Borradores de ECONOMÍA

Firm Productivity and Cities: The Case of Colombia

Por: Jorge Balat
Camila Casas

Núm. 1032
2018

tá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Col

BANCO DE LA REPÚBLICA
COLOMBIA

Firm Productivity and Cities: The Case of Colombia*

Jorge Balat† Camila Casas‡
The University of Texas at Austin Banco de la República

Abstract

In this paper we study the determinants of firm-level productivity in Colombia. We are interested in the effects of agglomeration forces that explain why manufacturing economic activity tends concentrate geographically as well as the effect of forces that can hurt productivity in high-density areas. An unfortunate major example of the latter forces in Colombia is its high levels of crime and terrorist attacks. We carry out this study by exploiting two very rich data sources: a firm-level panel which allows us to estimate firm-level productivity and a panel of municipality characteristics. To address selection and endogeneity issues in the estimation of firm-level productivity we use a control function approach. Our main findings are the following. First, scale economies do not seem to affect firm-level productivity. Second, we do find evidence of location economies (industrial specialization has a positive effect on productivity). Industrial variety, on the other hand, lowers productivity. It seems that firms in Colombia benefit from forming clusters and locating in cities with less industrial variety. We also find non-trivial effects of city fiscal performance, education level and quality, and crime and violent attacks.

Keywords: Cities; Colombia; Productivity; Total Factor Productivity; Agglomeration.
JEL Classification: R00, R12, R32, O14, L6, C14.

---

*We thank Gilles Duranton, María Marta Ferreyra, W. Walker Hanlon, Mark Roberts, Carlos Vegh and participants at the World Bank author's workshop for very useful comments and discussions. We also want to thank Stefany Moreno, who provided excellent research assistance. All remaining errors are our own.

†[Corresponding author] `jbalat@utexas.edu`
‡`mcasaslo@banrep.gov.co`

1 Introduction

In this paper we study the determinants of firm-level productivity in Colombia.[1] In particular, we are interested in the effects of agglomeration forces that explain why manufacturing economic activity tends to be concentrated in rather few cities as well as the effect of forces that can hurt productivity in high-density areas. While the former set of forces has received extensive attention in the urban economics literature, to the best of our knowledge this is the first study for Colombian manufacturing firms.[2] On the other hand, previous work has not focused on the effect of the second set of forces —partly due to a lack of detailed data. In Colombia, an unfortunate major example of these harmful forces is violence —from minor crimes to terrorist attacks which have been very prevalent in recent Colombian history. We carry out this study by exploiting two rich data sources. The first one is a firm-level panel which contains input and output data that will allow us to estimate firm-level productivity. The second one is a panel of municipalities containing very detailed information on city characteristics over time.

The work pioneered by Weber (1929), Hoover (1937), and Isard (1956) has highlighted the importance of agglomeration economies in explaining the concentration of economic activity in metropolitan areas. The literature in urban economics documenting agglomeration economies puts its focus on forces that are external to the firm. On the other hand, the empirical industrial organization literature estimating firm-level productivity implicitly assumes that economies are realized within a plant and is not interested in its determinants. This paper intends to bridge the two literatures by making the two sources of economies explicit and incorporating them simultaneously into the analysis. Our objective is to uncover systematic relationships between firm-level productivity and the characteristics of the cities in which they locate.

The first step in our investigation of the determinants of firm-level productivity is to get a clear picture of firms' location. This will help us to identify the geograph-

[1] Throughout this paper, productivity refers to total factor productivity (henceforth, TFP), or Hicks-neutral productivity, and we will use these terms interchangably. See Section 4 for a formal definition.

[2] In the case of Colombia a series of papers by Gilles Duranton (2015a,b) has looked at agglomeration effects on wages and the determinants of city growth. The effects of agglomeration forces on firm-level productivity has been addressed for cases other than Colombia —e.g., Cingano and Schivardi (2004) and Di Giacinto, Gomellini, Micucci, and Pagnini (2014) looked at Italian cities, Lall, Shalizi, and Deichmann (2004) looked at Indian cities, Saito and Gopinath (2009) and Lopéz and Südekum (2009) looked at the case of Chile.

ical distribution of industries across cities in Colombia and the differences in the degree of geographical concentration across industries. While Colombia is the fourth largest country in South America and highly urbanized (about 75% of the Colombian population live and work in cities), most of the population and economic activity is concentrated in a few areas. In broad terms, Colombia is composed of five regions. It is crossed longitudinally by the Andes Mountains. The Andean region occupies 30% of the total area and is the most populated region (about 70% of the country's population). To the East of the Andes there are two big regions accounting for 60% of the total area but less than 5% of the total population —the Orinoco lowlands located in the Northeast (18% of the area and 4% of the population) and the Amazonian lowlands in the Southeast (41% of the area and 0.6% of the population). To the West of the Andes there is the Pacific coast region (7% of the area and 2.5% of the population), and to the North, the Caribbean coast region (12% of the area and 22% of the population).

Most of the industrial activity is located in the Andean and Caribbean regions. The main cities in the Andean region, are Bogotá (the capital), Medellín, Cali, Bucaramanga, and Manizales, and the main cities in the Caribbean region are Cartagena and Barranquilla. More than 70% of the industrial establishments are located in these seven cities (and more than 90% if we consider the respective seven metropolitan areas). Our main preliminary findings when looking at the geographical distribution of firms are the following: (i) while Bogotá has the highest concentration of firms (44.7% of total establishments in the country) among the main seven cities, it has the lowest level of concentration across industries, meaning that it has the greatest diversity of industries; (ii) Cartagena, on the other hand, has the highest concentration, with only three industries (food, chemicals, rubber and plastic products) dominating its manufacturing production; (iii) in all of the main cities except for Medellín, the highest concentration of firms is in the food and beverages industry; (iv) the textiles and apparel industries appear to have two geographical clusters: Medellín and Bogotá; (v) the chemicals and rubber and plastic products industries seem to be clustered in Bogotá and to a lesser extent in Medellín and Cali; and (vi) manufacturing of office and computing machines is done exclusively in Medellín and Bogotá.

We then proceed to estimate firm-level productivity and provide a descriptive analysis of the estimates by industrial sector across geographical regions. Our estimation procedure relies on a control function approach to handle both endogeneity (since input choices are likely to be a function of productivity shocks observed by the firms but not by the econometrician) and selection bias (arising from firms entry and exit decisions which are partly determined by their productivity shocks). Our implementation follows that of Ackerberg, Caves, and Frazer (2016), which extends the seminal work by Olley and Pakes (1996).

This high-level analysis serves as preliminary evidence of how intraindustry and interindustry clustering affect firm productivity. Our key findings are the following. First, at the aggregate level, Bogotá shows the highest level of productivity, followed —in order— by Medellín, Cali, Barranquilla, Manizales, Cartagena, and Bucaramanga. Second, the most productive sector is apparel. Not only that, but the apparel industry is also the one which shows the lowest level of geographical dispersion in terms of productivity. On the contrary, manufacturing of motor vehicles shows the highest geographical variation. Third, while Bogotá shows the highest concentration of firms in every industry and the broadest scope of industries among the main cities, this does not translate into higher productivity for each and every industry. For example, Cali is the most productive city in apparel and metal products; Medellín is the most productive in wood products and computing machinery; Barranquilla in textiles and printing; Manizales in machinery; Bucaramanga in rubber and plastic products and non-metallic mineral products; and Cartagena in food and paper products. Bogotá remains the most productive city in leather products, chemicals, electrical machinery, motor vehicles, and furniture.

The heart of the paper is in the exploration of the determinants of firm-level productivity vis-à-vis city characteristics. We can consider the formation of cities as the outcome of a trade-off between agglomeration economies or localized aggregate increasing returns and the costs of urban congestion (see, e.g., Duranton and Puga (2004)). As we have mentioned above, these are the two sources of forces that we relate to firm-level productivity. On the one hand, we focus on the effect of the forces behind agglomeration that allow firms to attain higher levels of productivity as a result of their geographical location. On the other hand, we also investigate the centrifugal forces that hinder firms from attaining higher levels of productivity which come about in highly concentrated areas. We define 'agglomeration economies' in a broad sense as any force that increases firms' output when the size of the local economy grows (see, e.g., Combes and Gobillon (2015)). In particular, we investigate the effect of different measures of *agglomeration* commonly used in the literature. Our first set of basic measures are related to the scale of local economic activity which we define by aggregating (excluding firm i) either the number of manufacturing establishments, manufacturing employment, manufacturing capital stock, or manufacturing production. Our second set of measures intend to capture the composition of the local industrial structure and are aimed to quantify inter- and intra-industry types of externalities. On the one hand, we construct a measure of the degree of sector specialization to capture *localization economies* and, on the other, we construct a measure of industrial heterogeneity (or variety) to capture *urbanization economies*. To capture the innovation and technological diffusion through which the local industrial structure might affect productivity, our third set of variables measure the local level of competition and the average size of firms in a given sector.

There are two major struggles that we need to overcome in the estimation of the effects of agglomeration economies and other city characteristics on firm-level productivity. The first one is selection. Firms make entry and exit choices based on their productivity shocks and city characteristics. Even if the two are not correlated in the population, they might be correlated in the sample due to selection, and our estimate of the effect will pick this up. We solve this problem by including a selection correction term in the estimation of TFP and allowing city characteristics to be part of the firm's state space as proposed by Olley and Pakes (1996) and more recently by De Loecker, Goldberg, Khandelwal, and Pavcnik (2015). The second problem is the omitted variable bias that might arise when regressing TFP on city characteristics. We overcome this problem by exploiting the panel structure of our data and including city fixed effects.

Our main findings are the following. First, scale economies —i.e., the size of the city under alternative measures— do not seem to affect firm-level productivity. Second, we do find supporting evidence for *location economies*. That is, industrial specialization has a positive effect on productivity. Our point estimates indicate that a one standard deviation increase in sector specialization produces an increase in productivity ranging from 2 to 4%. In terms of cross-industry spillovers, we find evidence that they also matter (but the evidence is less robust): a one std. dev. increase in variety implies a decrease in productivity on the order of 2 to 3%. It seems then that firms in Colombia benefit from forming clusters and locating in cities with less industrial variety.

Extending our analysis to other city characteristics we do find non-trivial economic effects of other city 'amenities' (or lack of them). The fiscal performance of the city, its education level and quality, and more importantly crime and violent attacks have sizable effects on productivity. For example, a one std. dev. increase in the theft rate, the fraction of taxes on manufacture, or the Saber 11 test scores translates, respectively, into a decrease in productivity of 4 to 5%, a decrease in productivity of 3,3 to 3.9%, and an increase in productivity of 2 to 3.2%, all else equal.

Our paper is closely related to the urban economics literature that looks at the empirical determinants of firm productivity. A key interest in the urban economics literature regards the mechanisms that encourage productive activity to be concentrated. The seminal work of Marshall (1890) proposed three main sources of agglomeration economies: (i) knowledge or technological spillovers, (ii) labour-market interactions, and (iii) input-output linkages. The mechanisms behind each of the three sources of agglomeration are different. The technological spillovers involve the process by which workers acquire job-specific skills that can then be transferred to a new firm when workers change jobs and also workers in a given firm might learn about new technologies from workers in other firms through constant interaction due to proximity. The labour-market interactions operate by reducing the cost of matching between employers and employees in the labor market. Finally, input-output linkages are related to the benefits of sharing intermediate suppliers producing under increasing returns.

More recently, the literature has classified the mechanisms behind spatial concentration into sharing, matching, and learning effects (see Duranton and Puga (2004) for a great survey of the recent literature). This classification of the sources of agglomeration is also based on different mechanisms. *Sharing effects* are the result of sharing indivisible goods and facilities, also the gains from a greater variety of inputs suppliers that can be sustained by a large final goods industry, the gains of industry specialization and the pooling of risk. *Matching effects* refer to mechanisms that increase either the expected quality of matches or the probability of matching between workers and firms. Lastly, *learning effects* correspond to mechanisms based on the generation, diffusion and accumulation of knowledge. The empirical literature has lagged since assessment of these different mechanisms requires data that are not easily available. The progress in this area has mainly focused on evaluating the impact of local characteristics that shape agglomeration economies on local outcomes. Some classic papers that assess the impact of variables at the city-industry level on employment growth are Glaeser, Kallal, Scheinkman, and Shleifer (1992) and Henderson, Kuncoro, and Turner (1995). Some more recent papers are closely related to our paper in that they also relate measures of spatial concentration and productivity measured at the firm level.

Henderson (1986) analyses the economies of location and specialization in 243 metropolitan areas of the United States and 126 metropolitan areas in Brazil. He finds evidence in favor of economies of location (specialization), however there is no evidence of benefits of urbanization economies. Similarly, Henderson (2003) finds that specialization has a positive effect on the productivity level of high-tech firms in the United States but has no effect on the machinery industry. He also finds that urbanization economies have no effect in either sector. Cingano and Schivardi (2004) evaluate the impact of agglomeration economies on TFP using data from Italy and find positive effects on TFP from specialization economies and the size of the city. Urban density, competition, or average firm size do not affect TFP. Di Giacinto, Gomellini, Micucci, and Pagnini (2014) quantify productivity differences between Italian firms located in urban areas and firms located in industrial districts and find that firms clustered in urban areas have higher productivity gains over those in industrial districts. Lopéz and Südekum (2009) estimate the impact of agglomeration economies and vertical relations on firms' productivity using data form Chile. They find evidence supporting intraindustry productivity spillovers, but not urbanization economies effects (cross-industry effects). Saito and Gopinath (2009) also look at Chilean firms. They find positive effects of agglomeration, diversification, and market size.

The remainder of the paper is organized as follows. We begin by presenting and describing our two data sources in Section 2. In Section 3 we provide a preliminary analysis of the geographical structure of manufacturing industries in Colombia. Section 4 discusses the estimation approach to recover firm-level productivity and presents a preliminary analysis of the geographical distribution of productivity. In Section 5 we investigate the geographical determinants of productivity and present the main results of the paper. Finally, Section 6 concludes.

2 Firm-Level and City-Level Data

Our main empirical analysis requires us to obtain measures of firm-level TFP and relate them to city characteristics. To do so, we rely on two data sources. The first dataset comprises firm-level input and output data on manufacturing firms from Colombia. These data allow us to estimate firm-level TFP. More importantly, the dataset contains the location of the firms so we can then relate our TFP measure to the corresponding city characteristics. Our second dataset contains data at the city-level on an unusually rich set of variables. In what follows, we discuss our two datasets in detail and provide preliminary summary statistics.

2.1 Superintendencia de Sociedades

For our productivity analysis, we use a firm-level dataset which contains detailed balance sheet and operational information. Our data on firms' production and input use come from "Superintendencia de Sociedades," the agency in charge of supervising corporations. Specifically, the data come from the "Sistema de Información y Riesgo Empresarial" (SIREM) database.[3] The data are at an annual frequency and our dataset covers the period 2005–2013. We have access to public information such as balance sheets, as well as to confidential data included in the annexes filed by the firms.[4] These variables include the revenues obtained from the sales of each product, the use of raw materials, investments, the capital stock, and the number of employees and payroll, broken down by type (executive, administrative, and production workers) and tenure (permanent or temporary). What is key for our later analysis, we observe the firm's location.[5]

The data from SIREM include information on firms from several industries. In general, we focus only on manufacturing firms, excluding manufacturers of coke, refined petroleum products, nuclear fuel, and basic metals (which include metals such as gold, silver, platinum, and nickel). We exclude the firms classified in these two manufacturing sectors because they are commodity producers, and therefore their dynamics are probably different from those of the other manufacturing firms. We classify firms into industries following the International Standard Industrial Classification (ISIC), Revision 3.1. (see Table 1 for the list of manufacturing sectors and their corresponding ISIC codes).

Given our focus on the manufacturing sector, the first step prior to estimation is to precisely define which firms are considered manufacturers. This step is relevant

[3]The SIREM includes information for relatively large firms, and for firms in financial trouble. In particular, firms must report their financial data if their assets and/or income (adjusted by inflation) are grater than 30,000 times the current legal monthly minimum wage, if their external liability is grater than the total assets, if the financial expenditures are at least 50% of their income, if their cash flow is negative, or if their losses reduce the net equity below 70% of the social capital.

[4]We obtained access to the confidential data through Banco de la República.

[5]Ideally, we would have preferred to use in our analysis census data on manufacturing establishments such as the Encuesta Anual Manufacturera (EAM) conducted by the Departamento Administrativo Nacional de Estadística (DANE). Unfortunately, the EAM does not include the firms' locations but only indicators for the main seven metropolitan areas which severely hinders the spirit of our analysis. While the main seven metropolitan areas concentrate most of the manufacturing activity, metropolitan area is too coarse of a unit of analysis which would introduce significant noise into our analysis. At the same time, when comparing the two datasets, SIREM accounts for more than 90% of the employment reported in the EAM in the manufacturing sector (see **?** for more details). This means that when using SIREM we are not losing much of the representativeness of the EAM while retaining a finer unit of analysis.

Table 1: ISIC (Rev. 3.1) Industry Classification

Code	Description
15	Manufacture of food products and beverages
16	Manufacture of tobacco products
17	Manufacture of textiles
18	Manufacture of apparel; dressing and dyeing of fur
19	Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear
20	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
21	Manufacture of paper and paper products
22	Publishing, printing and reproduction of recorded media
23	Manufacture of coke, refined petroleum products and nuclear fuel
24	Manufacture of chemicals and chemical products
25	Manufacture of rubber and plastics products
26	Manufacture of other non-metallic mineral products
27	Manufacture of basic metals
28	Manufacture of fabricated metal products, except machinery and equipment
29	Manufacture of machinery and equipment n.e.c.
30	Manufacture of office, accounting and computing machinery
31	Manufacture of electrical machinery and apparatus n.e.c.
32	Manufacture of radio, television and communication equipment and apparatus
33	Manufacture of medical, precision and optical instruments, watches and clocks
34	Manufacture of motor vehicles, trailers and semi-trailers
35	Manufacture of other transport equipment
36	Manufacture of furniture; manufacturing n.e.c.

for multi-product firms that are not limited to manufacturing activities. For our estimations, we consider as manufacturing firms only firms that report having positive revenues from manufacturing products in all the years they appear in the sample. In applying this definition, we took advantage of the rich data on revenues which are reported by the firms at the product level.[6]

For multi-product manufacturing firms, the second step is to decide how to allocate each firm to a specific manufacturing sector.[7] Once again, we used the information on revenues by product, and we assigned each firm to the sector that includes the product that generated the most income throughout the sample period. Specifically, we added up the (deflated) revenues per product for 2005–2013, and assigned the firm to the manufacturing sector with the highest share.

With the subset of manufacturing firms clearly defined, the final step is the cleaning of the raw data. The cleaning process included removing observations with implausible annual growth rates (perhaps confusing thousands with millions of Colombian pesos, or number of employees with payroll), as well as occasional value interpolation when a particular variable was missing for a single year. The resulting dataset contains 26,131 firm-year observations, corresponding to over 4,000 firms in 154 municipalities. This is the baseline sample we use in our estimations.

Table 2 presents some basic statistics of our SIREM sample. In the first column we observe that, on average, we have around 3,000 manufacturing firms per year. In the remaining columns we report, for the average firm in our sample, the revenues from sales, the value of the firm's capital stock, value of raw materials used, number of workers employed, and the share of these that were production (blue-collar) workers. We see that the average firm had annual revenues, on average, of 30.6 billion Colombian pesos, an average capital stock of 18 billion, used raw materials worth 14 billion, and employed about 160 workers, of whom 55 percent were production workers.[8]

In Table 3 we present analogous statistics, broken down by industry and averaged across time.[9] From the table it is clear that there is a great level of heterogeneity

[6]In the revenues annex, products are defined at the 4-digit level according to the ISIC classification.

[7]By sector or industry we mean, specifically, a 2-digit industry based on the ISIC classification.

[8]Throughout the paper monetary figures are expressed in billions of Colombian pesos of 2005. Each variable was deflated using a variable-specific deflator.

[9]There are three industries for which we only have very few observations: ISIC 16 (tobacco), ISIC 30 (office and computing machinery), and ISIC 32 (radio, television and communication equipment). In order to avoid disclosing confidential information, we do not report statistics for these industries.

Table 2: Firms' Summary Statistics: All Manufacturing Sectors

	Firms (#)	Sales ($)	Capital ($)	Materials ($)	All Workers (#)	Production Workers (%)
2005	2,914	25.9	12.1	11.9	145.4	59.0
2006	3,373	26.0	11.4	11.6	141.1	59.2
2007	2,939	30.4	14.6	13.6	161.7	57.9
2008	2,882	31.3	17.1	14.8	166.2	55.1
2009	3,086	27.9	16.4	12.8	148.8	55.0
2010	2,974	30.5	19.3	13.7	153.5	55.4
2011	3,067	31.8	20.4	14.4	159.0	54.6
2012	2,928	33.5	22.7	15.6	169.4	52.9
2013	2,726	37.5	26.8	17.2	177.7	51.4
Average	2,988	30.6	17.9	13.9	158.1	55.6

Source: Authors' calculations based on data from SIREM and DIAN/DANE.

Notes: This sample includes manufacturers of coke, refined petroleum products, nuclear fuel, and basic metals. The sign '$' corresponds to billions of Colombian pesos of 2005.

across sectors. For instance, manufacturers of food products and beverage (ISIC 15) and motor vehicles (ISIC 34) have similar average revenues, however, the average number of employees is 20% larger in the former industry. Moreover, the distribution between production and white-collar workers is quite dissimilar across the two sectors. In a similar fashion, sectors ISIC 15 (foods and beverage) and ISIC 18 (apparel) have work forces of similar size, despite the latter using significantly less amounts of other inputs (both capital and materials) than the former.

We also observe non-trivial heterogeneity in industry composition and firm characteristics across cities. We defer this discussion to Section 3 in which we explore the geographical dimension of the firm-level data and show the distribution of manufacturing firms and their characteristics across the country.

2.2 Municipal Panel – CEDE

Our second dataset is a comprehensive panel of municipal variables including data on more than 1,300 variables. The Municipal Panel is overseen and continuously updated by the Center of Economic Development Studies at Universidad de los Andes (CEDE).

Table 3: Firms' Summary Statistics: by Sector

Sector	15	16	17	18	19	20	21	22	23	24	25
Firms (#)	500	2	154	252	76	44	61	220	24	349	330
Sales ($)	50.94	–	17.82	14.01	12.22	6.68	63.98	13.87	119.28	49.43	16.88
Capital ($)	29.00	–	11.31	5.11	3.56	6.83	52.84	9.06	198.06	22.90	11.35
Materials ($)	26.83	–	6.59	3.87	3.49	1.90	27.05	3.44	148.49	18.50	6.85
Workers (#)	226.14	–	170.18	199.91	155.99	81.53	227.83	117.12	80.26	163.65	114.00
Production Workers (%)	51.69	–	71.02	59.67	62.77	65.88	57.41	36.62	52.27	39.86	64.88

Sector	26	28	27	29	30	31	32	33	34	35	36
Firms (#)	137	239	60	94	3	53	2	15	105	11	257
Sales ($)	30.20	17.13	47.28	20.01	–	30.64	–	8.40	51.79	161.27	14.35
Capital ($)	34.66	8.98	34.20	7.83	–	11.54	–	7.33	10.81	25.15	5.79
Materials ($)	8.39	9.33	22.71	8.28	–	11.47	–	2.42	26.64	53.09	6.28
Workers (#)	165.73	100.50	130.16	152.14	–	173.87	–	108.56	181.92	489.38	102.47
Production Workers (%)	64.30	63.06	68.95	64.46	–	50.40	–	66.81	69.01	45.93	56.15

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: The sign '$' corresponds to billions of Colombian pesos of 2005.

The unit of observation is a *municipio*, which is the fundamental administrative unit in Colombia. The definition of a municipality in Colombia is political and each municipality maintains political, fiscal, and administrative autonomy. While technically a municipality might refer to a city, town, or village, throughout the paper we use the terms municipality and city interchangeably. The dataset consolidates municipal-level data from a myriad of sources —including Departamento Administrativo Nacional de Estadística (DANE), Instituto Geográfico Agustín Codazzi (IGAC), Departamento Nacional de Planeación, Ministerio de Agricultura, Red de Información y Comunicación del Sector Agropecuario Colombiano (AGRONET), Ministerio de Salud, Ministerio de Minas y Energía, among others— on 1,122 Colombian municipalities from 1993 to 2014. The dataset contains six main modules: general municipal characteristics, fiscal variables, conflict and violence, agricultural sector and land, education, and health and public services.

To save space, we briefly discuss some key features of the data in what follows and we reserve the discussion of these data in more detail to the Appendix, including summary statistics.

Cities' General Characteristics The key variables we use from this module in our later analysis are: year of incorporation, population (total, rural, and urban), area (km^2), distance to Department's capital (km), distance to main wholesale food market (km), distance to Bogota (km), GDP (total, agriculture, manufacture, services), a measure of income inequality as given by the Gini coefficient, poverty rate, unsatisfied basic needs rates, and a multidimensional measure of poverty.

Table 20 in the Appendix shows averages (across years) of select variables for the main seven cities in Colombia (all remaining cities are included in Other). In terms of population, Bogotá —the capital city— is the main city with an average of over 7 million inhabitants. It is followed by Medellín and Cali (over 2 million inhabitants each), and Barranquilla and Cartagena (each with about 1 million inhabitants). Finally, Bucaramanga follows with a population of half a million and Manizales with close to 400,000 inhabitants. In terms of GDP per capita, the ranking changes slightly with Bucaramanga at the top, followed by Bogotá, Medellín, Barranquilla, Cartagena, Cali, and Manizales. Poverty rates are relatively high in the seven main cities. Bogotá exhibits the lowest rate at 17%; Cali, Bucaramanga, Barranquilla have poverty rates of about 33%, followed by Cartagena (40%), Medellín (46%), and Manizales (48%).

Manufacturing as a whole accounts for about over 30% of total economic activity in Medellín, Cartagena and Bucaramanga; around 28% in Manizales, about 24% in Barranquilla and Cali, and 22% in Bogotá.

Fiscal Variables We selected the following key variables from this module: total revenue, tax revenue, tax on manufacturing, total expenditure, expenditure on infrastructure, primary deficit, deficit, and borrowing. We also include a fiscal performance index (Indice de Desarrollo fiscal) that ranges from 0 to 100 (lowest to highest) and provides a measure of the city's fiscal health. Each city is then ranked at the national and state level. Other measures of fiscal health include: the ratio of operating expenditure to ordinary revenues (tax and non-tax revenues), which quantifies to what extent regular operating costs (public sector wage-bill and general operating costs) are paid from city's own discretion revenues; the ratio of debt interest payments to total revenues, which quantifies the debt burden and to what extent it might affect other expenditures; the ratio of tax revenues to ordinary revenues, which quantifies to what extent the city can generate its own resources (the complement would be transfers from the federal and state level); and the ratio of investment to total expenditure, where investment not only includes infrastructure but also 'social investment' (e.g., public sector wage-bill for doctors and teachers and some subsidies). Finally, we also include some measures of expenditure on key areas (as a fraction of total expenditure) like transportation infrastructure, education, institutional development, economic development, public services, and water and sanitation.[10] Table 21 in the Appendix shows averages of select fiscal variables for the main seven cities in Colombia. Most notably is the Fiscal Performance Index. According to this index, Medellín ranks highest, followed by Bogotá, Bucaramanga, Manizales, Barranquilla, Cartagena, and Cali.

Conflict and Violence The main variables we include in our analysis are dummies for presence of terrorist groups (e.g., FARC, ELN, AUC), population displaced

[10]For these measures, expenditure is understood in a broad sense, e.g., it also includes investment expenditure. Transportation infrastructure includes expenditure on construction, repairs, and maintenance of roads, ports, and airports. Education includes expenditure on subsidies, especial students, and training. Institutional development includes expenditure on institutional evaluation, administrative reorganization, training, registries updates, among other items. Economic development includes expenditure on industrial and business promotion, labor training programs, technical assistance to firms among other items. Public services includes subsidies for low income people, maintenance and expansion of street lighting, expansion of electric and gas service.Water and sanitation includes expenditures (maintenance and capital expenditures) on the drinking water system, sewage system, and trash collection system.

by terrorism (in- and out-flows), arrests, terrorist surrenders, thefts, homicides, kidnappings for ransom, terrorist attacks, and a dummy variable for violent activity from 1948 to 1953.[11]

Table 22 in the Appendix presents summary statistics. There are a few things worth noting. First, there is a high number of terrorist attacks: about 49 attacks on average per year in Medellín, about 38 per year in Barranquilla, 32 in Cali, 15 in Cartagena, about 9 in Bogotá, and about 3 in Manizales and Bucaramanga. Second, there is also a high level of kidnapping events. Bogotá tops with 36 kidnappings on average per year, followed by 15 in Cali, 7 in Medellín, 4 in Barranquilla. Third, there is a high level of FARC (Fuerzas Armadas Revolucionarias de Colombia) guerrilla groups presence in all 7 cities (measured by a dummy variable which takes a value of 1 if the group is present in a given city-year). Similarly, the AUC (Autodefensas Unidas de Colombia) has a rather homogeneous geographical presence but at a lower level compared to the FARC. On the other hand, there is some regional variation in the presence of ELN (Ejército de Liberación Nacional). Specifically, the ELN has a constant presence in Bogotá, Medellín, and Cali, and less so in the other districts. Fourth, homicides rates are high, ranging from 19 (per 100,000 people) in Bogotá to 74 in Cali. Fifth, there is a very large number of displaced people (resulting from terrorist groups seizing their lands). These people are forced out of their homes in smaller more distant districts, but also in Medellín, and they migrate to the main seven cities.

Education The key variables that we use in our study are the following: average years of schooling, historic alphabetism rates (at years 1918, 1938, 1951, 1964, 1985), number of establishments, teachers and students, grade retention, students receiving subsidies, number of higher education students, number of university professors, and test scores from the SABER 11 test (a standardized test similar to the SAT in the United States). The test score is used to rank schools and we construct two measures of school quality: percentage of low- and high-ranking schools. We show summary statistics of these variables by cities in Table 23 in the Appendix. While enrollment rates are homogeneous across cities, and the number of establishments, teachers, and students seem to be proportional to each city's population, measures of performance

[11]This period, called "la Violencia" was one of the most cruel periods in the Colombian history, comparable to a civil war. This conflict caused between 200,000 and 300,00 deaths and the forced displacement of about 2 million people (almost a fifth of the population at the time) from their and seizure of their lands by terrorist groups.

offer a different picture. Students in Cartagena have the lowest test scores, and students in Bucaramanga, the highest. Bogotá has the lowest fraction of low-ranking schools (less than 1%) and the highest fraction of high-ranking school (more than 60%). Cartagena and Barranquilla are the flip-side, with high-ranking schools account for about one third of the schools and the low-ranking for about 16 to 18%.

Health and Public Services The main variables we use in our study include: health insurance enrollment (by insurance type), number of births, number of low-weight births, infant mortality, access to potable water, access to sewage, access to electricity, and access to natural gas network. Table 24 in the Appendix presents summary statistics of select variables. While health insurance coverage is universal in Medellín, Barranquilla, and Bucaramanga, it attains 90% of the population in Cali and about 92% in Bogotá and Manizales. There seems to be a wide disparity in terms of drinking water coverage ranging from 30% in Cartagena to 100% in Barranquilla; with Medellín at about 90% coverage, and Bogotá Bucaramanga, and Cali with about 76 to 78% coverage. A similar picture arises in terms of trash collection services coverage, and sewage coverage. On the other hand, access to electricity seems universal. The lower access to basic public services in Cartagena is translated into a high infant mortality rate. Finally, access to the natural gas network, ranges from 54% in Medellín to more than 90% in Bogotá, Manizales, and Bucaranga, while Barranquilla has 88% access, Cartagena 84%, and Cali 73%.

3 Firm Location and Industrial Clustering

In this section we start to explore the geographical dimension of the firm-level data. We begin by showing the distribution of manufacturing firms across the country and then show the geographical distribution of firms' characteristics. The idea of this section is to get a high-level picture of where firms in different industries cluster, the differences across industries, and the characteristics of the firms in each cluster. For presentational purposes, in the discussion that follows we consider only the main seven cities in Colombia (and in some cases their corresponding metropolitan areas) but it is worth to emphasize that in our main analysis of the determinants of firm productivity we include data on firms from all of the municipalities covered by our sample.

Figure 1: Geographical Distribution of Manufacturing Firms

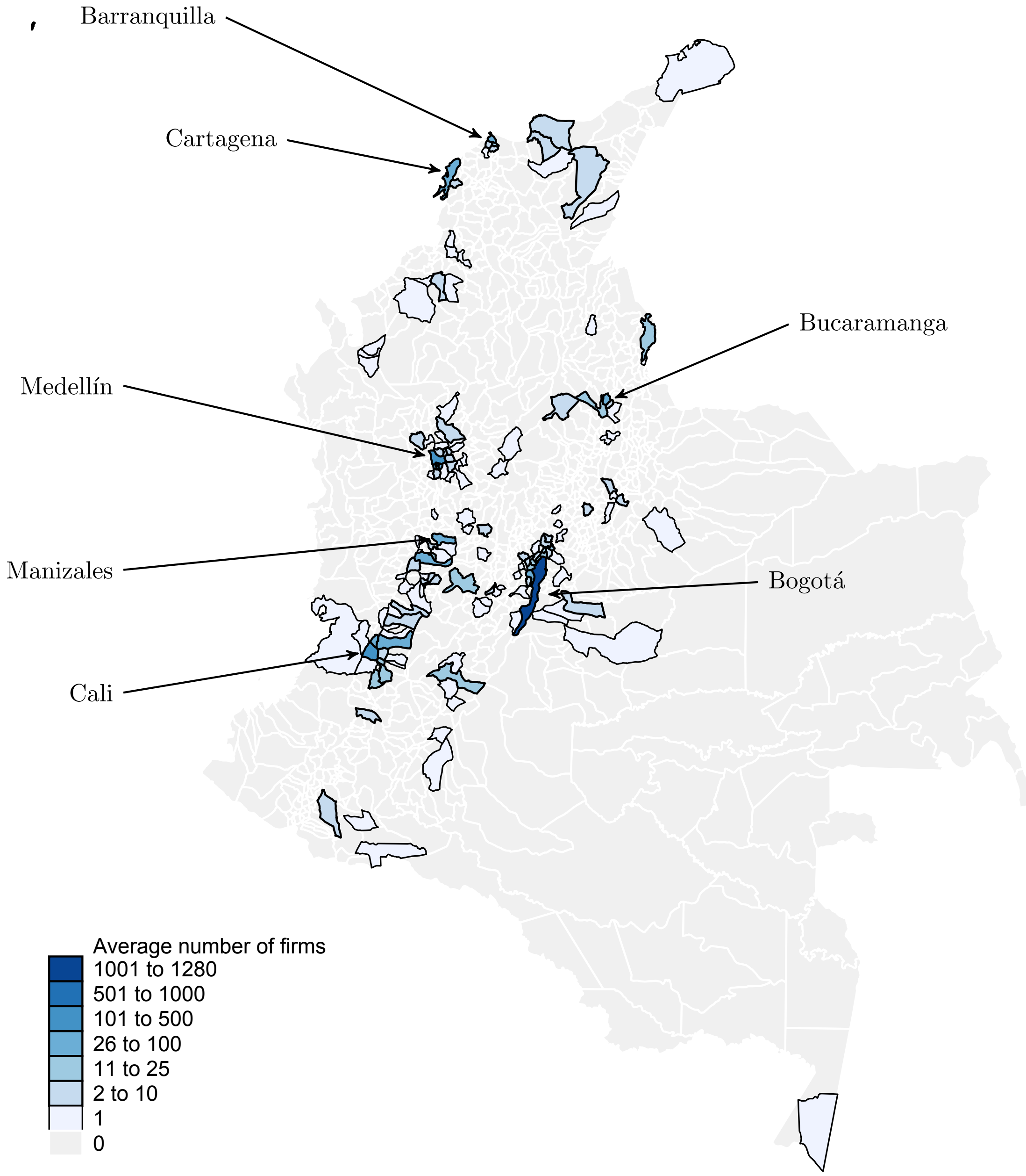
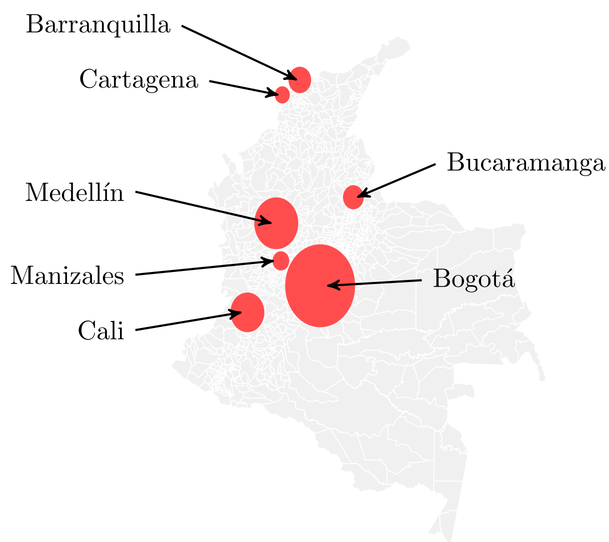


Figure 2: Geographical Distribution of Manufacturing Firms



Number of Firms The manufacturing firms in our sample are highly concentrated in a few regions. Figure 1 shows the distribution of firms across all municipalities in our data. Most of the firms are located along the Andean Mountains region and the North coast by the Caribbean Sea. In particular, while manufacturers in our data are located in 156 different cities, seven of those cities concentrate 70% of the manufacturing firms (and account for 67% of total revenues). These seven cities are Medellín, Barranquilla, Bogotá, Cartagena, Manizales, Bucaramanga, and Cali. The next group of cities that also concentrate a relatively high share of manufacturers is located within the metropolitan areas of the main three cities: Yumbo and Palmira (Cali metropolitan area), Itagüí and Envigado (Medellín metropolitan area), and Chía and Tocancipá (Bogotá metropolitan area). When we consider the main seven metropolitan areas, they account for 90% of the industrial establishments in Colombia.

To get a more clear picture of the geographical distribution of firms, in Figure 2, the area of each circle is proportional to the number of firms in each of the seven main cities. We can see that most firms are concentrated in the largest three cities. Bogotá is the city with the largest number of firms (44.7%), and it is followed by Medellín (9.9%) and Cali (6.2%).

We next explore the within-city distribution of firms across industries and the within-industry geographical distribution of firms. To do so, in Table 4 we present the number of firms by city and 2-digit industry, averaged across years —we include

Table 4: Number of Firms by City and Industry (Average)

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	38.4	18.4	173.6	8.3	12.3	20.0	34.0	184.4
17	26.4	5.4	70.4	1.0	1.3	1.0	6.7	42.6
18	57.6	4.9	97.6		1.9	12.0	22.2	55.7
19	5.0	2.4	38.7		1.0	4.1	4.7	19.6
20	4.8	2.1	18.2	1.2	1.0	1.0	1.4	16.1
21	5.3	1.0	18.1	1.0		1.0	5.4	31.0
22	21.1	8.9	122.3	1.0	3.8	6.0	23.2	32.8
24	28.8	14.1	174.6	8.4	3.2	2.2	20.0	96.6
25	24.2	13.3	158.3	8.3	3.4	9.7	14.7	97.3
26	14.6	7.0	39.1	1.6	4.8	5.0	4.9	60.3
28	23.7	5.4	110.4	3.1	7.1	3.3	14.0	70.8
29	4.6	2.9	49.4	1.4	1.1	3.7	3.3	27.1
30	1.2		1.8					2.0
31	4.7	2.0	27.9	1.0			5.0	13.6
34	5.2	1.3	58.2	1.0	2.3	1.8	3.9	31.9
36	18.2	10.0	121.8	1.9	5.0	4.8	15.1	79.4
Overall	283.2	97.4	1280.4	35.6	47.6	74.0	178.3	860.0

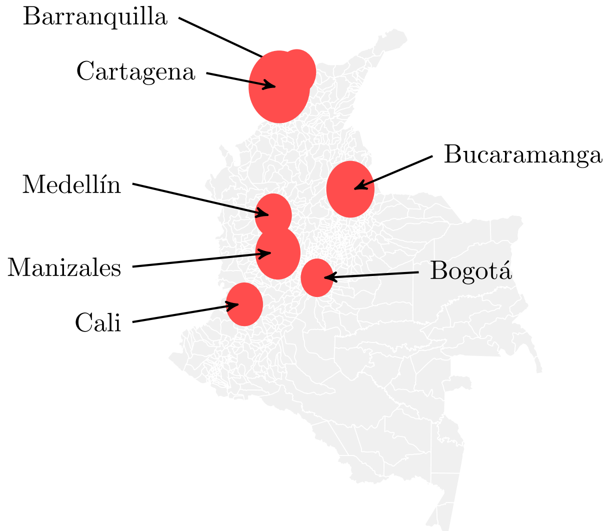
Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín; Barr.=Barranquilla; Bog.=Bogotá; Cart.=Cartagena; Man.=Manizales; and Buc.=Bucaramanga. 'Other' aggregates over the remaining municipalities.

the seven main cities and group together all the remaining cities in "Other." In Table 25 in the Appendix we present the share of firms by city and industry: panel (a) shows the distribution of industries within cities and panel (b) the geographical distribution within industries. Tables 30 and 31 in the Appendix present similar information for the main seven metropolitan areas. First, we observe that Bogotá and Medellín are the only two locations that have firms in all manufacturing sectors. More formally, to describe the diversity of industries within cities we construct an industry concentration index similar to a Herfindahl index.[12] Figure 3 shows the concentration index of the main seven cities (the area of each circle is proportional to the city's concentration index; the higher the index the more concentrated). Bogotá has the lowest sectoral concentration index meaning that the distribution of firms across industries is closest to a uniform distribution, that is, all industries are most closely to be uniformly represented. In terms of concentration, Bogotá is followed by Medellín and Cali. On the other hand, Cartagena exhibits the highest level of sectoral concentration, with food, chemicals, rubber and plastics accounting for three quarters of the total number of firms. Barranquilla, Manizales, and Bucaramanga exhibit an intermediate level of sectoral concentration with one dominating industry (the food industry, with at least 20% of the firms) followed by two or three industries that account for more than 30% of the firms.

If we now look at the distribution of industries within cities, the food industry tops the ranking in all cities, except in Medellín, accounting for a fifth or more of the firms in each of the cities (except in Medellín and Bogotá where food products account for about 14% of the firms). In Bogotá, the chemical and rubber and plastic products sectors are about the same size as the food industry. The main industry in Medellín is apparel (20% of firms), being the second largest cluster after Bogotá with around 20% of the firms in the sector. It is followed by food and beverages sector (14%) and chemicals (10%). At a lower level, apparel has a significant presence in Bucaramanga and Cali. Cartagena has the highest concentration of firms in the chemicals and rubber and plastics industries. The main sectors in Cali are food products, apparel, publishing, and chemical products.

[12]The Herfindahl index is used in the industrial organization literature to measure the level of concentration in a market. The index is simply given by $H \equiv \sum_i s_i^2$, where i indexes firms in the market and s_i is the market share of firm i. Here we adapt the index in two ways. First, we construct the concentration index for city c, $H_c \equiv \sum_s \alpha_{cs}^2$, where s indexes industries and α_{cs} is the fraction of firms in city c from sector s to the total number of firms in city c. Second, we construct the concentration index for industry s, $H_s \equiv \sum_c \gamma_{cs}^2$, where c indexes the main seven cities industries and γ_{cs} is the fraction of firms in city c from sector s to the total number of firms in sector s.

Figure 3: Sectoral Concentration Index



To sum up our preliminary findings so far: (i) Bogotá has the lowest level of sectoral concentration —i.e., the most diverse— among the main seven cities; (ii) Cartagena has the highest concentration with food, chemicals, rubber and plastic products dominating its manufacturing production; (iii) in all of the main cities except for Medellín, the highest concentration of firms is in the food and beverages industry; (iv) the textile and apparel industry appears to be clustered in Medellín and Bogotá; (v) the chemicals, rubber and plastic products industries seem to be clustered in Medellín and Bogotá; (vi) manufacturing of office and computing machines is done exclusively in Medellín and Bogotá.

We now extend our analysis of the geographical distribution of manufacturing activities by looking at firms' characteristics in different locations and industries. Specifically, we are going to look at the spatial distribution of revenues, skilled and unskilled labor, and capital stocks.

Revenues from Sales As we saw in the discussion above, manufacturing in Colombia is very concentrated in a few cities and a few sectors. This can also be seen by looking at revenues from sales. In Table 5, we show the average value of revenues (in billions of 2005 Colombian pesos) of manufacturing firms, broken down by industry and location.[13] Similar tables for the metropolitan areas can be found in the

[13] Again, Table 26 in the Appendix shows the revenue shares by city and industry (panel (a) shows the distribution of revenues across industries within cities, and panel (b) the geographical distribution of revenues within industries).

Table 5: Revenues by City and Industry (Average)

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	3,343	732	8,689	203	729	454	2,197	9,064
17	250	155	1,765	4	1	0	32	549
18	1,189	14	677		4	26	260	1,263
19	71	70	285		144	6	72	227
20	14	21	188	9	3	0	3	72
21	743	5	210	4		10	351	2,378
22	207	40	1,755	19	21	39	665	257
24	1,792	895	6,957	1,548	102	111	2,318	3,601
25	349	183	1,974	436	65	77	267	2,176
26	122	192	2,096	95	219	38	36	1,350
28	300	83	1,289	138	95	24	98	1,990
29	92	24	563	19	472	25	9	726
30	3		6					1
31	80	30	1,163	10			30	568
34	75	1	2,802	14	14	128	60	2,291
36	205	237	1,625	29	38	23	198	1,300
Overall	8,834	2,677	32,046	2,497	1,855	951	6,594	27,810

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín; Barr.=Barranquilla; Bog.=Bogotá; Cart.=Cartagena; Man.=Manizales; and Buc.=Bucaramanga. 'Other' aggregates over the remaining municipalities. Monetary figures in billions of constant Colombian pesos of 2005.

Appendix. Also, as a visual aid, Figure 4 shows the geographical distribution of revenues for the main seven cities (the area of each circle being proportional to the city's share in total revenues). Concentration is more accentuated when looking at revenues figures compared to what we found when looking at the number of firms. In other words, it appears that cities that concentrate many firms tend to have larger firms. The top industry in terms of revenues is food products in all of the main cities —expect for Barranquilla, Cartagena, and Cali— generating between 30 to 50% of each city's manufacturing revenues. The second industry is chemicals, which happens to be the main industry in Barranquilla, Cartagena, and Cali. These two industries together generate more than 50% of revenues in all cities.

Similarly to what we have done above, we can compute a concentration index now based on the revenues shares of the different industry-city pairs. Bogotá is still the least concentrated city, while Cartagena is the most concentrated one. In Cartagena, chemicals alone generate above 60% of the city's revenues, and the top two industries (chemicals and rubber and plastic products) generate 80% of the city's revenues.

Next we look at the geographical distribution within industries. Bogotá is the main manufacturing cluster in Colombia, attracting firms from all industries and

Figure 4: Geographical Distribution of Revenues

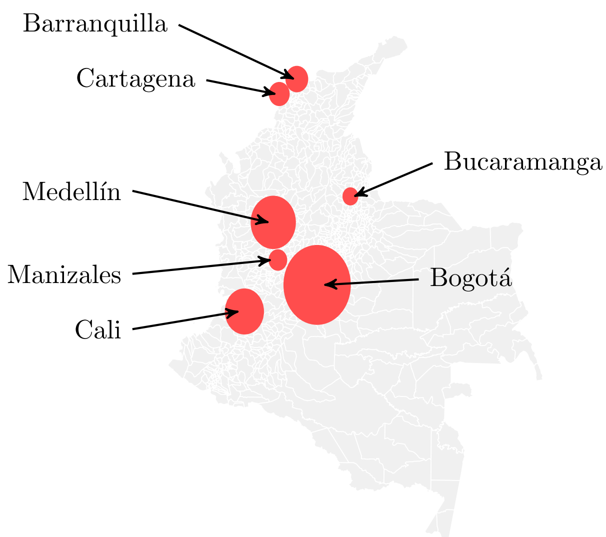


Figure 5: High Concentration Industries (Dist. of Revenues)

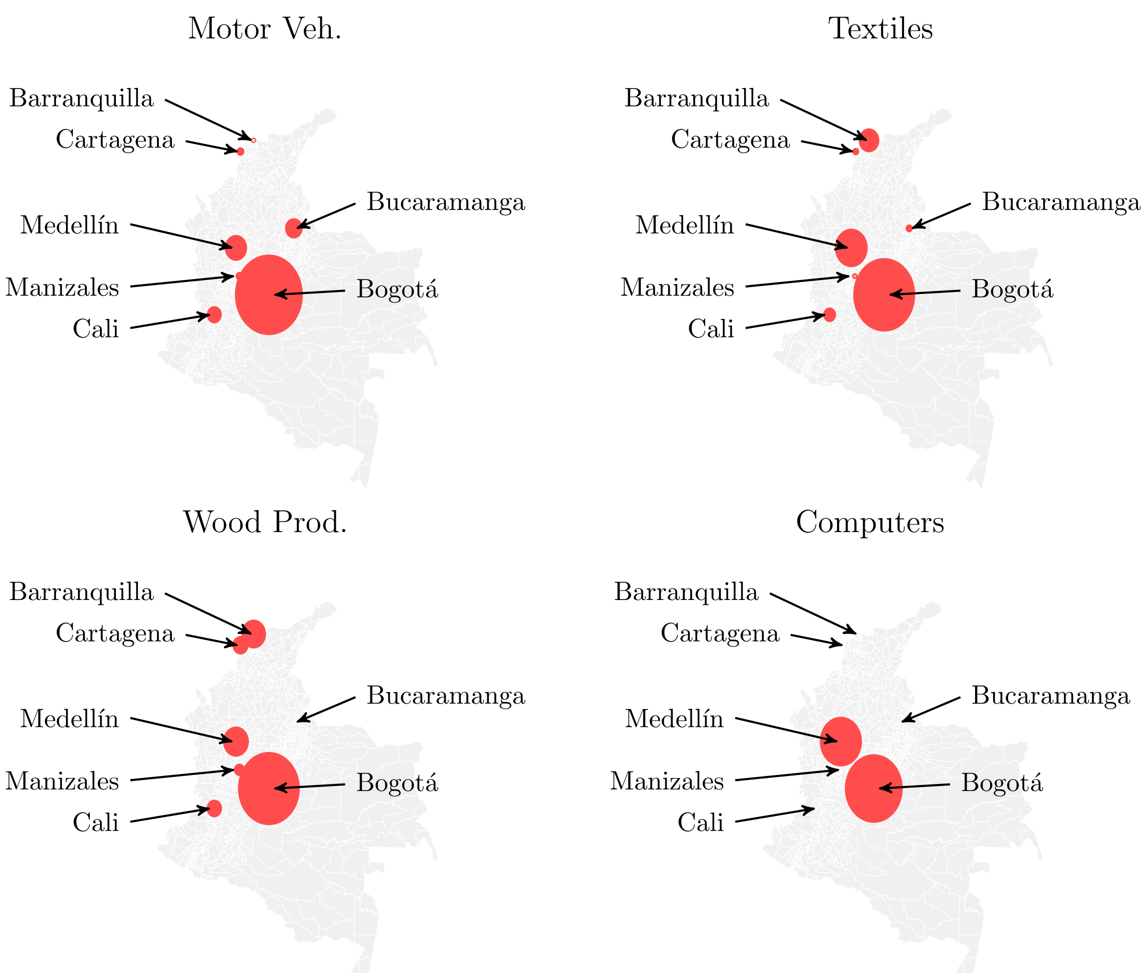


Figure 6: Medium Concentration Industries (Dist. of Revenues)

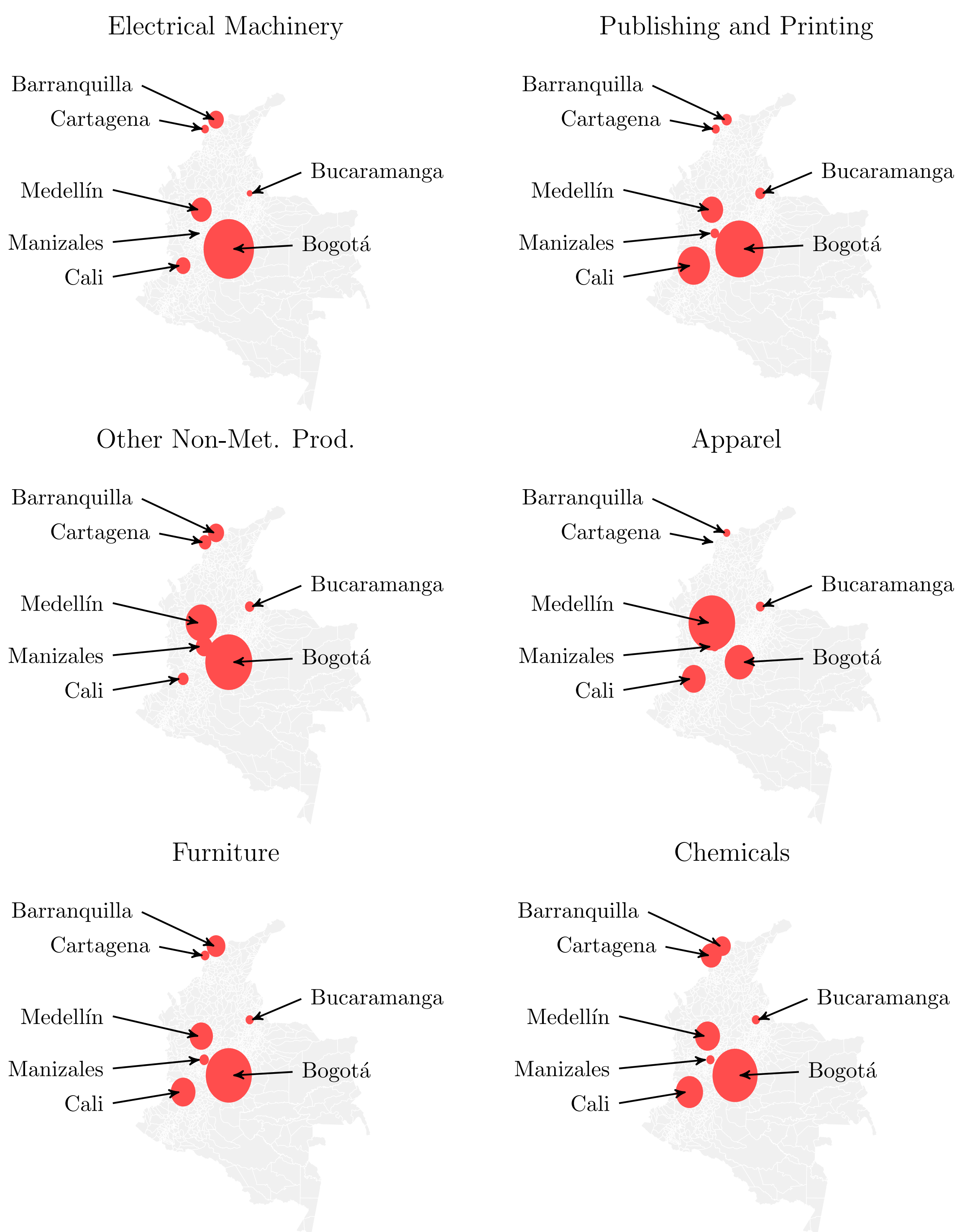
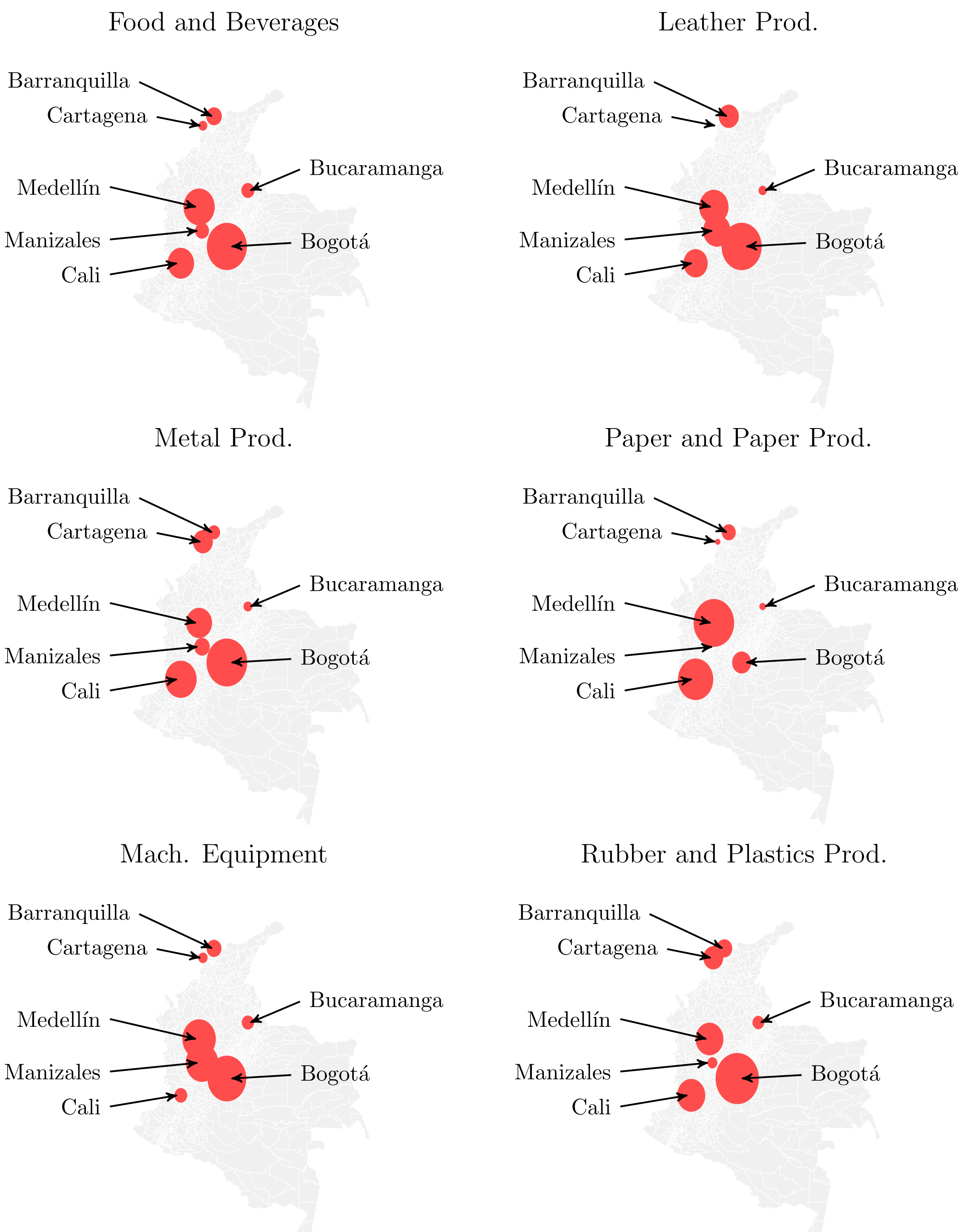


Figure 7: Low Concentration Industries (Dist. of Revenues)



generating a large fraction of the revenues in each sector, in many cases reaching more than 50% of nationwide revenues (like, e.g., in textiles, wood products, printing, rubber and plastics, office and computing machines, electrical machines, and motor vehicles). Apparel and paper products are the only two instances in which Bogotá is not the main producer. In both cases Medellín is the top location. To investigate the within industry spacial distribution, Figures 5 to 7 show the distribution of revenues by industry (the area of the circles being proportional to each city's share in revenues). We can broadly classify industries into highly spatially concentrated (motor vehicles, textiles, wood products, and computers), medium level of concentration (electrical machinery, publishing and printing, other non-metallic mineral products, apparel, furniture, and chemicals), and low concentration industries (food and beverages, leather products, metal products, paper and paper products, machinery equipment, and rubber and plastics products).

In particular, in addition to Bogotá, the food industry has two additional clusters in Medellín and Cali, with 13 and 9% of revenues, respectively. While Bogotá dominates the textiles industry there are also two smaller clusters: Medellín and Barranquilla. As we mentioned before, the apparel industry has its main cluster in Medellín (with 35% of nationwide revenues), followed by a cluster in Bogotá (about 20% of revenues) and a smaller one in Cali (about 8% of revenues). Leather products seem to be more geographically dispersed with a big cluster in Bogotá, but also sizable clusters in Manizales, Medellín, Barranquilla, and Cali. Paper products firms are clustered in Medellín and Cali. The printing industry is highly concentrated in a big cluster in Bogotá (with about 60% of total industry revenues) and a second, smaller one in Cali (22% of revenues). Chemical products have one big cluster of firms in Bogotá and three smaller ones (each about the same size) in Medellín, Cartagena, and Cali. Similarly for rubber and plastic products. Metal products has a big cluster in Bogotá and a smaller one in Medellín. Machinery and equipment has two cluster of about the same size in Bogotá and Manizales. Similarly, manufacturing of office and computing machines is concentrated almost exclusively in two clusters: Medellín (30% of revenues) and Bogotá (60% of revenues).

On the other hand, we observe very little geographical dispersion in a few industries. Wood products is highly concentrated in a unique cluster in Bogotá (more than 60% of revenues). We find a similar picture for non-metallic mineral products, electrical machinery, motor vehicles, and furniture with one big cluster of firms in Bogotá (50%, 60%, 50% and 45% of each sector's revenues, respectively).

Table 6: Number of Production Workers by City and Industry (Average)

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	7,864	2,452	16,139	697	1,751	1,123	5,961	21,582
17	1,893	1,493	9,466	14	24	8	440	5,230
18	8,579	392	6,925		83	578	2,266	11,572
19	603	555	2,372		438	50	884	2,354
20	152	113	1,187	32	6	4	37	854
21	879	26	1,124	0		342	670	5,050
22	810	180	4,872	34	88	24	2,059	1,301
24	2,659	1,490	7,053	324	232	597	3,843	6,490
25	1,798	1,527	9,572	712	485	217	1,078	8,945
26	1,241	925	4,916	283	577	170	186	5,650
28	1,521	527	4,925	335	607	165	506	6,494
29	436	111	3,470	78	747	221	68	4,175
30	4		15					10
31	392	192	1,858	45			333	1,844
34	338	8	5,526	39	269	603	303	6,103
36	1,509	1,191	5,600	145	155	49	883	5,167
Overall	30,677	11,144	85,021	2,633	5,371	3,845	19,514	92,814

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín; Barr.=Barranquilla; Bog.=Bogotá; Cart.=Cartagena; Man.=Manizales; and Buc.=Bucaramanga. 'Other' aggregates over the remaining municipalities.

Skilled and Unskilled Labor In Table 6, we report the number of production workers (i.e., blue-collar workers, our measure of unskilled labor) by city and industry (averaged across years).[14] Similar tables for the metropolitan areas can be found in the Appendix.

The first striking feature we observe from these tables is that —with a few exceptions— the distribution of unskilled workers within cities is less concentrated than what we original obtained by looking at either the number of firms or revenues. For example, in Bogotá, while the top two industries, food and chemical products, account for more than 27% and 20% of the city aggregate revenues, respectively, they only account for 19% and 8% of the unskilled labor force in the city. It is still true that Bogotá is the city with the lowest concentration index when we look at production workers, followed by Barranquilla, and all of the remainder cities have a higher, though similar, concentration index.

As expected, Bogotá accounts for the largest share of unskilled workers, followed by Medellín, Cali, and Barranquilla, and to a lesser extent by Manizales, Bucara-

[14]In Table 27 in the Appendix we present the share of unskilled labor by city and industry (panel (a) shows the distribution of unskilled labor within cities and panel (b) the geographical distribution within industries).

Table 7: Number of White Collar Workers by City and Industry (Average)

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	5,838	1,084	19,284	629	4,452	1,425	5,776	15,876
17	747	273	4,499	10	7	4	304	1,770
18	5,615	316	5,449		17	298	2,117	6,062
19	272	161	1,980		785	76	397	636
20	63	79	575	51	5	2	16	490
21	1,058	22	556	12		42	407	3,831
22	1,132	191	9,311	363	190	563	2,890	1,609
24	3,420	1,137	14,611	509	262	235	6,060	8,061
25	1,131	481	6,319	355	157	246	342	4,178
26	304	332	2,747	77	226	148	127	4,572
28	609	219	3,792	152	198	87	379	3,429
29	231	89	2,677	73	217	256	46	1,494
30	4		78					22
31	483	50	2,282	15			625	1,167
34	242	17	2,835	22	86	434	199	1,991
36	803	739	4,636	77	84	107	1,168	3,838
Overall	21,949	5,152	81,630	2,271	6,660	3,885	20,851	59,015

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín; Barr.=Barranquilla; Bog.=Bogotá; Cart.=Cartagena; Man.=Manizales; and Buc.=Bucaramanga. 'Other' aggregates over the remaining municipalities.

manga, and Cartagena. The food industry has the highest share of unskilled workers in all cities expect for Medellín and Cartagena, where apparel and rubber and plastic products have the highest share, respectively. In fact, in Medellín, food and apparel account for more that 50% of the unskilled labor force. Similarly, food and rubber and plastics account for over 50% of Cartagena's labor force.

In terms of the geographical distribution of the unskilled labor force within the different industries, the food industry has a major cluster in Bogotá with 28% of the industry's production workers followed by two clusters in Medellín and Cali (14% and 10%, respectively). Textiles is more concentrated, with a big cluster in Bogotá (51%) and two smaller ones in Medellín (10%) and Barranquilla (8%). We observe a similar picture for wood products and furniture. The apparel industry has two clusters of similar size in Medellín and Bogotá. Paper products has three cluster of about the same size in Medellín, Bogotá, and Cali.

Table 7 (and Table 28 in the Appendix) report similar statistics for white-collar workers (our definition of skilled workers). Qualitatively, we obtain similar results as we have seen for unskilled workers. Skilled labor seems to be more concentrated across industries within cities, but only slightly more than unskilled labor.

Table 8: Capital Stock by City and Industry (Average)

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	2,080	271	4,116	74	391	244	2,009	5,141
17	126	88	1,146	1	1	1	7	361
18	302	7	310		1	12	43	569
19	15	35	63		26	3	11	103
20	11	9	199	4	1	0	1	75
21	280	2	124	0		3	524	2,061
22	135	15	1,164	14	13	38	385	192
24	887	560	2,641	884	51	104	1,129	1,719
25	184	110	1,263	433	43	35	311	1,355
26	75	202	2,642	404	154	16	11	1,251
28	136	59	620	102	43	14	43	1,117
29	30	5	211	6	114	13	6	364
30	1		3					1
31	42	14	308	4			5	250
34	22	1	550	3	7	33	18	486
36	103	125	510	21	8	5	73	635
Overall	4,429	1,499	15,871	1,940	840	518	4,578	15,679

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín; Barr.=Barranquilla; Bog.=Bogotá; Cart.=Cartagena; Man.=Manizales; and Buc.=Bucaramanga. 'Other' aggregates over the remaining municipalities. Monetary figures in billions of constant Colombian pesos of 2005.

Capital Stock In Table 8, we report capital stocks by city and industry and, once again, in Table 29 in the Appendix we report shares across industries within cities and across cities within industries. Bogotá tops the ranking with a capital stock which is more than three times that of the two cities that follow it (Cali and Medellín). The remaining cities have a much lower stock of capital. In contrast with what we have observed for the labor measures, industrial concentration indices constructed using capital stock shares are significantly higher —even higher than those constructed using revenue shares. Capital appears to be more concentrated in fewer industries within cities. In relative terms, it is still the case that Bogatá is the city where capital is least concentrated across industries and Cartagena the city where capital is most concentrated. Barranquilla ranks second, and Medellín, Manizales, Bucaramanga, and Cali have similar concentration indices which are not very far from that of Cartagena. To illustrate, the level of concentration of the top two industries in Medellín, Cartagena, Manizales, Bucaramanga, and Cali account for almost 70% of each city's capital stock. In Barranquilla the latter figure goes down to 55% and in Bogotá to about 40%.

Similar to our previous findings, the food industry is the most important one in all cities except for Barranquilla and Cartagena, making up for about 45% of the capital stock in many cities (Medellín, Manizales, Bucaramanga, and Cali). The second industry in terms of capital stock is chemicals, except in Barranquilla and Cartagena where it ranks first, and Manizales where chemicals has very little presence. In terms of geographical concentration of capital stock our findings are the following. The food industry has three clusters: one relatively bigger in Bogotá and two of similar size in Medellín and Cali. Textiles, wood products, metal products, electrical machines, and motor vehicles are highly concentrated in one cluster in Bogotá. Apparel exhibits two clusters of similar size in Medellín and Bogotá. Printing has a big cluster in Bogotá and a second smaller one in Cali. In the case of chemicals, there is a relatively big cluster in Bogotá and three smaller ones (of about the same size) in Medellín, Cartagena, and Cali. Rubber and plastics has a big cluster in Bogotá and two smaller ones in Cartagena and Cali. Machinery has a big cluster in Bogotá and a smaller one in Manizales.

4 Productivity and Cities in Colombia

Productivity is generally understood as a source of heterogeneity across firms in the measure of output per inputs used. Hence, the first step in estimating firm-level TFP is to estimate the firm's production function which relates inputs to output. However, as pointed out by Marschak and Andrews (1944), the estimation of a production function has a fundamental difficulty: if the unobserved productivity shocks are correlated with the firm's input choices, then standard econometric techniques will yield biased estimates of the production function coefficients, affecting the resulting TFP estimates as well. In this section, we describe the methodology that we follow in order to account for the potential simultaneity bias, and we present the estimated production function coefficients and the corresponding TFP estimates.

4.1 Methodology

There are different ways in which the empirical literature has dealt with the endogeneity problem caused by the unobserved productivity shock. Initially, two traditional methods to address the endogeneity problem have been suggested in the production function estimation literature: (firm) fixed effects and instrumental variables. These approaches, however, rely on implausible assumptions and thus have not yielded satisfactory results in empirical applications (see Ackerberg, Benkard, Berry, and Pakes

(2005) for a review). Another approach, that has proven to be empirically successful, relies on structural timing assumptions that determine the relationship between input choices and unobserved productivity shocks. This approach, called the *control function* approach, leverages an observable endogenous variable to proxy for productivity, thus allowing the researcher to *control* for it, eliminating the endogeneity problem. This idea was originally presented in the context of production functions in the seminal paper by Olley and Pakes (1996) and later extended by Levinsohn and Petrin (2003) and Ackerberg, Caves, and Frazer (2016). Intuitively, if the demand for some input is strictly increasing in the productivity shock, that input demand function can be inverted and the unobservable productivity shock can be written as an unknown function of observables. The latter function of observables is included now in the estimating regression of the production function as another regressor to control for the TFP shock. For our analysis, we follow the methodology proposed by Ackerberg, Caves, and Frazer (2016) and use the firm's raw materials consumption as the observable proxy for its TFP. We briefly describe the methodology next.[15]

We observe an unbalanced panel of firms $j \in \{1, \ldots, J\}$ over periods $t \in \{1, \ldots, T\}$ located in cities $c \in \{1, \ldots, C\}$.[16] The relevant variables for our purpose are the firm's output (Y_{jt}), and its inputs: labor (L_{jt}), capital (K_{jt}), and intermediate inputs (M_{jt}). We follow the convention in the literature and use lowercase letters, y_{jt}, l_{jt}, k_{jt} and m_{jt}, to denote the logarithms of their uppercase counterparts. It is assumed that period t's capital is chosen in $t-1$ (investments take a full period to become productive, i.e., there is time-to-build), labor is chosen sometime between $t-1$ and t, and materials are fully flexible and are chosen at t once the productivity shock is observed by the firm.

The value-added production function that describes the firms' technology can be expressed as

$$Y_{jt} = F_t(L_{jt}, K_{jt}) e^{\nu_{jt}} \ , \tag{1}$$

where ν_{jt} is a Hicks-neutral efficiency term given by the sum of a productivity shock, ω_{jt}, observed by the firm at the beginning of every period, plus an idiosyncratic ex-post shock ϵ_{jt} which is unforeseen by the firm when making its input choices, is assumed to be i.i.d., and without loss of generality can be normalized to $E[\epsilon_{jt}] = 0$. The latter can also be thought as measurement error. Under this definition, it is the

[15] For more details, see Levinsohn and Petrin (2003) and Ackerberg, Caves, and Frazer (2016).

[16] In our sample, while we observe firms entering and exiting, we do not observe firms changing their location.

productivity term ω_{jt} that causes the endogeneity issue.

The observed (by the firm) productivity shock is assumed to be persistent, evolving over time following a first-order Markov process. Important to us, and differently from what is usual in the literature, we allow city characteristics to directly affect the process that governs the evolution of TFP. We can then write the process as

$$\omega_{jt} = h(\omega_{jt-1}, x_{ct}) + \eta_{jt} \ , \tag{2}$$

where x_{ct} is a vector of city c's characteristics at time t and η_{jt} can be interpreted as the innovation to the firm's persistent productivity in each period, and is orthogonal to ω_{jt-1}. The vector x_{ct} is assumed to follow an exogenous process.

Ackerberg, Caves, and Frazer (2016) propose a multi-stage, semiparametric estimation algorithm to recover the parameters of the production function. The first stage's purpose is to separate the productivity term ω from the ex-post shock ϵ, while all the production function coefficients are estimated in a second stage. For estimation, we assume that the production function is Cobb-Douglas as is common practice in the literature. Taking logs, the production function can then be written as

$$y_{jt} = \beta_l l_{jt} + \beta_k k_{jt} + \omega_{jt} + \epsilon_{jt}. \tag{3}$$

Since intermediate inputs are assumed to be fully flexible and chosen after the firm observes ω_{jt}, these are used as the proxy variable. This means that materials are treated like a function of the firm's state variables, assumed strictly increasing in ω for any realization of (k, l, x), and that for strictly positive values of m this function can be inverted to express productivity as a function of materials and these state variables—all observable variables.

More formally, in the first stage, m_{jt} is expressed as a function —denoted $m(\cdot)$— of productivity, labor, capital, and city characteristics. This function can be inverted to express productivity as a function of observable variables $(m_{jt}, l_{jt}, k_{jt}, x_{ct})$, that is,

$$\omega_{jt} = m^{-1}(m_{jt}, l_{jt}, k_{jt}, x_{ct}).$$

This proxy function is then replaced in the production function, such that output can be written as

$$y_{jt} = \beta_l l_{jt} + \beta_k k_{jt} + m^{-1}(m_{jt}, l_{jt}, k_{jt}, x_{ct}) + \epsilon_{jt} \ . \tag{4}$$

Now, grouping all terms that involve the inputs l, k, and m, we rewrite the production function as

$$y_{jt} = \phi(l_{jt}, k_{jt}, m_{jt}, x_{ct}) + \epsilon_{jt} \tag{5}$$

where $\phi(\cdot)$ is an unknown function. We can estimate $\phi(\cdot)$ using any consistent nonparametric estimator, allowing one to isolate the ex-post shock ϵ, and to get an estimated value of the expected outcome $\widehat{\phi}$. In the second stage, we use the fact that TFP follows a first-order Markov process and hence we can write

$$\omega_{jt} = E[\omega_{jt}|\omega_{jt-1}, x_{ct}] + \xi_{jt}$$

where ξ_{jt} is the innovation in the evolution process of ω. Using our definition of $\phi_t(l_{jt}, k_{jt}, m_{jt})$ we can then write

$$\begin{aligned}\xi_{jt} &= \omega_{jt} - E[\omega_{jt}|\omega_{jt-1}, x_{ct}] \\ &= \phi(l_{jt}, k_{jt}, m_{jt}, x_{ct}) - \beta_l l_{jt} - \beta_k k_{jt} - \Lambda\left(\phi(l_{jt-1}, k_{jt-1}, m_{jt-1}, x_{ct-1}) - \beta_l l_{jt-1} - \beta_k k_{jt-1}\right)\end{aligned}$$

where $\Lambda(\cdot) \equiv E[\omega_{jt}|\cdot]$ is an unknown function that can be estimated nonparametrically using the estimates of ω_{jt} obtained in the first-stage and candidate values for β_l and β_k. Fixing the values of the parameters β_l and β_k, we can obtain an estimate of ξ simply by plugging in estimates of ϕ and Λ:

$$\hat{\xi}_{jt}(\beta_l, \beta_k) = \hat{\phi}_t(l_{jt}, k_{jt}, m_{jt}) - \beta_l l_{jt} - \beta_k k_{jt} - \hat{\Lambda}\left(\hat{\phi}(l_{jt-1}, k_{jt-1}, m_{jt-1}, x_{ct-1}) - \beta_l l_{jt-1} - \beta_k k_{jt-1}\right)$$

Given the timing assumptions in the choices of inputs, the current level of capital and the lagged level of labor are valid instruments in that they are orthogonal to ξ_{jt}. Then, we can estimate β_l and β_k by GMM using the moment conditions

$$E\left[\xi_{jt}\begin{pmatrix} k_{jt} \\ l_{jt-1} \end{pmatrix}\right] = 0$$

Once we obtain estimates of β_l and β_k we can then recover the TFP shock for firm j at time t as

$$\hat{\omega}_{jt} = \hat{\phi}_t(l_{jt}, k_{jt}, m_{jt}) - \hat{\beta}_l l_{jt} - \hat{\beta}_k k_{jt}.$$

We compute standard errors via block-bootstrap.

4.2 Results

We first report our estimates of the production function coefficients and the corresponding TFP for the period 2005–2013. Since the production technology might differ across industries, we estimate separate production functions for each industry and allow for geographical and temporal variation in the evolution of productivity. In our implementation we break down labor into blue- and white-collar. The coefficients we present below were obtained by pooling all manufacturing firms within a 2-digit industry together to estimate sector-specific production functions. We include the largest industries (in terms of the number of firms), which represent around 98% of the observations included in our sample.[17] While sectors 29, 30, and 31 are relatively small if we consider them separately, we can group them together into a 'machinery' sector and we estimate a single production function for it. The estimates of the production function coefficients for each 2-digit sector are reported in Table 9 (standard errors were obtained via block-bootstrap). We also include in the table the estimated returns to scale (RTS).

In general, most sectors have close to constant returns to scale. Printing, chemicals, and the furniture industry seem to have moderate increasing returns to scale. In terms of input-output elasticities, we do find heterogeneity across industries. The unskilled labor elasticity is about .18 in many industries, with leather products having the highest elasticity (.23), and chemicals having the lowest (.06). In terms of skilled labor elasticities, there seems to be two groups of industries. The first one has an elasticity between .45 to .5 (food, tobacco, textiles, apparel, leather, rubber and plastics, metal products, machinery, motor vehicles) and the second one an elasticity of about .6 (wood products, printing, chemicals, and furniture). Finally, in terms of capital elasticity, the most productive industries are paper products, chemicals, metal products, and motor vehicles with an elasticity of about .5; the least productive industries are apparel and wood products with an elasticity of about .25; there is also a set of industries with an intermediate level of elasticities ranging from .35 to .45 (food, textiles, printing, rubber and plastics, machinery, and furniture).

Productivity and Cities Now, turning to our estimates of TFP, given the estimated production function coefficients, we can recover the productivity shock for

[17] As we have mentioned above, we do not include sectors 23 and 27 since these are closely related to commodities and hence their dynamics are likely different from other manufacturers. Sectors 16, 32, 33 and 35 contain only a few firms each and therefore are excluded.

Table 9: Estimated Production Function Coefficients

2-digit Sector	Blue Collar	White Collar	Capital	RTS	Observations
15	0.1580***	0.4818***	0.4558***	1.0956	4,500
	(0.0133)	(0.0271)	(0.0328)		
17	0.1873***	0.4657***	0.4052***	1.0582	1,390
	(0.0280)	(0.0486)	(0.0640)		
18	0.1811***	0.5248***	0.2549***	0.9608	2,270
	(0.0255)	(0.0441)	(0.0497)		
19	0.2325***	0.4607***	0.3727***	1.0659	684
	(0.0618)	(0.1219)	(0.1411)		
20	0.1864***	0.6052***	0.2583***	1.0499	395
	(0.0502)	(0.1083)	(0.0780)		
21	0.1453**	0.3979***	0.4814***	1.0246	550
	(0.0764)	(0.1120)	(0.1543)		
22	0.1791***	0.5903***	0.3508***	1.1202	1,975
	(0.0262)	(0.0566)	(0.0687)		
24	0.0634***	0.5911***	0.4870***	1.1415	3,135
	(0.0142)	(0.0385)	(0.0361)		
25	0.1917***	0.4654***	0.3793***	1.0364	2,972
	(0.0239)	(0.0468)	(0.0511)		
26	0.1685***	0.3559***	0.5435***	1.0679	1,235
	(0.0220)	(0.0677)	(0.0850)		
28	0.1700***	0.4411***	0.4679***	1.079	2,151
	(0.0234)	(0.0534)	(0.0688)		
29–31	0.1475***	0.5241***	0.4070***	1.0786	1,358
	(0.0279)	(0.0849)	(0.0859)		
34	0.1598***	0.4406***	0.4977***	1.0981	938
	(0.0359)	(0.0758)	(0.1199)		
36	0.1796***	0.5649***	0.4070***	1.1515	2,316
	(0.0212)	(0.0507)	(0.0544)		

every firm-year observation. To calculate annual averages for the manufacturing industry as a whole, we compute a weighted average where the weights are given by firm-level output. Then, we normalize the resulting estimates so that the value of the TFP index for the overall manufacturing sector equals 100 in 2005.

We begin by showing the geographical dispersion of TFP. We compute the weighted average of TFP of all firms in a given city across all years in our sample. The results are shown in Figure 8. While the analysis is done on all of the 154 cities, for presentational purposes we focus on the main seven cities. We see that, on average, Bogotá shows the highest level of productivity, followed in order by Medellín, Cali, Barranquilla, Manizales, Cartagena, and Bucaramanga. To also look at the time dimension of the geographical distribution of the TFP index, in Figure 9 we depict three snapshots in which the areas of the circles are proportional to TFP. To complement this, in Figure 10 we plot the evolution of the estimated aggregate TFP during 2005–2013 for each one of the seven main cities. The thick, black, solid line shows the national TFP index and the colored, patterned lines show the average TFP for the main cities. We can see that there is great deal of heterogeneity across locations. Barranquilla and Bucaramanga are consistently showing low levels of productivity, with a flat trend throughout our sample period. On the other hand, Bogotá and Medellín show the highest levels of productivity across time with a slight downward general trend. In the case of Medellín, average productivity increases from 2005 to 2007 to then decrease until the end of our sample period. Manizales follows a similar time pattern. The case of Cartagena is surprising. Cartagena started with the lowest productivity level of our main cities in 2005, and by the year 2013 it attained the highest level, with an increase of 250%, or about a 12% average annual increase.

To further investigate the heterogeneity in productivity, we now look at city-industry averages. In Table 10, we average across years and we break TFP across cities and industries. The table shows that the most productive sector is apparel. Not only that, but the apparel industry is also the one which shows the lowest level of geographical variation. Within this industry, Cali is the most productive city, followed by Bogotá (10% less productive than Cali), Barranquilla (16% less productive), Medellín (24% less productive), Manizales (31% less productive), and Bucaramanga (35% less productive). Bogotá is the most productive city in leather products, chemicals, electrical machinery, motor vehicles, and furniture. Cali is the most productive in the apparel and metal products industries. Medellín is the most productive in the wood products and computing machines; Barranquilla in textiles and printing indus-

Figure 8: Average TFP by City

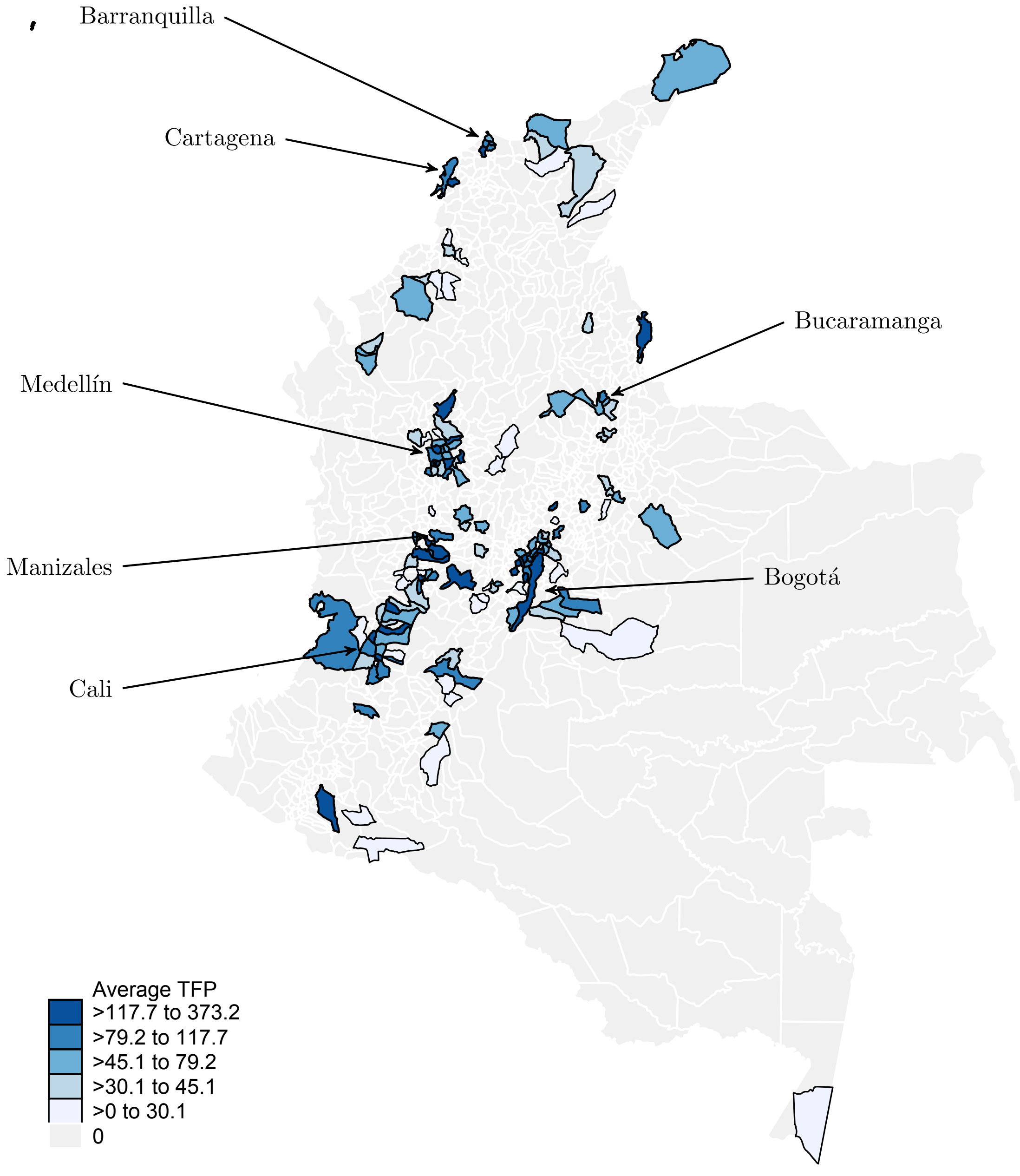


Figure 9: Geographical Distribution of the TFP Index

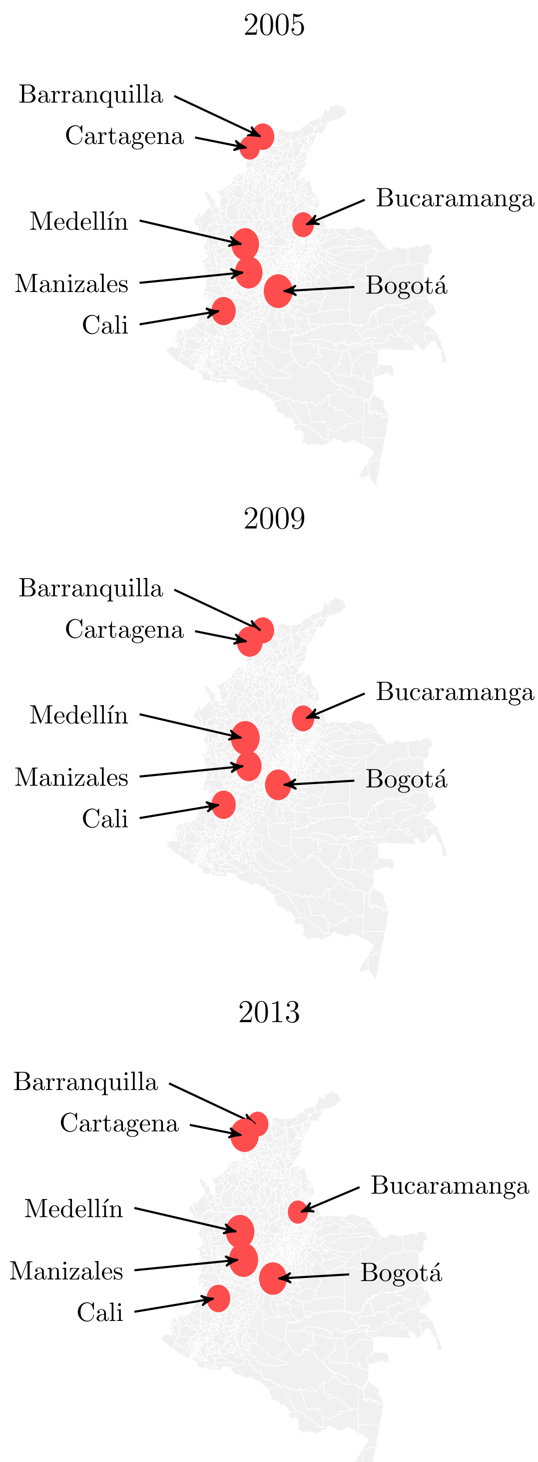


Figure 10: TFP Index

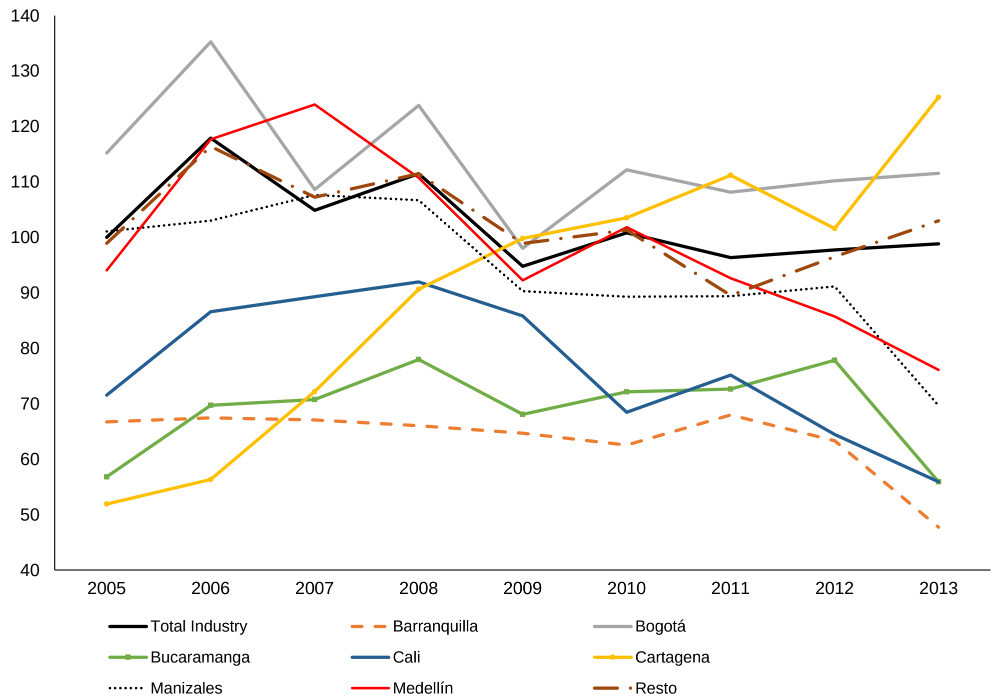


Table 10: Average TFP by city and industry

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	53.7	40.0	73.7	78.9	54.3	39.8	43.4	59.7
17	93.2	116.8	87.9	101.8	40.6	16.2	86.0	132.0
18	334.3	340.3	380.9		289.4	266.9	436.4	573.6
19	184.8	96.4	197.5		199.6	92.6	170.1	204.7
20	205.5	171.9	158.6	141.0	166.7	90.3	161.2	111.8
21	66.6	30.1	58.3	157.2		32.1	50.5	64.1
22	84.0	141.4	119.2	36.9	70.4	56.6	87.4	97.9
24	72.8	41.9	94.3	68.7	30.3	19.3	38.3	45.5
25	134.3	128.5	133.5	176.2	95.1	215.8	146.9	167.3
26	31.0	19.0	23.1	9.6	38.8	46.7	32.9	29.4
28	43.9	45.4	66.0	110.3	52.5	48.5	171.4	56.8
29	161.8	176.4	83.7	95.0	197.3	59.4	56.6	83.5
30	192.0		67.3					35.7
31	74.6	78.6	200.1	45.1			94.4	108.0
34	64.3	30.0	216.5	34.9	27.5	62.8	56.2	118.1
36	84.1	81.7	197.3	52.9	71.8	200.4	84.5	132.2
All	91.4	60.6	113.0	88.2	97.2	66.8	71.2	98.7

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín; Barr.=Barranquilla; Bog.=Bogotá; Cart.=Cartagena; Man.=Manizales; and Buc.=Bucaramanga.

tries; Cartagena in food and paper products; Manizales in machinery; Bucaramanga in rubber and plastic products.

5 The Determinants of Productivity

In this section we address the main objective of the paper: uncovering the determinants of productivity. Cities, and the firms that locate in them, are the outcome of a trade-off between agglomeration economies or localized aggregate increasing returns and the costs of urban congestion (see, e.g., Duranton and Puga (2004)). As we have mentioned before, these are the two sources of forces that we relate to firm-level productivity. On the one hand, we focus on the effect of the forces behind agglomeration that allow firms to attain higher levels of productivity as a result of their geographical location. On the other hand, we also investigate the centrifugal forces that hinder firms from attaining higher levels of productivity which come about in

Table 11: Scale by City (Average)

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
1	269	97	1280	36	48	74	178	858
L	50,407	16,296	166,651	4,904	12,031	7,730	40,366	151,457
K	4,393	1,499	15,871	1,940	840	518	4,578	15,617
Y	8,367	2,677	32,046	2,497	1,855	951	6,594	27,810

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín; Barr.=Barranquilla; Bog.=Bogotá; Cart.=Cartagena; Man.=Manizales; and Buc.=Bucaramanga. 'Other' aggregates over the remaining municipalities. Monetary figures in billions of constant Colombian pesos of 2005.

highly concentrated areas.

We proceed in two parts. We first look at traditional measures of economies of agglomeration that have been proposed in the literature and relate them to our estimates of firm-level productivity. We then augment our analysis by incorporating very detailed information at the city level and further investigate other drivers of productivity that are new in this type of studies.

5.1 Economies of Agglomeration

We define 'agglomeration economies' in a broad sense as any force that increases firms' output when the size of the local economy grows (see, e.g., Combes and Gobillon (2015)). In particular, we investigate the effect of different measures of *agglomeration* commonly used in the literature. Formally, let i denote a firm, s denote a sector, and c denote a city. In an abuse of notation, we write '$i \in c$' to denote whether firm i is located in city c or not and '$i \in c, s$' to denote whether firm i is located in city c and belongs to sector s. Also, let S denote the set of all sectors, C the set of all cities, and let $\mathcal{N}_c$ denote the set of city c's neighboring cities.

Our first set of basic measures are related to the *scale* of local economic activity which we define by aggregating different variables. Specifically, we measure the scale of city c by

$$\mathrm{sc}_c^x = \sum_{i \in c} x_i$$

where x_i is either the constant 1 (so that the scale variable is equal to the number of manufacturing establishments in city c), firm i's employment, firm i's capital stock, or firm i's production. Table 11 presents summary statistics broken down by cities.

Our second set of measures intend to capture the composition of the local industrial structure and are aimed to quantify inter- and intra-industry types of externalities. We construct measures of the degree of sector specialization to capture *localization economies* or intraindustry spillovers. Formally, our specialization variables for sector s in city c are given by

$$\mathrm{sp}^{x}_{cs} = \frac{\sum_{i \in c,s} x_i}{\sum_{i \in c} x_i} \times 100 = \frac{\sum_{i \in c,s} x_i}{\mathrm{sc}^{x}_{c}} \times 100$$

where, again, x_i is either the constant 1, firm i's employment, firm i's capital stock, or firm i's production. In words, these measures quantify the size of a sector in a given city relative to the overall industrial activity in the city. These variables range from 0 to 100, where 0 denotes that the industry has no participation in the city and 100 denotes that the industry dominates all of the manufacturing activity in the city. Table 12 presents summary statistics by cities and industries.

We also construct measures of industrial heterogeneity (or variety) to capture *urbanization economies* or cross-industry spillovers. Formally, our variety variables for sector s in city c are given by pseudo-Herfindahl indices

$$\mathrm{va}^{x}_{cs} = \sum_{s' \neq s} \left(\frac{\sum_{i \in c,s'} x_i}{\sum_{i \in c, i \notin s} x_i} \times 100 \right)^2$$

where, again, x_i is either the constant 1, firm i's employment, firm i's capital stock, or firm i's production. These variables range from 0 to 10,000, where the closer to 0 (10,000) means that the industry under consideration faces a very diverse (concentrated) set of industries in the city. Table 13 presents summary statistics.

To capture the innovation and technological diffusion through which the local industrial structure might affect productivity, our third set of variables measure the local level of *competition* and the average *size* of rival firms in a given sector. Formally, our competition variables for firm i in sector s and city c are given by

$$\mathrm{co}^{x}_{ics} = \sum_{j \neq i, j \in c,s} \left(\frac{x_j}{\sum_{j \neq i, j \in c,s} x_j} \times 100 \right)^2$$

where, x_i is either firm i's employment, firm i's capital stock, or firm i's production. Again, these Herfindahl indices range from 0 to 10,000, where the closer to 0 (10,000) means that the firm faces a very competitive (concentrated) set of rivals in its industry. Table 14 presents summary statistics.

Table 12: Specialization by City and Industry (Averages)

		Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	1	14.35	18.79	13.59	23.39	26.04	26.98	19.05	52.48
	L	27.18	21.67	21.26	26.80	46.85	33.23	29.02	55.38
	K	45.43	18.33	25.24	3.97	46.64	47.16	43.72	58.61
	Y	40.14	27.17	26.99	8.46	39.59	47.63	33.41	61.10
17	1	9.83	5.51	5.49	2.91	2.63	1.36	3.64	15.05
	L	5.27	10.89	8.41	0.56	0.30	0.16	1.89	14.20
	K	3.18	5.63	7.43	0.08	0.14	0.12	0.17	13.72
	Y	3.03	5.83	5.55	0.16	0.06	0.02	0.49	11.63
18	1	15.94	4.96	7.59		3.94	16.04	12.38	26.90
	L	23.69	4.14	7.40		0.85	11.40	10.77	38.85
	K	6.64	0.51	1.97		0.14	2.48	0.94	27.99
	Y	8.68	0.53	2.12		0.24	2.81	3.89	30.59
19	1	1.84	2.51	2.99		2.11	5.45	2.60	23.62
	L	1.77	4.35	2.60		10.95	1.67	3.26	24.40
	K	0.34	2.41	0.41		3.16	0.54	0.25	19.70
	Y	0.85	2.63	0.90		7.89	0.65	1.10	16.60
20	1	1.74	2.21	1.42	3.38	2.07	1.36	0.75	24.85
	L	0.43	1.22	1.06	1.97	0.10	0.07	0.14	21.46
	K	0.26	0.58	1.32	0.38	0.08	0.06	0.03	20.35
	Y	0.18	0.86	0.59	0.49	0.14	0.04	0.05	19.11
21	1	1.99	1.17	1.42	2.94		1.35	3.07	20.58
	L	3.85	0.31	1.01	0.26		4.94	2.58	22.09
	K	6.21	0.14	0.78	0.02		0.75	10.36	27.54
	Y	8.60	0.19	0.66	0.17		1.28	5.16	24.46
22	1	7.86	9.21	9.56	2.82	8.02	8.14	12.92	15.38
	L	3.87	2.32	8.52	8.31	2.41	7.64	12.13	9.58
	K	3.08	1.06	7.38	0.82	1.63	7.31	8.46	7.74
	Y	2.44	1.61	5.49	0.81	1.16	4.18	9.97	6.07
24	1	10.68	14.51	13.66	23.76	6.80	2.97	11.32	28.47
	L	12.02	16.03	13.02	17.23	4.50	10.67	24.90	28.05
	K	20.86	36.33	16.45	51.32	6.26	20.21	25.22	28.48
	Y	21.39	32.55	21.57	62.72	5.46	12.13	35.41	29.19

		Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
25	1	8.99	13.68	12.36	23.47	7.22	13.25	8.34	19.15
	L	5.82	12.44	9.53	22.17	5.85	5.99	3.55	14.43
	K	4.39	7.56	8.05	22.03	4.93	6.60	7.34	15.29
	Y	4.19	6.95	6.17	16.69	3.58	8.14	3.99	14.17
26	1	5.48	7.24	3.06	4.40	10.06	6.86	2.75	34.63
	L	3.07	7.71	4.60	7.28	7.21	4.12	0.79	33.39
	K	1.73	13.85	16.95	14.86	17.62	3.15	0.26	35.25
	Y	1.46	7.66	6.57	3.22	11.90	4.01	0.54	29.48
28	1	8.89	5.70	8.61	8.70	14.90	4.56	7.85	19.64
	L	4.24	4.53	5.21	9.13	7.34	3.28	2.23	18.80
	K	3.31	3.82	3.90	5.17	5.23	2.57	1.01	18.48
	Y	3.67	3.00	4.02	5.53	5.16	2.37	1.47	18.93
29	1	1.72	2.91	3.87	4.04	2.32	4.88	1.94	11.93
	L	1.32	1.21	3.68	3.02	9.30	6.06	0.29	13.61
	K	0.72	0.32	1.30	0.36	13.90	2.56	0.14	13.07
	Y	1.10	0.93	1.75	0.77	24.81	2.68	0.14	12.50
30	1	0.44		0.14					5.26
	L	0.02		0.06					0.59
	K	0.02		0.02					0.17
	Y	0.03		0.02					0.13
31	1	1.72	2.09	2.18	2.63			2.85	14.13
	L	1.73	1.52	2.48	1.10			2.24	14.21
	K	1.00	0.94	1.99	0.25			0.12	15.58
	Y	0.94	1.18	3.68	0.40			0.46	19.26
34	1	1.97	1.26	4.55	2.89	4.91	2.40	2.12	16.04
	L	1.16	0.14	5.02	1.43	3.39	13.22	1.25	20.84
	K	0.53	0.04	3.51	0.24	0.87	6.28	0.42	17.83
	Y	0.91	0.04	8.77	0.86	0.76	12.99	0.90	19.65
36	1	6.78	10.26	9.52	5.24	10.52	6.51	8.50	26.94
	L	4.57	11.98	6.15	4.60	2.08	1.99	4.97	22.46
	K	2.32	8.68	3.31	1.25	1.00	0.94	1.58	20.11
	Y	2.41	9.14	5.14	1.25	2.07	2.25	3.03	22.74

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín; Barr.=Barranquilla; Bog.=Bogotá; Cart.=Cartagena; Man.=Manizales; and Buc.=Bucaramanga. 'Other' aggregates over the remaining municipalities.

Table 13: Variety by City and Industry (Average)

		Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	1	1,086	1,177	1,002	2,229	1,231	1,333	1,092	4,472
	L	1,679	1,370	984	2,129	1,438	1,589	1,979	5,491
	K	1,977	2,743	1,434	4,336	2,170	2,094	3,107	6,021
	Y	1,914	2,583	1,475	5,183	2,311	1,934	3,313	5,728
17	1	1,118	1,240	1,011	1,933	1,385	1,483	1,147	2,129
	L	1,802	1,503	1,186	2,098	3,009	1,862	1,916	3,553
	K	2,882	2,418	1,637	5,108	2,855	2,931	3,001	4,247
	Y	2,448	2,335	1,668	4,641	2,467	2,864	2,655	4,148
18	1	1,051	1,233	1,025		1,451	1,671	1,205	2,805
	L	1,796	1,400	1,178		2,877	2,172	2,167	3,947
	K	3,042	2,204	1,518		2,806	3,066	3,048	4,997
	Y	2,678	2,130	1,584		2,441	3,031	2,823	4,546
19	1	1,040	1,192	982		1,413	1,581	1,130	2,798
	L	1,700	1,415	1,118		3,337	1,913	1,960	3,915
	K	2,743	2,282	1,475		2,972	2,955	3,006	4,888
	Y	2,349	2,216	1,550		2,791	2,901	2,687	4,601
20	1	1,039	1,149	959	1,895	1,395	1,472	1,098	3,312
	L	1,659	1,333	1,090	2,057	2,609	1,959	1,799	4,692
	K	2,737	2,207	1,499	4,797	2,738	2,936	2,902	5,265
	Y	2,319	2,129	1,540	4,542	2,300	3,149	2,549	5,208
21	1	1,043	1,122	959	2,158		1,436	1,139	2,421
	L	1,756	1,351	1,089	2,624		2,594	1,937	3,647
	K	3,057	2,459	1,486	4,184		3,203	3,498	4,182
	Y	2,666	2,237	1,543	4,359		3,280	2,878	3,958
22	1	1,111	1,276	1,030	1,980	1,528	1,634	1,201	3,248
	L	1,764	1,371	1,190	2,267	2,959	2,105	2,187	4,340
	K	2,888	2,227	1,641	4,072	2,888	3,329	3,457	4,891
	Y	2,420	2,171	1,670	4,481	2,485	3,094	3,100	4,520
24	1	1,117	1,268	1,002	2,253	1,506	1,524	1,208	3,121
	L	1,935	1,483	1,189	2,436	3,043	2,141	2,160	4,238
	K	3,577	1,892	1,704	4,431	3,138	3,794	4,125	5,037
	Y	2,975	2,243	1,721	3,266	2,680	3,413	3,180	4,802
25	1	1,116	1,271	1,017	2,233	1,516	1,676	1,200	2,909
	L	1,816	1,507	1,192	2,433	3,110	2,052	1,971	4,183
	K	2,950	2,473	1,651	5,669	3,067	3,279	3,332	4,930
	Y	2,498	2,372	1,686	5,923	2,601	3,278	2,824	4,621
26	1	1,090	1,258	984	2,033	1,553	1,608	1,133	3,439
	L	1,741	1,467	1,149	2,235	3,176	1,996	1,877	4,623
	K	2,816	2,597	1,678	4,558	3,653	3,102	3,006	5,358
	Y	2,377	2,374	1,692	4,691	2,948	3,089	2,657	5,302
28	1	1,114	1,238	1,029	2,159	1,559	1,559	1,198	2,930
	L	1,774	1,414	1,158	2,178	3,139	1,967	1,927	4,121
	K	2,893	2,345	1,568	4,360	3,086	3,072	3,047	4,798
	Y	2,476	2,235	1,635	4,894	2,674	2,995	2,704	4,748
29	1	1,038	1,199	994	2,013	1,407	1,569	1,118	2,077
	L	1,688	1,344	1,137	2,089	2,843	2,038	1,859	3,066
	K	2,761	2,196	1,501	4,031	3,423	3,072	3,000	4,041
	Y	2,361	2,146	1,575	4,475	3,079	3,016	2,636	3,479
30	1	1,015		936					1,451
	L	1,640		1,070					2,274
	K	2,658		1,464					3,246
	Y	2,366		1,523					2,766
31	1	1,038	1,182	971	1,673			1,134	2,942
	L	1,701	1,352	1,116	1,606			1,922	4,479
	K	2,776	2,222	1,518	3,414			2,999	5,689
	Y	2,353	2,156	1,623	4,010			2,653	5,313
34	1	1,042	1,161	1,002	1,858	1,473	1,510	1,122	2,794
	L	1,683	1,334	1,153	1,950	2,966	2,151	1,892	3,776
	K	2,752	2,314	1,556	5,430	2,845	3,273	3,014	4,475
	Y	2,352	2,376	1,704	4,801	2,465	3,528	2,675	4,265
36	1	1,103	1,278	1,030	2,032	1,556	1,604	1,202	3,053
	L	1,783	1,506	1,171	2,071	2,943	1,923	2,009	4,400
	K	2,848	2,503	1,551	4,087	2,854	2,982	3,091	5,125
	Y	2,418	2,434	1,658	4,530	2,527	2,995	2,788	4,921

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín; Barr.=Barranquilla; Bog.=Bogotá; Cart.=Cartagena; Man.=Manizales; and Buc.=Bucaramanga. 'Other' aggregates over the remaining municipalities.

Table 14: Competition by City and Industry (Average)

		Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	L	1,201	1,201	297	3,431	3,565	1,337	1,033	4,393
	K	2,366	1,297	1,025	4,379	2,550	1,711	2,670	5,195
	Y	1,579	1,411	422	4,223	3,335	1,321	1,552	4,773
17	L	963	4,422	719	0	4,883	0	4,469	3,421
	K	1,392	6,305	902	0	4,883	0	4,251	3,755
	Y	992	5,233	1,101	0	4,883	0	3,997	3,358
18	L	1,457	4,806	413		6,267	1,747	1,685	4,578
	K	3,101	4,787	835		5,207	2,739	1,666	4,885
	Y	1,527	3,798	529		5,697	1,310	2,201	4,513
19	L	6,398	8,787	1,830		0	4,888	5,528	3,199
	K	6,670	9,346	1,378		0	5,366	5,723	3,796
	Y	6,495	9,336	2,048		0	4,551	5,897	3,037
20	L	4,183	7,101	2,266	2,046	0	0	7,010	1,049
	K	4,714	7,071	4,020	2,046	0	0	7,010	1,075
	Y	3,993	7,611	2,757	2,046	0	0	7,010	974
21	L	4,931	0	1,605	0		0	4,031	6,451
	K	5,360	0	2,449	0		0	6,824	6,686
	Y	5,978	0	2,156	0		0	6,291	6,296
22	L	1,571	1,915	459	0	5,749	3,356	2,010	4,857
	K	2,130	1,982	979	0	7,193	3,981	2,732	5,526
	Y	2,145	2,304	815	0	5,547	3,275	2,614	4,939
24	L	1,095	1,953	251	3,596	6,384	6,358	1,609	4,995
	K	1,955	4,180	978	7,094	5,652	5,921	2,736	5,428
	Y	1,879	2,873	517	5,825	6,552	6,216	2,121	5,330

		Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
25	L	1,215	1,638	401	2,058	5,829	2,432	2,051	3,550
	K	1,860	2,233	675	3,100	6,510	2,913	3,173	4,514
	Y	1,505	1,547	454	3,007	6,462	2,830	2,875	4,157
26	L	1,328	2,899	1,305	9,660	3,994	3,676	3,870	6,334
	K	1,453	5,006	3,778	9,660	5,678	3,262	4,385	7,473
	Y	1,423	4,574	2,506	9,660	5,640	4,300	3,815	7,358
28	L	1,323	5,002	375	6,539	3,994	5,363	1,437	3,658
	K	2,745	5,252	1,536	6,757	5,696	6,189	2,046	4,791
	Y	2,290	4,246	1,097	7,726	5,118	5,206	2,048	4,552
29	L	5,060	6,281	900	5,773	1,255	5,993	5,497	1,908
	K	6,016	6,172	1,318	5,840	1,255	5,674	5,824	1,913
	Y	5,866	5,877	1,004	6,360	1,255	5,526	5,826	1,906
30	L	3,370		6,429					0
	K	3,370		6,393					0
	Y	3,370		6,254					0
31	L	6,791	8,481	1,382	0			5,902	2,832
	K	8,454	7,388	2,717	0			5,068	2,947
	Y	6,338	8,460	2,574	0			6,281	2,838
34	L	3,283	3,416	602	0	8,033	8,921	4,908	7,711
	K	4,030	3,416	1,326	0	7,799	8,682	5,705	7,710
	Y	4,573	3,416	2,474	0	7,960	8,880	4,982	7,684
36	L	1,823	2,179	402	7,065	4,899	4,256	2,124	5,207
	K	2,737	3,300	1,412	7,473	5,342	5,322	2,374	5,450
	Y	2,194	2,496	1,489	7,388	5,472	4,851	2,210	5,725

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín; Barr.=Barranquilla; Bog.=Bogotá; Cart.=Cartagena; Man.=Manizales; and Buc.=Bucaramanga. 'Other' aggregates over the remaining municipalities.

Our size variables for firm i in sector s and city c are given by

$$\mathrm{sz}^x_{ics} = \frac{\sum_{j\neq i, j\in c,s} x_j}{\sum_{j\neq i, j\in c,s} 1}$$

where, x_i is either firm i's employment, firm i's capital stock, or firm i's production. These variables simply quantify the average size of firm i's rivals in its sector. Table 15 presents summary statistics.

We also define scale, specialization, and variety measures for city c's neighboring cities. Formally,

$$\begin{aligned}
\mathrm{sc}^x_{-c} &= \sum_{i\in\mathcal{N}_c\backslash c} x_i \\
\mathrm{sp}^x_{-cs} &= \frac{\sum_{i\in\mathcal{N}_c\backslash c,s} x_i}{\sum_{i\in\mathcal{N}_c\backslash c} x_i} \times 100 \\
\mathrm{va}^x_{-cs} &= \sum_{s'\neq s} \left(\frac{\sum_{i\in\mathcal{N}_c\backslash c,s'} x_i}{\sum_{i\in\mathcal{N}_c\backslash c, i\notin s} x_i} \times 100 \right)^2
\end{aligned}$$

In our application we use two definitions for $\mathcal{N}_c$. The first one is broad and includes all cities in the department in which c is located.[18] The second is narrow and considers cities/municipalities in a metropolitan area.

To explore the effect of the agglomeration economies on firm-level productivity we estimate the following empirical model:

$$\omega_{isct} = \gamma_0 + \gamma_{\mathrm{sc}}\mathrm{sc}^x_{ct} + \gamma_{\mathrm{sp}}\mathrm{sp}^x_{cst} + \gamma_{\mathrm{va}}\mathrm{va}^x_{cst} + \alpha_t + \alpha_c + \alpha_s + \epsilon_{isct} \qquad (6)$$

where ω_{isct} is firm i's (log) TFP at time t that we estimated in Section 4 and α_t, α_c, and α_s are year, city, and industry fixed effects. While the year fixed effects should control for aggregate macro shocks, it is of special interest to us the inclusion of the city fixed effects to control for city-level unobservables that might confound the effect of the agglomeration regressors. In this sense, the effects of the agglomeration regressors are identified out of within city variation. Table 16 presents the regression estimates. Each column corresponds to the different definitions of the agglomeration measures. In all cases we cluster standard errors at the city level. We observe two pretty robust

[18]A department is the equivalent of a state in the US. Colombia has 32 departments plus the capital district of Bogotá.

Table 15: Size by City and Industry

		Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	L	326	182	203	127	449	125	343	230
	K	48.7	14.0	23.6	6.4	28.0	11.9	56.8	38.1
	Y	77.7	36.9	49.6	19.5	40.0	22.0	61.8	62.3
17	L	94	186	183		9		76	179
	K	4.4	13.0	15.2		0.2		0.8	7.4
	Y	8.9	16.8	22.5		0.3		3.2	9.7
18	L	241	120	125		40	74	178	329
	K	5.0	1.4	3.2		0.5	1.0	1.8	8.2
	Y	14.8	2.6	6.9		1.5	2.2	9.8	20.6
19	L	301	198	104			28	158	163
	K	5.8	8.0	1.5			0.5	1.6	2.1
	Y	26.2	14.6	6.4			1.3	8.6	7.8
20	L	40	57	75	15			43	41
	K	1.8	3.0	5.2	2.6			0.7	0.8
	Y	2.6	7.1	6.0	4.3			2.9	2.0
21	L	284		79	0			163	338
	K	45.2		5.6				65.9	80.3
	Y	100.1		9.2				44.1	95.7
22	L	81	41	111		47	35	185	70
	K	5.9	1.8	9.0		1.8	1.8	14.8	5.2
	Y	8.8	4.4	13.3		3.4	2.1	24.3	6.0
24	L	204	168	125	75	134	17	441	195
	K	29.4	26.8	15.1	58.8	14.8	2.2	49.3	28.7
	Y	54.0	42.7	39.5	114.2	27.5	1.9	99.3	61.7

		Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
25	L	113	142	98	116	106	46	76	148
	K	7.0	7.7	7.8	32.3	5.7	3.2	9.2	14.5
	Y	13.3	12.9	12.1	32.8	10.0	6.9	9.7	22.3
26	L	104	137	176	214	139	59	59	205
	K	5.0	18.6	52.6	45.1	28.7	3.2	2.1	62.9
	Y	8.1	16.5	43.8	17.3	30.0	6.8	6.9	73.0
28	L	85	105	78	105	91	64	57	177
	K	4.9	4.9	5.0	15.8	4.7	3.6	2.5	22.9
	Y	10.6	8.5	10.5	28.5	10.4	5.8	5.6	40.8
29	L	115	63	116	61	13	117	33	128
	K	3.2	1.8	4.0	3.2	0.3	2.7	1.8	4.4
	Y	10.6	7.2	10.7	6.9	1.2	5.4	2.6	10.3
30	L	6		39					
	K	0.2		1.4					
	Y	0.7		2.7					
31	L	137	101	125				130	128
	K	6.5	6.8	8.7				0.9	7.9
	Y	12.2	12.7	30.2				4.7	15.9
34	L	101	15	139		47	497	115	1,484
	K	3.8	0.3	7.8		1.4	16.0	4.0	50.0
	Y	11.1	0.6	35.8		2.2	50.9	13.4	219.1
36	L	120	181	80	79	44	27	123	115
	K	5.0	11.8	3.7	7.4	1.4	0.6	4.6	13.4
	Y	10.4	21.5	11.1	8.6	4.0	2.6	11.1	25.0

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín; Barr.=Barranquilla; Bog.=Bogotá; Cart.=Cartagena; Man.=Manizales; and Buc.=Bucaramanga. 'Other' aggregates over the remaining municipalities. Monetary figures in billions of constant Colombian pesos of 2005.

Table 16: Firm-Level TFP and Agglomeration Economies

		Labor			
	Num. Firms	Total	White	Blue	Revenues
	(1)	(2)	(3)	(4)	(5)
Scale	−0.000	0.000	0.000	−0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Specialization	0.273***	0.148***	0.098**	0.149***	0.390***
	(0.077)	(0.043)	(0.043)	(0.041)	(0.039)
Variety	0.104	0.138**	0.117*	0.089	0.179***
	(0.086)	(0.067)	(0.061)	(0.068)	(0.062)
City FE	yes	yes	yes	yes	yes
Industry FE	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes
Obs.	23581	23581	23579	23487	23581

results. First, scale economies do not seem to affect firm-level productivity. One has to be careful not to jump to the conclusion that there are no scale economies present. It could well be the case that there is not enough (within city) variation in the data to detect any effect (recall that we are including city fixed effects). Second, we do find supporting evidence that industrial specialization has a positive effect on productivity. Our point estimates indicate that a one standard deviation increase in sector specialization (effectively more than doubling the average sector's share in the city) produces an increase in productivity ranging from 2 to 4% (and up to 8% when using revenues to construct our variables). In terms of cross-industry spillovers, we find evidence that they also matter (but the evidence is less robust): greater sectoral variety in a city leads to lower productivity in all of our specifications. Moreover, a one std. dev. increase in variety implies a decrease in productivity on the order of 2 to 3%. It seems then that firms in Colombia benefit from forming bigger clusters and locating in cities with less industrial variety. This is in line with our preliminary findings of the geographical distribution of firms in Section 3.

Given that in many cases cities are very close to other cities —such as in metropolitan areas— the scale, specialization, and variety effects might extend the boundaries of a city. To investigate this, we augment our empirical model to include the effects of agglomeration economies in neighboring cities. Table 17 presents the regression

Table 17: Firm-Level TFP: Neighbor Effects (Metropolitan Area)

		Labor			
	Num. Firms	Total	White	Blue	Revenues
	(1)	(2)	(3)	(4)	(5)
(a) City					
Scale	0.000	−0.000	0.000	−0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Specialization	0.316***	0.167***	0.110*	0.157***	0.355***
	(0.114)	(0.059)	(0.057)	(0.056)	(0.050)
Variety	−0.041	0.174*	0.084	0.156*	0.123
	(0.133)	(0.090)	(0.080)	(0.090)	(0.078)
(b) Neighbor					
Scale	0.000*	−0.000	−0.000	0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Specialization	0.141	0.311***	0.283***	0.295***	0.245***
	(0.093)	(0.062)	(0.054)	(0.063)	(0.054)
Variety	0.137	0.121	0.089	0.054	0.088
	(0.149)	(0.125)	(0.086)	(0.117)	(0.096)
City FE	yes	yes	yes	yes	yes
Industry FE	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes
Obs.	21450	21450	21450	21435	21450

estimates, where again each column corresponds to the different definitions of the agglomeration measures. We use our preferred definition of neighbor cities given by metropolitan areas. We still find robust supporting evidence in favor of location economies: a one std. dev. increase in own-city specialization results in an increase in productivity ranging from 2 to 7% on average, while a one std. dev. increase in the industry specialization in the neighboring cities results in an additional 3% increase in productivity. We also find less robust evidence of urbanization economies. Firms seem to benefit from less diverse industrial composition. Finally, city and neighbor scales seem not to have a statistically significant effect. As a robustness check we include the results using our alternative, broader notion of neighbor (all other cities in a Department) in the Appendix; results are qualitatively similar (see Table 41).

We next try including our variables of competition and rival firms' average size. Table 18 shows the regression estimates. We observe, once again, that (i) location economies have a statistically significant effect with economic magnitudes similar to the ones found before; (ii) city scale has no statistically significant effect on productivity; and (iii) that there is mild evidence that less industrial variety increases productivity. More importantly, while the average size of the firm's rivals seems not to have an effect on productivity, firms located in cities with more competition attain lower productivity.

5.2 Extending the Analysis

We now extend our analysis to include detailed information at the city level from the municipal panel dataset we introduced in Section 2.2. The objective is to try to uncover systematic relationships between city characteristics and firm-level productivity. We look at several dimensions that we believe might affect firm productivity including cities' fiscal performance, education, health, and violence. The municipal panel contains very rich data on all of these dimensions including over a thousand variables. The natural hurdle we thus face is how to select which variables to include into our analysis. We tackle this problem in the following way. In a first step we identify viable variables by using basic metrics like coverage both in the time and geographical dimensions and discard variables that have a significant amount of missing values. We then transform some variables to make them suitable for the analysis. For example, when dealing with variables related to violence, we take variables like number of homicides, thefts, kidnappings, etc., and transform them into 'per 100,000 people' rates so that they are comparable across cities of different scale. In the Ap-

Table 18: Firm-Level TFP: Competition

	Labor			
	Total	White	Blue	Revenues
	(1)	(2)	(3)	(4)
Scale	0.000	0.000	−0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Specialization	0.177**	0.178***	0.111*	0.411***
	(0.071)	(0.067)	(0.064)	(0.059)
Variety	0.155*	0.122	0.035	0.161**
	(0.084)	(0.075)	(0.088)	(0.073)
Size	0.000	−0.000	0.000	−0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
Competition	0.068**	0.056*	0.094***	0.062**
	(0.033)	(0.032)	(0.028)	(0.031)
City FE	yes	yes	yes	yes
Industry FE	yes	yes	yes	yes
Year FE	yes	yes	yes	yes
Obs.	22042	22042	21986	22042

pendix we provide a full list of the candidate variables we selected and how they were defined. In the second step, we now face a model selection problem. While we did use economic reasoning in the first step to transform the variables, in the second step there is little a priori economic reasoning that we can use —in most cases— to select one variable over another. Instead, we opt to use the LASSO for model selection. We provide a brief description of the method in the following section.

5.2.1 The LASSO

The LASSO (least absolute shrinkage and selection operator) method (see, e.g., Tibshirani (1996)) is appropriate for sparse regression problems, i.e., when there is a large number of potential regressors but only a small (unknown) subset of them is important in capturing the regression function accurately. In other words, it is suitable in situations in which we believe that many of the slope coefficients should be 0. The objective of LASSO is to identify the relevant regressors by performing a modified OLS procedure which penalizes a large number of nonzero coefficients through regularization by a penalty based on the $\mathcal{L}_1$ norm. Formally, say you want to run a linear regression of an outcome y_i on a vector of covariates x_i (where the dimension of x_i, denoted p, is generally large and can even be larger than the number of observations).[19] Then, LASSO solves the following problem

$$\min_{\beta_0,\beta}\left\{\frac{1}{N}\sum_{i=1}^{N}(y_i-\beta_0-x_i'\beta)^2\right\} \text{ subject to } \sum_{j=1}^{p}|\beta_j|\leq t$$

which can be rewritten as a penalized OLS problem

$$\min_{\beta_0,\beta}\frac{1}{N}\sum_{i=1}^{N}(y_i-\beta_0-x_i'\beta)^2+\lambda\sum_{j=1}^{p}|\beta_j|.$$

LASSO requires us to choose the *tuning* or penalty parameter, $\lambda\geq 0$, which controls the amount of regularization and has a one-to-one correspondence with the threshold parameter t. Clearly, if $\lambda=0$, the LASSO and OLS would provide the same results. By increasing λ we effectively control the degree of model selection and the "size" of the resulting model, since LASSO will force some of the β coefficients to be equal

[19] In practice, y is centered —so that $\beta_0=0$— and the covariates are typically standardized so that the solution does not depend on the scale of the different covariates.

to 0.[20] We choose λ by cross-validation. Since we want to force the agglomeration economy variables and the fixed effects into the model (i.e., we don't want the LASSO to exclude them), we can either set variable specific penalty parameters or simply regress the left-hand-side variable on the agglomeration variables and fixed effects and perform LASSO with the remaining variables on the residuals (see Efron, Hastie, Johnstone, and Tibshirani (2004) for details). In our application we do the latter. The resulting model estimates are obtained by the Post-LASSO method (see Belloni and Chernozhukov (2011)), which simply performs OLS on the model chosen by the LASSO. The estimation results are shown in Table 19.

5.2.2 Results

The results in terms of our 'traditional' agglomeration variables remain qualitatively unchanged after augmenting the model with the city level characteristics. Out of the fiscal variables the LASSO selected three variables: revenue from taxes on businesses and manufacture (as a fraction of total tax revenue); a fiscal performance indicator that measures the ratio of debt interest payments to total revenues (this variable quantifies the debt burden and to what extent it might compromise other expenditures); and expenditure on transportation infrastructure.[21] The first two variables can be interpreted as capturing the business environment in the city: the fraction of revenues from taxes on manufacturing firms should pick up how 'friendly' the city is to businesses and the debt ratio should pick up how well the city is being managed. It is very reassuring that the LASSO selected the fraction of expenditure on transportation infrastructure since, a priori, one would expect that transportation infrastructure to be a main determinant of productivity (both from the point of view of moving raw inputs and the final product and from the point of view of facilitating the movement of the labor force). The estimated coefficients have the expected signs: a worse business environment has a negative effect on productivity and more expenditure on transportation infrastructure has a positive effect on productivity. To help interpret the magnitudes of the effects of the different variables, in what follows we quantify the effect of a one standard deviation change in a given variable. In particular, a one std. dev. increase in the fraction of taxes on manufacture (equivalent

[20] In the extreme, when $\lambda \to \infty$ all of the slope coefficients are set to 0, i.e., we obtain an intercept-only model.

[21] Where expenditure is interpreted in a broad sense, i.e., it also includes investment expenditure. Transportation infrastructure includes expenditure on construction, repairs, and maintenance of roads, ports, and airports.

Table 19: Firm-Level TFP

	Labor			
	Total	White	Blue	Revenues
	(1)	(2)	(3)	(4)
Scale	−0.0000	−0.0000	−0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Specialization	0.2045**	0.1794**	0.1397*	0.3872***
	(0.0889)	(0.0859)	(0.0773)	(0.0731)
Variety	0.2678**	0.2084**	0.1611	0.2521***
	(0.1113)	(0.0981)	(0.1183)	(0.0972)
Size	0.0000	−0.0001	0.0000	−0.0000
	(0.0001)	(0.0001)	(0.0001)	(0.0000)
Competition	0.0956**	0.0493	0.1242***	0.0767**
	(0.0398)	(0.0378)	(0.0330)	(0.0377)
Fiscal:				
Tax on Manuf./Tax Rev.	−0.2153**	−0.2304**	−0.2380**	−0.2007***
	(0.0966)	(0.0955)	(0.1185)	(0.0732)
Fiscal Perf. Debt	−0.0002*	−0.0003*	−0.0000*	−0.0003*
	(0.0001)	(0.0002)	(0.0000)	(0.0002)
Exp. Transp. Infrast. (%)	0.0014***	0.0012**	0.0013**	0.0007**
	(0.0005)	(0.0005)	(0.0006)	(0.0003)
Violence:				
Thefts to Individuals (1)	−0.0003*	−0.0003*	−0.0004**	−0.0004**
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Homicides (1)	0.0002*	0.0002*	0.0002*	0.0001*
	(0.0001)	(0.0001)	(0.0001)	(0.0000)
Kidnappings (1)	−0.0002**	−0.0003**	−0.0014**	−0.0001***
	(0.0001)	(0.0001)	(0.0005)	(0.0000)
AUC Presence	−0.0316*	−0.0317*	−0.0304*	−0.0322*
	(0.0184)	(0.0173)	(0.0183)	(0.0189)
Terrorist Attacks	−0.0004*	−0.0004*	−0.0004	−0.0005*
	(0.0002)	(0.0002)	(0.0003)	(0.0003)
Education:				
Students per Teacher	−0.0114**	−0.0112*	−0.0095***	−0.0096***
	(0.0050)	(0.0065)	(0.0023)	(0.0024)
Univ. Professors (1)	−12.9552*	−11.4512*	−15.8864**	−11.2245*
	(7.8464)	(6.3582)	(7.2853)	(5.9014)
Saber 11	0.0109***	0.0097**	0.0075**	0.0121***
	(0.0036)	(0.0048)	(0.0033)	(0.0045)
Low-Rank Schools	−0.0591*	−0.0460*	−0.0095*	−0.0114*
	(0.0343)	(0.0269)	(0.0051)	(0.0068)
Health:				
IMR	−0.0380*	−0.0321*	−0.0323*	−0.0315*
	(0.0223)	(0.0186)	(0.0250)	(0.0190)
City FE	yes	yes	yes	yes
Industry FE	yes	yes	yes	yes
Year FE	yes	yes	yes	yes
Obs.	16427	16427	16396	16427

to a 50% increase over the mean) translates into a decrease in productivity ranging from 3.3 to 3.9% depending on the specification, on average; a one std. dev. increase in the debt indicator has a small economic effect on productivity (about .2 to .3%); and an increase in one std. dev. in the fraction of total expenditure devoted to transportation infrastructure brings about a .5 to .9% increase in firm-level productivity, on average.

In terms of the education variables, the LASSO selected the following set: the ratio of students per teacher, the test-score Saber 11, the fraction of low-rank schools out of all schools, and the number of university professors per 100,000 people. Intuitively, the first three education variables are capturing the 'quality' level of the schools in the city which in turn might reflect on the quality of the labor force in the city (in a direct way if the workers themselves have gone to the schools in the city, or in an indirect way if more skilled workers prefer to locate in cities with better schools). Signs are as expected: better quality schools in the city, everything else equal, have a positive effect on firm productivity. In terms of the magnitudes of the effects, we find that a decrease in the students per teacher ratio of one std. dev. (equivalent to a 14% decrease with respect to the mean level) implies a 3 to 3.6% increase in firm productivity. A one std. dev. increase in the Saber 11 test-score (equivalent to a 5% increase on the mean level) translates into 2 to 3.2% higher productivity. Finally, a decrease in the fraction of low-ranking schools in the city of one std. dev. implies an increase in productivity ranging from 0.2 to 0.7%. The LASSO also selected the number of university professors per 100,000 people. Its effect on productivity is negative (a one std. dev. increase in university professors in the city translates into a 2.6 to 3.7% decrease in productivity). One way of interpreting this result is the following. Job opportunities for highly-skilled workers in jobs other than manufacturing compete with the labor demand from manufacturing firms. This may make it harder for firms to hire or retain high-skilled workers and since the actual level of skill used by a firm is not reported in our data it will be picked up by the TFP term.

From the health and public services variables, only one variable made it to the model: the infant mortality rate (IMR). Most of these variables do not show a lot of time series variation, hence our model with city fixed-effects is unable to detect any effect from them. The effect of IMR is negative, as expected. Worse health conditions in a city are associated with lower productivity. In particular, a one std. dev. increase in the IMR turns into a decrease in firm-level TFP ranging from 18 to 22% . This is quite a big effect and we should interpret it with caution as it might be picking up poor health conditions other than IMR that are correlated with this measure.

Perhaps the most unfortunate dimension of city characteristics that is more prevalent in Colombia compared to other countries, is that related to crime and violent terrorist attacks. In recent history, Colombia has suffered the effect of various terrorists groups and their frequent attacks on civil population and businesses, including kidnappings for ransom. These events have the potential to generate big economic loses and here we quantify their effects on firm-level productivity.

Out of the several variables related to crime and violence, the LASSO has selected the following ones: number of thefts to individuals per 100,000 people, number of homicides per 100,000 people, number of kidnappings per 100,000 people, a dummy that takes value 1 if the paramilitary, terrorist, and drug-trafficking group AUC (Autodefensas Unidas de Colombia) is present in the city, and the number of terrorist attacks. Intuitively, violent attacks to the population and businesses create a climate which is not conducive to business and hence it is expected they will have an impact on productivity: firms have to devote resources out of production to protect their property hence reducing their output holding inputs constant. In fact, it is extremely rare for a business in Colombia, regardless of its size, not to have private security. At the same time, and even if the firm does not allocate resources to protect itself, violent episodes might disrupt everyday life for the workers; interfere in the transport of both inputs and output, disrupting supply chains; and generally disrupting social interaction. The effects of all of these channels will be picked up as a shock to TFP.

All the signs are as expected (except for the number of homicides): everything else equal, more violence in a city leads to lower productivity. A one std. dev. increase in the theft rate brings about a 3.9 to 5.2% decrease in firm productivity. Similarly, a one std. dev. increase in the number of terrorist attacks produces a decrease in productivity between 1.5 and 2%. On the other hand, a one std. dev. increase in the kidnapping rate has a negligible economic effect on productivity (about 0.4%). Also, negligible, though with the wrong sign, is the effect of an increase in one std. dev. in the number of homicides which produces an increase in productivity between 0.3 to 0.4%. Perhaps more striking is the effect of the presence of AUC in a given municipality which lowers productivity by 3 to 3.2%, on average. Notice that this effect is already conditional on the number of terrorist attacks and crime in general.

Discussion Our contribution follows in the tradition of Glaeser, Kallal, Scheinkman, and Shleifer (1992) and Henderson, Kuncoro, and Turner (1995). These papers assess the impact of variables at the city-industry level on employment growth. Our paper differs from them in that we try to understand how local industrial structure and city characteristics affect firm productivity. In terms of the former effects, in particular, we investigate how overall city scale, localization economies (specialization), urbanization economies (diversity) affect productivity; in terms of the latter ones, we quantify how other city amenities (or lack of them) affect productivity by looking at the city's fiscal performance, education, health, and security dimensions.

The richness of our dataset allows us to go beyond what traditional papers in the literature have done and, taken together, we found non-trivial effects of key city characteristics on firm-level TFP above and beyond the effects of local industrial structure. Many of the variables impacting productivity have clear and simple policy recommendations. Clearly, a cost-benefit analysis of such policy interventions is beyond the scope of this paper but quantifying the economic effects of both positive and negative forces that affect the productivity of firms is a necessary step.

We deliberately focus on firm productivity —which we estimate using both state-of-the-art techniques from the industrial organization literature and detailed firm-level data— because we believe that firms are the fundamental unit to develop a city. Understanding, how firms' operations are affected by city factors is thus key and often neglected in the urban economics literature. Of course, our approach is not free from criticisms. We now discuss them in turn.

The first set of disadvantages are data related. Due to data constraints, we need to restrict our analysis to manufacturing firms, leaving aside the services and agricultural sectors. We believe that there is no strong reason to think that our conclusions would not extrapolate, at the very least qualitatively, to the services or agricultural sectors. Also in terms of data, ideally, we would have preferred to use in our analysis census data on manufacturing firms such as the Encuesta Anual Manufacturera (EAM) conducted by the Departamento Administrativo Nacional de Estadística (DANE) in Colombia. Unfortunately, the EAM does not include the firms' locations but only indicators for the main seven metropolitan areas which severely hinders the implementation of our analysis.[22] While the main seven metropolitan areas concentrate most of the manufacturing activity, a metropolitan area is too coarse of a unit of analysis. The latter has two drawbacks: first, it will likely introduce significant noise into our analysis; second, were we to deal with metropolitan areas, we would need to find a way to aggregate the municipality-level variables, which is not a trivial

[22]Locations other than those seven metropolitan areas are left unspecified in the data.

task. Additionally, by focusing on the manufacturing sector our analysis only covers 156 out of the 1,122 municipalities in Colombia. Moreover, the municipalities in our sample tend to be bigger.[23] This is to be expected from agglomeration economies. We should then keep in mind that the analysis in this paper excludes cities with no manufacturing activity (or with small manufacturers not covered by SIREM). This is different from other studies on the effect of agglomeration economies that use, for example, wage data from (nationally representative) household surveys. Again, there is no reason to believe that our findings would not extend to municipalities in which services or agricultural activities dominate the economic scene. On the other hand, when comparing the EAM to our dataset, SIREM accounts for more than 90% of the employment reported in the EAM in the manufacturing sector (see **?** for more details). This means that when using the data from SIREM we are not losing much of the representativeness of the EAM while retaining the advantages of a finer unit of analysis.

The second set of concerns are related to our focus on productivity. Our analysis does not capture other important channels, which are complementary to the one we highlight. For example, the effects on productivity will have further effects, in turn, on output prices, labor demand, or the development of input-output networks across firms. In particular, high productivity firms will be able to charge lower prices for their output to consumers, demand more labor, and be able to pay higher wages. Quantifying these important effects is beyond the scope of this paper. We should treat then our findings as lower bounds on the overall effects of the various dimensions of city characteristics. A second, more technical point, has to do with the type of production function we estimate to recover firm-level productivity. Ideally, we would like to estimate a production function in physical units so that the estimated productivity shocks are not contaminated with unobserved differences in quality and demand shocks and firms' markups (see Klette and Griliches (1996), Foster, Haltiwanger, and Syverson (2008),and De Loecker (2011) for a discussion of this issue). There have been recent methodological advances to correct for these problems (see Balat, Brambilla, and Sasaki (2016) or De Loecker, Goldberg, Khandelwal, and Pavcnik (2015)) but they require data at the firm-product level which we do not have. What we do, instead, to mitigate the output price contamination of our measure of productivity is to deflate net revenues —our left-hand-side variable in equation (3)—

[23]Municipalities in our data range from 225 to 7 million inhabitants, with a median of 12,500. Municipalities with manufacturing activity range from 10,000 to 7 million inhabitants and a median of 63,000.

using industry-city specific price indices that we constructed.

Finally, a few comments about our empirical implementation. A major concern regarding the estimation of the effects of local industrial structure and city characteristics on productivity as in equation (6) is the endogeneity of the regressors. Clearly, if there is an omitted city characteristic that affects productivity and is correlated with our regressors, our estimates of the causal effects are biased. As we have mentioned before, this is the reason why we include city fixed effects in equation (6). Under the assumption that the unobservable characteristic is time invariant, the city fixed effects would control for the city-level unobservable. Therefore, the agglomeration and city characteristics effects on productivity are identified out of within city variation. In principle, a better way to solve the potential endogeneity problem would be to include city-year fixed effects. Of course, these fixed effects would absorb all of our regressors of interest that do not vary within a city-year (such as city scale and our municipal regressors) preventing us to get estimates of the effects of these variables. To recover the effect of the latter variables, we could run a second-stage regression of the city-year fixed effects on the variables of interest (note, though, that the number of observations in this regression is given by the number of city-year pairs). We have experimented with this implementation as a robustness check and the results do not change qualitatively nor quantitatively (although some variables lose their significance). There is one caveat, though. The city-year fixed effects estimated in the first-stage are not identified and inconsistent since the asymptotics is in the time dimension (this is known as the incidental parameter problem; see, e.g., **?** for more details). One final comment about our implementation. The causal effects we estimate from a model like (6) are average effects. An interesting extension to our analysis would like at distributional effects. That is, it would be interesting to investigate whether firms at different point in the distribution of productivity are impacted differentially by city characteristics. One should be able to look at this using quantile regression techniques but, to the best of our knowledge, we are not aware of model selection methods similar to the LASSO in the context of quantile regressions.

6 Conclusion

In this paper we study the determinants of firm-level productivity in Colombia. In particular, we are interested in the effects of agglomeration forces that explain why manufacturing economic activity tends to be concentrated in rather few cities and the effect of forces that can hurt productivity. We carry out this study by exploiting two rich data sources. The first one is a firm-level panel which contains input and output data that will allow us to estimate firm-level productivity. The second one, is a panel of municipalities containing very detailed information on city characteristics over time. In terms of our empirical implementation, in a first stage we estimate productivity for each firm in each time period using a control function approach to handle endogeneity concerns in the estimation of a production function. Our implementation follows that of Ackerberg, Caves, and Frazer (2016), which extends the seminal work by Olley and Pakes (1996).

Our key findings are the following. First, at the aggregate level, Bogotá shows the highest level of productivity, followed —in order— by Medellín, Cali, Barranquilla, Manizales, Cartagena, and Bucaramanga. Second, the most productive sector is apparel. Not only that, but the apparel industry is also the one which shows the lowest level of geographical dispersion in terms of productivity. On the contrary, manufacturing of motor vehicles shows the highest geographical variation. Third, while Bogotá shows the highest concentration of firms in every industry and the broadest scope of industries among the main cities, this does not translate into higher productivity for each and every industry. For example, Cali is the most productive city in apparel and metal products; Medellín is the most productive in wood products and computing machinery; Barranquilla in textiles and printing; Manizales in machinery; Bucaramanga in rubber and plastic products and non-metallic mineral products; and Cartagena in food and paper products. Bogotá remains the most productive city in leather products, chemicals, electrical machinery, motor vehicles, and furniture.

We then move to explore the determinants of productivity. Our focus is in the effect of local industrial structure measures (scale, specialization, and variety) and in the effect of other city characteristics which are novel in the literature. Our main findings are the following. First, scale economies —i.e., the size of the city under alternative measures— do not seem to affect firm-level productivity. Second, we do find supporting evidence for *location economies*. That is, industrial specialization has a positive effect on productivity. Our point estimates indicate that a one standard deviation increase in sector specialization produces an increase in productivity ranging from 2 to 4%. In terms of cross-industry spillovers, we find evidence that they also matter (but the evidence is less robust): a one std. dev. increase in variety implies a decrease in productivity on the order of 2 to 3%. It seems then that firms in Colombia benefit from forming clusters and locating in cities with less industrial variety.

Extending our analysis to other city characteristics we do find non-trivial economic effects of other city 'amenities' (or lack of them). The fiscal performance of the city, its education level and quality, and more importantly crime and violent attacks have sizable effects on productivity. For example, a one std. dev. increase in the theft rate brings about a 4 to 5% decrease in firm productivity. The presence of the paramilitary, terrorist, and drug-trafficking group AUC (Autodefensas Unidas de Colombia) in a city translates into 3.2% lower productivity, on average, everything else equal. Finally, a one std. dev. increase in the number of terrorist attacks implies a 1.5% decrease in productivity.

The richness of our dataset allowed us to go beyond what traditional papers in the literature have done and, taken together, we found non-trivial effects of key city characteristics on firm-level TFP above and beyond the effects of local industrial structure. Many of the variables impacting productivity have clear and simple policy recommendations. Clearly, a cost-benefit analysis of such policy interventions is beyond the scope of this paper but quantifying the economic effects of both positive and negative forces that affect the productivity of firms is a much needed necessary step to inform policy makers.

References

Ackerberg, D., L. Benkard, S. Berry, and A. Pakes (2005): "Econometric Tools for Analyzing Market Outcomes," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, vol. 6. Elsevier.

Ackerberg, D., K. Caves, and G. Frazer (2016): "Identification Properties of Recent Production Function Estimators," forthcoming in Econometrica.

Balat, J., I. Brambilla, and Y. Sasaki (2016): "Heterogeneous Firms: Skilled-Labor Productivity and the Destination of Exports," working paper, Johns Hopkins University.

Belloni, A., and V. Chernozhukov (2011): "High-Dimensional Sparse Econometric Models, an Introduction," *Springer Lecture Notes*.

Cingano, F., and F. Schivardi (2004): "Identifying the sources of local productivity growth," *Journal of the European Economic Association*, 2(4), 720–742.

Combes, P.-P., and L. Gobillon (2015): "The empirics of agglomeration economies," *Handbook of Regional and Urban Economics*, 5(1), 247–348.

De Loecker, J. (2011): "Product Differentiation, Multi-Product Firms and Estimating the Impact of Trade Liberalization on Productivity," *Econometrica*, 5, 1407–1451.

De Loecker, J., P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik (2015): "Prices, Markups and Trade Reform," forthcoming in Econometrica.

Di Giacinto, V., M. Gomellini, G. Micucci, and M. Pagnini (2014): "Mapping Local Productivity Advantages in Italy: Industrial Districts, Cities, or Both?," *Journal of Economic Geography*, 14, 365–394.

Duranton, G. (2015a): "Agglomeration Effects in Colombia," working paper.

——— (2015b): "Determinants of City Growth in Colombia," working paper.

Duranton, G., and D. Puga (2004): "Micro–foundations of urban agglomeration economies," *Handbook of Regional and Urban Economics*, 4(1), 2063–2117.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani (2004): "Least Angle Regression," *The Annals of Statistics*, 32(2), 407–499.

Foster, L., J. Haltiwanger, and C. Syverson (2008): "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?," *American Economic Review*, 98(1), 394–425.

Glaeser, E., H. Kallal, J. Scheinkman, and A. Shleifer (1992): "Growth in cities," *Journal of Political Economy*, 100(6), 1126–1152.

Henderson, J. V. (1986): "Efficiency of resource usage and city size," *Journal of urban economics*, 19(1), 47–70.

——— (2003): "Marshall's scale economies," *Journal of urban economics*, 53(1), 1–28.

Henderson, V., A. Kuncoro, and M. Turner (1995): "Industrial development in cities," *Journal of Political Economy*, 103(5), 1067–1090.

Hoover, E. M. (1937): *Location Theory and the Shoe Leather Industries*, vol. 55. Harvard University Press.

Isard, W. (1956): *Location and Space-Economy*. The Massachusetts Institute of Technology Press.

Klette, T. J., and Z. Griliches (1996): "The Inconsistency of Common Scale Estimators when Output Prices are Unobserved and Endogenous," *Journal of Applied Econometrics*, 11(4), 343–361.

Lall, S., Z. Shalizi, and U. Deichmann (2004): "Agglomeration Economies and Productivity in Indian Industry," *Journal of Development Economics*, 73, 643–673.

Levinsohn, J., and A. Petrin (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70(2), 317–342.

Lopéz, R., and J. Südekum (2009): "Vertical Industry Relations, Spillovers and Productivity: Evidence from Chilean Plants," *Journal of Regional Science*, 49(4), 721–747.

Marschak, J., and W. Andrews (1944): "Random Simultaneous Equations and the Theory of Production," *Econometrica*, 12, 143–205.

Marshall, A. (1890): *Principles of Economics*. Macmillan, London.

Olley, G. S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297.

Saito, H., and M. Gopinath (2009): "Plants' Self-selection, Agglomeration Economies and Regional Productivity in Chile," *Journal of Economic Geography*, 9(1), 539–558.

Tibshirani, R. (1996): "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society Series B*, 58(1), 267–288.

Weber, A. (1929): *Theory of the Location of Industries*. University of Chicago Press.

Appendix I: Municipal Panel – CEDE

For our city-level analysis we use information from the Municipal Panel, overseen and continuously updated by the Center of Economic Development Studies (CEDE) at Universidad de los Andes. The dataset is a comprehensive panel of municipal variables that consolidates municipal-level data from a myriad of sources —including Departamento Administrativo Nacional de Estadística (DANE), Instituto Geográfico Agustín Codazzi (IGAC), Departamento Nacional de Planeación, Ministerio de Agricultura, Red de Información y Comunicación del Sector Agropecuario Colombiano (AGRONET), Ministerio de Salud, Ministerio de Minas y Energía, among others— on 1,122 Colombian municipalities from 1993 to 2014. The dataset contains six main modules: general municipal characteristics, fiscal variables, conflict and violence, agricultural sector and land, education, and health and public services. We briefly discuss them in turn in what follows.

Cities' General Characteristics The general characteristics' section contains basic data on municipalities. There are a total of 81 variables in this module from which we select the following ones for our later analysis: year of incorporation, population (total, rural, and urban), area (km^2), distance to Department's capital (km), distance to main wholesale food market (km), distance to Bogota (km), GDP (total, agriculture, manufacture, services), a measure of income inequality as given by the Gini coefficient, poverty rate, unsatisfied basic needs rates, and a multidimensional measure of poverty. Table 20 shows averages (across years) of select variables for the main seven cities in Colombia (all remaining cities are included in Other).

Fiscal Variables The fiscal module contains 224 variables regarding revenues, expenditure, investment, and measures of fiscal performance. The source of municipal budget execution is Departamento Nacional de Planeación (DNP) which collects data reported by each local entity. The data on the fiscal performance index come from Procuraduría General de la Nación. For our purposes we selected the following variables: total revenue, tax revenue, tax on manufacturing, total expenditure, expenditure on infrastructure, primary deficit, deficit, and borrowing. We also include a fiscal performance index (Indice de Desarrollo fiscal) that ranges from 0 to 100 (lowest to highest) and provides a measure of the city's fiscal health. Each city is then ranked at the national and state level. Among other measures of fiscal health included in the panel, the ratio of operating expenditure to ordinary revenues (tax and non-tax

Table 20: City Characteristics: General Characteristics

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
Year of Incorporation	1,675	1,876	1,538	1,599	1,849	1,623	1,536	1,808
Area (km^2)	387	166	1,605	559	477	154	552	653
Altitude	1,486	30	2,700	2	2,216	959	995	1,204
Dist. to Bogota (km)	264	736	0	679	168	327	287	272
Dist. to wholesale food market	0	0	0	99	44	0	0	84
Population (total)	2,316,569	1,177,922	7,258,612	934,484	386,700	522,324	2,219,744	98,870
Population (urban)	2,281,856	1,173,797	7,242,692	888,184	359,673	515,390	2,183,074	82,194
Population (rural)	34,713	4,125	15,920	46,300	27,027	6,933	36,671	16,676
GDP ($)	24,808,869	10,190,163	99,965,997	7,920,328	2,098,613	10,033,064	17,416,734	783,266
GDP Agric. ($)	18,319	13,979	83,467	6,447	86,566	15,002	39,896	38,421
GDP Manuf. ($)	8,024,775	2,505,267	21,820,427	2,837,734	613,701	3,667,717	5,049,224	255,253
GDP Serv. ($)	13,052,944	6,034,627	68,405,570	3,356,087	1,016,707	3,809,570	10,182,833	381,954
GDP per cap. ($)	9.87	8.55	13.83	7.90	4.79	17.01	7.44	7.88
Poverty rate (2005)	0.46	0.33	0.17	0.40	0.48	0.33	0.33	0.42
Gini (2005)	0.40	0.46	0.44	0.47	0.41	0.41	0.45	0.42
Unsatisfied Basic Needs (%, 2005)	12.17	17.69	9.16	26.16	9.97	11.31	11.01	23.99
Multidim. Poverty Index (%, 2005)	32.40	37.20	24.30	42.58	29.56	28.49	31.97	48.32

Source: Authors' calculations based on data from CEDE-Panel municipios-Universidad de los Andes.
Notes: Med.=Medellín; Barr.=Barranquilla; Bog.=Bogotá; Cart.=Cartagena; Man.=Manizales; and Buc.=Bucaramanga. 'Other' averages over the remaining municipalities. Monetary figures in millions of constant Colombian pesos of 2005.

revenues) quantifies to what extent regular operating costs (public sector wage-bill and general operating costs) are paid from city's own discretion revenues; the ratio of debt interest payments to total revenues quantifies the debt burden and to what extent it might affect other expenditures; the ratio of tax revenues to ordinary revenues quantifies to what extent the city can generate its own resources (the complement would be transfers from the federal and state level); and the ratio of investment to total expenditure, where investment not only includes infrastructure but also 'social investment' (e.g., public sector wage-bill for doctors and teachers and some subsidies). Finally, we also include some measures of expenditure on key areas (as a fraction of total expenditure) like transportation infrastructure, education, institutional development, economic development, public services, and water and sanitation.[24] Table 21 shows averages of select fiscal variables for the main seven cities in Colombia.

Conflict and Violence The conflict and violence module contains 315 variables on historic violence, drug plantations, forced population displacement, public force defensive activity, and armed groups offensive activities. The sources for this module include: Sistema Integrado de Monitoreo de Cultivos Ilícitos (SIMCI), Policía Nacional, Ministerio de Defensa, Observatorio de Derechos Humanos (ODH), Centro de Estudios sobre Seguridad y Drogas (CESED) and several research studies by CEDE.

The main variables we include in our analysis are dummies for presence of terrorist groups (FARC, ELN, AUC), population displaced by terrorism (in- and out-flows), arrests, terrorist surrenders, thefts, homicides, kidnappings, and terrorist attacks, and a dummy variable for violent activity from 1948 to 1953.[25] Table 22 presents summary statistics on a selected group of variables.

[24]For these measures, expenditure is understood in a broad sense, e.g., it also includes investment expenditure. Transportation infrastructure includes expenditure on construction, repairs, and maintenance of roads, ports, and airports. Education includes expenditure on subsidies, especial students, and training. Institutional development includes expenditure on institutional evaluation, administrative reorganization, training, registries updates, among other items. Economic development includes expenditure on industrial and business promotion, labor training programs, technical assistance to firms among other items. Public services includes subsidies for low income people, maintenance and expansion of street lighting, expansion of electric and gas service.Water and sanitation includes expenditures (maintenance and capital expenditures) on the drinking water system, sewage system, and trash collection system.

[25]This period, called "la Violencia" was one of the most cruel periods in the Colombian history, comparable to a civil war. This conflict caused between 200,000 and 300,00 deaths and the forced displacement of about 2 million people (almost a fifth of the population at the time) from their and seizure of their lands by terrorist groups.

Table 21: City Characteristics: Fiscal Variables

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
Revenues (total)	2,658,591	1,021,473	7,861,688	829,959	252,245	378,036	1,430,385	62,889
Tax Revenues	924,701	405,488	4,000,373	282,634	82,935	151,739	577,042	18,451
Tax on Manuf.	347,252	169,115	2,064,710	122,757	22,747	66,868	211,859	6,931
Expenditure (total)	2,843,853	1,107,045	7,844,770	778,196	249,547	405,105	1,407,054	65,875
Infrastructure	918,890	336,938	2,241,493	212,267	59,529	130,599	337,948	19,818
Primary Deficit	901,924	296,908	3,298,635	185,172	59,009	102,144	398,456	12,052
Deficit	-185,261	-85,572	16,918	51,764	2,698	-27,069	23,331	-1,687
Borrowing	46,153	6,766	-94,612	-8,951	-1,158	1,414	-30,559	515
Fiscal Performance Index (IDF)	81	73	78	74	75	77	70	71
Operating Expend./Ordinary Rev. (%)	37.6	43.6	28.0	49.0	38.6	38.7	45.7	53.3
Debt Interest/Revenues (%)	7.1	26.2	20.7	14.7	12.6	7.7	31.6	8.9
Tax Rev./Ordinary Rev. (%)	57.1	61.1	64.3	57.9	58.0	62.7	58.4	52.6
Investment/Expenditure	85.4	83.0	77.8	82.4	84.0	82.8	73.4	80.5
IDF rank (national)	15	144	39	147	88	52	254	246
IDF rank (state)	4	3	1	2	2	5	13	17
Expend. Institutional Development (%)	4.49	13.48	3.64	1.85	3.15	2.16	2.81	3.68
Expend. Economic Development (%)	3.51	0.30	0.89	0.44	0.64	0.98	0.23	0.48
Expend. Public Services (%)	1.57	4.39	1.77	0.18	0.11	4.10	0.01	1.47
Expend. Transportation Infrast. (%)	11.97	5.76	10.58	5.10	5.59	9.79	9.26	8.28
Expend. Water & Sanitation (%)	3.27	3.48	5.43	6.70	2.23	2.34	2.91	10.27
Expend. Education (%)	23.73	28.60	22.67	35.45	38.97	36.28	24.37	16.92

Source: Authors' calculations based on data from CEDE-Panel municipios-Universidad de los Andes.
Notes: Med.=Medellín; Barr.=Barranquilla; Bog.=Bogotá; Cart.=Cartagena; Man.=Manizales; and Buc.=Bucaramanga. 'Other' averages over the remaining municipalities. Monetary figures in millions of constant Colombian pesos of 2005.

Table 22: City Characteristics: Violence

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
ELN presence	0.67	0.11	0.78	0.44	0.56	0.44	0.78	0.17
FARC presence	0.78	0.78	0.89	0.78	0.78	0.78	0.78	0.47
AUC presence	0.56	0.33	0.44	0.33	0.33	0.33	0.22	0.12
Displaced Pop. Out (*)	375.29	36.65	11.89	57.52	42.54	71.68	60.10	455.92
Displaced Pop. In (*)	979.53	312.80	483.50	469.78	205.87	499.42	421.21	625.75
Arrests (*)	428.31	71.14	22.42	175.79	628.17	292.61	210.70	146.14
common criminal (*)	427.10	68.87	22.01	172.45	627.22	291.21	209.73	140.72
terrorist (*)	1.21	2.27	0.41	3.34	0.94	1.40	0.98	5.42
Terrorist Surrenders (*)	3.32	0.57	4.72	0.88	4.43	3.21	2.75	3.56
Thefts (*)	183.25	415.30	404.03	251.21	644.37	647.45	474.66	186.37
Auto (*)	56.17	31.73	43.25	4.77	9.33	7.49	73.85	9.26
Businesses (*)	23.35	49.05	64.92	44.65	30.37	86.43	53.01	31.55
Individuals (*)	91.79	298.28	227.61	162.44	523.18	486.28	288.14	104.14
Residences (*)	11.93	36.24	68.25	39.35	81.49	67.25	59.66	41.41
Homicides (*)	45.64	29.59	19.31	24.34	37.04	24.30	74.27	39.77
common criminal (*)	3.68	0.03	0.43	0.00	0.17	0.00	0.07	1.02
terrorist (*)	41.96	29.56	18.89	24.34	36.87	24.30	74.21	38.75
Kidnappings (*)	0.32	0.37	0.50	0.26	0.46	0.47	0.70	1.11
common criminal (*)	0.28	0.28	0.43	0.23	0.29	0.36	0.58	0.47
terrorist (*)	0.04	0.10	0.07	0.02	0.18	0.11	0.12	0.65
Terrorist Attacks	49.33	37.89	9.11	15.33	3.44	3.56	32.33	3.05

Source: Authors' calculations based on data from CEDE-Panel municipios-Universidad de los Andes.
Notes: Med.=Medellín; Barr.=Barranquilla; Bog.=Bogotá; Cart.=Cartagena; Man.=Manizales; and Buc.=Bucaramanga. 'Other' averages over the remaining municipalities. (*) per 100,000 people. .

Education This module contains 640 variables on number of educational establishments, number of teachers, number of enrolled students (at the primary-, secondary-, and higher-education levels), and test-scores. The primary data source is Censo Anual de Instituciones Educativas carried out by DANE and Ministerio de Educación Nacional. Additional sources include: Sistema Nacional de Información de la Educación Superior (SNIES) from Ministerio de Educación Nacional, and test results from SABER test carried out by ICFES.

The key variables that we use in our study are the following: average years of schooling, historic alphabetism rates (at years 1918, 1938, 1951, 1964, 1985), number of establishments, teachers and students, grade retention, students receiving subsidies, number of higher education students, number of university professors, and test scores from the SABER 11 test (a standardized test similar to the SAT in the United States). The test score is used to rank schools and we construct two measures of school quality: percentage of low- and high-ranking schools. We show summary statistics of these variables by cities in Table 23.

Health and Public Services The health services module contains 41 variables on health and public services coverage, health insurance enrollment, and access to public health services. The primary sources of information are Ministerio de Salud (REPS, CUBOS and social protection), Estadísticas Vitales (DANE), Sistema Único de Información de Servicios Públicos SUI, Unidad de Planeación Minas y Energía (UPME) and Ministerio de Minas y Energía.

The main variables we use in our study include: health insurance enrollment (by insurance type), number of births, number of low-weight births, infant mortality, access to potable water, access to sewage, access to electricity, and access to natural gas network. Table 24 presents summary statistics of select variables.

Table 23: City Characteristics: Education

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
Primary-School Age Pop.	160,226	105,608	616,107	89,647	28,789	42,661	188,183	9,617
Secondary-School Age Pop.	218,643	127,306	763,333	107,135	39,445	55,235	238,034	11,702
Pop. 5–24yrs. Enrolled 2005 (%)	64.52	66.90	68.60	69.27	70.67	67.58	67.72	64.94
Avg. Years of Educ. 1993	7.74	8.12	8.52	7.94	7.78	7.72	7.62	6.24
Avg. Years of Educ. 2005	10.18	10.46	10.99	10.20	10.51	10.32	10.33	8.54
Literacy 1918 (%)	8.95	7.78	8.91	8.76	7.47	6.96	7.84	7.17
Literacy 1993 (%)	86.39	84.07	84.54	82.86	85.48	80.62	81.44	79.83
Literacy 2005 (%)	92.23	95.41	94.49	92.26	96.30	95.85	97.01	90.68
Establishments (total)	862	851	2,902	538	270	306	1,544	89
Public	408	220	765	185	165	130	353	55
Private	454	631	2,137	353	105	176	1,190	34
Teachers (total)	17,718	12,684	66,593	10,241	3,485	5,256	20,071	1,010
Public	10,147	6,779	31,430	5,122	2,344	2,863	6,476	637
Private	7,571	5,905	35,163	5,118	1,141	2,393	13,595	372
Students (total)	413,259	258,351	1,338,950	201,270	76,333	111,398	347,390	21,942
Public	320,109	172,756	829,204	137,709	60,439	83,857	186,388	17,352
Private	93,150	85,595	509,746	63,561	15,894	27,541	161,002	4,590
Retained Students (total)	16,832	4,309	49,434	5,635	3,512	3,206	6,458	706
Public	15,316	3,741	40,406	5,174	3,225	2,903	4,895	652
Private	1,515	568	9,028	461	287	303	1,563	54
Students with Subsidy (total)	63,259	8,670	126,839	36,503	274	3,391	93,634	1,179
University Professors	10,809	3,396	27,626	1,811	2,038	4,277	5,685	123
Higher Educ. Students (total)	340,106	130,523	1,589,034	77,791	53,712	148,687	175,777	4,007
Technical School (tech)	3,241	2,502	526,207	7,155	0	31,897	11,185	65
Technical School (prof)	826	2,317	58,386	0	0	378	8,016	63
Comm. College	153,934	18,904	302,100	25,941	0	5,838	17,818	637
University	182,104	106,800	702,340	44,695	53,712	110,575	138,758	3,242
SABER11 Test Score (total)	50.57	50.12	53.39	48.91	52.49	54.72	50.38	49.20
Language	51.38	50.19	52.84	49.08	52.12	53.47	50.62	49.38
Math	50.54	50.46	52.26	49.29	51.45	54.25	50.10	49.37
Low-Ranking Schools (%)	3.53	16.06	0.55	17.97	1.80	3.35	5.90	8.08
High-Ranking Schools (%)	40.09	33.50	61.36	30.62	42.64	54.64	42.92	31.08

Source: Authors' calculations based on data from CEDE-Panel municipios-Universidad de los Andes.
Notes: Med.=Medellín; Barr.=Barranquilla; Bog.=Bogotá; Cart.=Cartagena; Man.=Manizales; and Buc.=Bucaramanga. 'Other' averages over the remaining municipalities.

Table 24: City Characteristics: Health and Public Services

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
Health Insurance Enrollment (private)	1,662,602	767,098	5,383,141	507,313	271,311	455,643	1,393,314	48,409
Health Insurance Enrollment (public)	762,873	511,540	1,420,878	422,498	90,740	168,940	595,542	41,206
Births (per 1,000)	13.36	19.30	15.40	19.36	11.02	17.39	13.43	14.97
Low-Weight Births (per 1'000)	1.41	1.64	1.95	1.57	0.84	1.42	1.24	1.25
Deaths (per 1,000)	5.28	4.97	3.83	3.60	5.84	5.78	5.49	4.66
Deaths 1-4yrs. (per 1,000)	0.03	0.04	0.03	0.04	0.03	0.03	0.03	0.06
Deaths <1yr. (per 1,000)	0.14	0.32	0.20	0.27	0.13	0.17	0.15	0.20
Infant Mortality Rate	13.43	14.85	13.39	16.32	11.76	11.44	9.70	16.06
Drinking Water 2012 (%)	93.07	100.00	77.01	30.64	82.49	77.24	78.69	79.50
Trash Removal 2012 (%)	90.01	67.53	74.13	9.56	80.51	62.99	68.80	76.11
Sewage 2012 (%)	88.85	94.10	76.05	15.57	81.75	77.24	78.23	75.23
Electric 2012 (%)	100.0	100.0	100.0	100.0	100.0	99.9	99.9	97.0
Natural Gas 2012 (%)	53.66	88.11	92.29	83.81	90.83	92.21	73.45	68.85

Source: Authors' calculations based on data from CEDE-Panel municipios-Universidad de los Andes.
Notes: Med.=Medellín; Barr.=Barranquilla; Bog.=Bogotá; Cart.=Cartagena; Man.=Manizales; and Buc.=Bucaramanga. 'Other' averages over the remaining municipalities.

Appendix II: Firm Characteristics by City-Industry (Shares)

Table 25: Share of Firms by City and Industry (Average, Percent)

(a) Distribution of Industries within Cities

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	13.7	19.3	13.7	23.9	25.9	27.5	19.1	22.0
17	9.3	5.5	5.5	2.9	2.6	1.3	3.6	4.9
18	20.1	4.9	7.6		3.9	15.8	12.4	6.5
19	1.7	2.5	3.0		2.1	5.4	2.6	2.3
20	1.6	2.2	1.4	3.3	2.1	1.3	0.7	1.9
21	1.9	1.2	1.4	2.9		1.3	3.1	3.7
22	7.5	9.1	9.5	2.8	8.0	8.0	13.0	3.8
24	10.1	14.4	13.6	23.4	6.7	2.9	11.3	11.1
25	8.5	13.6	12.3	23.4	7.6	13.1	8.3	11.2
26	5.2	7.2	3.0	4.3	10.0	6.8	2.7	7.0
28	8.5	5.7	8.6	8.8	14.8	4.5	7.8	8.1
29	1.6	3.0	3.9	4.0	2.3	5.1	1.9	3.1
30	0.4		0.1					0.2
31	1.6	2.1	2.2	2.6			2.8	1.6
34	1.9	1.3	4.5	2.8	4.9	2.4	2.1	3.7
36	6.5	10.2	9.5	5.2	10.7	6.6	8.5	9.2
Total	100	100	100	100	100	100	100	100

(b) Geographical Distribution of Industries

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other	Total
15	7.9	3.8	35.5	1.7	2.5	4.1	6.9	37.7	100.0
17	17.1	3.5	45.5	0.6	0.8	0.6	4.3	27.5	100.0
18	22.9	1.9	38.7	0.0	0.8	4.8	8.8	22.1	100.0
19	6.6	3.2	51.3	0.0	1.3	5.4	6.2	25.9	100.0
20	10.4	4.6	39.8	2.6	2.2	2.2	3.0	35.2	100.0
21	8.5	1.6	28.8	1.6	0.0	1.6	8.7	49.3	100.0
22	9.6	4.1	55.8	0.5	1.7	2.7	10.6	15.0	100.0
24	8.3	4.1	50.2	2.4	0.9	0.6	5.7	27.8	100.0
25	7.4	4.0	48.1	2.5	1.0	2.9	4.5	29.6	100.0
26	10.6	5.1	28.5	1.1	3.5	3.6	3.6	44.0	100.0
28	9.9	2.3	46.4	1.3	3.0	1.4	5.9	29.8	100.0
29	4.9	3.1	52.8	1.5	1.2	3.9	3.6	29.0	100.0
30	24.1	0.0	35.7	0.0	0.0	0.0	0.0	40.2	100.0
31	8.6	3.7	51.5	1.8	0.0	0.0	9.2	25.1	100.0
34	4.9	1.3	55.1	0.9	2.2	1.7	3.7	30.2	100.0
36	7.1	3.9	47.5	0.7	2.0	1.9	5.9	31.0	100.0

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín; Barr.=Barranquilla; Bog.=Bogotá; Cart.=Cartagena; Man.=Manizales; and Buc.=Bucaramanga. 'Other' aggregates over the remaining municipalities.

Table 26: Revenues Share by City and Industry (percent)

(a) Distribution of Revenues within Cities

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	38.0	27.2	27.0	8.5	39.6	47.6	33.4	32.6
17	2.9	5.8	5.6	0.2	0.1	0.0	0.5	2.0
18	13.5	0.5	2.1		0.2	2.8	3.9	4.6
19	0.8	2.6	0.9		7.9	0.7	1.1	0.8
20	0.2	0.9	0.6	0.5	0.1	0.0	0.1	0.3
21	8.1	0.2	0.7	0.2		1.3	5.2	8.6
22	2.3	1.6	5.5	0.8	1.2	4.2	10.0	0.9
24	20.3	32.5	21.6	62.7	5.5	12.1	35.4	12.9
25	4.0	7.0	6.2	16.7	3.6	8.1	4.0	7.8
26	1.4	7.7	6.6	3.2	11.9	4.0	0.5	4.9
28	3.5	3.0	4.0	5.5	5.2	2.4	1.5	7.2
29	1.0	0.9	1.8	0.8	24.8	2.7	0.1	2.6
30	0.0		0.0					0.0
31	0.9	1.2	3.7	0.4			0.5	1.9
34	0.9	0.0	8.8	0.9	0.8	13.0	0.9	8.2
36	2.3	9.1	5.1	1.2	2.1	2.2	3.0	4.7
All	100	100	100	100	100	100	100	100

(b) Geographical Distribution of Revenues by Industry

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other	Total
15	13.2	2.9	34.2	0.8	2.9	1.8	8.6	35.7	100.0
17	9.1	5.6	64.0	0.1	0.0	0.0	1.2	19.9	100.0
18	34.6	0.4	19.7	0.0	0.1	0.8	7.6	36.8	100.0
19	8.1	8.0	32.6	0.0	16.5	0.7	8.2	25.9	100.0
20	4.5	6.8	60.6	2.9	0.8	0.1	1.1	23.2	100.0
21	20.1	0.1	5.7	0.1	0.0	0.3	9.5	64.3	100.0
22	6.9	1.3	58.4	0.6	0.7	1.3	22.1	8.6	100.0
24	10.3	5.2	40.2	8.9	0.6	0.6	13.4	20.8	100.0
25	6.3	3.3	35.7	7.9	1.2	1.4	4.8	39.4	100.0
26	2.9	4.6	50.5	2.3	5.3	0.9	0.9	32.5	100.0
28	7.5	2.1	32.1	3.4	2.4	0.6	2.4	49.5	100.0
29	4.8	1.2	29.2	1.0	24.5	1.3	0.5	37.6	100.0
30	29.0	0.0	57.9	0.0	0.0	0.0	0.0	13.1	100.0
31	4.3	1.6	61.8	0.5	0.0	0.0	1.6	30.2	100.0
34	1.4	0.0	52.0	0.3	0.3	2.4	1.1	42.5	100.0
36	5.6	6.5	44.5	0.8	1.0	0.6	5.4	35.6	100.0

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín; Barr.=Barranquilla; Bog.=Bogotá; Cart.=Cartagena; Man.=Manizales; and Buc.=Bucaramanga. 'Other' aggregates over the remaining municipalities.

Table 27: Share of Production Workers by City and Industry (Average, Percent)

(a) Distribution of Production Workers within Cities

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	25.8	22.0	19.1	26.5	32.6	29.3	30.7	23.2
17	6.1	13.6	11.1	0.7	0.4	0.2	2.3	5.6
18	27.5	3.5	8.1		1.6	14.8	11.4	12.5
19	2.0	4.9	2.8		8.2	1.4	4.6	2.5
20	0.5	1.0	1.4	1.5	0.1	0.1	0.2	0.9
21	2.9	0.3	1.3	0.0		8.5	3.3	5.4
22	2.6	1.6	5.7	1.4	1.6	0.6	10.4	1.4
24	8.8	13.2	8.3	12.1	4.4	15.5	19.9	7.0
25	5.8	13.8	11.3	27.7	9.0	5.7	5.5	9.6
26	4.1	8.3	5.8	10.6	10.8	4.5	1.0	6.1
28	5.0	4.6	5.8	11.7	11.1	4.3	2.6	7.0
29	1.4	0.9	4.1	2.8	14.0	5.8	0.4	4.5
30	0.0		0.0					0.0
31	1.3	1.7	2.2	1.5			1.6	2.0
34	1.1	0.1	6.4	2.0	4.7	15.7	1.6	6.5
36	4.9	10.8	6.6	5.8	3.0	1.3	4.6	5.6
All	100	100	100	100	100	100	100	100

(b) Geographical Distribution of Production Workers by Industry

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other	Total
15	13.7	4.3	28.0	1.2	3.0	2.0	10.4	37.5	100.0
17	10.2	8.0	51.0	0.1	0.1	0.0	2.4	28.2	100.0
18	28.2	1.3	22.8	0.0	0.3	1.9	7.5	38.1	100.0
19	8.3	7.7	32.7	0.0	6.0	0.7	12.2	32.4	100.0
20	6.4	4.7	49.8	1.4	0.2	0.2	1.6	35.8	100.0
21	10.9	0.3	13.9	0.0	0.0	4.2	8.3	62.4	100.0
22	8.6	1.9	52.0	0.4	0.9	0.3	22.0	13.9	100.0
24	11.7	6.6	31.1	1.4	1.0	2.6	16.9	28.6	100.0
25	7.4	6.3	39.3	2.9	2.0	0.9	4.4	36.8	100.0
26	8.9	6.6	35.2	2.0	4.1	1.2	1.3	40.5	100.0
28	10.1	3.5	32.7	2.2	4.0	1.1	3.4	43.1	100.0
29	4.7	1.2	37.3	0.8	8.0	2.4	0.7	44.9	100.0
30	12.8	0.0	52.5	0.0	0.0	0.0	0.0	34.7	100.0
31	8.4	4.1	39.8	1.0	0.0	0.0	7.1	39.5	100.0
34	2.6	0.1	41.9	0.3	2.0	4.6	2.3	46.3	100.0
36	10.3	8.1	38.1	1.0	1.1	0.3	6.0	35.2	100.0

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín; Barr.=Barranquilla; Bog.=Bogotá; Cart.=Cartagena; Man.=Manizales; and Buc.=Bucaramanga. 'Other' aggregates over the remaining municipalities.

Table 28: Share of White Collar Workers by City and Industry (Average, Percent)

(a) Distribution of White Collar Workers within Cities

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	26.4	21.1	23.4	25.0	57.6	37.1	27.3	27.1
17	3.4	5.0	5.5	0.5	0.1	0.1	1.5	3.0
18	25.7	5.4	6.7		0.3	8.0	10.1	10.2
19	1.2	3.1	2.4		14.7	2.0	2.0	1.1
20	0.3	1.7	0.7	2.6	0.1	0.0	0.1	0.8
21	4.7	0.4	0.7	0.6		1.1	1.9	6.6
22	5.1	3.9	11.4	16.6	3.4	14.6	13.8	2.7
24	15.6	22.0	17.9	24.0	5.2	6.0	29.7	13.7
25	5.3	9.6	7.7	16.1	3.2	6.3	1.7	7.1
26	1.4	6.6	3.4	3.6	4.3	3.8	0.6	7.5
28	2.8	4.4	4.7	6.2	3.8	2.3	1.9	5.9
29	1.1	1.8	3.3	3.1	4.5	6.3	0.2	2.6
30	0.0		0.1					0.0
31	2.2	1.0	2.8	0.6			2.8	1.9
34	1.1	0.3	3.5	1.2	1.8	10.8	1.0	3.4
36	3.6	14.6	5.7	3.5	1.5	2.6	5.3	6.5
Total	100	100	100	100	100	100	100	100

(b) Geographical Distribution of White Collar Workers by Industry

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other	Total
15	10.7	2.0	35.5	1.2	8.2	2.6	10.6	29.2	100.0
17	9.8	3.6	59.1	0.1	0.1	0.1	4.0	23.2	100.0
18	28.3	1.6	27.4	0.0	0.1	1.5	10.7	30.5	100.0
19	6.3	3.7	46.0	0.0	18.2	1.8	9.2	14.8	100.0
20	4.9	6.1	44.9	4.0	0.4	0.1	1.3	38.3	100.0
21	17.8	0.4	9.4	0.2	0.0	0.7	6.9	64.6	100.0
22	7.0	1.2	57.3	2.2	1.2	3.5	17.8	9.9	100.0
24	10.0	3.3	42.6	1.5	0.8	0.7	17.7	23.5	100.0
25	8.6	3.6	47.8	2.7	1.2	1.9	2.6	31.6	100.0
26	3.6	3.9	32.2	0.9	2.7	1.7	1.5	53.6	100.0
28	6.9	2.5	42.8	1.7	2.2	1.0	4.3	38.7	100.0
29	4.5	1.7	52.7	1.4	4.3	5.0	0.9	29.4	100.0
30	4.1	0.0	75.2	0.0	0.0	0.0	0.0	20.8	100.0
31	10.4	1.1	49.4	0.3	0.0	0.0	13.5	25.2	100.0
34	4.2	0.3	48.7	0.4	1.5	7.4	3.4	34.2	100.0
36	7.0	6.5	40.5	0.7	0.7	0.9	10.2	33.5	100.0

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín; Barr.=Barranquilla; Bog.=Bogotá; Cart.=Cartagena; Man.=Manizales; and Buc.=Bucaramanga. 'Other' aggregates over the remaining municipalities.

Table 29: Share of Capital Stock by City and Industry (Average, Percent)

(a) Distribution of Capital Stock within Cities

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	45.1	18.3	25.2	4.0	46.6	47.2	43.7	32.7
17	3.2	5.6	7.4	0.1	0.1	0.1	0.2	2.4
18	7.3	0.5	2.0		0.1	2.5	0.9	3.7
19	0.3	2.4	0.4		3.2	0.5	0.2	0.7
20	0.3	0.6	1.3	0.4	0.1	0.1	0.0	0.4
21	6.2	0.1	0.8	0.0		0.8	10.4	13.4
22	3.1	1.1	7.4	0.8	1.6	7.3	8.5	1.2
24	20.7	36.3	16.5	51.3	6.3	20.2	25.2	10.9
25	4.4	7.6	8.0	22.0	4.9	6.6	7.3	8.5
26	1.7	13.8	16.9	14.9	17.6	3.1	0.3	8.1
28	3.3	3.8	3.9	5.2	5.2	2.6	1.0	7.1
29	0.7	0.3	1.3	0.4	13.9	2.6	0.1	2.4
30	0.0		0.0					0.0
31	1.0	0.9	2.0	0.2			0.1	1.4
34	0.5	0.0	3.5	0.2	0.9	6.3	0.4	3.1
36	2.3	8.7	3.3	1.2	1.0	0.9	1.6	4.1
Total	100	100	100	100	100	100	100	100

(b) Geographical Distribution of Capital Stock by Industry

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other	Total
15	14.5	1.9	28.7	0.5	2.7	1.7	14.0	35.9	100.0
17	7.3	5.1	66.2	0.1	0.1	0.0	0.4	20.9	100.0
18	24.3	0.6	24.9	0.0	0.1	1.0	3.4	45.7	100.0
19	5.9	13.7	24.8	0.0	10.0	1.0	4.3	40.2	100.0
20	3.6	2.9	66.1	1.4	0.2	0.1	0.5	25.1	100.0
21	9.4	0.1	4.1	0.0	0.0	0.1	17.5	68.8	100.0
22	6.9	0.8	59.5	0.7	0.7	1.9	19.7	9.8	100.0
24	11.1	7.0	33.1	11.1	0.6	1.3	14.2	21.5	100.0
25	4.9	2.9	33.8	11.6	1.1	0.9	8.3	36.3	100.0
26	1.6	4.3	55.6	8.5	3.2	0.3	0.2	26.3	100.0
28	6.4	2.8	29.1	4.8	2.0	0.7	2.0	52.3	100.0
29	4.0	0.7	28.2	0.8	15.2	1.7	0.8	48.6	100.0
30	17.7	0.0	57.9	0.0	0.0	0.0	0.0	24.4	100.0
31	6.7	2.2	49.5	0.6	0.0	0.0	0.9	40.1	100.0
34	2.0	0.0	49.1	0.3	0.6	2.9	1.6	43.4	100.0
36	7.0	8.4	34.5	1.4	0.6	0.3	4.9	42.9	100.0

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín; Barr.=Barranquilla; Bog.=Bogotá; Cart.=Cartagena; Man.=Manizales; and Buc.=Bucaramanga. 'Other' aggregates over the remaining municipalities.

Appendix III: Summary Statistics of Metropolitan Areas

Table 30: Number of Firms by Metropolitan Areas and Industry (Average)

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	66.4	25.1	199.0	8.3	13.6	25.9	56.6	94.7
17	50.3	5.4	78.4	1.0	1.3	2.0	8.4	7.9
18	82.7	4.9	98.2		2.8	12.9	26.4	23.9
19	12.3	2.6	41.7		1.0	4.1	8.7	5.1
20	8.1	2.3	21.2	1.2	1.0	1.0	3.9	6.7
21	15.7	2.9	19.4	1.0		1.0	11.1	11.0
22	31.8	8.9	123.7	1.0	4.2	6.2	30.4	12.9
24	62.2	15.3	199.1	8.9	3.7	2.4	34.2	22.0
25	61.4	14.4	177.2	8.3	3.4	11.1	33.8	19.6
26	27.1	7.1	51.8	2.7	4.9	5.2	8.0	30.4
28	49.0	6.0	124.0	4.1	8.1	6.0	24.3	16.3
29	11.9	4.1	55.9	1.4	1.1	5.9	6.3	6.9
30	1.4		1.8					1.0
31	9.1	2.9	29.6	1.0		1.0	7.1	4.2
34	10.1	1.3	71.7	1.0	2.3	3.6	6.3	9.3
36	39.8	10.8	142.3	2.2	5.0	5.2	26.2	24.4
Overall	539.1	113.2	1435.0	38.7	51.7	91.2	291.9	295.8

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín, Caldas, La Estrella, Sabaneta, Itagüí, Envigado, Bello, Copacabana, Barbosa, Girardota, Guarne; Barr.=Barranquilla, Soledad, Malambo, Galapa; Bog.=Bogotá, Tocancipá, Soacha, Mosquera, Cajicá, Sopó, Madrid, Zipaquirá, Bojacá, Funza, Chía, Sibate, Cota, Facatativá; Cart.=Cartagena, Turbaco; Man.=Manizales, Villamaría, Chinchiná; and Buc.=Bucaramanga, Girón, Piedecuesta, Floridablanca; Cali=Cali, Yumbo, Jamundí, Palmira, Candelaria. 'Other' aggregates over the remaining municipalities.

Table 31: Share of Firms by Metropolitan areas and Industry (Average, Percent)

(a) Distribution of Industries within Metropolitan areas

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	12.3	22.1	13.9	21.5	26.3	28.4	19.4	31.9
17	9.3	4.7	5.5	2.7	2.4	2.2	2.8	2.7
18	15.4	4.3	6.8		5.4	14.0	9.0	8.0
19	2.3	2.2	2.9		1.9	4.4	2.9	1.7
20	1.5	2.1	1.5	3.1	1.9	1.1	1.3	2.3
21	2.9	2.6	1.4	2.6		1.1	3.8	3.7
22	5.9	7.9	8.6	2.6	8.2	6.8	10.4	4.4
24	11.5	13.5	13.9	22.9	7.1	2.7	11.8	7.5
25	11.4	12.7	12.3	21.6	6.6	12.3	11.7	6.6
26	5.0	6.3	3.6	6.9	9.5	5.8	2.7	10.3
28	9.1	5.4	8.6	10.6	15.7	6.6	8.3	5.6
29	2.2	3.6	3.9	3.7	2.1	6.4	2.2	2.3
30	0.3		0.1					0.3
31	1.7	2.6	2.1	2.5		1.1	2.5	1.4
34	1.9	1.1	5.0	2.7	4.5	3.9	2.2	3.1
36	7.4	9.5	9.9	5.8	9.7	5.8	9.0	8.3
Total	100	100	100	100	100	100	100	100

(b) Geographical Distribution of Industries

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other	Total
15	13.6	5.1	40.6	1.7	2.8	5.3	11.6	19.3	100.0
17	32.5	3.5	50.7	0.6	0.8	1.3	5.5	5.1	100.0
18	32.8	1.9	39.0	0.0	1.1	5.1	10.5	9.5	100.0
19	16.3	3.4	55.2	0.0	1.3	5.4	11.5	6.8	100.0
20	17.9	5.1	46.7	2.6	2.2	2.2	8.6	14.7	100.0
21	25.2	4.7	31.3	1.6	0.0	1.6	17.9	17.7	100.0
22	14.5	4.1	56.4	0.5	1.9	2.8	13.9	5.9	100.0
24	17.9	4.4	57.2	2.6	1.1	0.7	9.8	6.3	100.0
25	18.7	4.4	53.8	2.5	1.0	3.4	10.3	5.9	100.0
26	19.8	5.2	37.7	1.9	3.6	3.8	5.8	22.2	100.0
28	20.6	2.5	52.1	1.7	3.4	2.5	10.2	6.9	100.0
29	12.7	4.4	59.7	1.5	1.2	6.3	6.8	7.4	100.0
30	34.0	0.0	42.3	0.0	0.0	0.0	0.0	23.8	100.0
31	16.6	5.3	53.8	1.8	0.0	1.8	13.0	7.7	100.0
34	9.6	1.3	67.8	0.9	2.2	3.4	6.0	8.8	100.0
36	15.5	4.2	55.6	0.9	2.0	2.0	10.2	9.5	100.0

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín, Caldas, La Estrella, Sabaneta, Itagüí, Envigado, Bello, Copacabana, Barbosa, Girardota, Guarne; Barr.=Barranquilla, Soledad, Malambo, Galapa; Bog.=Bogotá, Tocancipá, Soacha, Mosquera, Cajicá, Sopó, Madrid, Zipaquirá, Bojacá, Funza, Chía, Sibate, Cota, Facatativá; Cart.=Cartagena, Turbaco; Man.=Manizales, Villamaría, Chinchiná; and Buc.=Bucaramanga, Girón, Piedecuesta, Floridablanca; Cali=Cali, Yumbo, Jamundí, Palmira, Candelaria. 'Other' aggregates over the remaining municipalities.

Table 32: Revenues by Metropolitan areas and Industry (Average)

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	5,703	1,073	9,892	203	793	665	3,880	3,204
17	468	155	1,914	4	1	5	43	166
18	1,887	14	681		25	29	417	380
19	169	70	343		144	6	109	33
20	30	27	212	9	3	0	8	19
21	1,479	117	257	4		10	1,089	741
22	322	40	1,766	19	23	39	747	46
24	2,366	928	8,883	1,548	109	112	2,901	478
25	951	239	2,554	436	65	107	967	208
26	950	194	2,324	101	219	42	67	250
28	586	89	1,616	314	165	33	925	290
29	505	73	702	19	472	43	48	70
30	3		6					1
31	169	81	1,230	10		3	68	323
34	367	1	4,548	14	14	201	129	110
36	427	249	1,961	29	38	23	483	442
Overall	16,380	3,347	38,889	2,684	2,018	1,307	11,878	6,761

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín, Caldas, La Estrella, Sabaneta, Itagüí, Envigado, Bello, Copacabana, Barbosa, Girardota, Guarne; Barr.=Barranquilla, Soledad, Malambo, Galapa; Bog.=Bogotá, Tocancipá, Soacha, Mosquera, Cajicá, Sopó, Madrid, Zipaquirá, Bojacá, Funza, Chía, Sibate, Cota, Facatativá; Cart.=Cartagena, Turbaco; Man.=Manizales, Villamaría, Chinchiná; and Buc.=Bucaramanga, Girón, Piedecuesta, Floridablanca; Cali=Cali, Yumbo, Jamundí, Palmira, Candelaria. 'Other' aggregates over the remaining municipalities. Monetary figures in billions of constant Colombian pesos of 2005.

Table 33: Revenues Share by Metropolitan areas and Industry (percent)

(a) Distribution of Revenues within Metropolitan areas

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	34.8	32.1	25.4	7.7	39.5	50.7	32.8	47.5
17	2.9	4.6	5.0	0.1	0.1	0.4	0.4	2.5
18	11.6	0.4	1.8		1.2	2.3	3.5	5.8
19	1.0	2.1	0.9		7.2	0.5	0.9	0.5
20	0.2	0.8	0.6	0.4	0.1	0.0	0.1	0.3
21	9.2	3.5	0.7	0.1		1.1	9.2	11.0
22	1.9	1.2	4.6	0.7	1.2	3.0	6.3	0.7
24	14.5	27.1	22.7	58.1	5.4	8.3	24.4	6.9
25	5.8	7.3	6.6	15.5	3.3	8.1	8.1	3.0
26	5.9	6.0	6.0	3.3	10.9	3.1	0.6	3.7
28	3.6	2.6	4.2	12.2	8.1	2.5	7.7	4.2
29	3.0	2.2	1.8	0.7	23.1	3.2	0.4	1.1
30	0.0		0.0					0.0
31	1.0	2.5	3.2	0.4		0.2	0.6	4.4
34	2.0	0.0	11.7	0.8	0.7	16.0	1.1	1.7
36	2.6	7.6	5.1	1.2	1.9	1.8	4.0	6.7
All	100	100	100	100	100	100	100	100

(b) Geographical Distribution of Revenues by Industry

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other	Total
15	22.4	4.2	38.9	0.8	3.1	2.6	15.3	12.6	100.0
17	17.0	5.6	69.4	0.1	0.0	0.2	1.6	6.0	100.0
18	55.0	0.4	19.8	0.0	0.7	0.8	12.1	11.1	100.0
19	19.3	8.0	39.2	0.0	16.5	0.7	12.5	3.8	100.0
20	9.8	8.8	69.0	2.9	0.8	0.1	2.4	6.2	100.0
21	40.0	3.2	7.0	0.1	0.0	0.3	29.5	20.0	100.0
22	10.7	1.3	58.8	0.6	0.8	1.3	24.9	1.5	100.0
24	13.7	5.4	51.3	8.9	0.6	0.6	16.7	2.8	100.0
25	17.2	4.3	46.2	7.9	1.2	1.9	17.5	3.8	100.0
26	22.9	4.7	56.0	2.4	5.3	1.0	1.6	6.0	100.0
28	14.6	2.2	40.2	7.8	4.1	0.8	23.0	7.2	100.0
29	26.1	3.8	36.3	1.0	24.4	2.2	2.5	3.6	100.0
30	28.6	0.0	65.7	0.0	0.0	0.0	0.0	5.6	100.0
31	9.0	4.3	65.3	0.5	0.0	0.2	3.6	17.1	100.0
34	6.8	0.0	84.5	0.3	0.3	3.7	2.4	2.0	100.0
36	11.7	6.8	53.7	0.8	1.0	0.6	13.2	12.1	100.0

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín, Caldas, La Estrella, Sabaneta, Itagüí, Envigado, Bello, Copacabana, Barbosa, Girardota, Guarne; Barr.=Barranquilla, Soledad, Malambo, Galapa; Bog.=Bogotá, Tocancipá, Soacha, Mosquera, Cajicá, Sopó, Madrid, Zipaquirá, Bojacá, Funza, Chía, Sibate, Cota, Facatativá; Cart.=Cartagena, Turbaco; Man.=Manizales, Villamaría, Chinchiná; and Buc.=Bucaramanga, Girón, Piedecuesta, Floridablanca; Cali=Cali, Yumbo, Jamundí, Palmira, Candelaria. 'Other' aggregates over the remaining municipalities.

Table 34: Number of Production Workers by Metropolitan areas and Industry (Average)

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	10,586	3,767	18,838	697	1,816	1,484	10,673	9,708
17	4,176	1,493	10,306	14	24	8	556	1,990
18	13,019	392	6,958		379	589	3,494	5,565
19	1,737	555	2,957		438	50	1,208	311
20	271	139	1,472	32	6	4	129	315
21	1,970	621	1,262	0		342	2,180	1,695
22	1,493	180	4,929	34	106	32	2,378	217
24	4,206	1,512	10,290	325	246	598	4,535	977
25	5,170	1,606	11,647	712	485	433	3,459	823
26	3,233	932	6,001	335	577	170	329	2,369
28	3,722	548	5,899	542	695	219	2,145	1,309
29	2,805	550	4,079	78	747	414	289	344
30	4		15					8
31	1,005	341	1,989	45		0	481	803
34	793	8	9,491	39	269	777	641	1,172
36	2,938	1,201	6,987	133	155	49	1,949	1,269
Overall	57,126	13,840	103,120	2,898	5,853	4,862	34,448	28,871

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín, Caldas, La Estrella, Sabaneta, Itagüí, Envigado, Bello, Copacabana, Barbosa, Girardota, Guarne; Barr.=Barranquilla, Soledad, Malambo, Galapa; Bog.=Bogotá, Tocancipá, Soacha, Mosquera, Cajicá, Sopó, Madrid, Zipaquirá, Bojacá, Funza, Chía, Sibate, Cota, Facatativá; Cart.=Cartagena, Turbaco; Man.=Manizales, Villamaría, Chinchiná; and Buc.=Bucaramanga, Girón, Piedecuesta, Floridablanca; Cali=Cali, Yumbo, Jamundí, Palmira, Candelaria. 'Other' aggregates over the remaining municipalities.

Table 35: Share of Production Workers by Metropolitan areas and Industry (Average, Percent)

(a) Distribution of Production Workers within Metropolitan areas

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	18.6	27.3	18.4	23.9	30.9	30.5	31.0	33.5
17	7.3	10.9	10.0	0.6	0.4	0.2	1.6	7.0
18	22.6	2.8	6.7		6.6	12.4	10.1	19.2
19	3.0	4.0	2.9		7.5	1.1	3.5	1.1
20	0.5	1.0	1.4	1.4	0.1	0.1	0.4	1.1
21	3.5	4.4	1.2	0.0		7.4	6.3	5.9
22	2.6	1.3	4.8	1.2	1.8	0.7	6.9	0.7
24	7.4	10.8	10.0	10.8	4.2	12.0	13.2	3.4
25	9.1	11.7	11.3	25.1	8.2	9.0	10.0	2.9
26	5.7	6.7	5.8	11.4	9.9	3.6	1.0	8.2
28	6.5	3.9	5.7	18.2	11.7	4.6	6.2	4.5
29	4.9	4.0	3.9	2.7	12.9	8.4	0.8	1.2
30	0.0		0.0					0.0
31	1.7	2.5	1.9	1.5		0.0	1.4	2.8
34	1.4	0.1	9.2	1.6	4.4	15.8	1.9	4.0
36	5.2	8.8	6.8	4.9	2.7	1.1	5.7	4.5
All	100	100	100	100	100	100	100	100

(b) Geographical Distribution of Production Workers by Industry

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other	Total
15	18.4	6.5	32.7	1.2	3.2	2.6	18.5	16.9	100.0
17	22.5	8.0	55.5	0.1	0.1	0.0	3.0	10.7	100.0
18	42.8	1.3	22.9	0.0	1.2	1.9	11.5	18.3	100.0
19	23.9	7.7	40.7	0.0	6.0	0.7	16.6	4.3	100.0
20	11.4	5.9	62.1	1.4	0.2	0.2	5.5	13.3	100.0
21	24.4	7.7	15.6	0.0	0.0	4.2	27.0	21.0	100.0
22	15.9	1.9	52.6	0.4	1.1	0.3	25.4	2.3	100.0
24	18.5	6.7	45.4	1.4	1.1	2.6	20.0	4.3	100.0
25	21.2	6.6	47.9	2.9	2.0	1.8	14.2	3.4	100.0
26	23.2	6.7	43.0	2.4	4.1	1.2	2.4	17.0	100.0
28	24.7	3.6	39.1	3.6	4.6	1.5	14.2	8.7	100.0
29	30.1	5.9	43.8	0.8	8.0	4.5	3.1	3.7	100.0
30	13.6	0.0	56.1	0.0	0.0	0.0	0.0	30.4	100.0
31	21.5	7.3	42.6	1.0	0.0	0.0	10.3	17.2	100.0
34	6.0	0.1	72.0	0.3	2.0	5.9	4.9	8.9	100.0
36	20.0	8.2	47.6	0.9	1.1	0.3	13.3	8.6	100.0

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín, Caldas, La Estrella, Sabaneta, Itagüí, Envigado, Bello, Copacabana, Barbosa, Girardota, Guarne; Barr.=Barranquilla, Soledad, Malambo, Galapa; Bog.=Bogotá, Tocancipá, Soacha, Mosquera, Cajicá, Sopó, Madrid, Zipaquirá, Bojacá, Funza, Chía, Sibate, Cota, Facatativá; Cart.=Cartagena, Turbaco; Man.=Manizales, Villamaría, Chinchiná; and Buc.=Bucaramanga, Girón, Piedecuesta, Floridablanca; Cali=Cali, Yumbo, Jamundí, Palmira, Candelaria. 'Other' aggregates over the remaining municipalities.

Table 36: Number of White Collar Workers by Metropolitan areas and Industry (Average)

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	8,402	1,575	21,320	629	4,542	1,883	8,434	7,579
17	1,637	273	4,772	10	7	236	352	327
18	8,033	316	5,460		118	336	3,435	2,178
19	454	162	2,125		785	76	521	186
20	184	95	700	51	5	2	139	95
21	2,196	287	623	12		42	1,602	1,150
22	1,531	191	9,407	363	208	567	3,281	701
24	4,938	1,178	18,486	513	290	235	6,913	1,741
25	2,883	523	7,297	355	157	294	1,158	540
26	1,067	341	5,629	138	227	159	212	761
28	1,588	254	4,419	277	241	135	1,288	662
29	854	168	3,174	73	217	346	134	116
30	9		78					10
31	707	119	2,444	15		53	713	611
34	409	17	3,966	22	86	512	316	497
36	1,863	788	5,930	81	84	113	1,689	895
Overall	36,753	6,276	95,831	2,473	6,939	4,910	30,187	18,044

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín, Caldas, La Estrella, Sabaneta, Itagüí, Envigado, Bello, Copacabana, Barbosa, Girardota, Guarne; Barr.=Barranquilla, Soledad, Malambo, Galapa; Bog.=Bogotá, Tocancipá, Soacha, Mosquera, Cajicá, Sopó, Madrid, Zipaquirá, Bojacá, Funza, Chía, Sibate, Cota, Facatativá; Cart.=Cartagena, Turbaco; Man.=Manizales, Villamaría, Chinchiná; and Buc.=Bucaramanga, Girón, Piedecuesta, Floridablanca; Cali=Cali, Yumbo, Jamundí, Palmira, Candelaria. 'Other' aggregates over the remaining municipalities.

Table 37: Share of White Collar Workers by Metropolitan areas and Industry (Average, Percent)

(a) Distribution of White Collar Workers within Metropolitan areas

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	22.6	25.1	22.2	23.3	56.7	38.9	27.8	42.2
17	4.5	4.2	5.0	0.4	0.1	4.6	1.2	1.8
18	22.1	4.6	5.7		1.8	7.2	11.2	12.2
19	1.2	2.6	2.2		13.9	1.5	1.8	1.0
20	0.5	1.6	0.7	2.5	0.1	0.0	0.5	0.5
21	6.2	4.5	0.6	0.5		1.0	5.4	6.3
22	4.1	3.1	9.9	15.1	3.6	11.6	11.0	3.9
24	13.4	18.7	19.3	21.8	5.4	4.7	23.0	9.7
25	7.8	8.5	7.6	14.4	3.0	6.0	3.9	3.0
26	2.9	5.5	5.8	5.9	4.1	3.1	0.7	4.3
28	4.3	4.2	4.6	11.2	4.3	2.8	4.3	3.6
29	2.3	2.7	3.3	2.8	4.3	6.9	0.4	0.7
30	0.0		0.1					0.1
31	2.0	2.0	2.6	0.6		1.0	2.3	3.3
34	1.1	0.3	4.1	1.1	1.7	10.2	1.1	2.8
36	5.0	12.7	6.2	3.4	1.4	2.2	5.5	4.9
Total	100	100	100	100	100	100	100	100

(b) Geographical Distribution of White Collar Workers by Industry

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other	Total
15	15.5	2.9	39.2	1.2	8.4	3.5	15.5	13.9	100.0
17	21.5	3.6	62.7	0.1	0.1	3.1	4.6	4.3	100.0
18	40.4	1.6	27.5	0.0	0.6	1.7	17.3	11.0	100.0
19	10.5	3.8	49.3	0.0	18.2	1.8	12.1	4.3	100.0
20	14.5	7.5	55.1	4.0	0.4	0.1	10.9	7.4	100.0
21	37.2	4.9	10.5	0.2	0.0	0.7	27.1	19.4	100.0
22	9.4	1.2	57.9	2.2	1.3	3.5	20.2	4.3	100.0
24	14.4	3.4	53.9	1.5	0.8	0.7	20.2	5.1	100.0
25	21.8	4.0	55.2	2.7	1.2	2.2	8.8	4.1	100.0
26	12.5	4.0	66.0	1.6	2.7	1.9	2.5	8.9	100.0
28	17.9	2.9	49.9	3.1	2.7	1.5	14.5	7.5	100.0
29	16.8	3.3	62.5	1.4	4.3	6.8	2.6	2.3	100.0
30	9.7	0.0	79.8	0.0	0.0	0.0	0.0	10.5	100.0
31	15.2	2.6	52.4	0.3	0.0	1.1	15.3	13.1	100.0
34	7.0	0.3	68.1	0.4	1.5	8.8	5.4	8.5	100.0
36	16.3	6.9	51.8	0.7	0.7	1.0	14.8	7.8	100.0

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín, Caldas, La Estrella, Sabaneta, Itagüí, Envigado, Bello, Copacabana, Barbosa, Girardota, Guarne; Barr.=Barranquilla, Soledad, Malambo, Galapa; Bog.=Bogotá, Tocancipá, Soacha, Mosquera, Cajicá, Sopó, Madrid, Zipaquirá, Bojacá, Funza, Chía, Sibate, Cota, Facatativá; Cart.=Cartagena, Turbaco; Man.=Manizales, Villamaría, Chinchiná; and Buc.=Bucaramanga, Girón, Piedecuesta, Floridablanca; Cali=Cali, Yumbo, Jamundí, Palmira, Candelaria. 'Other' aggregates over the remaining municipalities.

Table 38: Capital Stock by Metropolitan areas and Industry (Average)

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	2,799	382	4,680	74	399	400	3,657	1,935
17	280	88	1,190	1	1	2	22	146
18	666	7	319		4	14	74	160
19	47	35	110		26	3	17	17
20	48	11	213	4	1	0	6	16
21	589	84	143	0		3	1,399	775
22	221	15	1,168	14	15	38	441	44
24	1,081	586	3,376	884	53	104	1,604	287
25	490	145	1,641	433	43	67	759	156
26	866	206	2,837	412	154	17	32	232
28	244	64	838	225	62	18	520	163
29	243	32	272	6	114	28	15	40
30	1		3					0
31	82	29	325	4		2	30	153
34	70	1	899	3	7	66	38	36
36	185	128	705	20	8	5	209	218
Overall	7,913	1,812	18,718	2,073	874	764	8,821	4,379

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín, Caldas, La Estrella, Sabaneta, Itagüí, Envigado, Bello, Copacabana, Barbosa, Girardota, Guarne; Barr.=Barranquilla, Soledad, Malambo, Galapa; Bog.=Bogotá, Tocancipá, Soacha, Mosquera, Cajicá, Sopó, Madrid, Zipaquirá, Bojacá, Funza, Chía, Sibate, Cota, Facatativá; Cart.=Cartagena, Turbaco; Man.=Manizales, Villamaría, Chinchiná; and Buc.=Bucaramanga, Girón, Piedecuesta, Floridablanca; Cali=Cali, Yumbo, Jamundí, Palmira, Candelaria. 'Other' aggregates over the remaining municipalities. Monetary figures in billions of constant Colombian pesos of 2005.

Table 39: Share of Capital Stock by Metropolitan areas and Industry (Average, Percent)

(a) Distribution of Capital Stock within Metropolitan areas

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other
15	34.0	21.3	24.5	3.7	46.0	50.1	41.1	44.1
17	3.7	4.8	6.5	0.1	0.1	0.3	0.3	3.4
18	8.6	0.4	1.7		0.5	2.0	0.8	3.8
19	0.6	2.0	0.6		3.0	0.4	0.2	0.4
20	0.5	0.6	1.2	0.4	0.1	0.0	0.1	0.4
21	8.0	4.6	0.8	0.0		0.6	15.8	18.1
22	2.7	0.9	6.3	0.8	1.7	5.2	5.0	1.0
24	13.9	31.7	17.8	47.4	6.2	13.8	18.3	6.3
25	6.2	8.2	8.8	20.3	4.7	9.1	8.8	3.5
26	11.4	11.4	15.4	14.7	16.9	2.3	0.4	5.3
28	3.2	3.5	4.5	11.4	7.0	2.4	5.9	3.6
29	3.1	1.8	1.5	0.3	13.4	3.8	0.2	1.0
30	0.0		0.0					0.0
31	1.1	1.6	1.8	0.2		0.2	0.3	3.1
34	0.8	0.0	4.9	0.2	0.8	9.8	0.4	0.8
36	2.2	7.3	3.9	1.1	1.0	0.7	2.4	5.0
Total	100	100	100	100	100	100	100	100

(b) Geographical Distribution of Capital Stock by Industry

	Med.	Barr.	Bog.	Cart.	Man.	Buc.	Cali	Other	Total
15	19.5	2.7	32.7	0.5	2.8	2.8	25.5	13.5	100.0
17	16.2	5.1	68.8	0.1	0.1	0.1	1.3	8.5	100.0
18	53.5	0.6	25.6	0.0	0.4	1.1	5.9	12.8	100.0
19	18.5	13.7	43.1	0.0	10.0	1.0	6.7	6.8	100.0
20	15.9	3.8	71.1	1.4	0.2	0.1	2.0	5.4	100.0
21	19.7	2.8	4.8	0.0	0.0	0.1	46.7	25.9	100.0
22	11.3	0.8	59.7	0.7	0.8	2.0	22.5	2.3	100.0
24	13.6	7.3	42.3	11.1	0.7	1.3	20.1	3.6	100.0
25	13.1	3.9	43.9	11.6	1.1	1.8	20.3	4.2	100.0
26	18.2	4.3	59.6	8.7	3.2	0.4	0.7	4.9	100.0
28	11.4	3.0	39.3	10.5	2.9	0.9	24.3	7.7	100.0
29	32.4	4.2	36.3	0.8	15.2	3.8	2.0	5.4	100.0
30	23.2	0.0	65.1	0.0	0.0	0.0	0.0	11.7	100.0
31	13.2	4.7	52.0	0.6	0.0	0.3	4.8	24.4	100.0
34	6.3	0.0	80.3	0.3	0.6	5.9	3.4	3.2	100.0
36	12.5	8.6	47.7	1.4	0.6	0.3	14.2	14.7	100.0

Source: Authors' calculations based on data from SIREM and DIAN/DANE.
Notes: Med.=Medellín, Caldas, La Estrella, Sabaneta, Itagüí, Envigado, Bello, Copacabana, Barbosa, Girardota, Guarne; Barr.=Barranquilla, Soledad, Malambo, Galapa; Bog.=Bogotá, Tocancipá, Soacha, Mosquera, Cajicá, Sopó, Madrid, Zipaquirá, Bojacá, Funza, Chía, Sibate, Cota, Facatativá; Cart.=Cartagena, Turbaco; Man.=Manizales, Villamaría, Chinchiná; and Buc.=Bucaramanga, Girón, Piedecuesta, Floridablanca; Cali=Cali, Yumbo, Jamundí, Palmira, Candelaria. 'Other' aggregates over the remaining municipalities.

Appendix IV: Productivity and Agglomeration Economies

Table 40: Firm-Level TFP - Main Cities Only

		Labor			
	Num. Firms	Total	White	Blue	Revenues
	(1)	(2)	(3)	(4)	(5)
Scale	0.000	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Specialization	0.804***	0.278**	0.147	0.357***	0.215**
	(0.196)	(0.116)	(0.109)	(0.112)	(0.103)
Variety	-1.213	0.108	0.177	-0.409	-0.120
	(0.776)	(0.197)	(0.157)	(0.268)	(0.166)
City FE	yes	yes	yes	yes	yes
Industry FE	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes
Obs.	16766	16766	16766	16766	16766

Table 41: Firm-Level TFP: Neighbor Effects (Department)

		Labor			
	Num. Firms	Total	White	Blue	Revenues
	(1)	(2)	(3)	(4)	(5)
(a) City					
Scale	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Specialization	0.155*	0.098**	0.054	0.101**	0.368***
	(0.082)	(0.045)	(0.044)	(0.043)	(0.041)
Variety	0.157*	0.172**	0.129**	0.136*	0.202***
	(0.091)	(0.068)	(0.064)	(0.070)	(0.064)
(b) Neighbor					
Scale	0.000*	0.000	−0.000	0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Specialization	0.153*	0.192***	0.169***	0.215***	0.148***
	(0.078)	(0.053)	(0.049)	(0.053)	(0.048)
Variety	0.011	0.080	0.022	0.124	0.042
	(0.102)	(0.088)	(0.074)	(0.083)	(0.081)
City FE	yes	yes	yes	yes	yes
Industry FE	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes
Obs.	23358	23358	23356	23266	23358

Table 42: Firm-Level TFP: Neighbor Effects (Metropolitan Area) - Main Cities Only

		Labor			
	Num. Firms	Total	White	Blue	Revenues
	(1)	(2)	(3)	(4)	(5)
(a) City					
Scale	0.000	−0.000	−0.000	0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Specialization	0.835***	0.250**	0.116	0.320***	0.167
	(0.208)	(0.117)	(0.110)	(0.114)	(0.105)
Variety	−1.244	0.204	0.238	−0.389	−0.108
	(0.793)	(0.205)	(0.165)	(0.272)	(0.168)
(b) Neighbor					
Scale	0.000	0.000	−0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Specialization	−0.040	0.246***	0.230***	0.250***	0.227***
	(0.106)	(0.067)	(0.058)	(0.070)	(0.059)
Variety	0.182	0.215	0.144	0.037	0.123
	(0.154)	(0.138)	(0.100)	(0.126)	(0.103)
City FE	yes	yes	yes	yes	yes
Industry FE	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes
Obs.	16734	16734	16734	16728	16734

Table 43: Firm-Level TFP: Competition - Main Cities Only

	Labor			
	Total	White	Blue	Revenues
	(1)	(2)	(3)	(4)
Scale	−0.000	−0.000	−0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Specialization	0.521***	0.370**	0.527***	0.422***
	(0.164)	(0.150)	(0.153)	(0.143)
Variety	0.001	0.076	−0.525*	−0.129
	(0.204)	(0.163)	(0.274)	(0.169)
Size	−0.000***	−0.000***	−0.000**	−0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Competition	0.110*	0.083*	0.155***	0.069
	(0.057)	(0.050)	(0.050)	(0.050)
City FE	yes	yes	yes	yes
Industry FE	yes	yes	yes	yes
Year FE	yes	yes	yes	yes
Obs.	16666	16666	16666	16666


BANCO DE LA REPÚBLICA
COLOMBIA