown via the two dark vertical lines. The top right subfigure shows the corresponding R-squared, which tends to increase until 75 minutes, or 43 cities. The bottom left figure repeats the top left, but instead the coefficient varies with the number of distinct cities. Similarly, on the bottom right we show the corresponding R-squared.

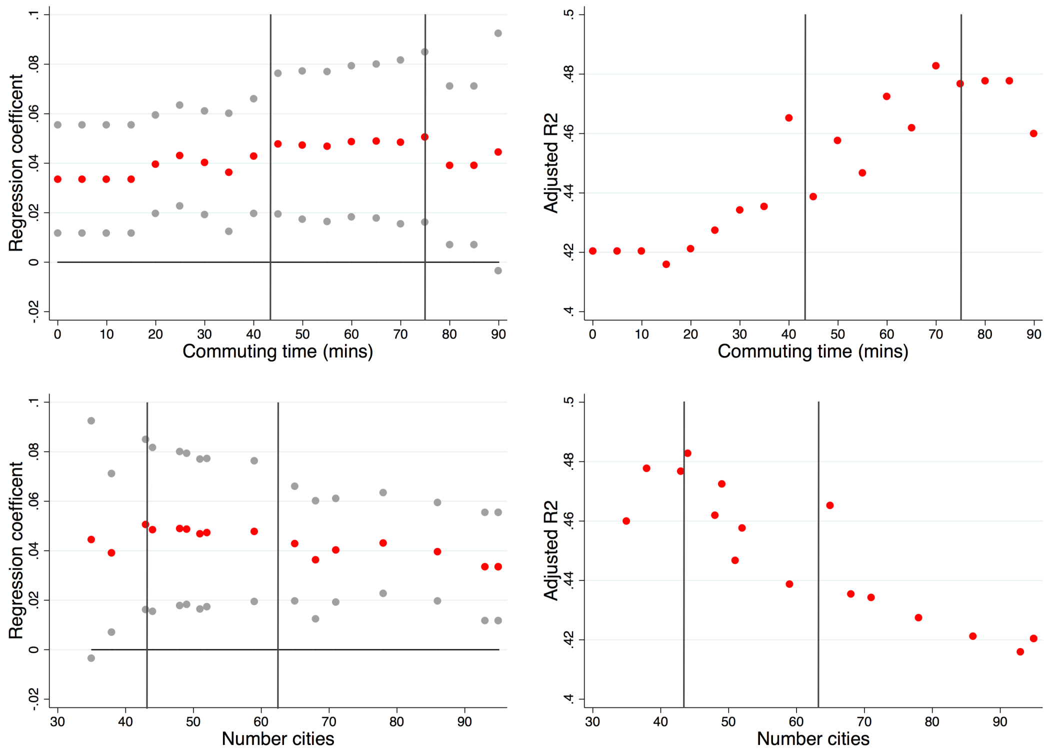

Figure 4: The left-side panels show the coefficient of complexity potential (including all controls of column 5 of Table 2) for different city definitions based on commuting-time aggregates (as described in main text). The coefficient is shown in red, with the confidence intervals shown in grey. We observe that the coefficient is significant up to about 90 minutes (35 cities) - with a high plateau between 45 and 75 minutes (or 62 to 43 cities, delimited by dark lines). The right-side panels show the corresponding R-squared.

Figure 5 illustrates the connectivity between the 96 urban municipalities (nodes) to form 62 cities, the largest number of cities within the high plateau mentioned (this corresponds to edges between nodes for which the commuting time is less than 45 minutes). Nodes are colored corresponding to the resulting 62 cities, i.e, the distinct connected components in the network. As expected, we observe several larger agglomerations including many proximate municipalities.

Table 3 presents our full set of regression specifications for these 62 cities. Table 4 presents an identical set of regressions for the 48 single-municipality cities that are part of our 62 cities (the remaining 48 urban municipalities being part of the 14 multiple-municipality cities).

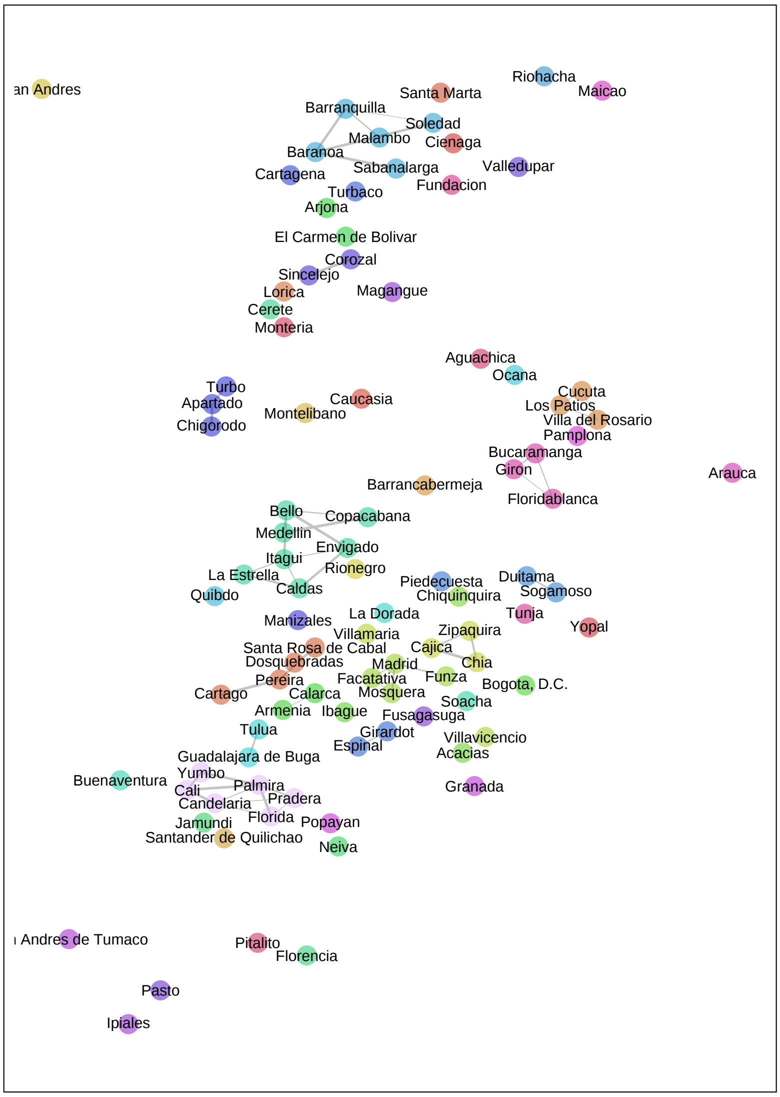

Figure 5: Here we show edges corresponding to commuting times less than 43 minutes. Nodes are colored corresponding to the resulting 62 cities (connected components).

VARIABLES	(1)	(2)	(3)	(4)	(5)
Log Complexity Potential	0.052***	0.048***	0.052***	0.050***	0.049***
	(0.015)	(0.014)	(0.015)	(0.014)	(0.014)
Log working age pop 2008	-0.016**	-0.011**	-0.014***	-0.014***	-0.011**
	(0.006)	(0.004)	(0.005)	(0.005)	(0.005)
Log GDP per capita 2011	0.017**	-0.004	0.009	0.014*	-0.007
	(0.007)	(0.009)	(0.007)	(0.007)	(0.010)
Binary oil: one well/1000		0.054**			0.054**
		(0.027)			(0.022)
Govt spending pc			0.018		0.014
			(0.015)		(0.013)
Sectoral demand				0.551	-0.349
				(0.529)	(0.446)
Formality rate 2008	0.131	0.147*	0.147	-0.023	0.249
	(0.086)	(0.075)	(0.091)	(0.144)	(0.157)
Constant	0.220**	0.336***	0.257**	0.223**	0.352***
	(0.105)	(0.123)	(0.118)	(0.107)	(0.125)
Observations	62	62	61	62	61
R-squared	0.428	0.502	0.463	0.441	0.519

Table 3: This table presents regressions for the 62 cities shown in Figure 5. As before, the dependent variable is change in formal occupation rates and the main explanatory variable is complexity potential, which is highly significant in all regressions.

VARIABLES	(1) Single	(2) Single	(3) Single	(4) Single	(5) Single
Log Complexity Potential	0.042***	0.038***	0.041***	0.042***	0.037***
	(0.014)	(0.013)	(0.014)	(0.014)	(0.013)
Log working age pop 2008	-0.008	-0.005	-0.007	-0.008	-0.002
	(0.007)	(0.005)	(0.006)	(0.007)	(0.006)
Log GDP per capita 2011	0.017**	-0.002	0.011	0.017**	-0.006
	(0.007)	(0.009)	(0.007)	(0.007)	(0.010)
Binary oil: one well/1000		0.049*			0.061**
		(0.027)			(0.024)
Govt spending pc			0.015		0.016
			(0.016)		(0.014)
Sectoral demand				0.105	-1.011
				(0.657)	(0.656)
Formality rate 2008	0.154	0.167*	0.166	0.124	0.453**
	(0.097)	(0.085)	(0.103)	(0.171)	(0.215)
Constant	0.103	0.215*	0.132	0.106	0.214*
	(0.110)	(0.123)	(0.122)	(0.112)	(0.122)
Observations	48	48	47	48	47
R-squared	0.533	0.597	0.559	0.534	0.632

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: This table replicates the same set of regressions for the 48 urban municipalities that are not part of multi-municipality cities in our initial set of 62 cities. We observe similar results, which confirm the influence of complexity potential in formal employment creation, even after controlling for a diversity of shocks.

Table 3 confirms the findings based on the individual 96 urban municipalities. Complexity potential is highly significant and its coefficient is stable, but higher (0.048-0.06 compared with 0.028-0.039). A new result, however, is that city population size *reduces* the ability of cities to create formal employment. The coefficient is highly significant and robust to all the controls and alternative specifications.

Table 4 replicates the same set of regressions for the 48 single-municipality cities that are part of our set of 62 cities in Table 3 (the remaining 14 cities being formed by the other 48 municipalities for a total of 96 urban municipalities). The results provide additional support to our main conclusions, and especially to the importance of skill availability in formal employment creation. It must be noticed that the coefficients of complexity potential are lower than in the set of 62 cities (moving in a range between 0.037 and 0.054 versus 0.048 and 0.06).

We use the coefficient estimates of complexity potential in regression 5 of Table 4, and regression 5 in Table 3 (0.037 and 0.049 respectively), to calculate the possible range of gains in formal employment creation in the 48 municipalities that benefit from being within a radius of 45 minutes from another urban municipality (thus being part of the 14 multi-municipality cities). In order to do this we calculate the difference between the complexity potential of the municipality taken in isolation and that of the group of municipalities that form the multi-municipality city. Then we multiply that value by the coefficient estimates just mentioned. The average gain range is 1.98-2.62 percent points, depending on the coefficient. This is a substantial gain for a five-year period, considering that the average increase in the formal occupation rates of the 96 urban municipalities between 2008 and 2013 was 4 percent points. Therefore, for the average municipality that is part of a multi-municipality city, a substantial part of the increase in formal employment between those years was a result of the possibility of mobilizing the wider availability of skills available in the neighboring municipalities. As Figure 6 shows, smaller municipalities gained more than larger ones, as expected given their lower complexity potential (calculations are made with a coefficient of 0.037). If anything, these calculations underestimate the effect of distance on employment creation because the initial complexity potential of a municipality that is part of an urban conglomerate is taken as exogenous, when in fact it is probably higher than it would be if the municipality were not already close to other municipalities.

As a means to further testing the robustness of the results, in Table 5, we randomly allocate the 96 urban municipalities to form sets of 62 cities, and re-compute the necessary variables as above. We present five random draws, all of which include the controls and the three shock variables. As expected, in all regressions complexity potential is not significant, which is reassuring, and suggests that the availability of skills for aggregates of disconnected regions has no impact on the growth of the formal occupation rate of the composite regions.

Finally, in Figure 7, we re-compute complexity potential (for cities defined for each commuting time radius as above) for subsets of industries, which can be compared with the base case, which includes all industries (top left panel). In the first case, shown top right, we include only low complexity industries (i.e. the lower half of all industries ranked by complexity). The coefficients are somewhat lower but similar to the base case. In

Figure 6: Here we show the gain in formal employment that results from urban municipalities being part of multi-municipality cities. Smaller municipalities gain more than larger ones because their own complexity is much lower than that of the urban agglomeration of which they are part.

VARIABLES	(1)	(2)	(3)	(4)	(5)
Log Complexity Potential	0.003	0.007	0.010	0.004	0.019*
	(0.009)	(0.008)	(0.010)	(0.008)	(0.010)
Log working age pop 2008	-0.002	-0.004	-0.000	0.000	-0.006
	(0.006)	(0.006)	(0.007)	(0.004)	(0.006)
Log GDP per capita 2011	0.010	-0.008	0.003	0.005	-0.004
	(0.010)	(0.007)	(0.019)	(0.009)	(0.016)
Binary oil: one well/1000	0.011	0.012	0.005	-0.004	0.002
	(0.014)	(0.013)	(0.031)	(0.022)	(0.019)
Govt spending pc	0.030	0.007	0.017	0.019	0.027
	(0.021)	(0.016)	(0.017)	(0.017)	(0.017)
Sectoral demand	2.152**	0.841	1.496*	1.149*	1.828
	(0.809)	(0.636)	(0.857)	(0.638)	(1.434)
Formality rate 2008	-0.450*	-0.055	-0.335	-0.145	-0.388
	(0.250)	(0.223)	(0.275)	(0.203)	(0.409)
Constant	-0.059	0.150	0.023	-0.038	0.172
	(0.146)	(0.126)	(0.221)	(0.122)	(0.199)
Observations	62	61	61	62	61
R-squared	0.453	0.258	0.310	0.451	0.413

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5: Here we assign municipalities randomly to 62 cities in five successive placebo cases and replicate the last regression of the previous tables in each case. Reassuringly, the coefficient of complexity potential is very low and never robustly significant, suggesting that information on the diversity and sophistication of the skills in the constructed disconnected cities bears no relation with the growth of the formal occupation rate.

the second and third cases (bottom left and right panels) we present the results for high complexity industries and for manufacturing industries only. For these two groups the coefficients are much lower and in general less significant. For high complexity industries, the highest significant coefficients occur in the range between 20 and 55 minutes. For manufacturing industries, the highest significant coefficient occurs at 20 minutes of commute time. These results suggest that highly complex sectors and manufacturing industries can only benefit from skills available within a radius shorter than that of low complexity industries. However, further work may be necessary to test the robustness of this conclusion.

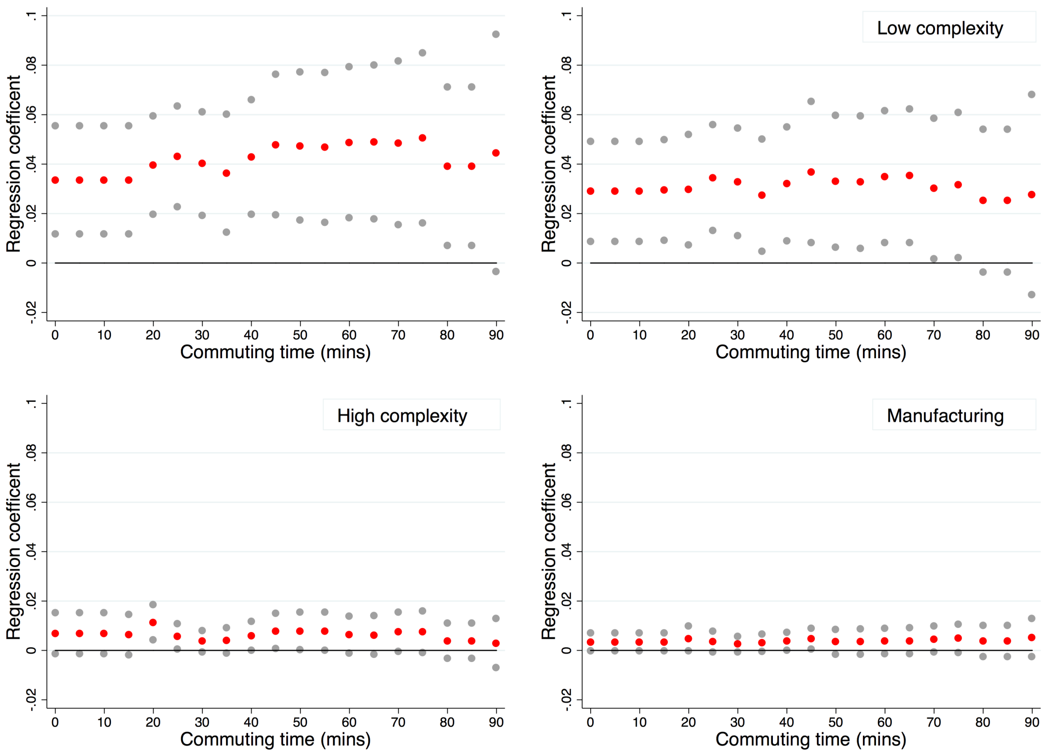

Figure 7: Here we show the coefficient of complexity potential (with the same controls of column 5 of Table 2) for the whole sample (as above), computing for low and high complexity, and manufacturing, industries only. The two latter show lower coefficients, which are significant within a shorter radius range.

5 Conclusion and Policy Implications

This paper has examined the relationship between formal employment creation, city size and traveling times within urban agglomerations. In a cross section of urban municipalities in Colombia, a strong relationship between city size and formal occupation rates is found. However, that relation is not stable through time, suggesting the presence of underlying forces other than city size. We argue that formal employment creation is mainly influenced by the availability of diverse and sophisticated skills. Formality grows as workers become absorbed into increasingly complex industries in large cities. We show that a metric, termed complexity potential, which captures a city's skill-proximity to complex industries in terms of its current labor resources and diversity, can explain the growth in the formal occupation rate for a set of 96 urban municipalities. Furthermore, after controlling for complexity potential, initial city size or productivity do not contribute to generate additional formal employment.

Political or bureaucratic definitions of municipalities, however, are not adequate to describe today's metropolitan sprawls as cities absorb workers from surrounding regions. While previous attempts [13] have been made to delineate Colombia's cities based on commuting flows, none has used a criterion based on economic outcomes (such as employment) to identify distinct cities as composites of groups of municipalities. Considering cities defined by aggregating urban municipalities within a given commuting time of each other, we show that a radius of between 45 and 75 minutes, or 62 to 43 cities, provides the optimum scale at which complexity potential explains the growth in the formal occupation rate. It is within this radius range that cities are optimally positioned to take advantage of the diversity in their labor force. Beyond a radius of about 90 minutes, the relationship between complexity potential and employment growth is no longer statistically significant, suggesting that firms cannot effectively make use of labor skills beyond that radius. Our results are robust to a host of tests, such as the inclusion of various demand shocks, the exclusion of multi-municipality cities, placebo tests and splitting the data by industry complexity and restricting it to manufacturing industries. Being connected to an urban agglomeration within a radius of 45 minutes explains a substantial part of the increase in formal employment observed between 2008 and 2013. The effect is higher in smaller municipalities, whose own skill availability is much lower than that of the agglomeration.

The most important policy implication of these results is that in order for larger cities to take advantage of the greater diversity of labor skills that comes with size, adequate transportation means are necessary to limit traveling times (across the constituent municipalities). There is no reason to discourage cities from expanding or absorbing neighboring municipalities, provided commuting times are kept within limits. Since existing political borders across municipalities may make decisions difficult to coordinate, some external mechanism or institution could be created to encourage coordinated infrastructure investments, urban planning and industry re-localization programs. Similarly, isolated mid- or small size cities may be able to generate more formal employment if adequate investments are made to connect them to larger cities. While the objective in transportation investments is usually reducing cargo transportation costs, our results suggest that passenger transportation is probably more important. The obsession with roads may be misplaced, therefore, because fast passenger transportation may be better achieved by investments in trains, dedicated bus lines and passenger stations, as well as by discouraging the use of private cars.

6 Acknowledgements

We would like to acknowledge CAF for their financial support and Manuel Toledo for thoughtful comments on an earlier draft. Andres Gomez at the Center for International Development played a critical role in earlier work leading to this research focus. Michele Coscia provided assistance in accessing data on commuting distance. Gilles Duranton, Pablo Sanguinetti, Juan Vargas and other participants of the Workshop on Urban Economics held in Buenos Aires in July, 2016, are also acknowledged for many useful suggestions. All errors are our own.

References

[1] GS Fields. Rural-urban migration, urban unemployment and underemployment, and job-search activity in ldcs. *Journal of development economics*, 2(2):165--187, 1975.

[2] E Ghani and R Kanbur. Urbanization and (in) formalization. *World Bank Policy Research Working Paper*, (6374), 2013.

[3] EL Glaeser. Are cities dying? *The Journal of Economic Perspectives*, pages 139--160, 1998.

[4] EL Glaeser. Triumph of the city: how our greatest invention makes us richer, smarter, greener, healthier, and happier. 2011.

[5] RE Lucas. On the mechanics of economic development. *Journal of Monetary Economics*, 22(1):3--42, July 1988.

[6] J Jacobs. *The death and life of great American cities*. Vintage, 1961.

[7] W Rybczynski. *Makeshift metropolis: ideas about cities*. Simon and Schuster, 2010.

[8] L Bettencourt, H Samaniego, and H Youn. Professional diversity and the productivity of cities. *Scientific reports*, 4, 2014.

[9] O'Clery N, Gomez A, and Lora E. The path to labour formality: Urban agglomeration and the emergence of complex industries. *Under Review*, 2015.

[10] Hidalgo CA, Klinger B, Barabasi AL, and Hausmann R. The product space conditions the development of nations. *Science*, 317:482--487, 2007.

[11] Frank Neffke, Martin Henning, and Ron Boschma. How do regions diversify over time? industry relatedness and the development of new growth paths in regions. *Economic Geography*, 87(3):237--265, 2011.

[12] G Duranton and HG Overman. Testing for localization using micro-geographic data. *The Review of Economic Studies*, 72(4):1077--1106, 2005.

[13] Duranton G. Delineating metropolitan areas: Measuring spatial labour market networks through commuting patterns. *Processed, Wharton School, University of Pennsylvania*, 2013.

[14] AJ Venables. Evaluating urban transport improvements: cost-benefit analysis in the presence of agglomeration and income taxation. *Journal of transport economics and policy*, pages 173--188, 2007.

[15] G Duranton and MA Turner. Urban growth and transportation. *The Review of Economic Studies*, 79(4):1407--1440, 2012.

[16] M Porter. *Competitive advantage of nations: creating and sustaining superior performance*. Simon and Schuster, 2011.

[17] Daron Acemoglu, Camilo García-Jimeno, and James A Robinson. State capacity and economic development: A network approach. *The American Economic Review*, 105(8):2364--2409, 2015.

[18] John Luke Gallup, Alejandro Gaviria, and Eduardo Lora. *Is geography destiny?: lessons from Latin America*. Stanford University Press, 2003.

[19] Hidalgo CA and Hausmann R. The building blocks of economic complexity. *Proc Natl Acad Sci USA*, 106(26):10570--10575, 2009.

[20] Hausmann R, Hidalgo CA, Bustos S, Coscia M, Chung S, Jimenez J, Simoes A, and Yildirim MA. *The atlas of economic complexity: Mapping paths to prosperity*. Harvard University Center for International Development, MIT Media Lab, 2011.

[21] Center International Development Harvard. www.datlascolombia.com.

[22] F Neffke and M Henning. Skill relatedness and firm diversification. *Strategic Management Journal*, 34(3):297--316, 2013.

[23] Hausmann R and Klinger B. *The evolution of comparative advantage: the impact of the structure of the product space*. Center for International Development and Kennedy School of Government, Harvard University, 2006.

[24] Ricardo Hausmann, Cesar Hidalgo, Daniel P. Stock, and Muhammed Ali Yildirim. Implied comparative advantage. *HKS Working Paper No. RWP14-003*, 2014.

[25] CEDE. https://datoscede.uniandes.edu.co/.

[26] Therese J McGuire and Timothy J Bartik. *Who Benefits From State and Local Economic Development Policies?* JSTOR, 1991.