Division of Labor and Productivity Advantage of Cities: Theory and Evidence from Brazil*

Lin Tian†

January 18, 2019

Click here for the latest version

Abstract

Firms are more productive in larger cities. This paper investigates a potential explanation that was first proposed by Adam Smith: Larger cities facilitate greater division of labor within firms. Using a dataset of Brazilian firms, I first document that division of labor is indeed robustly correlated with city size, controlling for firm size. To quantify the importance of division of labor in explaining productivity advantages of cities, I propose and estimate a quantitative model that embeds a theory of firms' choice of the optimal division of labor in a spatial equilibrium framework. In the model, the observed correlation between firm's division of labor and city size is generated by both a selection effect—firms endogenously sort across space, choosing different extents of division of labor—and a treatment effect—larger cities increase division of labor for all firms, by reducing the costs associated with greater division of labor. Exploiting a quasi-experiment that changes the cost of division of labor within cities—the gradual roll-out of broadband internet infrastructure—I validate key model assumptions and structurally estimate model parameters. Through a counterfactual analysis, I estimate that division of labor contributes to 15% of the productivity advantages of larger cities in Brazil, half of which is due to firm sorting and the other half to the treatment effect of larger city size.

*Previously circulated as "Division of Labor and Extent of the Market." I am grateful to Donald Davis, Jonathan Vogel, Réka Juhász, Jonas Hjort and David Weinstein for their invaluable guidance and continued support. I thank Nate Baum-Snow, Andy Bernard, Lorenzo Caliendo, Davin Chor, Victor Couture, Rebecca Diamond, Jingting Fan, Cecile Gaubert, Cecilia Fieler, Teresa Fort, François Gerard, Jessie Handbury, Yang Jiao, Amit Khandelwal, Brian Kovak, Nandita Krishnaswamy, Ildikó Magyari, Ferdinando Monte, Dávid Nagy, Fernando Parro, Stephen Redding, Tobias Salz, Rodrigo Soares, Meredith Startz, Adrian Tang, Danna Thomas, Sharon Traiberman, Eric Verhoogen, Mike Waugh, Shang-Jin Wei, Scott Weiner, Yue Yu, Ding Yuan, Jing Zhou, and numerous participants at New Faces of International Trade Conference, NBER Summer Institute ITI Workshop, and other conferences, seminars and workshops for helpful discussions and comments. All errors are mine.

†INSEAD, 1 Ayer Rajah Avenue, Singapore 138676. Email: lin.tian@insead.edu. Website: https://lin-tian.github.io.

"The greatest improvement in the productive powers of labour ... seem(s) to have been the effects of division of labour."

– Adam Smith, The Wealth of Nations, 1776

1 Introduction

Firms are more productive in larger cities. Numerous theories have been put forth to explain this stylized fact, including knowledge spillover, sharing of indivisible public facilities, and availability of intermediate inputs such as labor.[1] However, empirical literature that quantifies the importance of these mechanisms is limited. Understanding and identifying key sources of agglomeration forces is important, as different mechanisms may generate different productivity and welfare implications for a given policy. As Lucas (1976) points out, without knowing how policy affects the behavior of private agents such as firms, it is unwise to predict the effects of a new policy based on past data.

This paper investigates one potential mechanism for the city-size-productivity relation: *division of labor within firms*. The idea that division of labor may contribute to spatial productivity difference was first discussed by Smith (1776), who proposes that firms in larger cities adopt greater division of labor, thereby raising local productivity. However, there is little modern theory and no empirical work that studies the importance of this force for the productivity advantages of larger cities. In this paper, I investigate this problem using a combination of empirical, theoretical and structural analyses, and show that division of labor within firms is an important source for the productivity advantage in larger cities.

I construct a unique dataset on firm-level division of labor using a sample of matched employer-employee records of Brazilian firms.[2] The dataset allows me to document two new stylized facts on firms' division of labor: Within a sector, there is greater division of labor inside firms in larger cities; within a city, there is greater division of labor within firms that produce more complex goods.[3] The correlations remain largely unchanged when controlling for characteristics such as firm size and skill intensity.

Motivated by the stylized facts, I develop a theoretical framework in which the spatial distributions of firms' division of labor and productivity are determined jointly. The model embeds firms' endogenous decisions on division of labor into a spatial equilibrium model. Through the model, I propose potential mechanisms that generate the observed correlation between division of labor and city size. Firms, exogenously

[1] See Duranton and Puga (2004) for a review of this literature.

[2] The main dataset used is the confidential micro-level data from the Annual Social Report of Brazil (*Relação Anual de Informações*, or RAIS). The RAIS dataset covers all registered firms in Brazil and contains comprehensive information on firm and worker characteristics. The RAIS data classify workers into 6-digit CBO codes, each of which is accompanied by a detailed description of the tasks involved. In contrast, most other matched employer-employee datasets, such as the Portuguese *Quadros de Pessoal* and French *Déclarations Annuel des Données Sociales*, only provide 4-digit occupation classifications. I provide more details on the data and construction of division of labor within firms in Sections 2.1 and 2.2, and Online Appendix.

[3] For all empirical exercises, a firm is defined as an establishment for multi-establishment firms.

heterogeneous in the *complexity* of their products, choose division of labor and size of the city in which to locate to maximize profit. As in Becker and Murphy (1992), the optimal division of labor depends on two forces: gains from labor specialization and costs of hiring more specialized workers.[4] The model makes two key reduced-form assumptions, each of which is microfounded in the appendix: First, firms producing more complex products benefit relatively more from labor specialization; second, the costs of division of labor are lower in larger cities.[5] Since in equilibrium, more complex firms—i.e., firms with more complex products—choose greater division of labor, and firms with greater division of labor benefit more from being in larger cities, there is positive assortative matching between firm complexity and city size. Firms in larger cities exhibit a greater division of labor for two reasons: (i) the direct effect of city size, i.e., larger cities make it less costly for all firms to increase the level of worker specialization; and (ii) sorting of firms, i.e., larger cities are also occupied by more complex firms that choose to have a deeper division of labor. In equilibrium, the model generates results that are consistent with salient features of the Brazilian economy: Firms tend to be bigger and more productive in larger cities; and more complex sectors tend to have more firms in larger cities.

I next present empirical evidence that supports the proposed theory, in particular the two reduced-form assumptions. To do so, I focus on one possible channel that generates the complementarity between division of labor and city size: Larger cities provide better ICT infrastructure, which increases firms' division of labor. Modern ICT technologies, such as fast internet, can facilitate greater division of labor within firms through a number of mechanisms, e.g., by improving communications efficiencies, enhancing information storage and sharing, or reducing coordination frictions within firms (e.g., Borghans and Weel, 2006; Varian, 2010; McElheran, 2014; and Bloom et al., 2014).[6] Given an exogenous improvement in ICT infrastructure in certain areas, the model generates three predictions. First, in response to an exogenous improvement in ICT infrastructure in a given city, firms in that city increase their division of labor. Second, in response to an exogenous improvement in ICT infrastructure in a given city, the increase in the extent of division of labor is greater for firms in more complex sectors. Third, in response to an exogenous improvement in ICT

[4]Examples of the costs include training costs of specialists (Kim, 1989), monitoring costs (Holmstrom, 1982), coordination costs (Garicano, 2000), and the time lost in combining the output of specialized workers (Becker and Murphy, 1992).

[5]I microfound the first assumption following closely the argument in Costinot (2009), as detailed in Appendix A. I microfound the second assumption in two distinct ways. First, following the Henry George Theorem (Arnott and Stiglitz, 1979), larger cities spend more on non-rival public infrastructure (such as ICT infrastructure) and this infrastructure helps lower the cost of division of labor, e.g., by reducing information or communication frictions within firms. Second, following Marshall (1890), larger cities facilitate learning, inducing workers to pursue a more specialized set of skills that reduces the cost of training. While the model remains agnostic on the precise mechanisms at work, in the empirical analysis I provide reduced-form evidence for the importance of one particular channel: ICT infrastructure. See Section 3 for more detailed discussions.

[6]Borghans and Weel (2006) find, using a sample of Dutch establishment data, that adoption of computer technology enhances communication within the firm and leads to greater worker specialization. Varian (2010) proposes that computers can mediate transactions among workers and lead to productivity gains through improvements in coordination. Using US plant-level data, McElheran (2014) documents that IT purchasing and adoption are associated with reduction in within-firm coordination costs. Lastly, Bloom et al. (2014) argue that information technologies improve plant managers' span of control, while communication technologies improve coordination efficiencies within the firm and decrease the autonomy of plant managers.

infrastructure in a set of cities, the increase in the extent of division of labor is greater for firms in larger cities.

I confront the model's predictions for how variables respond to changes in ICT infrastructure with data, by leveraging a quasi-experiment in Brazil. I exploit the expansion of broadband infrastructure as part of the Brazilian National Broadband Plan. The new ICT infrastructure was implemented gradually between 2012 and 2014, creating a quasi-experiment that allows me to identify its effects using a difference-in-differences method. To identify the impact of improved ICT infrastructure on firms' division of labor, I compare establishments in locations that received new internet infrastructure to those that did not during the gradual roll-out of the broadband infrastructure. That the alignment of the infrastructure was predetermined and implementation followed a geographically determined order reduces concerns about nonparallel trends in the outcome of interest for locations on and off the new infrastructure network. I find evidence that verifies the model's three predictions: (i) division of labor increases in areas that receive faster internet connections relative to those in other areas; and (ii) the relative increases are greater for establishments producing in more complex sectors and (iii) for establishments located in larger cities.

Finally, relying on the reduced-form evidence from the quasi-experiment and cross-sectional microdata, I bring the model to data to recover estimates of the parameters. I parameterize an extended version of the model, which incorporates the standard reduced-form agglomeration externalities in the urban literature (e.g., Allen and Arkolakis, 2014), the standard spatial sorting of firms (e.g., Gaubert, Forthcoming), imperfect spatial sorting, and a discrete set of cities. To quantify the contribution of division of labor to productivity difference across cities, I perform a counterfactual analysis in which I shut down productivity improvement through division of labor. I find that division of labor accounts for 15% of the relationship between productivity and city size—roughly comparable to the importance of natural advantage and the labor-market-based knowledge spillover estimated in previous literature.[7] I further disentangle the roles played by spatial sorting of firms and the direct effect of city size in another counterfactual experiment, in which I shut down the systematic sorting of firms. I estimate that each channel contributes approximately half of the 15% productivity advantage through division of labor.

The paper connects several strands of literature. First, it is related to studies on agglomeration externalities. The productivity advantage of larger cities has been studied extensively on the empirical front (e.g., Rosenthal and Strange, 2004; and Melo, Graham and Noland, 2009) and theoretically (e.g., Eeckhout and Kircher, 2011; Davis and Dingel, 2012; Behrens, Duranton and Robert-Nicoud, 2014; and Gaubert, Forthcoming). My theoretical framework is most closely related to the one developed by Gaubert (Forthcoming), in which sorting of firms is generated by a reduced-form assumption that more productive firms benefit more

[7] Ellison and Glaeser (1999) find that natural advantage contributes to approximately 20% of productivity gains in larger cities. Serafinelli (2015) shows that firm-to-firm worker flows explain about 10% of agglomeration advantages in higher-density areas.

from being in a larger city. My model builds on her framework by putting forth a microfounded theory for the reduced-form assumption. This microfoundation allows me to both empirically identify a specific mechanism that generates the complementarity between firm technology and city size, and to derive a set of auxiliary predictions consistent with the data on several margins. More generally, by offering a closer look at firms' internal organization, the paper proposes theoretically and identifies empirically, a previously under-explored channel that explains the productivity advantage of larger cities, further opening up the "black box" of agglomeration externalities.

My paper complements recent work by Caliendo and Rossi-Hansberg (2012) and Caliendo, Monte and Rossi-Hansberg (2015), which examines the productivity impacts of firm organization, defined by a firm's vertical hierarchical layers. I focus on a distinct yet equally important dimension of firm organization, i.e., horizontal specialization by means of division of labor. Theoretically, I build on ideas introduced by Becker and Murphy (1992), who argue that division of labor is a tradeoff between gains from worker specialization and coordination costs, and by Costinot (2009), who finds that the gains from division of labor are related to the complexity of the production process.[8] I enrich these theoretical discussions by developing a spatial equilibrium framework that links a firm's decision on division of labor to its location choice, to study the relationship between division of labor and city size and determine how firms' organization decisions contribute to spatial productivity differences.[9]

My paper also contributes to a small empirical literature on division of labor. To my knowledge, my work is the first comprehensive empirical study of division of labor within firms. Previous literature tends to focus on particular industries, such as physicians (Baumgardner, 1988) and lawyers (Garicano and Hubbard, 2009). The results of these studies support my stylized fact that division of labor increases with city size. However, these detailed case studies, despite their advantage of offering precise measurements within the relevant industries, may not be representative of the wider economy and are thus unsuitable for assessing the general equilibrium effects of division of labor on productivity. A notable exception is Duranton and Jayet (2011), who study the whole of the manufacturing sector using French census data, and find that scarce specialist occupations are overrepresented in larger cities. My dataset allows me to go beyond this by observing the extent of division of labor within firms, which motivates my fully specified model of firm behavior with underlying heterogeneity. It is important to consider the role firms play in division of labor: As Garicano (2000) convincingly argues, organizations exist, to a large extent, to solve coordination problems in

[8] In a related empirical work, Boning, Ichniowski and Shaw (2007) document, using detailed panel data on production lines in U.S. minimills, that the adoption of a more effective organization structure is strongly influenced by the complexity of the production process, which suggests the presence of such a complementarity.

[9] Chaney and Ossa (2013) extend Krugman (1979)'s "new trade model" by allowing for an explicit decision regarding firms' division of labor. They show that an exogenous increase in the aggregate number of consumers induces a deeper division of labor due to increase in the residual demand for the firm. My model differs in two ways. First, it incorporates the direct effect of city size on division of labor, i.e., two firms facing the same residual demand may adopt different extents of division of labor if they are located in cities of different sizes. Second, my theory is a full spatial equilibrium model with endogenously determined city sizes.

the presence of specialization. Incorporating heterogeneous firms is also essential to study how firm sorting affects division of labor across different cities.

Lastly, I provide new evidence on the impact of ICT infrastructure. There is growing consensus that the adoption of ICT is associated with improvements in productivity.[10] While important, these studies almost exclusively present correlation results. Significant recent progress is made by Hjort and Poulsen (2016), who exploit the gradual arrival of submarine internet cables in African coastal cities as an exogenous shock to provide direct evidence on how access to modern ICT technology affects job creation, job inequality, and income in African countries. My work focuses on the impact of ICT infrastructure at the firm level and explores a new outcome, i.e., firms' division of labor. I demonstrate how access to faster internet affects productivity by increasing division of labor within firms, thus expanding the body of evidence on the productivity impact of new technologies.

The remainder of the paper is organized as follows. Section 2 describes the data and definitions, and documents stylized facts about firms' division of labor. Section 3 develops a spatial equilibrium model with endogenous firm organization and presents descriptive evidence that is consistent with the equilibrium characteristics in the model. Section 4 details results from a quasi-experiment, which provide empirical support for the model. Section 5 summarizes the quantitative framework and the estimation process. Section 6 presents results from the counterfactual exercises. Section 7 concludes.

2 Data and Stylized Facts

In this section, I first describe the data sources and definitions used in the empirical analysis. Using the dataset constructed, I then document two new stylized facts that motivate the theoretical framework: Division of labor within firms is positively correlated with both city size and production complexity.

2.1 Data

The primary data source is the Brazilian Annual Social Information Report (*Relação Anual de Informações*, or RAIS), spanning the period from 2006 to 2014. Constructed annually by the Ministry of Labor and Employment (*Ministerio do Trabalho e Emprego*, or MTE), this administrative dataset provides a high-quality census of all establishments operating in the formal labor market (Saboia and Tolipan, 1985; De Negri et al., 2001). RAIS data contain linked employer-employee records. Importantly, both employers and employees have an incentive to accurately report relevant information: The former are liable for fines if they fail to

[10] In developing countries, Clarke and Wallsten (2006) find that a 1% increase in the number of internet users is correlated with a 3.8% increase in exports from low-income to high-income countries. Qiang and Rossotto (2009) present cross-country evidence that a 10% increase in the broadband penetration ratio is associated with a 1.38% increase in GDP per capita growth rate. Commander, Harrison and Menezes-Filho (2011) show that adoption of the internet correlates positively with firm performance in Brazil and India. See Draca, Sadun and Reenen (2009) for a review of studies on developed countries.

report, and the latter are required to provide accurate information in RAIS to receive payments for several government benefit programs. Also, the MTE conducts frequent checks on establishments across the country to verify the accuracy of information reported.

The dataset has recently been used extensively (e.g., Dix Carneiro and Kovak, forthcoming; Helpman et al., forthcoming). The scope of RAIS includes almost all formally employed workers, i.e., workers who have signed work cards that give them access to the benefits and labor protections afforded by legal employment systems. The data contain unique, anonymized, and time-invariant establishment identifiers that allow me to track establishments over time. I also use the establishment's geographic location (municipality) and sector, and worker-level information including occupation, hours and days worked, and December earnings.[11]

These data have various advantages over other datasets used in previous studies. First, RAIS is a census rather than a sample, so it is representative at a fine geographic level. Second, relative to Duranton and Jayet (2011), the matched employer-employee records available in RAIS allow me to study division of labor within establishments and, in turn, develop a theory that models establishment-level decision regarding division of labor. Third, I can analyze adjustments in establishments' division of labor in response to shocks using a difference-in-differences (DiD) method, as the data is panel in nature and available every year. This allows me to control for both observable and unobservable establishment characteristics. Fourth, there has been considerable concern about the accuracy of self-declared occupations in the population census data.[12] Worker information in RAIS, in contrast, is provided by the employer (typically the human resources department). Hence, information on worker occupation is more accurate and reliable. Fifth, RAIS data offer detailed occupation codes at the 6-digit level (the Brazilian CBO-02 codes), with a total of more than 2,500 occupation codes, and each accompanied by detailed task descriptions.[13] The richness of the data allows me to chart out, in a precise manner, an establishment's internal organization structure and construct a measure for establishment-level division of labor.

I supplement the main dataset with other types of survey data. For information on local population and land area, I use the Brazilian National Household Sample Survey (PNAD). I rely on the Brazilian Annual Industry Survey (PIA) for sector-level data on firm revenue, value-added, and the number and value of intermediate inputs. For all empirical and structural analyses, I limit the sample of firms to only tradable sectors, to be consistent with the assumptions of the model.[14]

[11] RAIS reports earnings for December and average monthly earnings during employed months in the reference year. Following Dix Carneiro and Kovak (forthcoming), I use December earnings to avoid seasonal variation or month-to-month inflation.

[12] For example, Sullivan (2009) estimated that 9% of occupation choices in the National Longitudinal Survey of Youth are misclassified.

[13] In contrast, most other matched employer-employee datasets, such as the Portuguese *Quadros de Pessoal* and French *Déclarations Annuel des Données Sociales*, only provide 4-digit occupation classifications.

[14] I define tradable sectors as agriculture, mining, and manufacturing sectors, corresponding to Brazilian Industry Codes CNAE20 01113-33295. 90% of the establishments are in tradable sectors, amounting to about 80% of the total employment.

2.2 Definitions

Division of labor is defined as the extent of worker specialization within a firm. To produce anything, a number of different tasks need to be completed. A firm organizes its production process by partitioning these tasks into subsets and assign to different workers. The more partitions there are within the production process, the fewer the number of tasks that each worker specializes in, and the greater the division of labor. Empirically, each partition can be proxied by a job (or an occupation). The more occupations there are within the firm boundary, the greater the division of labor.[15]

More precisely, I construct a measure of within-firm division of labor based on the heterogeneity of 6-digit occupation codes within an establishment. I first remove occupation codes that involve primarily managerial or supervisory tasks. Managers play a coordinating role within an organization (e.g., Bloom et al., 2014), and therefore excluding them would allow me to more accurately measure the extent of task division involved in the actual production process.[16] I then use the remaining codes to construct two measures of division of labor. The first is a simple count of the number of nonmanagerial / nonsupervisory occupation codes within an establishment.[17] I consider an alternative measure, the specialization index, which is defined as one minus the Herfindahl index across occupations within an establishment (e.g., Ciccone, 2002; Duranton and Jayet, 2011). For robustness tests, I use the more aggregate 4-digit CBO codes.[18]

I define cities by "microregions," which are formally defined geographic unit constructed by Brazilian Statistical Agency (*Instituto Brasileiro de Geografia e Estatística*, or IBGE). A microregion is a cluster of economically integrated and geographically contiguous municipalities with similar geographic and productive characteristics (IBGE, n.d.). For my analysis, I use all 558 microregions. To compare city sizes, I use a normalized measure based on the population density.[19]

Lastly, I construct two measures of product complexity at the sector level. The first uses Brazilian Input-Output data and computes the number of intermediate inputs used by each sector in producing the sector-level outputs. The intuition is that a more diverse input structure may lead to a more complex output (see, e.g., Levchenko, 2007). The second focuses on the dimension of product sophistication. Following Hausmann, Hwang and Rodrik (2007) and Wang and Wei (2010), I measure sector-level complexity using the export share of goods by G3 economies (i.e., U.S., European Union, and Japan).[20]

[15] In Section 3, I develop a theory that maps division of labor directly into the number of occupations within a firm.

[16] I identify all 6-digit CBO occupation codes that are related to supervisory or managerial functions using a machine-learning method, as explained in Online Appendix. All empirical results are robust to keeping all occupations codes.

[17] For simplicity of exposition, I henceforth refer to this measure as the number of occupations within an establishment.

[18] See Online Appendix for a more detailed discussion on construction of measures for division of labor.

[19] Density is defined by microregion population size over the geographic area of the microregion. Standard urban models typically imply that both the density and the level of city population may generate agglomeration externalities. I follow Ciccone and Hall (1996) and use density as my primary agglomeration measure. Since microregion population and density are strongly positively correlated, the choice of measure matters little for my analyses.

[20] The key insight is that due to comparative advantage, goods exported by these advanced economies tend to be more technologically sophisticated. As a result, they also tend to involve more complex production processes.

2.3 Stylized facts

Using the dataset, I document new stylized facts on division of labor within firms. I find that within sectors, there is greater division of labor within firms in larger cities; and within cities, there is greater division of labor within firms producing in more complex sectors.

Fact 1: Positive correlation between division of labor within firms and city size

I investigate the relationship between division of labor and city size using the following OLS regression:

$$\log N_j = \alpha_0 + \alpha_1 \log L_{m(j)} + \delta_{s(j)} + \mathbf{X}_{m(j)} + \varepsilon_j$$

where N_j is the number of occupations within an establishment j (i.e., the empirical measure for an establishment's division of labor), $L_{m(j)}$ is the size of city m in which establishment j is located, $\delta_{s(j)}$ is the sector fixed effect, and X_j is a set of controls.[21]

Table 1 summarizes the relationship between division of labor and city size. Column (2) shows that even after conditioning on establishment size and other controls, division of labor is strongly and positively correlated with city size. Column (3) shows the results for the subset of establishments in export-intensive sectors.[22] Column (4) considers only mono-establishment firms to account for firms' endogenous allocation of different organizational functions to different locations. Column (5) uses only data from sectors that produce homogeneous products (Foster, Haltiwanger and Syverson, 2008) to account for possible spatial variation in the diversity of establishment outputs.

As an alternative way of controlling for establishment size, I divide establishments into deciles based on their sizes and find strong positive correlations between city size and division of labor across all groups. This test would also partially address the problem of not observing informal workers within formal establishments. Based on ECINF (the Urban Informal Economy Survey), the share of informal workers is negatively correlated with firm size. The positive correlations across all deciles suggest that the result is unlikely driven by differences in informal employment across space. The results are also robust to using the alternative definition of firms' division of labor—the specialization index—as well as the more aggregate 4-digit CBO codes. See Online Appendix for regression results and more details.

This analysis assumes that within a sector, the set of tasks performed within firms is the same across all cities. However, there is a literature arguing that firm boundaries tend to be narrower in larger cities, since

[21] The controls include establishment-level controls (establishment employment sizes and skill intensities within firms) and city-level controls (state fixed effects, share of high-skilled workers, average wage, sector diversity, and the total employment of sector s in city m).

[22] These establishments tend to rely less on the local demand compared to the rest of the economy. The elasticity of division of labor with respect to city size remains high, suggesting that the mechanism through which the city size affects firms' division of labor may go beyond the size of the local market.

Dependent variable	Log no of occupations within an establishment				
	All tradable		Export intensive	Mono-estb firms	Homogeneous
	(1)	(2)	(3)	(4)	(5)
Log (city size)	.0501***	.0214***	.0219***	.0195***	.0173***
	(.0032)	(.0038)	(.0037)	(.0029)	(.0082)
Controls	No	Yes	Yes	Yes	Yes
Obs	304503	304503	115449	284592	34058
R-sq	.13	.842	.836	.853	.821

Standard errors clustered by city in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include state and sector FEs. Establishment-level controls are establishment size and skill intensity within the firm. City-level controls are share of high-skilled workers, average wage, sector diversity, and the size of local sectoral employment. Occupations are measured by 6-digit Brazilian CBO codes. Sectors are measured by 5-digit Brazilian CNAE codes. Homogeneous sectors include corrugated and solid fiber boxes, white pan bread, carbon black, roasted coffee beans, ready-mixed concrete, oak flooring, motor gasoline, block ice, processed ice, hardwood plywood, and raw cane sugar (Foster, Haltiwanger and Syverson, 2008).

Table 1: Correlation of establishment's division of labor and city size

it is easier to outsource some peripheral tasks (for example business services) when there is an abundance of such providers in the same location (Duranton and Puga, 2005); or when these providers are more efficient (Akerman and Py, 2010). This implies there could be fewer number of occupation codes inside an establishment in larger cities. To the extent that this effect is present, my estimated elasticity can be considered a lower bound of the actual value.[23]

Fact 2: Positive correlation between division of labor within firms and complexity

I next document the correlation between division of labor and sector-level product complexity, using:

$$\log N_j = \alpha_0 + \alpha_1 \log c_{s(j)} + \delta_{m(j)} + \mathbf{X}_{s(j)} + \varepsilon_j$$

where $c_{s(j)}$ is the complexity of sector s in which establishment j produces (measured by number of intermediate inputs or export share by G3 economies), $\delta_{m(j)}$ is a city fixed effect, and X_j is a set of controls.[24]

Table 2 summarizes the relationship between division of labor and complexity. In particular, Column (2) shows that within a city, division of labor is strongly and positively correlated with product complexity, i.e., firms producing in more complex sectors tend to have a greater extent of division of labor.[25]

Summary: The positive correlations documented above, though robust, cannot be interpreted as causal

[23]Though establishment boundaries are not directly observed in the data, I consider two additional checks to account for potential biases introduced by systematic variation in the boundary. In the first one, I construct a measure of the total number of distinct tasks performed within the establishment, using a machine learning method based on the task descriptions from the official CBO-02 code book. In the second one, I categorize establishments into two groups, based on their likelihood to break up their production process, which is measured at the industry level using "fragmentation index" (Fort, 2017). Reassuringly, correlation results under these further checks remain both qualitatively and quantitatively close to the baseline estimates.

[24]The controls include establishment-level controls (establishment employment sizes and skill intensities within firms) and the total employment of sector s in city m.

[25]The results are robust when I consider different subsets of firms and across both measures of complexities, and to using the alternative definition of firms' division of labor—the specialization index—as well as the more aggregate 4-digit CBO codes (see Online Appendix for more details).

Dependent variable	Log no. of occupations					
	No. of intermediate inputs			G3 export share		
	All tradable		Mono-estb firms	All tradable		Mono-estb firms
	(1)	(2)	(3)	(4)	(5)	(6)
Log (complexity)	.0423***	.0363***	.0372***	5.481***	.5388***	.632***
	(.0145)	(.0043)	(.0043)	(.5432)	(.1756)	(.1376)
Controls	No	Yes	Yes	No	Yes	Yes
Obs	304503	304503	284592	304503	304503	284592
R-sq	.035	.787	.79	.039	.787	.79

Standard errors clustered by sector in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a city FE and a 2-digit industry FE. Occupations are measured by 6-digit Brazilian CBO codes. Sectors are defined at 4-digit Brazilian CNAE codes.

Table 2: Correlation of establishment's division of labor and sector complexity

relationships. Instead, these are general equilibrium observations since both division of labor and production location are endogenous to firms. To formally investigate what drives these relationships, I develop a model in the next section.

3 Theory

The theory embeds a model of firms' internal organizations in a standard spatial sorting model with heterogeneous firms. My theory builds, in part, on the insights of Becker and Murphy (1992) and Costinot (2009). Like Becker and Murphy (1992), a firm's optimal division of labor is driven by the tradeoff between productivity gains and costs. Akin to Costinot (2009), the magnitude of these gains crucially depends on the complexity of firms' products. In my model, the costs of division of labor vary with city size. The theory's basic logic can be sketched as follows. First, given city size, a firm determines its optimal division of labor based on the firm's complexity. Second, larger cities reduce the costs of division of labor but have higher factor prices. In equilibrium, more complex firms choose greater division of labor. Since firms with greater division of labor benefit more from being in larger cities, there is, in equilibrium, positive assortative matching between firm complexity and city size.

3.1 Set-up and agent's problem

The economy consists of a continuum of homogeneous sites that may be developed into cities. The number of cities and their corresponding population sizes are endogenous. Each site is endowed with a fixed stock of housing land, which is normalized to 1. The fixed land constraint acts as the congestion force in the economy. I use L to index both the city and its size, as it is the sufficient statistic that summarizes all economic characteristics within a city.[26] The economy has a continuum of heterogeneous firms producing in

[26] I focus on the tradable sector in my model. Under the further assumption that goods are costlessly traded across space, distance between cities plays no part in the model. I make the assumption for zero trade costs largely because it is convenient

cities using local labor. City size grows with increases in local labor demand. I further assume that each firm produces only one good and that labor is the only factor in production.

There is a mass of $\bar{L}$ of agents in the economy. Agents are homogeneous, with perfect mobility across cities. Each individual is endowed with 1 unit of labor supply, which they supply inelastically. Agents consume both housing, h, and a bundle of freely traded goods, X, according to a Cobb-Douglas utility function:[27]

$$U = \left(\frac{X}{\eta}\right)^{\eta} \left(\frac{h}{1-\eta}\right)^{1-\eta}. \tag{1}$$

The bundle of tradable goods X is a Cobb-Douglas combination of goods over $s = \{1, \dots, S\}$ sectors.

$$X = \prod_{s=1}^{S} X_s^{\xi_s}, \quad \text{with} \sum_{s=1}^{S} \xi_s = 1.$$

Within a sector s, consumers choose varieties according to a CES aggregator:

$$X_s = \left[\int x_s(z)^{\frac{\sigma_s - 1}{\sigma_s}} dz\right]^{\frac{\sigma_s}{\sigma_s - 1}}, \tag{2}$$

where $\sigma_s > 1$ is the elasticity of substitution within sector s.

Given the Cobb-Douglas preference, equilibrium housing rents are given by:

$$p_h(L) = \frac{(1-\eta)w(L)L}{H} = (1-\eta)w(L)L, \tag{3}$$

where $w(L)L$ is the total income in city L, and the last equality relies on my assumption that total housing supply in each city is 1.

The spatial mobility assumption ensures that homogeneous agents' utility is equalized across space in equilibrium. The equilibrium level of utility, $\bar{U}$, is obtained by substituting Equation (3) into the utility function:

$$\bar{U} = \left[\frac{w(L)}{P}\right]^{\eta} \left[\frac{L^{-1}}{1-\eta}\right]^{1-\eta}, \tag{4}$$

where P is an aggregate price index for X.[28] Since goods are freely traded, P is same in all cities.

to derive my analytical results. In Online Appendix, I provide proof to show that all theoretical results hold with costly trade.

[27] We should think of h as representing all nontradable goods and services.

[28] P is an aggregate price index that summarizes the price indexes P_s for all tradable sectors. Formally, P is defined as

$$P = \left[\prod_{s=1}^{S} \left(\frac{P_s}{\xi_s}\right)^{-\xi_s}\right]^{-1}.$$

Additionally, I derive the equilibrium income of an agent in city L using Equation (4);

$$w(L) = \bar{w}\left((1-\eta)L\right)^{\frac{1-\eta}{\eta}}, \tag{5}$$

where $\bar{w} = \bar{U}^{1/\eta}P$ is a variable to be pinned down in general equilibrium.[29]

3.2 Firms and Production

I turn now to the production side of the economy. Firms differ exogenously in the complexity of their production technology, denoted by z. Firms choose their division of labor, production scale, and production location to maximize profits. Within a sector, firms engage in monopolistic competition, and outputs produced by firms are freely traded across space. The sectoral price index P_s is thus constant across space.

3.2.1 Production Technology

Like Smith (1776), I observe that in any firm, production of a good requires combining a collection of tasks.[30] A firm organizes its production process by partitioning these tasks into subsets of tasks and assigning them to workers. The more partitions there are, the narrower the range of tasks that each worker specializes in, the greater division of labor. In the model, a firm chooses the optimal level of division of labor, based on the complexity of its production process. I follow Costinot (2009) to interpret complexity of a production process as the total number of tasks involved in producing the firm's output. The more tasks there are, the more complex the firm's production process.

In the model, complexity of a firm has two dimensions: a sector-level parameter and a firm-specific parameter. All firms in a sector share a common sector-level complexity measure, denoted by c_s. The sector-level complexity summarizes the average complexity of the production technologies within the sector. Consider the aircraft engine manufacturing sector versus the sports shoes production sector. Producing aircraft engines clearly involves more tasks than producing sports shoes. Hence, in my model, the former has a higher c_s parameter. Within a sector, firms also differ in their production technology, denoted by z. Between a global company such as NIKE and a local Brazilian shoe factory, even though both firms produce sports shoes, the production process in NIKE likely consists of more tasks. In the model, while these two firms share a common sector complexity parameter c_s, NIKE has a higher firm-specific parameter z.[31]

[29] Given Equation (5), local housing price can be written as

$$p_h(L) = \bar{w}\left[(1-\eta)L\right]^{\frac{1}{\eta}}.$$

Under a standard model of the internal structure of a monocentric city in which commuting costs increase with population size (see, e.g., Behrens, Duranton and Robert-Nicoud 2014), $p_h(L)$ can also be interpreted as comprising both land rents and commuting costs.

[30] Smith (1776) notes that there are at least 18 tasks involved in making a pin.

[31] To derive all theoretical results, having the firm-specific parameter z is sufficient. I consider an economy with multiple sectors for three reasons. First, it allows the model to generate a richer set of results that match the salient features of Brazilian

Within sector s, firm z produces its output using the following technology:

$$Q = A(N, z, c_s)H(N, L)l, \tag{6}$$

where N denotes division of labor in firm z and l denotes the number of workers within the firm.

Firm productivity, defined as output per worker, is given by $A(N, z, c_s)H(N, L)$, which depends on the key endogenous variable N, i.e., division of labor within the firm. The first term, $A(N, z, c_s)$, characterizes individual worker productivity. $A(N, z, c_s)$ increases in N, reflecting *gains* from worker specialization. The marginal gain from N, however, depends on *complexity* of firms' production technology, z and c_s, which I will discuss below in Section 3.3.2.[32]

The second term, $H(N, L)$, denotes *costs* associated with greater division of labor. These costs can be attributed to a multitude of factors: It may be more expensive to train workers who are very specialized (Kim, 1989); hiring more specialized workers may make it harder to enforce contracts (Costinot, 2009); greater specialization may enable workers to more easily shirk responsibilities and free ride off others (Holmstrom, 1982); and it may also incur overhead for communication and coordination, and/or require workers to spend more time away from production (Garicano, 2000). The cost is increasing N and also related to city size L. I defer the discussion on assumptions adopted for $H(N, L)$ to Section 3.3.2.

3.2.2 Market structure

There is an infinite supply of potential entrants who can enter the market. Firms pay a sunk cost f_E in final good X to enter, then draw a complexity parameter z from a distribution $F(\cdot)$. Once firms discover z, they choose the size of the city in which they want to produce, the size of the firm, and the optimal division of labor.

3.2.3 The firm's problem

The firm maximizes its profit by choosing the optimal division of labor, firm size, price, and production location, given the demand and local labor costs. Given the isoelastic preferences in Equation (2), the demand schedule faced by firm z in sector s is:

$$p_s(z) = Q_s(z)^{-\frac{1}{\sigma_s}} R_s^{\frac{1}{\sigma_s}} P_s^{\frac{\sigma_s - 1}{\sigma_s}},$$

economy. Second, it facilitates the empirical analysis of the quasi-experiment. Third, it gives flexibility to the structural estimation as I can allow for sector-specific parameters.

[32] I further assume that $A(N, z, c_s)$ is increasing in z and c_s, so that productivity is higher for a more complex firm, i.e. $\frac{\partial A}{\partial z} > 0$, $\frac{\partial \log A}{\partial c_s} > 0$, and $\frac{\partial^2 \log A}{\partial c_s \partial z} \geq 0$. The first two assumptions imply that high-z and high-c_s firms will never choose to produce below their exogenously given level of complexity, whereas the last one assumes that the benefit of having a more complex firm-specific production process is (weakly) more important within more complex sectors.

where R_s denotes total sectoral revenue, P_s denotes the sectoral price index, and $Q_s(z) = \bar{L}x_s(z)$, since quantity produced equals the product of the quantity demanded by each agent and the number of agents. I can rewrite the firm's problem as follows:

$$\max_{N,l,p,L} pQ - w(L)l, \tag{7}$$

subject to:

$$p_s(z) = Q_s(z)^{-\frac{1}{\sigma_s}} R_s^{\frac{1}{\sigma_s}} P_s^{\frac{\sigma_s - 1}{\sigma_s}}, \tag{8}$$

and

$$Q_s(z) = A(N, z, c_s)H(N, L)l. \tag{9}$$

I adopt a recursive process to solve the profit maximization problem in two steps. First, I fix the location choice and compute the local labor market equilibrium with the optimal division of labor, firm size and price, taking the size of the city and local labor costs as given. Second, the firms make their location choices to maximize the optimized profit.

Consider a firm of product complexity draw z in city of size L. Given the CES preferences and the monopolistic competition, firms set constant markups over their marginal costs. For each firm z, the firm's profit can be written as a function of division of labor N and city size L,

$$\max_{N,L} \pi_s(z, L, N) = \max_{N,L} \frac{(\sigma_s - 1)^{\sigma_s - 1}}{\sigma_s^{\sigma_s}} \left(\frac{A(N, z, c_s)H(N, L)}{w(L)}\right)^{\sigma_s - 1} R_s P_s^{\sigma_s - 1}. \tag{10}$$

Based on Equation (10), product complexity, z, and city size, L, determine firms' profit as a function of division of labor N. Worker productivity, $A(N, z, c_s)$, is increasing in division of labor, N. At the same time, $H(N, L)$, goes down as workers become more specialized.

3.2.4 Optimal firm organization

The profit function given by Equation (10) is multiplicatively separable in $A(N, z, c_s)H(N, L)$. Hence, the optimal division of labor in a given city, $N_s(z, L)$, can be calculated by the following first-order condition:

$$\frac{A_N}{A} = -\frac{H_N}{H}, \tag{11}$$

where A_N and H_N denote the partial derivatives of $A(N, z, c_s)$ and $H(N, L)$ with respect to N, respectively.

In Equation (11), $\frac{A_N}{A}$ corresponds to the marginal benefit of increasing division of labor, which is equal to the additional worker productivity yielded by greater worker specialization. $-\frac{H_N}{H}$, on the other hand, corresponds to the marginal cost of increasing division of labor, and is equal to the extra units of labor lost as N increases. Equation (11) states that when N is chosen optimally, marginal gains from division of labor are equal to the marginal costs they create.

Substituting $N_s(z,L)$ into the profit function (10), I get the optimal profit of firm z in city L

$$\pi_s^*(z,L) \equiv \frac{(\sigma_s-1)^{\sigma_s-1}}{\sigma_s^{\sigma_s}} \left(\frac{A(N_s(z,L),z,c_s)H(N_s(z,L),L)}{w(L)} \right)^{\sigma_s-1} R_s P_s^{\sigma_s-1}. \tag{12}$$

Lastly, firm employment, conditional on being in a city of size L, is given by

$$l_s(z,L) = (\sigma_s-1)\frac{\pi_s^*(z,L)}{w(L)}. \tag{13}$$

3.3 Spatial equilibrium

I characterize spatial equilibrium in this section. I show that under two simple assumptions, there is positive assortative matching between firms' complexity draw and city size. In equilibrium, the positive assortative matching generates the positive correlation between division of labor and city size.

3.3.1 Definition

Homogeneous workers are indifferent across locations, while firms choose their locations optimally based on their complexity draws. I choose the reference level of wages $\bar{w}$ defined in Equation (5) as the numeraire. An equilibrium for a population $\bar{L}$ and firm with product distribution $f_s(z)$, for $s \in \{1,\dots,S\}$, in a set of locations $\mathcal{L}$ is characterized by a set of prices $\{w(L), p_H(L)\}$; a city-size distribution $f_L(\cdot)$; an optimal division of labor function $N_s(z)$; a location matching function $L_s(z)$; an employment function $l_s(z)$; a production function $Q_s(z)$; and a set of price index P_s and mass of firms M_s for $s \in \{1,\dots,S\}$ such that:

1. Workers maximize their utilities according to Equation (1), given $w(L), p_H(L)$ and P_s.

2. Worker's utility is equalized across all cities.

3. The housing market clears according to Equation (3).

4. Firms maximize profits according to Equation (12), given $w(L)$ and P_s.

5. For $s = 1,\dots,S$, aggregate sectoral production must be equal to the sum of individual firms' production:

$$1 = \frac{(\sigma_s-1)^{\sigma_s-1}}{\sigma_s^{\sigma_s}} M_s P_s^{\sigma_s-1} \int_z \left(\frac{A(N_s(z),z,c_s)H(N_s(z,L),L)}{[(1-\eta)L_s(z)]^{\frac{1-\eta}{\eta}}} \right)^{\sigma_s-1} dF_s(z). \tag{14}$$

6. Firms earn zero profits. Using the free-entry condition, the following condition must be met, for

$s = 1, \ldots, S$:

$$f_E P = \frac{(\sigma_s - 1)^{\sigma_s - 1}}{\sigma_s^{\sigma_s}} R_s P_s^{\sigma_s - 1} \int_z \left(\frac{A(N_s(z), z, c_s) H(N_s(z, L), L)}{[(1-\eta) L_s(z)]^{\frac{1-\eta}{\eta}}} \right)^{\sigma_s - 1} dF_s(z). \tag{15}$$

7. The national labor market clears:

$$\bar{L} = \frac{(\sigma_s - 1)^{\sigma_s}}{\sigma_s^{\sigma_s}} M_s R_s P_s^{\sigma_s - 1} \int_z \frac{[A(N_s(z), z, c_s) H(N_s(z, L), L)]^{\sigma_s - 1}}{[(1-\eta) L_s(z)]^{\frac{1-\eta}{\eta}\sigma_s}} dF_s(z). \tag{16}$$

8. The local labor markets clear:

$$\int_{L_0}^{L} n f_L(n) dn = \sum_{s=1}^{S} M_s \int_0^{\infty} \mathbf{1}_s(L, z) l_s(z) dF_s(z) \quad \forall L > L_0, \tag{17}$$

where $L_0 \equiv \inf(\mathcal{L})$, i.e., the smallest city size in equilibrium, and $\mathbf{1}_s(z, L) = 1$ if firm z in sector s is in city L, and 0 otherwise.

Finally, note that by Walras's Law, the goods market clears.

3.3.2 Model assumptions

To fully analyze the characteristics of the equilibrium, I make the following assumptions:

Assumption 1 *$A(N, z, c_s)$ is twice-differentiable, and strictly log-supermodular in firms' complexity z and division of labor N, and log-supermodular in sector-level complexity c_s and division of labor N, i.e.,*

$$\frac{\partial^2 \log A(N, z, c_s)}{\partial N \partial z} > 0; \quad \frac{\partial^2 \log A(N, z, c_s)}{\partial N \partial c_s} \geq 0.$$

Assumption 2 *$H(N, L)$ is twice-differentiable, and strictly log-supermodular in city size L and firms' division of labor, i.e.,*

$$\frac{\partial^2 \log H(N, L)}{\partial N \partial L} > 0.$$

Assumption 1 states that there is complementarity between complexity z (and c_s) and division of labor N, i.e., a more complex production process benefits more from greater division of labor. In Appendix A, I present one microfounded production process that generates these results. As in Costinot (2009), production requires completing a continuum of complementary tasks. More tasks are involved in producing more complex (high-z or high-c_s) products. Before performing any task, workers must spend a fixed amount of training time learning it. The more complex a good is—that is, the more tasks involved in producing the good—the longer it takes to learn how to perform all tasks. To minimize training time, firms assign each worker to perform a specialized set of tasks. I refer to each set of tasks as an *occupation*. More occupations within a firm imply that there are fewer number of tasks within each occupation, i.e., greater division of labor. Since more complex products require more training time, the gains from worker specialization are higher for more complex firms. In what follows, I remain agnostic on the sources of the productivity benefit through division of labor and its specific functional form. This allows me to highlight the generic features of an economy with such complementarity.

Assumption 2 states that there is complementarity between city size L and division of labor N, i.e., larger cities lower costs associated with greater division of labor. I hypothesize that one channel that generates this is through provision of better ICT infrastructure in larger cities. Modern ICT technologies, such as fast internet, can facilitate greater division of labor within firms through a number of channels, e.g., by improving communications efficiencies, enhancing information storage and sharing, or allowing firms to employ more capable software applications (e.g., Borghans and Weel, 2006; Varian, 2010; McElheran, 2014; and Bloom et al., 2014). In equilibrium, larger cities, with their larger tax bases, provide better local infrastructure including ICT infrastructure. Therefore, larger cities foster greater division of labor, creating the complementarity between N and city size L. I provide a formal test of this hypothesis in Section 4.[33]

I highlight several noteworthy points before proceeding. First, it is important to note that while I could include all other cases of $A(N, z, c_s)$ and $H(N, L)$ in the current discussion, I choose to focus on the empirically relevant cases specified above to avoid a cumbersome taxonomy. Under the current set of assumptions, the model generates a positive correlation between division of labor and city size, as shown in Section 2. Additionally, I do not impose these restrictions in the estimation of the model in Section 5. Instead, I let the data to inform me the appropriate choices for the parameters' values.

Second, in Section 4, I present causal empirical evidence that supports the log-supermodularity assumptions between N and z (and c_s), and between N and L. I do so by focusing on the specific channel mentioned above, i.e., better ICT infrastructure in larger cities facilitates greater division of labor. The model generates specific predictions for changes in firms' division of labor in response to an exogenous improvement in ICT infrastructure. I test these predictions using a quasi-experiment in Brazil.

Third, the baseline model adopts a minimum set of assumptions necessary to obtain the general equilibrium outcome, by which firms in larger cities have greater division of labor. This generates productivity advantage for larger cities through a specific channel, i.e., their ability to foster greater worker specialization. In estimating the model, I include a term that summarizes all other channels that might also increase firm

[33] While I propose this specific channel that generates the complementarity between N and L, the model is general enough not to preclude the existence of other sources. In Appendix B, I propose another microfoundation for the complementarity between city size and division of labor. Workers acquire both extensive and intensive human capital, which correspond to the breadth and depth of their knowledge set, respectively. Knowledge acquisition is costly. Larger cities have a comparative advantage in acquiring intensive human capital. As a result, firms with a greater division of labor—in which the requirement for extensive knowledge set is lower—would benefit more by being in a larger city, leading to the complementarity between N and L when the level of intensive human capital is optimally chosen. More details are discussed in Appendix B.

productivity in larger cities. By separately identifying the two channels, I can calibrate the model and investigate the importance of division of labor in affecting productivity differences across cities. I discuss this in detail in Section 5.

3.3.3 Characteristics of the profit function

In my theoretical framework, there is complementarity between complexity and division of labor, i.e., the productivity benefit from worker specialization is higher for more complex products. There is also complementarity between city size and division of labor, i.e., firms with greater division of labor benefit more from being in larger cities. Combining these assumptions generates the following properties for firms' profit function and division of labor:

Lemma 1 *Suppose that Assumptions 1 and 2 both hold, firms' profit function, $\pi_s(z, L, N)$, displays log-supermodularity in (z, L, N).*

Lemma 2 *Denoted by $N_s(z, L) = \arg\max_N \pi_s(z, L, N)$, the optimal division of labor given z and L increases in (z, L).*

Division of labor depends on the trade-off between gains and costs of specialization. A larger city lowers the costs at the margin, and thus division of labor increases. Similarly, as complexity increases, marginal gains shift up, which increases division of labor. Using a classic result in monotone comparative statics (Topkis, 1978), since the profit function $\pi_s(z, L, N)$ is log-supermodular in (z, L, N), once the firm solves for its optimal division of labor, $N_s(z, L)$, the profit function $\pi_s^*(z, L)$ displays log-supermodularity in (z, L).

Lemma 3 *Denoted by $\pi_s^*(z, L) = \max_N \pi_s(z, L, N)$, the optimal firm profit given z and L is log-supermodular in (z, L), if both Assumptions 1 and 2 hold.*

3.3.4 Equilibrium systems of cities

Following the standard literature (e.g., Henderson and Becker, 2000; Behrens, Duranton and Robert-Nicoud, 2014), I assume that cities emerge endogenously as a result of "self-organization." A new city opens up when there is incentive for firms and/or workers to do so. This happens when there exists a set of firms and workers that would be better off with their choices of the city size. Cities are therefore the outcome of the mutually compatible optimal choices of a continuum of firms and workers. Recall that the optimal profit function of firm z in city L is

$$\pi_s^*(z, L) = \frac{(\sigma_s - 1)^{\sigma_s - 1}}{\sigma_s^{\sigma_s}} \left(\frac{A(N_s(z, L), z, c_s) H(N_s(z, L), L)}{w(L)} \right)^{\sigma_s - 1} R_s P_s^{\sigma_s - 1}.$$

Lemma 1 implies that the profit function shown in Equation (12) is log-supermodular in (z, L), suggesting that more complex firms benefit more from being located in larger cities. However, given symmetric fundamentals, this does not preclude the existence of a symmetric equilibrium, in which all types of firms are equally represented in all cities. I show in Online Appendix that such an equilibrium is stable only if the gains from worker specialization are too small to cause agglomeration. When worker specialization is sufficiently rewarding, a small perturbation in city size would push the symmetric equilibria into a heterogeneous equilibrium.

Symmetric equilibria are both empirically counterfactual and theoretically not very illuminating. Henceforth, I focus on heterogeneous equilibria. Given its complexity draw z, the firm's problem is to choose L to optimize its profit. Using Equation (12), the first order condition with respect to L is therefore:

$$\frac{H_L}{H} = \frac{1-\eta}{\eta}\frac{1}{L}, \tag{18}$$

where $\frac{H_L}{H} = \frac{\partial H(N,L)}{\partial L}$.

In Equation (18), $\frac{H_L}{H}$ corresponds to the marginal benefit of being in a larger city. It equates to the additional productivity advantage generated by lower coordination costs there; $\frac{1-\eta}{\eta}\frac{1}{L}$ corresponds to the marginal cost of being in a larger city. It is equal to the extra costs due to more expensive labor prices. When production location is optimally chosen, the marginal gains from being in a larger city are equal to the marginal costs.

Under regularity conditions, there is a unique profit-maximizing city size for a firm of type z with c_s. Define the solution to Equation (18) as

$$L_s^*(z) = \arg\max_{L \geq 0} \pi_s^*(z, L). \tag{19}$$

Under the self-organization assumption of cities, the set of city sizes $\mathcal{L}$ in heterogeneous equilibria is necessarily the outcome of the mutually compatible optimal choices of the continuum of individuals and firms (see, e.g., Henderson and Becker, 2000 and Behrens, Duranton and Robert-Nicoud, 2014). Assume that for some firm z, no city size of $L_s^*(z)$ exists; then there is a profitable deviation for these firms to coordinate and open up this city on an unoccupied site. It will attract the corresponding workers by offering them a wage marginally higher than $w(L_s^*(z))$. The number of such cities adjusts so that each city has the right size in equilibrium. Therefore, in a heterogeneous equilibrium, the set of city sizes available in equilibrium, $\mathcal{L}$, is the union of the sector-by-sector intervals of the optimal set of city sizes for firm distribution $f_s(z)$. Given

$\mathcal{L}$, the optimal location choice for each firm z in sector s is defined by the following matching function:

$$L_s(z) = \arg\max_{L \in \mathcal{L}} \pi_s^*(z, L). \tag{20}$$

Using the definition in Equation (20) and Lemma 3, I can invoke a classic theorem in monotone comparative statics (Topkis, 1978) and obtain the following key theoretical result.

Proposition 4 *Suppose that Assumptions 1 and 2 hold. In the heterogeneous equilibrium, within a sector, high-z firms sort into larger cities. More formally, given c_s, the matching function is increasing in z, or* $L_s'(z) > 0$.

The intuition for Proposition 4 is straightforward. Larger cities have higher housing prices due to congestion, so workers require higher wages in these locations. Larger cities attract firms because of the productivity advantage brought about by the lower costs of division of labor. More complex firms benefit more from being in larger cities. In equilibrium, these firms are willing to pay more to be in a larger city, thus outbidding less complex firms. There is therefore spatial sorting for firms, which supports the equilibrium differences in the extent of worker specialization.

In Online Appendix, I further detail the properties of the heterogeneous spatial equilibrium. I prove the existence of the city-size distribution $f_L(\cdot)$, and verify that $f_L(\cdot)$ is unique and stable.

3.4 Characterizing spatial equilibrium

In the heterogeneous spatial equilibrium, division of labor $N_s(z)$, profit $\pi_s(z)$, revenue $r_s(z)$, and size $l_s(z)$ are all determined by the matching function $L_s(z)$. The strict sorting of z within a sector generates the strict sorting of firm profits and revenue. I denote the equilibrium variables using the following expressions:

$$N_s(z) = N_s(z, L_s(z)), \tag{21}$$

$$\pi_s(z) = \frac{(\sigma_s - 1)^{\sigma_s - 1}}{\sigma_s^{\sigma_s}} \left(\frac{A(N_s(z), z, c_s) H(N_s(z), L_s(z))}{w(L_s(z))} \right)^{\sigma_s - 1} P_s^{\sigma_s - 1} R_s, \tag{22}$$

$$r_s(z) = \sigma_s \pi_s(z), \tag{23}$$

$$l_s(z) = \frac{\pi_s(z)}{(\sigma_s - 1) w(L_s(z))}. \tag{24}$$

3.4.1 Within-sector characterization

Given the results in Proposition 4, firm-level observables also exhibit complementaries between firm complexity and city size, as stated in the following result:

Proposition 5 *In equilibrium, within each sector, firms' division of labor, revenue, and profit all increase with city size. More formally, consider two firms z and z' within sector s. If $L_s(z) > L_s(z')$, then $N_s(z) > N_s(z')$, $\pi_s(z) > \pi_s(z')$, $r_s(z) > r_s(z')$, and $w_s(z) > w_s(z')$.*

In equilibrium, within a sector, high-z firms sort into larger cities. This generates the motivating fact presented in Section 2, i.e., firms' division of labor is greater in larger cities. Through the lens of my model, I show how the correlation is achieved through two channels. First, firms in larger cities produce more complex products and have greater division of labor. Second, larger cities facilitate greater division of labor for all firms by lowering the costs of division of labor. Furthermore, unlike previous literature, my model does not assume any direct relationship between firm characteristics and city size. It nonetheless predicts that firms located in larger cities are bigger (in revenue), consistent with well-known empirical regularity on spatial distribution of firms. This equilibrium result is consistent with well-known features of firm-size distribution across cities, documented by past literature (see, e.g., Glaeser and Maré, 2001; Glaeser and Gottlieb, 2009) and actual firm-size distribution in Brazil (see Online Appendix).

3.4.2 Cross-sector characterization

Previous theories in this literature (see, e.g., Abdel-Rahman and Anas, 2004; Helsley and Strange, 2014) overwhelmingly describe polarized sectoral compositions within cities.[34] In contrast, my model offers a theory to explain how firms from different sectors can coexist in equilibrium within each type of city size. Holding city sizes fixed, a variety of sectors may be present in some types of cities at equilibrium. This is possible because the matching function $L_s(z)$ varies by sector, making it possible for firms with different complexity draws to choose the same optimal city size.

Using this result, I derive pattern of sectoral geographic distribution across cities. I define the geographic distribution of firms in a sector as the probability that a firm from the sector is in a city of smaller than L_c, that is,

$$\tilde{F}(L_c; c_s) = Pr(\text{firm from sector } s \text{ is in a city of size smaller than } L_c).$$

Proposition 6 *In equilibrium, all else equal, the geographic distribution $\tilde{F}_s$ of a higher c_s sector first-order stochastically dominates that of firms in a lower $c_{s'}$ sector.*

From Proposition 4, within a sector, the matching function is always increasing in z. However, the marginal increase of the matching function with respect to z is sector-specific and determined by complexity c_s. In more complex sectors, firms benefit more from being in a larger city, pushing the matching function up for all firms. All else equal, a greater share of firms in more complex sectors locates in larger cities. In

[34]These include *specialized cities* that have only one tradable sector and *perfectly diversified cities* that have all of the tradable sectors. Two notable exceptions are Davis and Dingel (2014) and Gaubert (Forthcoming).

Online Appendix, I present evidence that this equilibrium result is consistent with geographic distribution of sectors in Brazil.

3.4.3 Impact of ICT infrastructure improvement

All equilibrium properties presented so far rely on the validity of Assumptions 1 and 2, i.e., the complementarity between complexity z (and c_s) and division of labor N; and the complementarity between city size L and division of labor N, possibly through better ICT infrastructure. I now derive a set of predictions that I can bring to data to test my assumptions. In particular, the model makes predictions for changes in firms' division of labor in response to an exogenous improvement in ICT infrastructure.

Proposition 7 *In equilibrium, an improvement in ICT infrastructure increases firms' division of labor. The increases are larger for firms in more complex (or high-c_s) sectors, and for firms located in bigger cities.*

A shock to ICT infrastructure reduces costs of worker specialization at the margin, therefore increasing division of labor for all firms experiencing the change. Due to the complementarity between complexity and division of labor, the reduction in marginal costs benefits more complex (i.e., high-c_s and high-z firms) more. Therefore, the increases are larger for high-c_s and high-z firms. Furthermore, since high-z firms sort into larger cities, the model also predicts that affected firms in larger cities would increase their division of labor to a greater extent. In Section 4, I test the model predictions formally using a quasi-experiment in Brazil.[35]

4 Empirical support

In this section, I use Brazilian micro-level data to validate the theoretical predictions. The theoretical framework presented in Section 3 makes two key assumptions, i.e., the log-supermodularities between N and z (and c_s) and between N and L. Given these assumptions, the model generates the three predictions in Proposition 7: In response to an improvement in ICT infrastructure, (i) all firms affected would increase their division of labor, and the increases are greater for the firms (ii) in more complex sectors and (iii) in larger cities. In practice, many factors potentially affect firms' decisions on division of labor. To establish the causal impact of better ICT infrastructure on division of labor, I need a plausibly exogenous variation in the ICT infrastructure that is unrelated to firms' division of labor. To do so, I rely on a quasi-experiment in Brazil from the National Broadband Plan (*Programa Nacional de Banda Larga*, PNBL henceforth) to identify the effects of better ICT infrastructure on division of labor within establishments, and examine whether there are heterogeneities in the treatment effects in accordance with the predictions in Proposition 7.

[35] The exogenous improvement in ICT infrastructure also affects other firm-level variables such as firm profit and revenue, as well as general equilibrium outcomes such as city sizes. I discuss these additional results in Online Appendix.

4.1 Additional data and sample

I assemble a set of geo-coded data to assess the impact of the new policy that expands broadband accessibility in Brazil. I download the alignment of existing broadband networks from the Brazilian National Telecommunications Agency (*Agencia Nacional de Telecommunicacoes*, or Anatel). Data on the new broadband network are collected from a number of decentralized sources, including the Brazilian National Teaching and Research Network (*Rede Nacional de Ensino E Pesquisa*), press releases and annual reports from the companies contracted to implement the relevant infrastructure (including Telebras, Oi, Vivo, and Nextel). Information on municipality boundaries is obtained from IBGE. Locations of the submarine cable landing points are obtained from TeleGeography.[36] I geo-code all the data into *shp* files, and process them using QGIS to construct a consistent dataset for the quasi-experiment. The most detailed geographic information I observe for establishments is at the municipality level. I thus measure the distance between establishments and the new broadband network, using the centroids of the municipalities in which the establishments are located. Both the centroids and the nearest distance are computed by QGIS using WGS 84 Projection. Following conventional literature (e.g., Banerjee, Duflo and Qian, 2012), I use geographic distance measured in kilometers rather than travel distance.

In testing the model predictions, I use a balanced panel of establishments for the period 2006 to 2014. To investigate the interaction of the new infrastructure with city size and sector complexity, I remove those establishments that relocate or change their sector classifications during the study period. This leaves 86,344 establishments over 9 years, or 777,096 establishment-year observations, for the empirical analysis.

4.2 Background

In Brazil, the availability of broadband access closely reflects the country's wide variation in city size, as illustrated in Figure 1.[37]

This uneven distribution of broadband access is a direct result of lack of infrastructure for private internet providers in remote and low-density areas. Before 2010, the government played a very limited role in broadband provision, leaving private operators to provide broadband infrastructure where they find it profitable to do so (Jensen, 2011; Knight, 2016).[38] The prohibitively high cost of installing new broadband backbones in remote and low-density areas had prevented more even distribution of broadband availability. As a result, smaller cities of Brazil had no access to fast internet connection. To address this problem, the

[36] Data can be downloaded from the following web sources: http://www.anatel.gov.br/dados/2015-02-04-18-36-10; https://www.rnp.br/en/search?words=rua&begin=1681; http://www.telebras.com.br; http://www.oi.com.br; https://www.vivo.com.br; http://www.nextel.com.br; http://www.ibge.gov.br/english/geociencias/default_prod.shtm; and https://www.submarinecablemap.com.

[37] According to the 2010 Census Survey, fixed broadband penetration rate was 11% in Sao Paolo but only 1.5% in the low-density northeastern region. The correlation between city size and broadband penetration ratio was 0.79 in 2010.

[38] This is unlike other developing countries in which national backbones are typically built by a national state-owned telecom (see, e.g., Hjort and Poulsen, 2016).

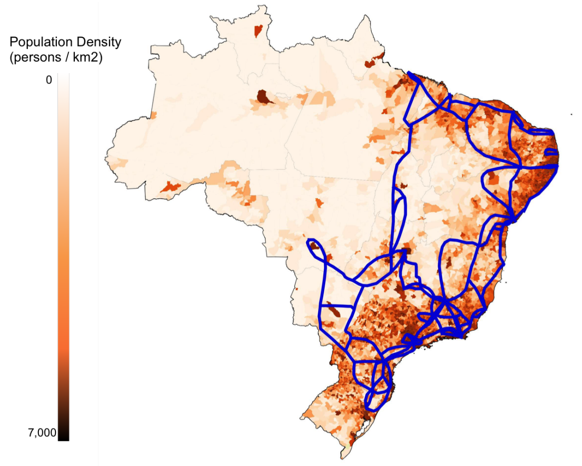


Figure 1: Broadband backbone and population density in 2010

federal government launched the largest ICT infrastructure project in 2010, i.e., PNBL.

The key objective of PNBL is to provide broadband access in poorly served areas, to trigger economic development and reduce regional inequalities (Knight, 2016). With a budget of \$600mil USD a year for four years, by 2014 the PNBL expanded broadband coverage from 681 to 2,930 municipalities; the increase amounted to 40% of the total population. I focus on a major initiative of PNBL that builds new national backbones extending to the remote areas of Brazil.[39] Between 2012 and 2014, PNBL added 48,000 km of new broadband backbone. Table 4 in Appendix C compares establishment characteristics between control and treatment groups. On average, treated establishments have greater division of labor (by both measures) and a higher share of managers, and are larger in size.[40]

4.3 Empirical strategy

The first empirical test is to investigate the relationship between division of labor within establishments in a time period and whether the establishments are connected to broadband backbone cables. I run:

$$\log N_{jt} = \alpha + \beta Backbone_{jt} + \delta_j + \delta_t + \varepsilon_{jt} \tag{25}$$

where $\log N_{jt}$ is the measured division of labor within establishment j at time t. $Backbone_{jt}$ is a dummy variable equal to one if establishment j is "connected" to the new backbone added in year t. All specifications

[39]"Backbones" are national trunk infrastructure that brings traffic from international submarine cables in coastal regions to inland parts of the country. Backbones consist of high-capacity fiber optic cables. See Online Appendix for more discussion.

[40]Even though a key policy objective for the PNBL is to expand the broadband network into the smaller, less developed areas in Brazil, the backbone infrastructure has to originate from the submarine cable landing points along the coast, and tend to pass through major cities before reaching the smaller cities. See Online Appendix for more extensive discussions on the background and details of the PNBL.

Figure 2: New broadband backbones implemented as part of PNBL: 2012-2014

include an establishment fixed effect, δ_j, that controls for any time-invariant differences across establishments, and a year fixed effect, δ_t, that controls for any establishment-invariant shocks to division of labor. Standard errors are clustered at the municipality level.[41] The key coefficient of interest here is β, which measures the effect of new broadband availability on division of labor within establishments. The model predicts that $\beta > 0$.

Following Hjort and Poulsen (2016), I determine whether an establishment is "connected" to broadband internet based on its geographic distance to the nearest backbone cable. From a technical perspective, connectivity decreases exponentially as one moves further away from the backbone network (Banerji and Chowdhury, 2013). Since I lack information on the middle and last-mile infrastructure, I cannot determine the actual adoption of broadband internet at the establishment level. Instead, I use its distance to the nearest backbone network to assess the feasibility that an establishment is connected to the backbone network.[42] The range that makes connecting to a broadband backbone cable feasible is between 100 km to 400 km. For baseline analysis, I define a location as connected to the new backbone if the distance to the nearest backbone cable is less than 250 km. I vary the radius for robustness tests.

The model also makes predictions regarding heterogeneities in the treatment effects, as stated in Proposition 7. Specifically, the impacts of the new ICT infrastructure are larger for establishments located in larger cities relative to smaller cities, and for establishments that produce in more complex sectors relative to less complex sectors. I test these predictions using Equations (26) and (27). The model predicts that $\gamma > 0$ and

[41] The results are also robust to using Conley standard errors to account for possible spatial correlations across locations.

[42] Essentially, I am defining an "intent to treat" variable, instead of the actual treatment. The estimate for β is, therefore, a lower bound of the actual effect of a faster internet connection on firms' division of labor. At the same time, using *intent to treat* also addresses the potential endogeneity in firms' decision to adopt new communications technologies.

$\omega > 0$.

$$\log N_{jt} = \alpha + \beta Backbone_{jt} + \gamma Backbone_{jt} \times \log L_{c(j),t_0} + \delta_j + \delta_t + \varepsilon_{jt}, \quad (26)$$

$$\log N_{jt} = \alpha + \beta Backbone_{jt} + \omega Backbone_{jt} \times \log c_{s(j),t_0} + \delta_j + \delta_t + \varepsilon_{jt}, \quad (27)$$

where $\log L_{c(j),t_0}$ is the size of the city c in which establishment j is located and $\log c_{c(j),t_0}$ denotes the complexity of sector s that establishment j produces in. I use both measures of sector-level complexity for the regressions.[43]

The identifying assumption is that establishments close to and farther away from new broadband backbones were on parallel trends in the outcome of interest prior to the completion of the new backbones, and did not experience systematically different idiosyncratic shocks after the new backbones arrived. Figure 3 plots the paths of the number of occupations within establishments in the treated and control groups before and after the completion of backbone cable in 2012. This enables me to inspect how the gap between the treated areas and control areas evolve after the new backbone cables arrive. More importantly, the plot allows me to check whether the identifying assumption of parallel trends holds. Indeed, while the average number of occupations within establishments is always the higher in the treated areas, shapes of the two graphs are virtually identical. The two lines seem to diverge after 2011, suggesting an increase in division of labor after the arrival of new broadband connections.[44] In Table 19 of Appendix C, I formally test the parallel-trends assumption by including two lead variables, which are two indicator functions taking the value of 1 in $t-2$ and $t-1$, respectively, if an establishment receives the treatment in t, and 0 otherwise. Coefficients on the lead variables are negative and insignificant, which supports the assumption of parallel trends.

Additionally, Figure 2, which shows the new broadband backbones that had been introduced at various times during the data period, illustrates three important aspects of the identifying variation I exploit. First, the new backbones were completed throughout the period I consider and were connected to different municipalities in time. This means that my DiD approach is dynamic in that I compare establishments in the treated and control groups across many points in time rather than on a single date. Second, alignment of the backbones was announced in 2010 and followed other infrastructures that had existed long before 2010, making it harder for policymakers to align the broadband cables in anticipation of economic changes in certain areas. Third, the order in which municipalities are connected is approximately geographically determined, according to their distances to the submarine cable landing points along the coast, as illustrated in Figure 2. It is thus a priori unlikely that the availability of the new backbones across different municipalities correlates with the temporal variation in the extents of firms' division of labor of areas on and off the new backbone cables in Brazil.

[43] I also include a specification with both interaction terms incorporated in a single regression. The specification and corresponding results are shown in Appendix C.

[44] Figure 4 in Appendix C shows the pre-trend graph for the specialization index. The two figures are very similar qualitatively.

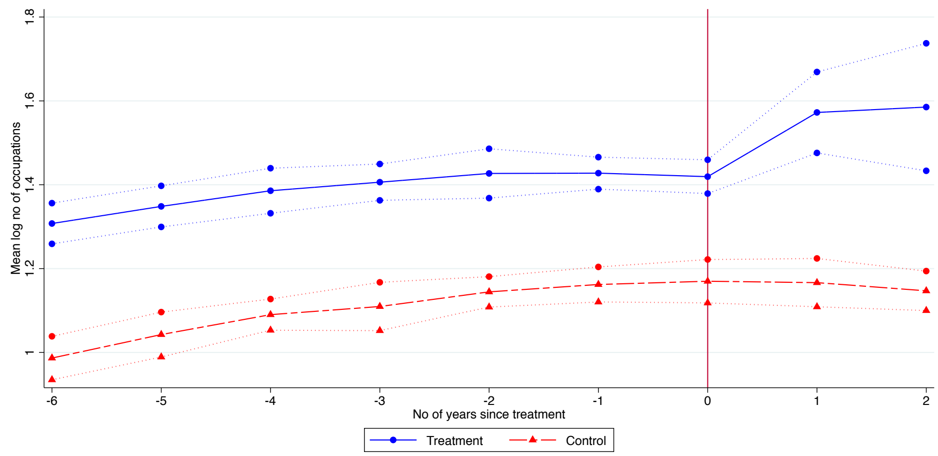


Figure 3: Log number of occs in treated versus control groups in Brazil
Dotted lines represent the 95% confidence bands.

4.4 Results

Table 3 reports the main findings: the estimated effects of new ICT infrastructure on establishment-level variables. Panel A shows from the baseline regression. Column (1) shows that establishments receiving fast internet access increase their number of occupations by 1.3 percentage point relative to other areas, whereas Column (3) shows that the specialization indices within these establishments increase by a relative value of 0.09. Columns (2) and (6) show the results for Equation (26).

Consistent with model predictions, the impacts of new ICT infrastructure are significantly greater for establishments located in larger cities. The estimated heterogeneity is substantial. A 1 percent increase in city size increases the estimated effects of new broadband connection by 0.8 percentage point and 0.01 when division of labor is measured by the number of occupation codes and the specialization index, respectively.

Next, I move on to the results for Equation (27). As explained in Section 2.2, I adopt two alternative measures to proxy sector-level complexity, i.e., the number of intermediate inputs to produce the sector output and the export share of G3 economies. Columns (3), (4), (7), and (8) illustrate results that are consistent with the model prediction that the impacts of new ICT infrastructure are greater for establishments that produce in more complex sectors.

Alternative interpretation and robustness checks

In my theory, improvement in ICT infrastructure increases firms' division of labor. Since the extent of worker specialization within firms is not directly observed, I use occupation codes to proxy firms' division of labor. I discuss alternative interpretations of the results and describe the tests in place to ensure the validity of my interpretation.

When the new broadband connection is introduced, establishments that adopt the new technology may need to hire new employees to work on IT-related jobs. If these occupations did not exist within the establishment before, this would lead to a mechanical increase in the number of occupations without changing division of labor within the establishment. To address this problem, I remove all IT-related occupations from the analysis, and re-estimate Equations (25), (26), and (27).[45] As shown in Panel B of Table 3, results are both qualitatively and quantitatively similar to the baseline results.

Faster internet may change the boundary of an establishment. If this happens, the increase in the number of occupation codes within an establishment would reflect an expansion of its boundary—for example, addition of a new department or a new product—instead of a greater extent of division of labor. Since I do not have the data for establishment-level product varieties or outsourcing decisions, I cannot test the alternative mechanisms directly. However, existing literature shows that modern communication technology is typically associated with a shrinkage in the establishment's boundary.[46] To the extent that this is true, my estimate presents a lower bound of the true effect of broadband connectivity on division of labor. I also derive a test to assess the possibility of changes in the establishment's boundary. To do so, I remove all occupation codes belonging to occupation categories that did not exist before the policy and re-estimate Equations (25), (26), and (27).[47] As shown in Panel C of Table 3, results are again similar to baseline results.

Additionally, I perform a comprehensive set of robustness tests. I show that my results are robust to varying the radius around the backbone network used to define connectivity status; to separating high and low-skilled occupations; to including only mono-establishment firms; to only including eventually-treated areas; to excluding municipalities very near or far from the backbone network from the sample; to excluding terminal locations along the new backbones; to excluding locations very close to submarine cable landing points; to excluding establishments already connected to the broadband network before PNBL; to excluding establishments located in rural areas or in very large cities; to removing firms in export-intensive sectors; and to controlling for location-specific linear trends in the outcomes. I also show that the p-values of the estimates are similar if I use a nonparametric permutation test for inferences. A detailed discussion of the robustness tests and results can be found in Appendix C.

In the empirical exercise, I focus on the total effect of the new broadband connection on firms' division of labor, without specifying the channels through which the faster internet can affect worker specialization within firms. Through my interactions with Brazilian establishments, the biggest changes after the advent

[45] IT-related occupations correspond to CBO codes 212205, 212210, 212215, 212305, 212310, 212315, 212320, 212405, 212410, 212420, 313220, 313305, 313310, 313320, 317205 and 317210. See http://www.mtecbo.gov.br/cbosite/pages/pesquisas/BuscaPorCodigo.jsf for more details on occupation codes.

[46] For example, Fort (2017) finds that communication technology lowers coordination costs, leading to more firm outsourcing or fragmentation.

[47] An occupation category is defined by the 3-digit CBO code. The assumption for this test is that addition or removal of occupation categories corresponds to changes in the boundary of an establishment.

of faster internet are two folds: (1) easier sharing of work output; and (2) easier monitoring of workers. With faster internet, file-sharing software like *Dropbox* and *Github* is now feasible to be implemented within-establishments. At the same time, managers can now better monitor and dedicate tasks to workers, often aided by HR software like *SAP*. Based on the anecdotal evidence, I hypothesize that broadband internet facilitates worker specialization through reduction in within-establishment coordination costs. To test this hypothesis, I present a supplementary test, which investigates changes in the share of managers within establishments. Managers play a coordinating role within an organization. Studies of the internal organization of firms confirm that a reduction in coordination costs within a firm would lead to greater centralization in the management structure—i.e., the share of managers would go down (see, e.g., Bloom et al., 2014 and McElheran, 2014). In Online Appendix, I show that an improvement in internet connectivity reduces the share of managers within establishments, consistent with a reduction in coordination frictions.

In sum, it appears that firms underwent organizational changes in response to improvements in ICT infrastructure. Workers become more specialized in areas that are now connected to fast internet, indicating that there is complementarity between division of labor and better ICT infrastructure in firms' production function. Additionally, the increases are higher for more complex firms and for firms in bigger cities, which are consistent model assumptions that there are complementarities between firms' division of labor and complexity, and between firms' division of labor and city size. These empirical results validate the theoretical predictions. In the next section, I use the reduced-form estimates to guide the structural estimation.

5 Estimation of the model

In this section, I structurally estimate the model, guided by results from the quasi-experiment in the pervious section. I follow a two-step procedure: In the first step, I estimate three sets of parameters that can be inferred directly from the data, and are separate from the rest of the system; in the second step, I estimate the six remaining parameters separately for each sector. I make parametric assumptions about firms' production function, simulate the profit-maximizing decisions of each firm, and estimate the remaining parameters using a method of simulated moments (MSM) approach (Gourieroux, Monfort and Renault, 1993). The main objects of interest are the complementarity between division of labor and city size, and the complementarity between division of labor and firm complexity. In the context of the parameterized version of my model, the first parameter controls the extent to which the cost of division of labor falls with city size and the second parameter controls the extent to which the benefits of division of labor rise with firm complexity.

Dependent variable	Log (No of occs)				Specialization index			
	(1)	(2)	(3) Interm. inputs	(4) G3 exp share	(5)	(6)	(7) Interm. inputs	(8) G3 exp share
Panel A: Baseline results								
$Backbone_{jt}$	.0127***	.0015	.0015	.0074**	.0855***	.0116	.0728***	.0805***
	(.0028)	(.003)	(.0038)	(.0032)	(.017)	(.0085)	(.014)	(.016)
$Backbone_{jt} \times \log L_{ct_0}$		.0077***				.0141***		
		(.0008)				(.0033)		
$Backbone_{jt} \times \log c_{st_0}$			.0139***	.004***			.0156***	.0064***
			(.0031)	(.0012)			(.0044)	(.0013)
Mean of outcome	1.45	1.45	1.45	1.45	.43	.43	.43	.43
Obs	777096	777096	777096	777096	777096	777096	777096	777096
R-sq	.853	.853	.853	.854	.717	.718	.717	.717
Panel B: Excluding IT-related occupations								
$Backbone_{jt}$	.0124***	.0020	.0019	.007*	.086***	.011	.0734***	.0819***
	(.0031)	(.0021)	(.0019)	(.0042)	(.0073)	(.0079)	(.0241)	(.0136)
$Backbone_{jt} \times \log L_{ct_0}$		.0067***				.0126***		
		(.0018)				(.0035)		
$Backbone_{jt} \times \log c_{st_0}$			.0123***	.0027*			.016***	.0038**
			(.0023)	(.0016)			(.0047)	(.0015)
Mean of outcome	1.42	1.42	1.42	1.42	.42	.42	.42	.42
Obs	721629	721629	721629	721629	721629	721629	721629	721629
R-sq	.851	.850	.850	.850	.714	.713	.713	.715
Panel C: Dropping occupation categories did not exist before								
$Backbone_{jt}$	.0124***	.001	.031	.008*	.076***	.012	.0743***	.0808***
	(.0037)	(.0021)	(.0029)	(.0032)	(.0063)	(.01)	(.0142)	(.013)
$Backbone_{jt} \times \log L_{ct_0}$		.0068***				.0126***		
		(.0018)				(.0035)		
$Backbone_{jt} \times \log c_{st_0}$			.0132***	.003**			.0143***	.0058***
			(.0053)	(.0015)			(.0057)	(.0015)
Mean of outcome	1.42	1.42	1.42	1.42	.42	.42	.42	.42
Obs	777096	777096	777096	777096	777096	777096	777096	777096
R-sq	.851	.850	.850	.851	.715	.715	.714	.715

Robust standard errors clustered by municipality in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term, establishment and year FEs.

Table 3: Impacts of fast internet on division of labor within establishments

The structural estimation uses data from RAIS and PIA in 2010. Using RAIS data, I construct establishment-level information on employment, labor payment, division of labor, location and industry classification. The PIA data report sector-level information on value-added, inputs, and production. I trim the bottom and top 1% of the data. This leaves me with 298,412 establishments. For the estimation, I aggregate establishments into 21 sectors.[48]

5.1 Step one: Direct calibration

I begin by estimating $2S+1$ parameters that can be extrapolated directly from the data without using the structure of the rest of the model. These are the elasticity of substitution σ_s and the Cobb-Douglas share ξ_s for each sector, and the Cobb-Douglas share of nontradable goods and services η in worker's utility function.

I assign values to parameters σ_s, ξ_s, and η as follows. The elasticity of substitution in the CES demand function is calibrated to match the sector-level markup charged to consumers, where $\frac{\sigma_s}{\sigma_s-1} = \frac{revenue_s}{cost_s}$. I then estimate the Cobb-Douglas share of each sector ξ_s by measuring its share of value-added produced. Lastly, $\frac{1-\eta}{\eta}$ corresponds to the elasticity of wages with respect to city size, from Equation (5). To account for heterogeneity of workers across space, I calculate the elasticity using residuals from a Mincerian wage regression and obtain an elasticity of 3.1%, which corresponds to $\eta = 0.97$.[49]

5.2 Step two: Method of simulated moments

In the second stage, I use MSM to estimate the remaining parameters. Given parameter estimates from the first step, and parametric assumptions on model specifications and distributions of the underlying firm heterogeneity, and idiosyncratic shocks to firms' location choices, I simulate the profit-maximizing decisions of each firm and calculate a set of non-parametric moments to characterize the economy. I then iterate over new choices of parameters and select the best set of parameters to minimize the distance between the simulated moments and their data analogs.

5.2.1 Model specification

To estimate the complementarity between firm complexity and city size, I need to fully characterize the features of firm production function. For ensuing discussions, it is useful to define a new term:

$$\psi_s(z) = A(N, z, c_s)H(N, L). \tag{28}$$

[48] Summary statistics are reported in Online Appendix.

[49] I first regress log hourly earnings of the workers in my sample on a gender dummy, a race dummy, a categorical variable for 10 levels of education attainment, a quartic in years of potential experience, and all pair-wise interactions of these values (where regressions are weighted by annual hours worked). I then take the residuals from the Mincerian regression and regress on log of city size to obtain the elasticity of wages to city size.

Given Equation (28), the profit function can be rewritten as a function of ψ,

$$\pi_s(z, L, N) = \frac{(\sigma_s - 1)^{\sigma_s - 1}}{\sigma_s^{\sigma_s}} \left(\frac{\psi_s(z)}{w(L)}\right)^{\sigma_s - 1} R_s P_s^{\sigma_s - 1}.$$

In each sector, given local labor costs $w(L)$, firms' profit increases with $\psi_s(z)$. I thus interpret $\psi_s(z)$ as the *productivity* of a firm with complexity z in sector s.

All propositions in Section 3 are derived based on Assumptions 1 and 2, i.e. complementarities in (N, z) and (N, L). In the structural estimation, however, I do not impose these relationships. Instead, I assume the following functional form for productivity, which allows for any sign of these effects:

$$\log \psi_s(z, N, L) \equiv \log A(N, z, c_s) + \log H(N, L) = (\log z)(1 + \log N)^{c_s} - \frac{\log N}{(1 + \log \tilde{L})^{\theta_s}} \tag{29}$$

where $\tilde{L} = \frac{L}{L_0}$, and L_0 is the smallest city size in the set of city size distribution $\mathcal{L}$.[50]

For worker productivity, I postulate that

$$\log A(N, z, c_s) = (\log z)(1 + \log N)^{c_s}.$$

It is straightforward to see that $(\log z)(1 + \log N)^{c_s}$ is strictly increasing in z, c_s, and N. The strength of complementarity between firm complexity and division of labor is determined by c_s. A positive value of c_s would confirm model assumption. When $c_s = 0$, there is no complementarity, and I obtain a model in which the worker productivity is solely determined by firms' complexity draws. The cost of division of labor takes the following functional form:

$$\log H(N, L) = -\log N(1 + \log \tilde{L})^{-\theta_s}.$$

The marginal cost of division of labor is given by $\frac{1}{N}\frac{1}{(1+\log \tilde{L})^{\theta_s}}$, which is decreasing in the normalized city size $\tilde{L}$, reflecting the complementarity between division of labor and city size.[51] To retain full flexibility, I allow the strength of the complementarity, θ_s, to vary across sectors. Similar to worker productivity, a positive estimate of θ_s would confirm model assumption. When $\theta_s = 0$, there is no complementarity, and the marginal cost of division of labor is constant across different city sizes. Finally, following the conventional literature, I assume that $\log z$ is distributed according to a normal distribution with variance ν_z, truncated at its mean to prevent $\log z$ from being negative.

[50] Under this specification, when N is optimally chosen, the productivity function in (29) can be mapped directly to the productivity function assumed in Gaubert (Forthcoming).

[51] I normalize city size by the minimum city size in Brazil L_0. The normalization does not affect the estimation, since the estimation relies on relative sizes of cities.

5.2.2 Model extensions

To bring the model to data, I incorporate three extensions: (i) spatial equilibrium with a discrete set of cities, (ii) imperfect sorting of firms, and (iii) other sources of agglomeration externalities. My extended model allows me to obtain results under less restrictive assumptions than Section 3, and to evaluate the contribution of division of labor to productivity differences across cities, on which my baseline model is silent.

First, I consider a discrete set of cities. In the baseline model, I assume that the whole economy consists of a continuum of identical sites. This assumption simplifies the theoretical analysis and generates the uniqueness of the heterogeneous equilibrium. For the quantitative exercise, I take the choice set of city sizes $\mathcal{L}$ as exogenously given.[52] Cities, indexed by n, are ordered by their city size L_n. Given the log-supermodularity of the profit function in (z, L), more complex firms still sort into larger cities. Within a sector, each city is occupied by a range of firms with different complexity draws, denoted by $[\underline{z}_s(n), \bar{z}_s(n)]$. Spatial equilibria are determined by the following indifference condition:

$$\pi_s(\bar{z}, n) = \pi_s(\underline{z}, n+1), \quad \forall L_n \in \mathcal{L}. \tag{30}$$

While the new spatial equilibria may no longer be unique, the equilibrium characteristics presented in Section 3.4 hold for both continuous and discrete cases.

Second, I introduce an error structure that allows firms' ex post productivity to vary within a city. In the baseline model, within a sector, there is strict sorting of firms across city sizes. As a result, within a city, all firms in the same sector share the same division of labor, productivity, revenue, and profit. In reality, there may be other factors that affect a firm's location choice, and there is great heterogeneity across firms within a city. To capture the imperfect sorting of firms, I add an error structure by assuming that each firm j draws an idiosyncratic shock ϵ_{jL} for each city size L, where ϵ_{jL} is i.i.d. across city size and firms. I further assume that these shocks follow a Type I Extreme Value distribution, with mean zero and variance ν_L. The shock captures idiosyncratic motives for firms' location choices. With the extension, in a sector, there is a distribution of complexities allocated to each city size. However, of the complexity level dominating each city, there is still positive assortative matching between the complexity and city size. Therefore, equilibrium characteristics in Section 3.4 still hold.[53]

Lastly, I include two terms in firms' productivity function that summarizes other sources of agglomeration externalities, which my model abstracts from. The first term incorporates productivity advantage of larger

[52] Note that the equilibrium distribution of cities is still endogeneoulsy determined as I do not impose restrictions on the existence or number of cities of any particular size in spatial equilibrium.

[53] I assume that ϵ_{jL} is city-size specific, rather than city-specific. If misspecified, these shocks can represent the maximum of shocks at a more disaggregate level, such as at the city level. See Gaubert (Forthcoming) for an excellent discussion of this.

cities, which are not correlated with firm complexity. This includes, but is not limited to, the sorting of skills among heterogeneous workers, knowledge spillover, and natural amenity differences. The second incorporates potential direct interaction between firm complexity and city size. The model assumes that firm complexity interacts with city size only through the proposed channel of division of labor. I do not impose this restriction in the structural estimation. Instead, this term allows more complex firms to sort into larger cities for reasons beyond division of labor. With these extensions, productivity of a firm j with complexity draw z in sector s takes the following form:

$$\log \psi_j(z, c_s, L, N) = \alpha_s \log L + (\log z_j)(1 + \log \tilde{L})^{\upsilon_s} + (\log z_j)(1 + \log N)^{c_s} - \frac{\log N}{(1 + \log \tilde{L})^{\theta_s}} + \epsilon_{jL}, \tag{31}$$

where α_s captures the standard reduced-form agglomeration externality and υ_s determines the strength of direct interaction between complexity and city size. When $\theta_s = 0$ or $c_s = 0$, I obtain a standard firm sorting model (see, e.g., Gaubert, Forthcoming). Additionally, when $\upsilon_s = 0$, I obtain a classic model of agglomeration externalities without division of labor or firm sorting (see, e.g., Allen and Arkolakis, 2014).

5.2.3 Estimation procedure and Moments

Given the distribution of firm complexities and idiosyncratic firm-city-size shocks, parametric assumptions, and the parameters estimated in the first stage, six parameters remain to be estimated for each sector: the reduced-form agglomeration externality (α), the interaction between firm complexity and city size (υ), the complementarity between firm complexity and division of labor (c), the complementarity between division of labor and city size (θ), the variance of complexity distribution (ν_z), and the variance of the firm-city-size shocks (ν_L). I use MSM to back out the six parameters, $\chi_s = (\alpha, \upsilon, c, \theta, \nu_z, \nu_L)_s$, for each $s = \{1, \ldots, S\}$.

I draw a sample of 100,000 firms for each sector and find the profit-maximizing division of labor, N^*, conditioning on city size, according to the following equation:

$$\log N_j^* \equiv \underset{N \in \mathbf{R}^+}{\arg\max} \; \log z_j (1 + \log N)^{c_s} - \frac{\log N}{(1 + \log \tilde{L})^{\theta_s}} \tag{32}$$

This gives me a firm productivity function conditioning on city size:

$$\log \psi_j(z, c_s, L) = \alpha_s \log L + \log z_j (1 + \log \tilde{L})^{\upsilon_s} + \log z_j (1 + \log N_j^*)^{c_s} - \frac{\log N_j^*}{(1 + \log \tilde{L})^{\theta_s}} + \epsilon_{jL}. \tag{33}$$

Based on $\log \psi_j(z, c_s, L)$, firms make a discrete choice of city size, according to the following equation:

$$\begin{aligned}\log \tilde{L}_s(z_j) &\equiv \underset{\tilde{L}\in\mathcal{L}}{\arg\max}\ \log \psi_j(z, c_s, L) - \log w(L)\\ &= \underset{\tilde{L}\in\mathcal{L}}{\arg\max}\ \log z_j(1+\log \tilde{L})^{\upsilon_s} + \log z_j(1+\log N_j^*)^{c_s} - \frac{\log N_j^*}{(1+\log \tilde{L})^{\theta_s}} + \left(\alpha_s - \frac{1-\eta}{\eta}\right)\log \tilde{L} + \epsilon_{jL}.\end{aligned} \tag{34}$$

To estimate the six parameters in χ_s, I match six sets of simulated and data moments for every sector. The primary set of moments relies on the reduced-form estimates from the quasi-experiment in Section 4. I first estimate the extent of changes in the cost of division of labor to match the average treatment effect of 1.27% (Table 3 of Section 4.4). I then feed the shocks to cost of division of labor into the model. Using simulated distribution of firms across space, I compute the average city-level changes in firms' division of labor in response to shocks, to match the corresponding moments observed from the quasi-experiment.[54] I further supplement the data with five additional sets of moment: geographic distribution of firms, firm-size distribution, increase in the average firm size across city sizes, increase in the average division of labor across city sizes, and within-city variations in division of labor.[55]

The identification of the two complementarity parameters, c and θ, is possible because I can observe division of labor within firms. In the empirical exercise, in response to the infrastructure improvement to a set of cities, establishments located in larger cities tend to undergo a greater increase in division of labor relative to those in smaller cities. Theoretically, this is driven by both complementarities, or the joint parameter $\frac{\theta}{1-c}$. More specifically, c determines the extent of increase in division of labor in response to the shock, given firm complexity distribution within a city; θ determines the firm complexity distribution for a given city size. Therefore, by observing variation in the changes of division of labor across city sizes, I can identify $\frac{\theta}{1-c}$. To separately identify c and θ, I consider within-city variations in firms' division of labor. Given a city size, the impact of city size on division of labor is the same for all firms located there. I can, therefore, identify the complementarity between division of labor and complexity—i.e., c—using the within-city variation in firms' division of labor, relative to that in firm complexities. Intuitively, all else equal, small changes in firm complexity would generate a huge variation in division of labor, if the complementarity is strong.[56]

[54] I use four moments, by calculating the average change in division of labor for a given sector across four quartiles of city sizes.

[55] I measure the geographic distribution of firms using the share of employment in a given sector that falls into one of the four bins of city sizes, in which the city-size bins are defined as threshold cities with less than 25%, 50%, and 75% of overall sectoral employment. To measure firm-size distribution, I use five moments that characterize nonparametrically the distribution. These bins are defined by the 25, 50, 75 and 90th percentiles of the distribution. On increases in average firm size and division of labor across city sizes, I use 8 moments summarizing the average labor payment and division of labor across four quartiles of city sizes. Lastly, I use the variance of firms' division of labor in each quartile of city sizes, to summarize variation in division of labor within cities.

[56] Please refer to Appendix D for further discussions on moments and identification.

MSM chooses parameters $\hat{\chi}_s$ to minimize the distance between simulated moments and targeted moments, using the criterion function:

$$\hat{\chi}_s = \arg\min\ (m_{s,data} - m_{s,sim}(\chi_s))' J_s (m_{s,data} - m_{s,sim}(\chi_s)) \tag{35}$$

where $m_{s,data}$ is the vector of empirical moments for sector s, and $m_{s,sim}$ is the vector of simulated moments calculated at χ_s.

I use the diagonal of the variance-covariance matrix of the moments as the weighting matrix J_s, rather than the optimal full variance-covariance matrix, due to concerns about bias raised by Altonji and Segal (1996).[57] I find the parameters that minimize the criterion function using the particle swarm optimization method (Kennedy and Eberhart, 1995). I provide more details on the estimation process in Online Appendix.

5.3 Structural results

In this section, I present results from the second-stage estimation. Estimated parameters by sector are reported in Table 22 of Appendix E. I first examine model fit for the set of targeted moments (reported in Figures 14 to 19 in Online Appendix). Overall, the estimated model captures well the cross-sectoral heterogeneities in treatment effects in response to the technology shock, location patterns, cross-city variations in firm sizes and division of labor, and within-city variations in division of labor. The fit for firm-size distribution in labor payment is better for the upper tail than the lower tail. The result is expected, since I target the upper-tail quantiles in the estimation.

I next move on to nontargeted moments. In particular, I consider three sets of nontargeted moments that combine the 21 sectoral estimation results. The first set considers the relative magnitude for the estimated sector-specific complexity parameter c across different sectors. The estimation is made for each sector separately. I make no assumption on the relative size of c—the complementarity between firm complexity and division of labor—across sectors. The theory, however, predicts that the complementarity is stronger in more complex sectors. I relate the estimates of c to the two empirical proxies of sector-level complexity, and estimate the rank correlations between them. Rank correlations are 0.68 and 0.62 for the measures using the number of intermediate inputs and the G3 export share, respectively. Figure 6 of Appendix F plots the rank of the estimates across sectors against the empirical measures.

[57] The variance-covariance matrix, Ω_s, is calculated from $m_{s,data}$, using a bootstrap procedure. Within each sector, I first sample, with replacement, firms from my data for 2,000 times. For each resampling b, I calculate m_s^b, the new moments generated from the bootstrap sample. I then calculate

$$\Omega_s = \frac{1}{2000} \sum_{b=1}^{2000} (m_s^b - m_{s,data})(m_s^b - m_{s,data})'.$$

The weight matrix J_s is simply the diagonal of Ω_s.

Next, I examine the simulated city-size distribution. The fact that city distribution follows Zipf's law is one of the most remarkable empirical facts in economics.[58] In estimating the model, I impose no restriction on the number of cities in each city-size bin, which defines the city-size distribution. Using the estimates, I can solve for the city-size distribution in equilibrium (see Online Appendix for detailed steps). As shown in Figure 7 of Appendix F the estimated city-size distribution adheres to Zipf's law and follows the actual city-size distribution well.

Lastly, I use the quasi-experiment to provide out-of-sample validation to the estimated model. I do not target sectoral distribution in the estimation. However, cities populated by more high-c_s firms would undergo a higher average city-level increase in firms' division of labor due to heterogeneity of the treatment across sectors with different complexities. I first calibrate the magnitude of the reduction in cost of division labor due to the rollout of the ICT infrastructure to match the average treatment effect, yielding a 5.6% reduction in costs. Using this, I then calculate the average change in division of labor across all sectors in a given city. The correlation between the average change in firms' division of labor within different cities predicted by the model and those in the data is high, at 0.73.[59] Looking at Figure 8 in Appendix F, one can see that the model accurately predicts that areas undergoing the highest increase are concentrated in the South, and that the increases tend to be smaller in the northern parts of the country.

6 Productivity impacts of division of labor

Productivity advantages in larger cities are well documented in the literature (see, e.g., Rosenthal and Strange, 2004; Melo, Graham and Noland, 2009). Unlike previous theories, in my model, the productivity distribution is determined not only by the standard agglomeration externalities, the variance of firm complexity distribution, and firm-city-size idiosyncratic shocks, but also by firms' endogenous decisions on the extent of division of labor. Using the estimated model, I next conduct a counterfactual exercise to quantify the contribution of division of labor to the productivity gains in larger cities.

I begin by computing the estimated elasticity of firm productivity to city size, using the following OLS regression on the simulated set of data:

$$\log \hat{\psi}_j = \beta_0 + \beta_1 \log L_j + \delta_{s(j)} + \epsilon_j \tag{36}$$

where $\hat{\psi}_j$ is the simulated firm productivity defined in Equation (33) for firm j, L_j is the optimal city size chosen by the firm according to Equation (34), and $\delta_{s(j)}$ is a sector fixed effect. β_1 denotes the elasticity

[58] According to Zipf's law, when we order cities in a country by size and regress the log of the rank against the log of the size, we get a straight line with a slope of -1.

[59] The benchmark correlation is 0.36, which is obtained by assuming a uniform sectoral distribution across cities.

of firm productivity to city size. Running the OLS regression in Equation (36), I get an OLS estimate of $\hat{\beta}_1 = 0.0826$.[60] This measure is consistent within the range of existing measures of agglomeration externalities, at 0.02–0.10 (Rosenthal and Strange, 2004; and Melo, Graham and Noland, 2009), providing another external validation for the estimation results.

To estimate productivity advantage of larger cities through division of labor, I conduct the following counterfactual analysis, in which I shut down productivity increase through division of labor. This is achieved by (i) allocating firms to city sizes according to the complexity draw and their firm-city-size specific shocks, instead of allowing firms to choose optimally based on considerations related to division of labor, and (ii) fixing firms' division of labor based on the average value within their sector. Under this counterfactual scenario, I re-estimate the model, which gives me a new set of productivities and their corresponding spatial distribution. Under the restriction, differences in firm productivity across space are only driven by firm complexity draws and the agglomeration externalities determined by the firm-city-size specific shocks. This counterfactual exercise allows me to identify what would be the realized productivities if division of labor did not affect the productivity and location choices of firms. Re-estimating Equation (36) using the new simulated data leads to an elasticity of firm productivity to city size of 0.0699. By this account, division of labor accounts for 15% of the productivity advantage in larger cities.[61] The estimated contribution is comparable to the importance of natural advantage and labor-market-based knowledge spillover estimated in previous literature (see, e.g., Ellison and Glaeser, 1999; Serafinelli, 2015).[62]

I further examine the importance of *firm sorting* to the 15% productivity contribution through division of labor. In the model, firms sort into larger cities because larger cities foster greater division of labor. To shut down the systematic sorting of firms, I allocate firms to city sizes according to the complexity draw and their firm-city-size specific shocks, similar to the first counterfactual exercise. However, firms are allowed to choose the optimal division of labor based on their complexity draws and the city in which they are allocated to. This counterfactual exercise thus allows me to study what would be the realized productivity if I only shut down one of the two channels through which division of labor could increase productivity.[63] I find, by re-estimating Equation (36), that the elasticity estimate drops to 0.0762. This suggests that firm sorting accounts for about half of the spatial productivity differences through division of labor.[64]

[60] This implies productivity goes up by 8.26% when city size is doubled.

[61] Without the endogenous choice of division of labor, the elasticity estimate goes down by 0.0127 (0.0826 - 0.0699), which is 15% (0.0127 / 0.0826) of the baseline elasticity.

[62] I also consider an alternative approach in which I re-estimate $\log \hat{\psi}_j$ by removing the standard agglomeration externalities, $\alpha \log L + \log z(1 + \log L)^{\upsilon}$, in Equation (33). This assumes that the productivity advantage in larger cities only comes from my proposed channel of division of labor. Re-estimating Equation (36) gives me similar results. I find that division of labor generates an elasticity estimate of 0.0122, which is 14.8% of the original value.

[63] I perform a robustness check in which I shut down the direct effect of city size on firm's division of labor while allowing for firm sorting, i.e., firms can endogenously sort into cities based on their complexities but have to choose a fixed level of division of labor. The results are similar to this approach.

[64] Without systematic firm sorting, the elasticity estimate goes down by 0.0064 (0.0826 - 0.00762), which is 50% (0.0064 / 0.0127) of the contribution of division of labor to the spatial productivity difference.

In addition to the counterfactual analysis, one can use the estimated model for policy evaluations. In Online Appendix, I provide one such illustration by estimating the short-run and long-run impact of the new ICT infrastructure in Brazil.

7 Conclusion

In this paper, I show that division of labor is an important contributing factor for the productivity advantage in larger cities. Using the unique data that measures division of labor at the firm level, I document a new empirical fact that firms adopt greater division of labor in larger cities. To explain this, I build a parsimonious model embedding firms' choices of the optimal division of labor into a spatial equilibrium framework, and propose mechanisms that generate the positive correlation between firms' division of labor and city size in equilibrium. Firms' optimal choices of division of labor drive sorting of firms across cities. This spatial sorting shapes the spatial distributions of division of labor and productivity jointly. Through a quasi-experiment, I provide causal empirical evidence that supports a set of theoretical predictions. Finally, the structure of the model, combined with the reduce-form estimates from the quasi-experiment and the detailed observables in the data, allows me to estimate the contribution of division of labor to productivity advantage in larger cities, and to separately identify the relative contributions of the different channels proposed in the model.

This project is a step toward further unpacking the black box of agglomeration externalities. Identifying the source of agglomeration externalities is important not only for our understanding of the regional productivity differences, but also matters for understanding aggregate productivity, which depends on the spatial distribution of firms and workers. The evidence on both the relationship between firms' division of labor and city size, and the underlying mechanisms driving this relationship has direct policy implications. In the quasi-experiment, the ICT infrastructure that improves coordination efficiencies within the firm may be an effective way of increasing labor productivity by enabling workers to be more specialized. Future works should evaluate the impact of other policy interventions related to reducing coordination costs, matching frictions, or learning and training costs associated with worker specialization.

References

Abdel-Rahman, Hesham M., and Alex Anas. 2004. "Theories of systems of cities." In Handbook of Regional and Urban Economics. Vol. 4 of Handbook of Regional and Urban Economics, , ed. J. V. Henderson and J. F. Thisse, Chapter 52, 2293–2339. Elsevier.

Acemoglu, Daron, and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." Handbook of Labor Economics, 4: 1043 – 1171.

Akerman, Anders, and Loriane Py. 2010. "Service Outsourcing and Specialization: A Theory on Endogeneous Task Scope." Stockholm University, Department of Economics Research Papers in Economics 2010:14.

Allen, Treb, and Costas Arkolakis. 2014. "Trade and the Topography of the Spatial Economy." The Quarterly Journal of Economics, 129(3): 1085–1140.

Altonji, Joseph G., and Lewis M. Segal. 1996. "Small-Sample Bias in GMM Estimation of Covariance Structures." Journal of Business & Economic Statistics, 14(3): 353–366.

Arnott, Richard J., and Joseph E. Stiglitz. 1979. "Aggregate Land Rents, Expenditure on Public Goods, and Optimal City Size." The Quarterly Journal of Economics, 93(4): 471–500.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian. 2012. "On the Road: Access to Transportation Infrastructure and Economic Growth in China." National Bureau of Economic Research, Inc NBER Working Papers 17897.

Banerji, Sourangsu, and Rahul Singha Chowdhury. 2013. "Wi-Fi & Wi-MAX: A Comparative Study." CoRR, abs/1302.2247.

Baumgardner, James R. 1988. "Physicians' Services and the Division of Labor Across Local Markets." Journal of Political Economy, 96(5): 948–982.

Becker, Gary S., and Kevin M. Murphy. 1992. "The Division of Labor, Coordination Costs, and Knowledge." The Quarterly Journal of Economics, 107(4): 1137–1160.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud. 2014. "Productive Cities: Sorting, Selection, and Agglomeration." Journal of Political Economy, 122(3): 507–553.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-In-Differences Estimates?" The Quarterly Journal of Economics, 119(1): 249–275.

Blei, David M, Andrew Y Ng, and Michael I Jordan. 2003. "Latent dirichlet allocation." Journal of machine Learning research, 3(Jan): 993–1022.

Bloom, Nicholas, Luis Garicano, Raffaella Sadun, and John Van Reenen. 2014. "The Distinct Effects of Information Technology and Communication Technology on Firm Organization." Management Science, 60(12): 2859–2885.

Boning, Brent, Casey Ichniowski, and Kathryn Shaw. 2007. "Opportunity Counts: Teams and the Effectiveness of Production Incentives." Journal of Labor Economics, 25(4): 613–650.

Borghans, Lex, and Bas Weel. 2006. "The Division of Labour, Worker Organisation, and Technological Change." Economic Journal, 116(509): 45–72.

Caliendo, Lorenzo, and Esteban Rossi-Hansberg. 2012. "The Impact of Trade on Organization and Productivity." The Quarterly Journal of Economics, 127(3): 1393–1467.

Caliendo, Lorenzo, Ferdinando Monte, and Esteban Rossi-Hansberg. 2015. "The Anatomy of French Production Hierarchies." Journal of Political Economy, 123(4): 809–852.

Caliendo, Lorenzo, Giordano Mion, Luca David Opromolla, and Esteban Rossi-Hansberg. 2015. "Productivity and Organization in Portuguese Firms." National Bureau of Economic Research, Inc NBER Working Papers 21811.

Cameron, A. Colin, and Douglas L. Miller. 2015. "A Practitioner?s Guide to Cluster-Robust Inference." Journal of Human Resources, 50(2): 317 – 372.

Chaney, Thomas, and Ralph Ossa. 2013. "Market size, division of labor, and firm productivity." Journal of International Economics, 90(1): 177–180.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." American Economic Review, 99(4): 1145–77.

Ciccone, Antonio. 2002. "Input Chains and Industrialization." Review of Economic Studies, 69(3): 565–587.

Ciccone, Antonio, and Robert E Hall. 1996. "Productivity and the Density of Economic Activity." American Economic Review, 86(1): 54–70.

Clarke, George, and Scott Wallsten. 2006. "Has the Internet Increased Trade? Evidence from Industrial and Developing Countries." Economic Inquiry, 44(4): 465–484.

Commander, Simon, Rupert Harrison, and Naercio Menezes-Filho. 2011. "ICT and Productivity in Developing Countries: New Firm-Level Evidence from Brazil and India." The Review of Economics and Statistics, 93(2): 528–541.

Conley, Timothy. 1999. "GMM estimation with cross sectional dependence." Journal of Econometrics, 92(1): 1–45.

Costinot, Arnaud. 2009. "On the origins of comparative advantage." Journal of International Economics, 77(2): 255–264.

Davis, Donald R., and Jonathan I. Dingel. 2012. "A Spatial Knowledge Economy." National Bureau of Economic Research, Inc NBER Working Papers 18188.

Davis, Donald R., and Jonathan I. Dingel. 2014. "The Comparative Advantage of Cities." National Bureau of Economic Research Working Paper 20602.

De la Roca, Jorge, and Diego Puga. 2017. "Learning by Working in Big Cities." The Review of Economic Studies, 84(1): 106.

De Negri, Joao Alberto, Paulo Furtado de Castro, Natalia Ribeiro de Souza, and Jorge Saba Arbache. 2001. "Mercado Formal de Trabalho: Comparaçao entre os microdados da RAIS e da PNAD."

Dix Carneiro, Rafael, and Brian K Kovak. forthcoming. "Trade Liberalization and Regional Dynamics." American Economic Review.

Draca, Mirko, Raffaella Sadun, and John Van Reenen. 2009. "Productivity and ICTs: A review of the evidence." Handbook of Information and Communication Technologies.

Duranton, Gilles, and Diego Puga. 2004. "Micro-Foundations of Urban Agglomeration Economies." Handbook of Regional and Urban Economics, 4: 2063 – 2117. Cities and Geography.

Duranton, Gilles, and Diego Puga. 2005. "From sectoral to functional urban specialisation." Journal of Urban Economics, 57(2): 343–370.

Duranton, Gilles, and Hubert Jayet. 2011. "Is the division of labour limited by the extent of the market? Evidence from French cities." Journal of Urban Economics, 69(1): 56–71.

Eaton, Jonathan, Samuel Kortum, and Francis Kramarz. 2011. "An Anatomy of International Trade: Evidence From French Firms." Econometrica, 79(5): 1453–1498.

Eeckhout, Jan, and Philipp Kircher. 2011. "Identifying Sorting–In Theory." Review of Economic Studies, 78(3): 872–906.

Ellison, Glenn, and Edward L. Glaeser. 1999. "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?" American Economic Review, 89(2): 311–316.

Fort, Teresa C. 2017. "Technology and Production Fragmentation: Domestic versus Foreign Sourcing." The Review of Economic Studies, 84(2): 650–687.

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

Fujita, Masahisa, and Jacques-François Thisse. 2013. Economics of Agglomeration: Cities, Industrial Location, and Globalization. . 2 ed., Cambridge University Press.

Garicano, L., and T. N. Hubbard. 2009. "Specialization, firms and markets: the division of labor between and within law firms." Journal of Law, Economics, and Organization, 25: 339–371.

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." Journal of Political Economy, 108(5): 874–904.

Gaubert, Cecile. Forthcoming. "Firm Sorting and Agglomeration." American Economic Review.

Glaeser, Edward L, and David C Maré. 2001. "Cities and Skills." Journal of Labor Economics, 19(2): 316–342.

Glaeser, Edward L, and Joshua D Gottlieb. 2009. "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States." Journal of Economic Literature, 47(4): 983–1028.

Gourieroux, C, A Monfort, and E Renault. 1993. "Indirect Inference." Journal of Applied Econometrics, 8(S): 85–118.

Hausmann, Ricardo, Jason Hwang, and Dani Rodrik. 2007. "What you export matters." Journal of Economic Growth, 12(1): 1–25.

Helpman, Elhanan, Oleg Itskhoki, Marc-Andreas Muendler, and Stephen Redding. forthcoming. "Trade and Inequality: From Theory to Estimation." Review of Economic Studies.

Helsley, Robert W., and William C. Strange. 2014. "Coagglomeration, Clusters, and the Scale and Composition of Cities." Journal of Political Economy, 122(5): 1064–1093.

Henderson, J V. 1974. "The Sizes and Types of Cities." American Economic Review, 64(4): 640–656.

Henderson, Vernon, and Randy Becker. 2000. "Political Economy of City Sizes and Formation." Journal of Urban Economics, 48(3): 453–484.

Hjort, Jonas, and Jonas Poulsen. 2016. "The Arrival of Fast Internet and Skilled Job Creation in Africa." Revise and Resubmit, American Economic Review.

Holmstrom, Bengt. 1982. "Moral Hazard in Teams." Bell Journal of Economics, 13(2): 324–340.

IBGE. n.d.. Instituto Brasileiro de Geografia e Estatística.

IGIC, Inc. Staff. 2004. Fiber Distribution Data Interface (FDDI). Information Gatekeepers Inc.

Jensen, M. 2011. Broadband in Brazil: A multipronged public sector approach to digital inclusion. Washington D.C: infoDev / World Bank.

Kennedy, J., and R. Eberhart. 1995. "Particle swarm optimization." Vol. 4, 1942–1948 vol.4.

Kim, Sunwoong. 1989. "Labor Specialization and the Extent of the Market." Journal of Political Economy, 97(3): 692–705.

Kline, Patrick, and Enrico Moretti. 2014. "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority *." The Quarterly Journal of Economics, 129(1): 275–331.

Knight, Peter. 2016. Broadband in Brazil: past, present and future. FGV Direito Rio.

Krugman, Paul R. 1979. "Increasing returns, monopolistic competition, and international trade." Journal of International Economics, 9(4): 469 – 479.

Levchenko, Andrei A. 2007. "Institutional Quality and International Trade." The Review of Economic Studies, 74(3): 791–819.

Lucas, Robert Jr. 1976. "Econometric policy evaluation: A critique." Carnegie-Rochester Conference Series on Public Policy, 1(1): 19–46.

Marshall, Alfred. 1890. Principles of Economics. MacMillan.

McElheran, Kristina. 2014. "Delegation in Multi-Establishment Firms: Evidence from I.T. Purchasing." Journal of Economics & Management Strategy, 23(2): 225–258.

Melo, Patricia, Daniel Graham, and Robert Noland. 2009. "A meta-analysis of estimates of urban agglomeration economies." Regional Science and Urban Economics, 39(3): 332–342.

Qiang, C., and C. Rossotto. 2009. "Economic impacts of broadband." Information and communication for development: Expanding reach and increasing impact.

Rosen, Sherwin. 1983. "Specialization and Human Capital." Journal of Labor Economics, 1(1): 43–49.

Rosenthal, Stuart S., and William C. Strange. 2004. "Evidence on the nature and sources of agglomeration economies." In Handbook of Regional and Urban Economics. Vol. 4 of Handbook of Regional and Urban Economics, , ed. J. V. Henderson and J. F. Thisse, Chapter 49, 2119–2171. Elsevier.

Saboia, Joao L. M., and Richardo M. L. Tolipan. 1985. "A relacao anual de informacoes sociais (RAIS) e o mercado formal de trabalho no Brasil: uma nota." Pesquisa e Planejamento Economico, 15(2): 447–456.

Serafinelli, Michel. 2015. "Good Firms, Worker Flows and Local Productivity." Verein für Socialpolitik / German Economic Association Annual Conference 2015 (Muenster): Economic Development - Theory and Policy 113063.

Smith, Adam. 1776. The Wealth of Nations.

Sullivan, Paul. 2009. "Estimation of an Occupational Choice Model When Occupations Are Misclassified." The Journal of Human Resources, 44(2): 495–535.

Topkis, D. M. 1978. "Minimizing a Submodular Function on a Lattice." Operations Research, 26(2): 305–321.

Varian, Hal R. 2010. "Computer Mediated Transactions." American Economic Review, 100(2): 1–10.

Wang, Zhi, and Shang-Jin Wei. 2010. "What Accounts for the Rising Sophistication of China's Exports?" In China's Growing Role in World Trade. NBER Chapters, 63–104. National Bureau of Economic Research, Inc.

A Microfounding the production function

In this section, I follow Costinot (2009) and develop a microfounded production function that generates log-supermodularities between division of labor N and firms' complexity draw z, and between division of labor N and sector-level complexity c_s.

Production technology

In sector s, for each firm z, a continuum of complementary tasks $t \in [0, c_s]$ must be performed to produce 1 unit of the output. c_s is an exogenously given parameter, shared by all firms producing in sector s. A higher value of c_s denotes a more complex production technology, as it requires more tasks to be completed before the output can be produced.

Production follows a Leontief production technology:

$$Q_z^s = \min_{t \in [0, c_s]} q^s(t, u) du \tag{37}$$

where $q^s(t, u)$ is the output for task $t \in [0, c_s]$.

There are increasing returns to scale in the performance of each task. Prior to any production, the worker would need to spend time learning how to perform the tasks assigned to her. The output of worker i performing task t is given by

$$Q(i, \tau, z) = l(i, \tau, z) - f(t, z),$$

where $l(i, t, z)$ is the time worker i allocates to perform task t, and $f(t, z)$ is the time necessary to learn how to perform task t. Note that $f(t, z)$ depends on task t and firms' complexity draw z.

I can normalize the tasks such that all tasks are identical. This implies that the learning cost per worker, $f(t, z)$, is also constant across all tasks and can thus be denoted by $f(z)$. I further normalize the training costs such that $f(z) = z$. The worker-level training cost involved in producing at least 1 unit of output in firm z is thus equal to

$$\int_0^{c_s} z \, dt = zc_s.$$

Across sectors, the more complex a production process (i.e., higher c_s), the more tasks someone must learn. Within a sector across firms, the more complex a product (i.e., higher z), the longer it takes for someone to learn any single task.

Optimal contracts

Given the costly training cost, workers who know how to perform a task should perform it as many times as possible. This implies that all workers specialize in one set of tasks, which I call an "occupation." Additionally, worker productivity depends on the number of tasks included in an occupation. More specifically, a marginal decrease in the number of tasks included in an occupation increases the time available for actual production. However, this increase is larger for occupations with more tasks. Since profit maximization requires that marginal changes in workers' productivity be equalized across occupations, it also requires that each occupation include the same number of tasks. Since each occupation consists of $\frac{c_s}{N}$ tasks, the training cost per worker is therefore $\frac{zc_s}{N}$. This leaves each worker i

$$1 - \frac{zc_s}{N}$$

units of time available for production. The worker productivity is therefore given by,

$$A(N, z, c_s) \equiv \frac{1}{c_s} - \frac{z}{N}. \tag{38}$$

The result in Equation (38) reflects the key argument of Rosen (1983). Worker productivity is maximized when N is infinite, and every worker only learns an infinitesimal task. In other words, if there is no coordination cost, efficiency requires that each skill be used as intensively as possible. Furthermore, it is straightforward to see that using this production process, $A(N, z, c_s)$ meets the conditions specified in Assumption 1—$A(N, z, c_s)$ is increasing in N, and is log-supermodular in (z, N) and (c_s, N).

B Microfoundation for the complementarity between N and L

In this part, I present two ways to microfound the complementarity between division of labor N and city size L. The first one argues that larger cities provide better infrastructure—in particular, ICT infrastructure—that reduces the costs of greater division of labor. The second focuses on the learning advantage in larger cities. It is relatively cheaper for firms with greater division of labor to train their workers in larger cities.

B.1 Local infrastructure provision

I first focus on the ability for larger cities to provide better public infrastructure. This is one of the most classic agglomeration externalities that justify the existence of cities (see Duranton and Puga, 2004, and Fujita and Thisse, 2013 for a review). Following Henderson (1974), I assume there is a class of local land developers. Land developers fully tax local landowners. They, in turn, invest the tax revenue in local infrastructure to attract firms. Land developers also play a coordination role, setting up cities on potential sites where they find profitable to do so, by announcing a city size L and a level of infrastructure investment, $\mathcal{I}$. Their revenues correspond to the profits made in the housing sector, i.e. $\pi_H(L) = (1 - \eta)Lw(L)$. Due to competition and free entry, land developers that invest less than $\pi_H(L)$ will not attract any firm to the city; whereas developers that invest more to attract firms will make negative profits. Therefore, in equilibrium, the optimal level of investment in L is

$$\mathcal{I}(L) \equiv (1 - \eta)Lw(L) = ((1 - \eta)L)^{\frac{1}{\eta}} \bar{w}, \tag{39}$$

where $\mathcal{I}(\cdot)$ denotes the optimal level of investment. Using Equation (39), it can be readily seen that $\mathcal{I}(\cdot)$ is an increasing function of the city size, L.

Note that the result in Equation (39) is stronger than the necessary condition to derive the positive correlation between division of labor and city size, which only requires that the aggregate level of infrastructure be greater in larger cities. However, under certain conditions, the provision of public infrastructure in Equation (39) is the socially optimal level.[65]

Next, I assume that there is complementarity between city infrastructure, $\mathcal{I}$, and firms' division of labor, N. Better infrastructure, e.g., ICT infrastructure such as faster internet, improves communication within a firm, making coordination among specialized workers more efficient. Since $\mathcal{I}$ is an increasing function of city size, L, the log-supermodularity between $\mathcal{I}$ and N implies the log-supermodularity between L and N.

B.2 Alternative microfoundation for the complementarity between N and L

I present an alternative way to microfound the complementarity between firms' division of labor N and city size L. The main idea follows Marshall (1890), who argues that a larger market facilitates learning, perhaps by providing better technologies or a better environment for knowledge sharing or idea exchange. This allows workers to pursue a more specialized set of skills that reduce the cost of training.

For simplicity, I discuss the single-sector case here.[66] I normalize sector-level complexity $c_s = 1$ for all firms. To produce any good, all tasks in $[0, 1]$ needs to be completed within a firm. Firms hire workers, whose productivity depends on their level of human capital.

Human capital of workers has two dimensions, intensive human capital b and extensive human capital $K = \frac{1}{N}$.[67] K is a measure of the breadth of a worker's skills, and b represents the depth of a worker's skills, which can be interpreted as the efficiency units supplied by a worker. Following Caliendo and Rossi-Hansberg (2012), I assume that the cost of acquiring human capital, $\gamma w(L)$, is proportional to the wage in the city, since learning requires teachers in the schooling sector who earn $w(L)$. Learning thus requires γ units of a teacher's time at wage $w(L)$. Since workers are ex ante identical, in equilibrium, the additional pay to workers over $w(L)$ must equal the learning costs. The total wage that workers receive from the firm is thus given by:

$$\text{worker wage} = (1 + \gamma)w(L).$$

Following conventional literature (see, e.g., Kim, 1989), I assume that the cost of acquiring human capital

[65]This is argued in Henry George Theorem (Arnott and Stiglitz, 1979), which claims that public expenditure on non-rival public infrastructure equals aggregate land rent when the population size of a city is optimal. Alternatively, the same outcome can be achieved using voting as an alternative decision-making mechanism to determine the location and the level of local public infrastructure. Given individual mobility within the city and competitive housing land prices, the optimal level of infrastructure provision $\mathcal{I}(L)$ is unanimously selected by consumers through voting if the local government implements a housing tax equivalent to housing rent. See Fujita and Thisse (2013) for details.

[66]Results with multiple sectors are very similar to single-sector case, though the derivation is messier and requires an additional assumption, which I discuss below.

[67]This assumption states that the more specialized workers are (i.e., larger N), the lower the level of extensive human capital.

is convex in both intensive and extensive human capital. Formally,

$$\gamma_b > 0,\ \gamma_K > 0,$$

$$\gamma_{bb} > 0,\ \gamma_{KK} \geq 0,\ \gamma_{bK} > 0.$$

where the subscripts refer to partial derivatives.[68]

The cost of knowledge acquisition also depends on the city-wide availability of intensive and extensive human capital, denoted by $\mathbf{b}(L)$ and $\mathbf{K}(L)$, respectively.[69] Importantly, $\mathbf{b}(L)$ is defined by the aggregate volume of intensive human capital available in city L, and $\mathbf{K}(L)$ is defined by the superset of the collection of extensive knowledge sets for all workers in the city. Formally,

$$\mathbf{b}(L) = \int_{i\in L} b(i)\ di;\ \ \mathbf{K}(L) = \sup\{K(i)\}_{i\in L},$$

where i denotes a worker living in city L.

To produce any good, all tasks must be completed. Therefore, the set of extensive human capital available, $\mathbf{K}(L)$, is the same everywhere, denoted by $\bar{\mathbf{K}}$. In other words, the marginal cost of pursuing extensive knowledge is unrelated to city size, i.e., $\gamma_{KL} = 0$.

On the other hand, the aggregate level of intensive human capital, $\mathbf{b}(L)$, is increasing in city size. In other words, all else equal, larger cities have a comparative advantage in pursuing intensive knowledge,[70]

$$\gamma_{bL} < \gamma_{KL} = 0.$$

With no search friction or information asymmetry in the model, I can combine the choice of human capital acquisition as part of the firm's problem, i.e. firms choose both N and b to maximize profits, given the learning costs γ associated with its choice of (N, b). The firms' production function is given by

$$Q = A(N, z, c_s)bl, \tag{40}$$

where $A(N, z, c_s)$ denotes worker productivity and b denotes the level of intensive human capital that a worker hired in z has.

[68] The first set of assumptions says that the cost of acquiring human capital is an increasing function of the level of both intensive and extensive human capital. The second set of assumptions says that the marginal costs are also increasing functions.

[69] The assumption builds on the idea that learning, in general, is more efficient when there is more knowledge available in the local labor market. See Davis and Dingel (2014) for theoretical discussion and De la Roca and Puga (2017) for empirical evidence on this assumption.

[70] The case for multiple sectors is slightly more complicated. In that case, it is possible that larger cities consist of firms producing in multiple sectors. Hence, $\mathbf{K}(L)$ may also vary by city size. However, so long as the elasticity of $\mathbf{K}(L)$ with respect to L is smaller than that of $\mathbf{b}(L)$—which can be proved true under regularity conditions—we still get back the same results.

The firm's problem is therefore

$$\max_{N,b,L} \pi_s(z,L,N,b) = \max_{N,b,L} \kappa \left(\frac{A(N,z,c_s)b}{(1+\gamma(N,b,L))w(L)} \right)^{\sigma-1} RP^{\sigma-1}. \tag{41}$$

It is straightforward to prove that the profit function is log-supermodular in (N,b,z,L). Using the classic theorem of monotone comparative statics in Topkis (1978), if the firm chooses b optimally, given (N,z,L), the resulting profit function would be log-supermodular in (N,z,L), and in $(N,\ L)$. The intuition is simple. Given $\gamma_{bK} > 0$, I have $\gamma_{bN} < 0$—i.e., the marginal cost of acquiring intensive human capital b for firms with greater division of labor is lower. Given, $\gamma_{bL} < 0$, the marginal cost of acquiring intensive knowledge is lower in larger cities. Combining these two assumptions, when b is optimally chosen, firms with higher N benefit more from being in larger cites due to the lower learning costs there, leading to the complementarity between N and L in the profit function. I can, therefore, define the cost function, H, in Equation (6) as

$$H(N,L,b) \equiv \frac{b}{1+\gamma(N,b,L)}.$$

When b is optimally chosen, $H(N,L)$ displays log-supermodularity in (N,L).

C Quasi-experiment

	Treatment	Control	All
Log (no of occupations)	1.41	1.16	1.33
	(.86)	(.85)	(.85)
Specialization index	.428	.379	.412
	(.264)	(.272)	(.268)
Share of managers	.105	.089	.100
	(.147)	(.140)	(.145)
Log (estab size)	2.64	2.32	2.53
	(1.32)	(1.28)	(1.30)
Log (city size)	5.52	4.52	5.18
	(1.76)	(1.05)	(1.79)
Total	.70	.30	1.00

Data source: RAIS 2010. Establishments are considered treated if distance to the nearest broadband backbone is less than 250 km. Standard deviations are shown in parentheses.

Table 4: Establishment-level characteristics before the treatment

C.1 Robustness tests

In this section, I detail the robustness tests I run for the regressions specified by Equations (25), (26) and (27) in Section 4. I start by showing the pre-trends using the specialization index as the alternative definition for division of labor. As shown in Figure 4, the paths of growth over time between the treatment and control groups are almost identical to each other before the treatment, similar to the trends depicted in Figure 3. The trends started to converge after the treatment, showing the effects of new broadband infrastructure on division of labor within establishments.

Next, I present the results when I change the radius around the backbone network used to define connection status. In Table 6, I show how the results change when the radius used to define connectivity is varied. In all cases, the results remain qualitatively and quantitatively similar to those in Table 3. This shows that the estimated effect of better ICT infrastructure is not sensitive to the definition of connectivity used. In addition to confirming that the estimated effect is not sensitive to varying the distance, the findings in Table 6 are useful because they reduce concerns one might have about potential violations of the Stable Unit Treatment Value Assumptions (SUTVA), which could lead to underestimation (if, e.g., establishments relocate) or overestimation (if, e.g., establishments in untreated areas suffer from fast internet access in the neighboring areas) of the effect of broadband internet access. Note that no significant effect of broadband internet access on the relocation of firms is found when I investigate this possibility directly.

While I focus on the impact of the ICT infrastructure on firms' division of labor in this paper, extensive literature has found that technological changes such as fast internet tend to benefit skilled workers and hurt low skilled workers, i.e., skill-biased technological change.[71] In Online Appendix, I show that faster internet connection indeed increases skill intensities within establishments in affected cities. If the codes for skilled occupations are more finely divided, then the increase in the total number of occupation codes in response to the new ICT infrastructure may simply reflect a shift towards more skilled occupations within an establishment. To investigate this, I separate occupation codes into two groups based on skill intensities of the workers within an occupation.[72] As shown in Table 30, the baseline results continue to hold when I estimate the impact of the ICT infrastructure for high and low skilled occupations separately.

Next, I vary the samples used for the regressions in several ways. I first exclude multiple-establishment firms to account for the possibility that firms relocate their resources across different establishments in response to the new ICT infrastructure. As shown in Table 8, The results are essentially unchanged.

While I argue in Section 4 that the alignment of the broadband backbones is exogenous conditional on observables as they follow the alignment of existing infrastructure, the locations for the origin and destination locations may be chosen endogenously, in anticipation to potential changes in certain economic outcomes in those locations. To account for the possible violation of the identifying assumption, I drop these terminal locations. The results, shown in Table 9, the results are not sensitive to excluding establishments in locations

[71] See Acemoglu and Autor (2011) for a review of the literature on skill-biased technological changes, and Hjort and Poulsen (2016) for direct evidence that impacts from improvement in ICT infrastructure is skill biased.

[72] A skilled occupation is one in which the share of high-skilled workers within that occupation is above the median of all occupation codes. Following conventional literature, I define high-skilled workers as those with some college degree or above.

where the new national backbone starts or ends.

Submarine cable landing points, in addition to being on the coast, are also typically in or near large cities. If such places were on a different trend in the outcomes of interest before the new backbones are introduced, I may incorrectly attribute an estimated treatment effect to the availability of the new broadband network. In Table 10, I exclude, from my sample, all establishments in locations closer than 100 kilometers from a landing point. The results remain robust.

Going by the similar logic, areas that had broadband access before PNBL tend to be larger or more densely populated cities. These places may also grow along a different path than other locations. To account for this, I drop all the establishments in locations that had already been connected to the broadband network before PNBL in 2009. As shown in Table 11, the baseline regression results continue to hold. The coefficients to test the heterogeneity in the treatment across cities and sectors are less precisely estimated, while still remaining positive. The latter regressions lack power because that more than half of the sample is dropped.

In Table 12, I restrict attention to connected locations and thus estimate the effect of better internet in the sub-sample consisting only to eventually treated establishments. In this case, the comparison group for establishments in a year when a location became connected to a new backbone cable consists of other establishments in the same year but in locations that did not have the new cable at that time. I thus prefer my baseline approach as outlined earlier to the one used in Table 12, but it is nonetheless reassuring that the estimated effect of access to broadband network to various establishment-level variables, if anything, is bigger in magnitude and remains significant when only establishments in connected locations are included in the sample.

For the next three robustness tests, I drop from the sample establishments located in areas that may grow on a different path from the other firms. In Table 13, I exclude establishments that are either very near (<10th percentile) or very far (>90th percentile) from the backbone network. In Table 14, I only consider establishments in urban areas by dropping establishments located in microregions with a density lower than 400 persons/km^2.[73] In Table 15, I drop establishments in very large cities.[74] The results are robust to all three tests.

In Table 16, I separate firms into two groups based on their sectoral share of exports. This is to account for the possibility that the results are driven by more export-oriented firms. As shown in the table, baseline results hold for both types of establishments.

The alignment of the new broadband backbones was announced in 2010. It is possible that establishments in the treated locations had anticipated the impending new infrastructure and started adjusting their organization structures prior to the actual implementation of the new backbones. If this was true, I may underestimate the true effects of the new infrastructure on division of labor. In Table 17, I drop the observations in 2010 and 2011. The estimates remain essentially the same as the baseline results.

Results in Table 18 control for municipality-specific linear trends. Including these restrictive controls have remarkably little effect on the magnitude and significance of the estimated effect of access to broadband

[73] This is based on World Bank's definition for urban *versus* rural areas

[74] "Very large cities" are defined as the top 10-percentile of the microregions in terms of city size.

on establishment-level variables of interest. In Table 19, I include two lead variables of $Backbone_{jt}$. The estimates on the two leads are zero, supporting the assumption of parallel trends.

There may also be potential concerns about spatial correlation in the error term. Cameron and Miller (2015) note that failure to account for such dependence may lead to over-rejection of the null hypothesis. To address this concern, I follow Conley (1999) to allow for serial correlation over all time periods, as well as spatial correlation among establishments that fall within 100km of each other. As shown in Table 20, the results are robust when I account for possible spatial correlation.

Lastly, a concern in DiD analysis is that serial correlation can bias standard errors, leading to over-rejection of the null hypothesis of no effect (Bertrand, Duflo and Mullainathan, 2004). I follow Chetty, Looney and Kroft (2009) to address this concern through a non-parametric permutation test for $\beta = 0$ in Equation (25). To do so, I sample from the set of true broadband backbone implementation years observed in the data and assign a randomly chosen "fake" treatment time to each municipality while maintaining the alignment of the new backbones thus keeping each observation's connectivity status. Defining $G(\beta^p)$ to be the empirical cumulative distribution functions of these placebo effects, the statistic $1 - G(\beta)$ gives a p-value for the hypothesis that $\beta = 0$. Intuitively, if the new broadband backbones had a significant effect on the number of occupations, the estimated coefficient for β should be in the upper tail of estimated placebo effects. Since this test does not make parametric assumptions about the error structure, it does not suffer from the over-rejection bias of the t-test. Figure 5 illustrates the empirical distributions of placebo effects G for $\hat{\beta}$ from performing the permutation tests 4000 times. The vertical lines denote the effect of new broadband backbone to treated areas. The implied p-values are 0.001 and 0.005, for division of labor measured by the log number of occupations and specialization index, respectively. These are very similar to the estimates from the t-tests as reported in Table 3.

Separately, I also consider the specification, in which I incorporate both interaction terms into a single regression equation. More specifically, I run

$$\log N_{jt} = \alpha + \beta Backbone_{jt} + \gamma Backbone_{jt} \times \log L_{c(j),t_0} + \upsilon Backbone_{jt} \times \log c_{s(j),t_0} + \delta_j + \delta_t + \theta_{m(j)} \times t + \varepsilon_{jt}. \quad (42)$$

As shown in Table 21, the results remain qualitatively very similar to the baseline estimates in which I separately identify the interaction effects of the treatment with city size, and with sector-level complexity.

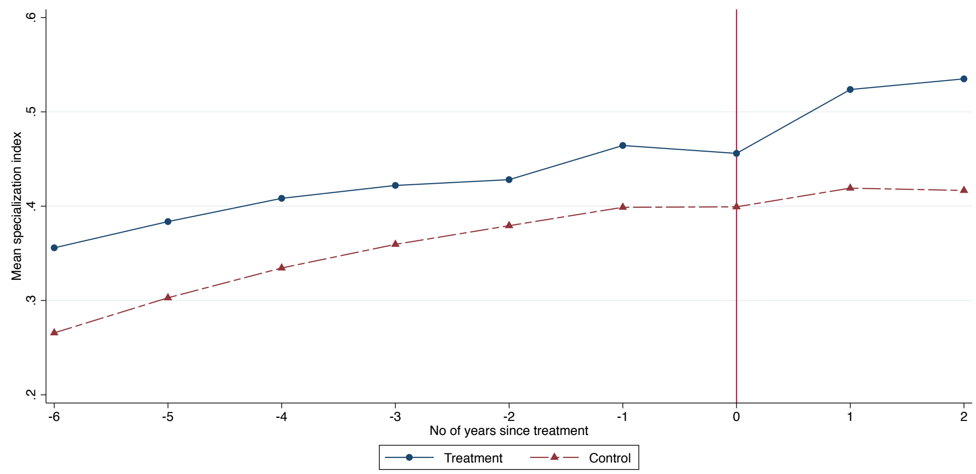


Figure 4: Specialization index in treated versus control groups in Brazil

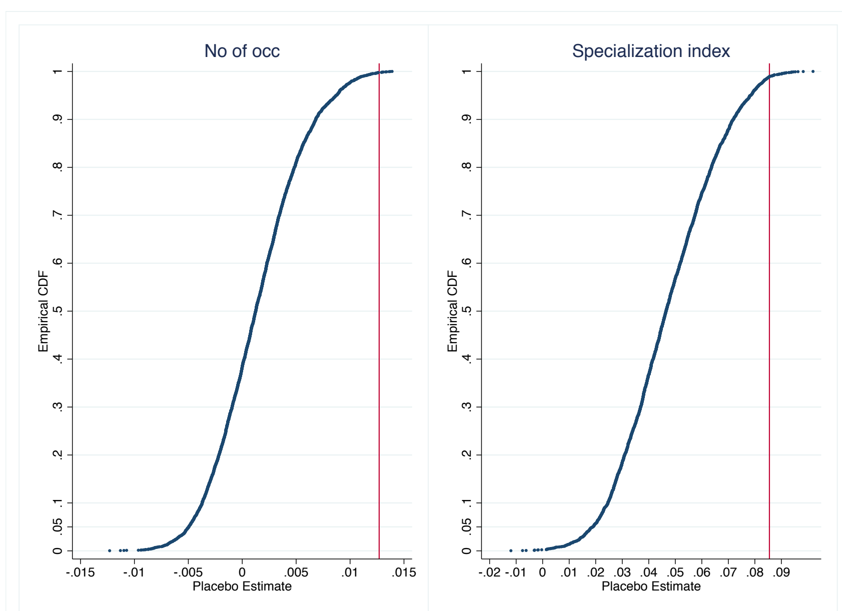


Figure 5: Distribution of placebo estimates

This figure shows a non-parametric permutation test of $\beta = 0$. I sample from the set of true broadband backbone implementation years observed in the data, assigning a randomly chosen "fake" time to each location with equal probability while maintaining each observation's backbone connectivity status. The figure depicts the empirical cdf of estimates resulting from permuting trajectories 4,000 times and running Equation (25) on the fake datasets. The vertical lines represent the true estimates; where these fall in empirical cdf of estimates from datasets with permuted trajectories implies their p-values. The implied p-values are 0.0022 for the log number of occupations and 0.011 for the specialization index. These can be compared to 0.007 and 0.000 from Table 3.

Dependent variable	Log (No of occs)				Specialization index			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Interm. inputs	G3 exp share		Interm. inputs	G3 exp share	
Radius: 100km								
$Backbone_{jt}$	.0072***	-.0009	-.0007	-.0029	.0562***	-.0003	.0447***	.0513***
	(.0025)	(.0027)	(.0037)	(.0029)	(.0158)	(.0107)	(.013)	(.0148)
$Backbone_{jt} \times \log L_{ct_0}$		.0095***				.014***		
		(.0009)				(.0035)		
$Backbone_{jt} \times \log c_{st_0}$			.0125***	.0046***			.0141***	.0059***
			(.0033)	(.0013)			(.0043)	(.0013)
Radius: 200km								
$Backbone_{jt}$	.0108***	-.0011	.0002	.0053*	.0722***	.0005	.0606***	.0674***
	(.0026)	(.0029)	(.0037)	(.003)	(.0173)	(.0099)	(.0144)	(.0164)
$Backbone_{jt} \times \log L_{ct_0}$		.0084***				.0136***		
		(.0008)				(.0035)		
$Backbone_{jt} \times \log c_{st_0}$			.0131***	.0043***			.0143***	.0061***
			(.0031)	(.0012)			(.0044)	(.0013)
Mean of outcome	1.45	1.45	1.45	1.45	.43	.43	.43	.43
Obs	777096	777096	777096	777096	777096	777096	777096	777096
R-sq	.853	.853	.853	.854	.716	.717	.716	.716

Robust standard errors clustered by municipality in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term, establishment and year FEs.

Table 5: Broadband connection and division of labor, connection radius 100km and 200km

Dependent variable	Log (No of occs)				Specialization index			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Interm. inputs	G3 exp share		Interm. inputs	G3 exp share	
Radius: 300km								
$Backbone_{jt}$	.0146***	.0033	.0041	.0096***	.0973***	.0267***	.0851***	.0925***
	(.0031)	(.0033)	(.004)	(.0034)	(.0191)	(.0084)	(.0162)	(.0181)
$Backbone_{jt} \times \log L_{ct_0}$		.0076***				.0136***		
		(.0008)				(.0035)		
$Backbone_{jt} \times \log c_{st_0}$			.0131***	.004***			.0151***	.0062***
			(.0031)	(.0012)			(.0043)	(.0013)
Radius: 400km								
$Backbone_{jt}$	.0098**	-.0037	-.0004	.005	.0869***	.0234**	.0749***	.082***
	(.0047)	(.005)	(.0053)	(.005)	(.0193)	(.0104)	(.0164)	(.0184)
$Backbone_{jt} \times \log L_{ct_0}$		.0081***				.0127***		
		(.0008)				(.0039)		
$Backbone_{jt} \times \log c_{st_0}$			.0128***	.004***			.015***	.0063***
			(.003)	(.0012)			(.0045)	(.0014)
Mean of outcome	1.45	1.45	1.45	1.45	.43	.43	.43	.43
Obs	777096	777096	777096	777096	777096	777096	777096	777096
R-sq	.853	.853	.853	.854	.716	.718	.717	.717

Robust standard errors clustered by municipality in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term, establishment and year FEs.

Table 6: Broadband connection and division of labor, connection radius 300km and 400km

Dependent variable	Log (No of occs)				Specialization index			
	(1)	(2)	(3) Interm. inputs	(4) G3 exp share	(5)	(6) Interm. inputs	(7) G3 exp share	(8)
				Low-skill occupations				
$Backbone_{jt}$	.0931***	.0069**	.003	.0016	.063***	.00536	.0621***	.0641***
	(.0027)	(.0029)	(.0035)	(.0031)	(.0109)	(.0077)	(.0114)	(.0106)
$Backbone_{jt} \times \log L_{ct_0}$		.0078***				.0117***		
		(.0007)				(.0023)		
$Backbone_{jt} \times \log c_{st_0}$			.0075***	.0033***			.01***	.0025***
			(.0028)	(.0012)			(.0022)	(.0009)
Mean of outcome	1.12	1.12	1.12	1.12	.56	.56	.56	.56
Obs	777096	777096	777096	777096	777096	777096	777096	777096
R-sq	.835	.835	.835	.835	.618	.618	.618	.618
				High-skill occupations				
$Backbone_{jt}$	.0131***	.0012	.0027	.0077**	.0905***	.0052	.0581***	.0478***
	(.0036)	(.0038)	(.0049)	(.0039)	(.0116)	(.0095)	(.0164)	(.0125)
$Backbone_{jt} \times \log L_{ct_0}$		.0093***				.02***		
		(.0009)				(.003)		
$Backbone_{jt} \times \log c_{st_0}$			.0193***	.0042***			.0198***	.0087***
			(.0037)	(.0012)			(.0064)	(.0012)
Mean of outcome	.88	.88	.88	.88	.44	.44	.44	.44
Obs	469224	469224	469224	469224	469224	469224	469224	469224
R-sq	.818	.818	.818	.819	.68	.68	.68	.681

Robust standard errors clustered by municipality in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term, establishment and year FEs.

Table 7: Impacts of fast internet on division of labor within establishments: separating high and low-skill occupations

Dependent variable	Log (No of occs)				Specialization index			
	(1)	(2)	(3) Interm. inputs	(4) G3 exp share	(5)	(6)	(7) Interm. inputs	(8) G3 exp share
$Backbone_{jt}$	.0128***	.0022	.0021	.0097***	.088***	.0114	.0744***	.0829***
	(.0029)	(.0031)	(.0039)	(.0033)	(.0173)	(.0089)	(.0141)	(.0163)
$Backbone_{jt} \times \log L_{ct_0}$		.0072***				.0147***		
		(.0008)				(.0035)		
$Backbone_{jt} \times \log c_{st_0}$			.0132***	.0021			.017***	.0068***
			(.0032)	(.0013)			(.0047)	(.0015)
Mean of outcome	1.43	1.43	1.43	1.43	.42	.42	.42	.42
Obs	721629	721629	721629	721629	721629	721629	721629	721629
R-sq	.851	.851	.851	.851	.713	.715	.713	.714

Robust standard errors clustered by municipality in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term, establishment and year FEs.

Table 8: Broadband connection and division of labor, only mono-establishment firms

Dependent variable	Log (No of occs)				Specialization index			
	(1)	(2)	(3) Interm. inputs	(4) G3 exp share	(5)	(6)	(7) Interm. inputs	(8) G3 exp share
$Backbone_{jt}$	.0107***	.0003	-.0003	.0051	.0846***	.0091	.0714***	.0793***
	(.0029)	(.0032)	(.0039)	(.0033)	(.0171)	(.0085)	(.014)	(.0161)
$Backbone_{jt} \times \log L_{ct_0}$		.0067***				.0147***		
		(.0008)				(.0034)		
$Backbone_{jt} \times \log c_{st_0}$			.0137***	.0041***			.0165***	.0069***
			(.0032)	(.0013)			(.0045)	(.0014)
Mean of outcome	1.45	1.45	1.45	1.45	.43	.43	.43	.43
Obs	738702	738702	738702	738702	738702	738702	738702	738702
R-sq	.853	.853	.853	.854	.715	.717	.715	.715

Robust standard errors clustered by municipality in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term, establishment and year FEs.

Table 9: Broadband connection and division of labor, excluding origin and destination cities

Dependent variable	Log (No of occs)				Specialization index			
	(1)	(2)	(3) Interm. inputs	(4) G3 exp share	(5)	(6)	(7) Interm. inputs	(8) G3 exp share
$Backbone_{jt}$	.0104***	.0026	-.0058	.0068*	.062***	.0113	.0548***	.0585***
	(.0032)	(.0041)	(.0042)	(.0037)	(.0072)	(.0074)	(.0065)	(.0069)
$Backbone_{jt} \times \log L_{ct_0}$		.0038***				.011***		
		(.0013)				(.0017)		
$Backbone_{jt} \times \log c_{st_0}$			.0207***	.0026*			.0092***	.0048***
			(.0035)	(.0014)			(.0028)	(.0009)
Mean of outcome	1.41	1.41	1.41	1.41	.43	.43	.43	.43
Obs	606294	606294	606294	606294	606294	606294	606294	606294
R-sq	.85	.85	.85	.85	.719	.72	.719	.72

Robust standard errors clustered by municipality in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term, establishment and year FEs.

Table 10: Broadband connection and division of labor, excluding locations within 100km of submarine cable landing points

Dependent variable	Log (No of occs)				Specialization index			
	(1)	(2)	(3) Interm. inputs	(4) G3 exp share	(5)	(6)	(7) Interm. inputs	(8) G3 exp share
$Backbone_{jt}$	.0159***	-.0022	-.0081	.0135***	.0456***	.0299***	.0429***	.0428***
	(.0044)	(.0052)	(.0056)	(.005)	(.0071)	(.008)	(.0066)	(.0068)
$Backbone_{jt} \times \log L_{ct_0}$		.0086***				.0034*		
		(.0014)				(.002)		
$Backbone_{jt} \times \log c_{st_0}$			.0314***	.002			.0036	.0037***
			(.0044)	(.0018)			(.0031)	(.001)
Mean of outcome	1.37	1.37	1.37	1.37	.41	.41	.41	.41
Obs	388539	388539	388539	388539	388539	388539	388539	388539
R-sq	.847	.847	.847	.847	.72	.72	.72	.72

Robust standard errors clustered by municipality in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term, establishment and year FEs.

Table 11: Broadband connection and division of labor, dropping establishments that were connected to the broadband network before PNBL

Dependent variable	Log (No of occs)				Specialization index			
	(1)	(2)	(3) Interm. inputs	(4) G3 exp share	(5)	(6)	(7) Interm. inputs	(8) G3 exp share
$Backbone_{jt}$	.0144***	.0031	.0032	.009***	.0923***	.0184**	.0797***	.0873***
	(.0027)	(.003)	(.0037)	(.0031)	(.0181)	(.0083)	(.0152)	(.0172)
$Backbone_{jt} \times \log L_{ct_0}$		.0077***				.0141***		
		(.0008)				(.0033)		
$Backbone_{jt} \times \log c_{st_0}$			.0139***	.004***			.0156***	.0064***
			(.0031)	(.0012)			(.0043)	(.0013)
Mean of outcome	1.46	1.46	1.46	1.46	.43	.43	.43	.43
Obs	764541	764541	764541	764541	764541	764541	764541	764541
R-sq	.854	.854	.854	.854	.717	.719	.717	.717

Robust standard errors clustered by municipality in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term, establishment and year FEs.

Table 12: Broadband connection and division of labor, including only establishments that were eventually treated

Dependent variable	Log (No of occs)				Specialization index			
	(1)	(2)	(3) Interm. inputs	(4) G3 exp share	(5)	(6)	(7) Interm. inputs	(8) G3 exp share
$Backbone_{jt}$	.0164***	.0067	.0037	.0104**	.1112***	.0068	.0935***	.1043***
	(.0041)	(.0043)	(.0052)	(.0045)	(.0223)	(.0134)	(.0196)	(.0215)
$Backbone_{jt} \times \log L_{ct_0}$		.0066***				.02***		
		(.001)				(.0014)		
$Backbone_{jt} \times \log c_{st_0}$			.0154***	.0043***			.0216***	.0085***
			(.0039)	(.0016)			(.0046)	(.0013)
Mean of outcome	1.51	1.51	1.51	1.51	.42	.42	.42	.42
Obs	450792	450792	450792	450792	450792	450792	450792	450792
R-sq	.862	.862	.862	.862	.716	.719	.716	.716

Robust standard errors clustered by municipality in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term, establishment and year FEs.

Table 13: Broadband connection and division of labor, excluding establishments that are very near or far from the backbones

Dependent variable	Log (No of occs)				Specialization index			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Interm. inputs	G3 exp share			Interm. inputs	G3 exp share
$Backbone_{jt}$	.0311***	.0311***	.0199***	.0278***	.1429***	-.0049	.1292***	.1394***
	(.0039)	(.0039)	(.0058)	(.0043)	(.0357)	(.0293)	(.0352)	(.0355)
$Backbone_{jt} \times \log L_{ct_0}$		.0114***				.0221***		
		(.0017)				(.0057)		
$Backbone_{jt} \times \log c_{st_0}$			.0123**	.0022			.015***	.0037***
			(.0048)	(.0016)			(.0042)	(.0011)
Mean of outcome	1.57	1.57	1.57	1.57	.46	.46	.46	.46
Obs	372726	372726	372726	372726	372726	372726	372726	372726
R-sq	.857	.857	.857	.857	.709	.711	.709	.71

Robust standard errors clustered by municipality in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term, establishment and year FEs.

Table 14: Broadband connection and division of labor, excluding establishments located in rural areas

Dependent variable	Log (No of occs)				Specialization index			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Interm. inputs	G3 exp share		Interm. inputs	G3 exp share	
$Backbone_{jt}$	.0124***	-.004	-.0019	.0086***	.0648***	.0226***	.0575***	.0614***
	(.0028)	(.0033)	(.0039)	(.0033)	(.0013)	(.0024)	(.0016)	(.0014)
$Backbone_{jt} \times \log L_{ct_0}$		.009***				.0086***		
		(.0011)				(.0004)		
$Backbone_{jt} \times \log c_{st_0}$			.0182***	.0028**			.0093***	.0046***
			(.0033)	(.0013)			(.0013)	(.0005)
Mean of outcome	1.41	1.41	1.41	1.41	.44	.44	.44	.44
Obs	705861	705861	705861	705861	705861	705861	705861	705861
R-sq	.85	.85	.85	.85	.72	.721	.72	.721

Robust standard errors clustered by municipality in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term, establishment and year FEs.

Table 15: Broadband connection and division of labor, excluding establishments located in very large cities

Dependent variable	Log (No of occs)				Specialization index			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Interm. inputs	G3 exp share		Interm. inputs	G3 exp share	
Export-intensive industries								
$Backbone_{jt}$	.0182***	.0146***	.0064	.023***	.0964***	.0212**	.0756***	.0864***
	(.0047)	(.005)	(.0058)	(.0051)	(.0177)	(.0093)	(.013)	(.0152)
$Backbone_{jt} \times \log L_{ct_0}$		.0027**				.014***		
		(.0013)				(.0028)		
$Backbone_{jt} \times \log c_{st_0}$			.0149***	.0051**			.0263***	.0092***
			(.004)	(.0021)			(.0064)	(.0027)
Mean of outcome	1.59	1.59	1.59	1.59	.4	.4	.4	.4
Obs	307872	307872	307872	307872	307872	307872	307872	307872
R-sq	.857	.857	.857	.857	.72	.722	.721	.721
Others								
$Backbone_{jt}$	.0131***	.0012	.0027	.0077**	.0905***	.0052	.0581***	.0478***
	(.0036)	(.0038)	(.0049)	(.0039)	(.0116)	(.0095)	(.0164)	(.0125)
$Backbone_{jt} \times \log L_{ct_0}$		.0093***				.02***		
		(.0009)				(.003)		
$Backbone_{jt} \times \log c_{st_0}$			.0193***	.0042***			.0198***	.0087***
			(.0037)	(.0012)			(.0064)	(.0012)
Mean of outcome	.88	.88	.88	.88	.34	.34	.34	.34
Obs	469224	469224	469224	469224	469224	469224	469224	469224
R-sq	.818	.818	.818	.819	.68	.68	.68	.681

Robust standard errors clustered by municipality in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term, establishment and year FEs.

Table 16: Broadband connection and division of labor, separating firms based on export intensity

Dependent variable	Log (No of occs)				Specialization index			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Interm. inputs	G3 exp share			Interm. inputs	G3 exp share
$Backbone_{jt}$	.0141***	.0016	.0022	.0081**	.0864***	.0092	.0734***	.0812***
	(.0031)	(.0034)	(.0043)	(.0036)	(.0176)	(.0084)	(.0145)	(.0166)
$Backbone_{jt} \times \log L_{ct_0}$		.0086***				.0147***		
		(.001)				(.0034)		
$Backbone_{jt} \times \log c_{st_0}$			.0148***	.0045***			.0161***	.0068***
			(.0036)	(.0014)			(.0046)	(.0014)
Mean of outcome	1.44	1.44	1.44	1.44	.44	.44	.44	.44
Obs	604408	604408	604408	604408	604408	604408	604408	604408
R-sq	.846	.846	.846	.846	.702	.704	.703	.703

Robust standard errors clustered by municipality in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term, establishment and year FEs.

Table 17: Broadband connection and division of labor, excluding observations in Year 2010 and 2011

Dependent variable	Log (No of occs)			Specialization index		
	(1)	(2)	(3)	(4)	(5)	(6)
		Interm. inputs	G3 exp share		Interm. inputs	G3 exp share
$Backbone_{jt}$	.0151***	.0021	.0106***	.0891***	.0789***	.0851***
	(.0026)	(.0036)	(.003)	(.0183)	(.0158)	(.0175)
$Backbone_{jt} \times \log c_{st_0}$		.0159***	.0033***		.0125***	.0051***
		(.0031)	(.0012)		(.0039)	(.0012)
Mean of outcome	1.45	1.45	1.45	.43	.43	.43
Obs	777096	777096	777096	777096	777096	777096
R-sq	.854	.854	.855	.718	.718	.719

Robust standard errors clustered by municipality in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term, establishment and year FEs.

Table 18: Broadband connection and division of labor, with microregion-specific trend

Dependent variable	Log (No of occs)			Specialization index		
	(1)	(2)	(3)	(4)	(5)	(6)
$Backbone_{jt}$	.0127***	.0122***	.0126***	.0855***	.0843***	.0849***
	(.0047)	(.0045)	(.0047)	(.017)	(.0169)	(.0171)
$Lead_{j,t-1}$		-.0043	-.004		.0098	.0094
		(.0029)	(.0027)		(.04)	(.039)
$Lead_{j,t-2}$			.0021			.0034
			(.0028)			(.0022)
Mean of outcome	1.45	1.45	1.45	.43	.43	.43
Obs	777096	777096	777096	777096	777096	777096
R-sq	.853	.853	.853	.717	.717	.717

Robust standard errors clustered by municipality in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term, establishment and year FEs.

Table 19: Broadband connection and division of labor, with lead controls

Dependent variable	Log (No of occs)				Specialization index			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Interm. inputs	G3 exp share		Interm. inputs	G3 exp share	
$Backbone_{jt}$	.0127***	.0015	.0015	.0074	.0855***	.0116	.0728***	.0805**
	(.0049)	(.006)	(.006)	(.005)	(.036)	(.0102)	(.034)	(.039)
$Backbone_{jt} \times \log L_{ct_0}$		.0077***				.0141***		
		(.0016)				(.0065)		
$Backbone_{jt} \times \log c_{st_0}$			.0139***	.004*			.0156*	.0064***
			(.0047)	(.0021)			(.0084)	(.0022)
Mean of outcome	1.45	1.45	1.45	1.45	.43	.43	.43	.43
Obs	777096	777096	777096	777096	777096	777096	777096	777096
R-sq	.853	.853	.853	.854	.717	.718	.717	.717

Conley standard errors in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term, establishment and year FEs.

Table 20: Broadband connection and division of labor, with Conley Standard Errors

Dependent variable	Log (No of occs)		Specialization index	
	(1) Interm. inputs	(2) G3 exp share	(3) Interm. inputs	(4) G3 exp share
$Backbone_{jt}$	-.0017	-.001	.0092	.0112
	(.0041)	(.0033)	(.0081)	(.0084)
$Backbone_{jt} \times \log L_{ct_0}$	.0089***	.0075***	.0138***	.0138***
	(.0008)	(.0008)	(.0034)	(.0034)
$Backbone_{jt} \times \log c_{st_0}$	.021***	.002*	.005**	.003***
	(.0032)	(.001)	(.002)	(.001)
Mean of outcome	1.45	1.45	.43	.43
Obs	777096	777096	777096	777096
R-sq	.854	.854	.718	.719

Robust standard errors clustered by municipality in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term, establishment and year FEs.

Table 21: Broadband connection and division of labor, combining two interactions

D Moments and identification

Average city-level change in division of labor:

The first set of moments is the average change in division of labor across each quartile of city sizes. I run the following regression for each quartile of city size, $q = 1 \dots 4$:

$$\log N_{jt} = \alpha_q + \beta_q Backbone_{jt} + \delta_{qj} + \delta_{qt} + \varepsilon_{jt},$$

I then use the simulated economy to match the regression estimates $\hat{\beta}_q$. To do so, I calibrate changes in cost of division of labor to match the average treatment effect of 1.27% in Table 3. I then feed the change in cost of division of labor into the model to predict the change in division of labor, assuming that firms do not relocate spatially. I finally compute the average change in division of labor within each quartile of city sizes. Due to complementarities between division of labor and complexity (c) and between division of labor and city size (θ), firms in larger cities would increase their division of labor to a larger extent in response to a reduction in the cost of division of labor. Therefore, this set of moments helps to identify c and θ.

Geographic distribution of firms:

The second set of moments I use is the share of sectoral employment that falls into one of the four city-size bins. City-size bins are obtained by ordering cities by their sizes and creating bins using the threshold cities with less than 25%, 50% and 75% of the overall sectoral workforce. They describe the geographic distribution of economic activities at the sector level and hence give information on the density of firms located in different city sizes. Therefore, they help to identify the distribution of firm complexities, i.e., ν_z.

Firm-size distribution:

The third set of moments is the share of firms that fall within the five bins of normalized firm labor payment.

These bins are defined by the 25, 50, 75 and 90th percentiles of the distribution of firm sizes measured in labor payment. The firm-size distribution is affected by the distributions of firm complexities and firm-city-size idiosyncratic shocks. These five moments allow me to identify ν_L and ν_z separately. Intuitively, ν_z affects the relative quantiles of the firm-size distribution both indirectly, through the matching function, and directly, through the distribution of firm complexity z. In contrast, ν_L only affects the relative quantiles directly, through the matching function.

Increases in the average division of labor across city sizes:
To measure increases in the average division of labor across city sizes, I consider 4 moments, i.e., the average firms' division of labor within each quartile of city size. These four moments, together with the first set of moments, contribute to the identification of (c, θ) separately from α—the reduced-form agglomeration externalities—and υ—the interaction between city size and complexity. As city size increases, firm productivity increases through α, υ and (c, θ). However, these channels differ importantly: (c, θ) can only increase firm productivity through division of labor, whereas α or υ increases firm productivity directly and does not affect firms' division of labor.

Increases in the average firm size across city sizes:
To measure increases in the average firm size across city sizes, I consider 4 moments, i.e., the average firm size measured in labor payment within each quartile of city size. This set of moments contribute to the identification of α—the reduced-form agglomeration externalities—separately from υ—the interaction between city size and complexity, which jointly determine the sorting of firms across space with (c, θ). As mentioned, α, υ and (c, θ) all affect how firm productivity increases as city size increases. Given (c, θ), we can separately identify α from υ, because there is an interaction between firm complexity and city size through υ, which pushes the productivity up more than linearly (through α), since the latter does not interact with firm complexities.

Within-city variations in firms' division of labor:
To summarize within-city variations in firms' division of labor, I use the variance of firms' division of labor in each quartile of city sizes. These four moments help to separately identify c and θ. Given a city size, the impact of city size on division of labor is the same for all firms there. I can, therefore, identify the complementarity between division of labor and complexity (i.e., c) using the within-city variation in firms' division of labor, relative to that in firm complexities. Intuitively, all else equal, small changes in firm complexity would generate large variation in division of labor, if the complementarity is strong.

E Estimation results

Sector	$\hat{\alpha}$	$\hat{\upsilon}$	$\hat{c}$	$\hat{\theta}$	$\hat{\nu_z}$	$\hat{\nu_L}$
Agriculture, and mining	0.133	n.d.	n.d.	n.d.	0.458	0.834
	(0.141)	(–)	(–)	(–)	(0.815)	(1.675)
Mfg of food products, beverages and tobacco products	0.167	0.001	0.296	0.8177	0.268	0.109
	(0.008)	(0.306)	(0.130)	(0.185)	(0.055)	(0.084)
Mfg of textiles	0.023	0.010	0.355	0.320	0.375	0.546
	(0.007)	(0.121)	(0.113)	(0.425)	(0.091)	(0.903)
Mfg of wearing apparel	0.063	0.027	0.208	0.412	0.744	0.426
	(0.015)	(0.154)	(0.162)	(0.274)	(0.820)	(0.035)
Mfg of leather goods and footwear, leather tanning	0.047	-0.050	0.118	0.316	0.399	0.298
	(0.124)	(0.978)	(0.204)	(0.198)	(0.141)	(0.098)
Mfg and products of wood, except furniture	0.001	0.014	0.058	0.716	0.540	0.430
	(0.014)	(0.088)	(0.066)	(0.246)	(0.146)	(0.278)
Mfg of pulp, paper and paper products	0.010	0.012	0.036	0.248	0.272	0.452
	(0.055)	(0.006)	(0.098)	(0.300)	(0.685)	(0.581)
Publishing, printing and reproduction of recorded media	0.048	0.012	0.371	0.584	0.607	0.762
	(0.023)	(0.007)	(0.233)	(0.352)	(0.412)	(0.451)
Mfg of chemicals and chemical products	0.015	0.072	0.401	0.772	0.475	0.113
	(0.008)	(0.351)	(0.398)	(0.721)	(0.619)	(0.012)
Mfg of pharmaceutical products	0.146	0.200	0.565	0.234	0.977	0.647
	(0.676)	(0.004)	(0.671)	(0.278)	(0.878)	(0.632)
Mfg of rubber and plastic products	0.034	0.005	0.423	0.130	0.813	0.224
	(0.021)	(0.687)	(0.141)	(0.073)	(0.243)	(0.015)
Mfg of glass, ceramic, brick and cement products	0.046	0.045	0.189	0.078	0.233	0.157
	(0.021)	(0.743)	(0.111)	(0.021)	(0.140)	(0.007)
Mfg of basic metals	0.014	-0.030	0.159	0.264	0.300	0.303
	(0.026)	(0.513)	(0.046)	(0.184)	(0.167)	(0.361)
Mfg of fabricated metal products, except machinery	0.094	0.024	0.340	0.532	0.399	0.707
	(0.019)	(0.320)	(0.049)	(0.116)	(0.088)	(0.147)
Mfg of computer and electronic products	0.073	0.080	0.612	0.252	0.637	0.401
	(0.084)	(0.973)	(0.922)	(0.392)	(0.326)	(0.422)
Mfg of electrical machines	0.090	0.081	0.509	0.178	0.401	0.125
	(0.166)	(0.052)	(0.200)	(0.299)	(0.356)	(0.667)
Mfg of other equipments and machines	0.067	0.023	0.453	0.119	0.239	0.400
	(0.042)	(0.089)	(0.310)	(0.062)	(0.056)	(0.091)
Mfg of automotive vehicles	0.002	0.203	0.601	0.724	0.275	0.139
	(0.010)	(0.076)	(0.099)	(0.366)	(0.084)	(0.564)
Mfg of other transport equipment	0.020	0.240	0.591	0.278	0.647	0.481
	(0.067)	(0.090)	(.911)	(0.986)	(1.246)	(0.988)
Mfg of furniture	0.017	0.011	0.441	0.628	0.827	0.538
	(0.019)	(0.315)	(0.041)	(0.285)	(0.566)	(0.073)
Mfg of miscellaneous products, other mfg activities	0.036	0.359	0.542	0.798	0.836	0.986
	(0.132)	(0.066)	(0.588)	(0.466)	(0.376)	(0.162)

α is the log-linear standard agglomeration coefficient; υ is the log-supermodularity coefficient on the complementarity between complexity and city size; c is the log-supermodulary coefficient on the complementarity between complexity and the division of labor; θ is the log-supermodulary coefficient on the complementarity between the division of labor and city size; ν_z is the variance of firm complexity draws; ν_L is the variance of firm-city size specific shocks.

Table 22: Estimated parameters

F Non-targetted Moments

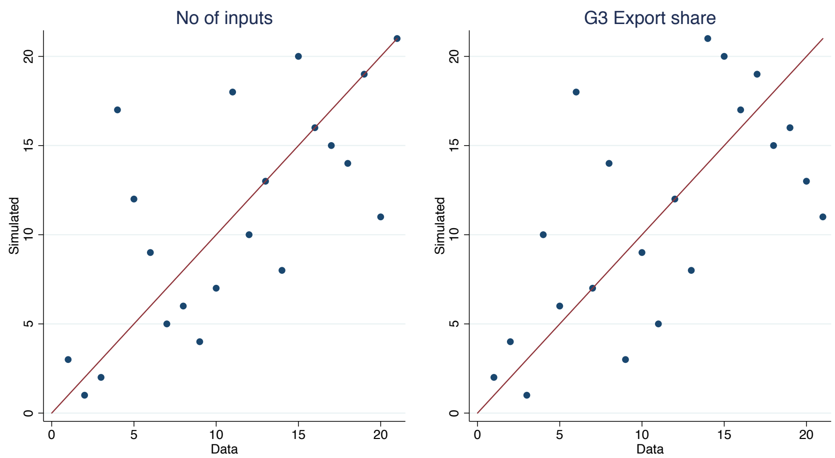


Figure 6: Rank correlations of complexity measures

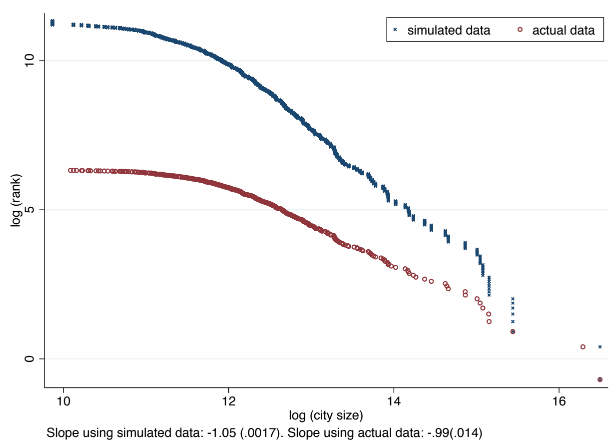
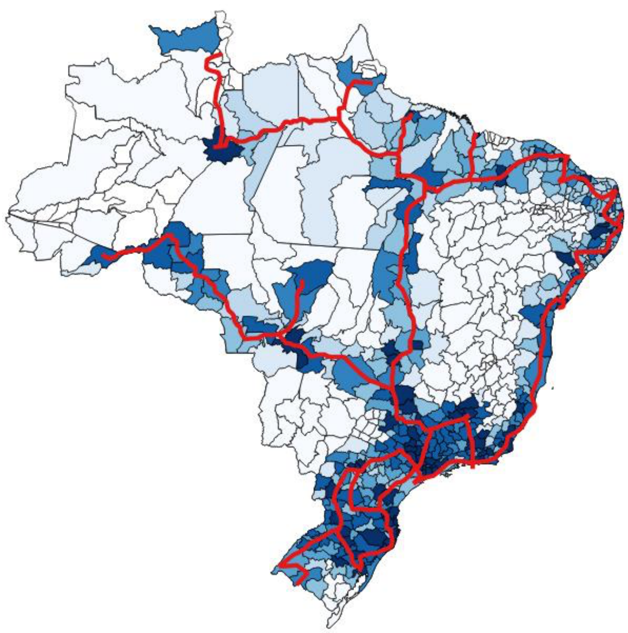
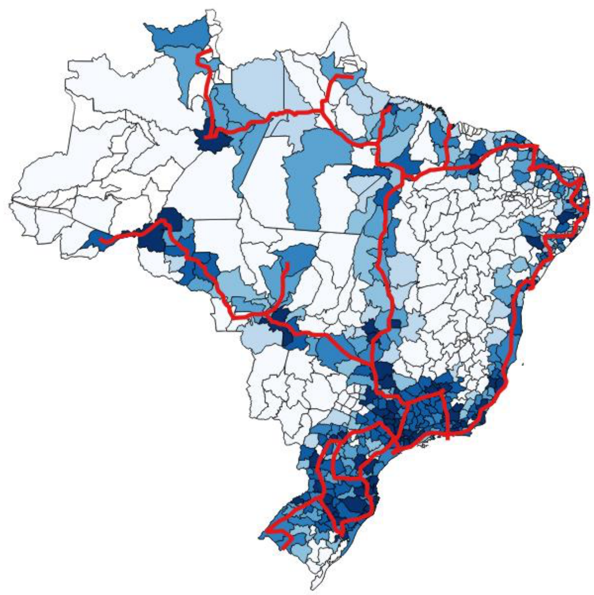


Figure 7: City size distribution



(a) Actual



(b) Predicted

Figure 8: Actual v.s. Simulated changes in firms' division of labor across cities

ONLINE APPENDIX

OA1 Data and stylized facts

OA1.1 Construction of measures for division of labor

In the data exercise, I measure division of labor by the heterogeneity of occupations that are involved in the actual production within an establishment. The baseline definition for division is labor is the number of non-managerial/supervisory occupations codes within an establishment. As an alternative definition, I also consider a normalized measure of the diversity of the occupation codes.

I construct the two measures by first removing occupation codes that are related to managerial or supervisory functions within an establishment.[75] My goal is to identify, out of the 2,544 6-digit CBO codes, the ones that most likely involve managerial or supervisory tasks, from the occupation descriptions.[76] To implement this in a principled manner, I leverage the Latent Dirichlet Allocation (LDA) method (Blei, Ng and Jordan, 2003), a widely-used topic modeling technique in machine learning, to infer a collection of "topics" or "themes" from the occupation descriptions. Using LDA, I first learn a list of "topics" across all code descriptions, where each "topic" can be represented with a collection of keywords. Next, I identify all "topics" that contain words that are derivatives of "manage" and "supervise". Finally, with each occupation code along with its description associated with as a mixture of underlying "topics", I remove all occupation codes that have a more than 50% distribution of identified "topics" related to "manage" and "supervise".[77] This leaves, in total, 1821 occupation codes in the dataset across all establishments.[78] For simplicity of exposition, I drop the adjectives and refer to these *non-managerial/supervisory occupations* as *occupations* henceforth.

For the alternative measure, I account for the difference in distribution of workers across occupations. To do so, I construct a "specialization index", which is defined as one minus the Herfindahl index across occupations within an establishment. Formally, let o represent an occupation at the 6-digit CBO level, the specialization index for establishment j with the set of occupation codes $\mathcal{O}$ is calculated as:

$$N_j = 1 - \sum_{o=1}^{\mathcal{O}} \left(\frac{l_j(o)}{l_j} \right)^2,$$

where $l_j(o)$ and l_j denote the number of workers employed in occupation o and the total number of workers in establishment j, respectively. Large values of N_j indicate higher degree of division of labor and small values of N_j indicate lower degree of division of labor.

[75] The purpose of this step is to identify occupations that are directly involved in the production process, so that the empirical measure is more consistent with the theory.

[76] The complete CBO 6-digit codes and the corresponding descriptions can be downloaded from the Brazilian Ministry of Labor website: http://www.mtecbo.gov.br/cbosite/pages/pesquisas/BuscaPorCodigo.jsf.

[77] See Figure 9 for an illustration of the procedure.

[78] As a robustness check, I follow Caliendo, Monte and Rossi-Hansberg (2015) and separate the employees within an establishment into four vertical hierarchical layers, based on their level of authority. I then remove all occupations codes at the top three layers (which correspond to *firm owners*, *senior management* and *supervisors*, respectively), and only consider the occupation codes at the bottom layer. All results are robust to this alternative construction.

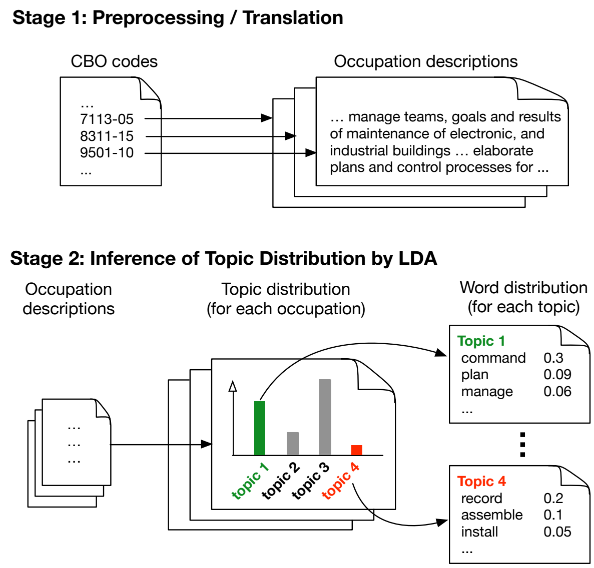


Figure 9: Removing managerial / supervisory occupations using the LDA technique

OA1.2 Additional results for stylized facts

Using the baseline measure of the number of occupations within an establishment, there may exist a mechanical relationship between division of labor and the size of the establishment, as larger establishments tend to have more occupations. To address this concern, I consider the alternative definition, i.e., the specialization index.[79] Using the measure that is not directly related to the establishment size, the results remain qualitatively similar to the baseline results, for both correlation analyses.

Dependent variable	Specialization index				
	All tradable		Export intensive	Mono-estb firms	Homogeneous
	(1)	(2)	(3)	(4)	(5)
Log (city size)	.02***	.0106***	.0106***	.0097***	.0103*
	(.0008)	(.0021)	(.002)	(.0016)	(.0061)
Controls	No	Yes	Yes	Yes	Yes
Obs	304503	304503	115449	284592	34058
R-sq	.095	.537	.539	.526	.504

Standard errors clustered by city in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include state and sector FEs. Establishment-level controls are establishment size and skill intensity within the firm. City-level controls are share of high-skilled workers, average wage, sector diversity, and the size of local sectoral employment. Occupations are measured by 6-digit Brazilian CBO codes. Sectors are measured by 5-digit Brazilian CNAE codes.Homogeneous sectors include corrugated and solid fiber boxes, white pan bread, carbon black, roasted coffee beans, ready-mixed concrete, oak flooring, motor gasoline, block ice, processed ice, hardwood plywood, and raw cane sugar (Foster, Haltiwanger and Syverson, 2008).

Table 23: Correlation of the establishment's normalized division of labor and city size

Next, I consider the possibility that establishments in larger cities are better at recording their employee's occupations accurately. To address this concern, I study the number of occupations within establishments

[79]Recall it is defined as one minus the Herfindahl index across occupations in an establishment.

Dependent variable	Log no. of occupations					
	No. of intermediate inputs			G3 export share		
	All tradable		Mono-estb firms	All tradable		Mono-estb firms
	(1)	(2)	(3)	(4)	(5)	(6)
Log (complexity)	.0388***	.0386***	.0382***	2.162***	.311***	.2207***
	(.0046)	(.0022)	(.0022)	(.1996)	(.0678)	(.0634)
Controls	No	Yes	Yes	No	Yes	Yes
Obs	304503	304503	284592	304503	304503	284592
R-sq	.044	.555	.561	.046	.553	.558

Standard errors clustered by city in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a city FE. Occupations are measured by 6-digit Brazilian CBO codes. Sectors are defined at 4-digit Brazilian CNAE codes.

Table 24: Correlation of the establishment's normalized division of labor and complexity

at the 4-digit level. As shown in Tables 25 and 26, though lower in the values of the estimates, the positive correlations remain strong.

Dependent variable	Log no of occupations within an establishment				
	All tradable		Export intensive	Mono-estb firms	Homogeneous
	(1)	(2)	(3)	(4)	(5)
Log (city size)	.0479***	.0197***	.0204***	.0188***	.0132**
	(.0032)	(.0034)	(.0034)	(.0028)	(.0066)
Controls	No	Yes	Yes	Yes	Yes
Obs	304503	304503	115449	284592	34058
R-sq	.132	.832	.825	.845	.807

Standard errors clustered by city in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include state and sector FEs. Establishment-level controls are establishment size and skill intensity within the firm. City-level controls are share of high-skilled workers, average wage, sector diversity, and the size of local sectoral employment. Occupations are measured by 6-digit Brazilian CBO codes. Sectors are measured by 5-digit Brazilian CNAE codes. Homogeneous sectors include corrugated and solid fiber boxes, white pan bread, carbon black, roasted coffee beans, ready-mixed concrete, oak flooring, motor gasoline, block ice, processed ice, hardwood plywood, and raw cane sugar (Foster, Haltiwanger and Syverson, 2008).

Table 25: Correlation of the establishment's division of labor (measured at 4-digit level) and city size

Lastly, I divide establishments into deciles and study the correlation between firms' division of labor and city size across different groups. This would partially address the problem of not observing informal workers within establishments. Based on ECINF (the Urban Informal Economy Survey), the share of informal workers is negatively correlated with firm size. As shown in Table 27, the correlation remains positive for all deciles, suggesting that the result is unlikely driven by differences in informal employment across space. Furthermore, the elasticity between division of labor and city size tends to be greater for larger firms, indicating that the interaction between firms' division of labor and city size may be stronger for larger firms. This pattern is consistent with the theoretical results.

Dependent variable	Log no. of occupations					
	No. of intermediate inputs			G3 export share		
	All tradable		Mono-estb firms	All tradable		Mono-estb firms
	(1)	(2)	(3)	(4)	(5)	(6)
Log (complexity)	.1065***	.1108***	.1105***	9.056***	1.488***	1.713***
	(.0181)	(.0041)	(.0042)	(.807)	(.16)	(.1612)
Controls	No	Yes	Yes	No	Yes	Yes
Obs	304503	304503	284592	304503	304503	284592
R-sq	.046	.847	.841	.053	.845	.839

Standard errors clustered by city in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a city FE. Occupations are measured by 6-digit Brazilian CBO codes. Sectors are defined at 4-digit Brazilian CNAE codes.

Table 26: Correlation of the establishment's division of labor (measured at 4-digit level) and complexity

Dependent variable:	Log no of occupations within an establishment		
1st decile	.0005***	6th decile	.0324***
	(.0001)		(.0026)
2nd decile	.0045***	7th decile	.0366***
	(.001)		(.0033)
3rd decile	.0145***	8th decile	.0472***
	(.0014)		(.0039)
4th decile	.0186***	9th decile	.0502***
	(.0018)		(.0046)
5th decile	.0253***	10th decile	.045***
	(.0022)		(.004)

Standard errors clustered by city in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include state and sector FEs, and city-level controls including share of high-skilled workers, average wage, sector diversity, and scale of the sector within cities. Occupations are measured by 6-digit Brazilian CBO codes.

Table 27: Correlation of the establishment's division of labor and city size, by decile

OA2 Theory Appendix

OA2.1 Proofs

This section presents the proofs to the propositions and lemmas discussed in the main text. The proof for Proposition 7 is included in Section F.

Lemma 1

Proof. Taking log of Equation (10),

$$\log \pi_s(z, L, N) = constant + (\sigma_s - 1)\left[\log H(N, L) + \log A(N, z, c_s) - \log w(L)\right]$$

Taking partial derivatives with respect to its arguments, I get

$$\frac{\partial \log \pi_s}{\partial z} = (\sigma_s - 1)\frac{\partial \log A}{\partial z};$$

$$\frac{\partial \log \pi_s}{\partial L} = (\sigma_s - 1)\left[\frac{\partial \log H}{\partial L} - \frac{\partial \log w(L)}{\partial L}\right];$$

$$\frac{\partial \log \pi_s}{\partial N} = (\sigma_s - 1)\left[\frac{\partial \log H}{\partial N} + \frac{\partial \log A}{\partial N}\right].$$

To prove supermodularity, cross-partials of $\log \pi_s(z, L, N)$ must be non-negative:

$$\frac{\partial^2 \log \pi_s}{\partial z \partial L} = 0;$$

$$\frac{\partial^2 \log \pi_s}{\partial z \partial N} = (\sigma - 1)\frac{\partial^2 \log A}{\partial N \partial z} > 0.$$

$$\frac{\partial^2 \log \pi_s}{\partial N \partial L} = (\sigma - 1)\frac{\partial^2 \log H}{\partial N \partial L} > 0;$$

The last two inequalities come from Assumptions 1 and 2. ∎

Lemma 2

Proof. Using the result from Lemma 1, applying the implicit function theorem to the first-order condition, $\frac{\partial \log \pi_s(z,L,N)}{\partial N} = 0$, and invoking the second-order condition, $\frac{\partial^2 \log \pi_s(z,L,N)}{\partial N^2} < 0$, I get

$$\frac{\partial N}{\partial z} = -\frac{\partial^2 \log \pi_s / \partial N \partial z}{\partial^2 \log \pi_s / \partial N^2} > 0$$

$$\frac{\partial N}{\partial L} = -\frac{\partial^2 \log \pi_s / \partial N \partial L}{\partial^2 \log \pi_s / \partial N^2} > 0$$

$$\frac{\partial N}{\partial c_s} = -\frac{\partial^2 \log \pi_s / \partial N \partial c_s}{\partial^2 \log \pi_s / \partial N^2} > 0$$

The last result shows that when $N_s(L)$ is optimally chosen, it is also increasing in the sector complexity c_s. ∎

Lemma 3

Proof. By Proposition 4.3 of Topkis (1978), I can invoke the property that supermodularity continues to hold when some arguments of a function are chosen optimally. That is, if $\pi_s(z, L, N)$ is log-supermodular in (z, L, N), then $\log \pi_s(z, L) \equiv \max_N \log \pi_s(z, L, N)$ is supermodular in (z, L). ∎

Proposition 4

Proof. By Lemma 3, $\log \pi_s(z, L)$ is supermodular in (z, L).

It then follows that for all $z_1 > z_2$ and $L_1 > L_2$ within sector s,

$$\frac{\pi_s(z_1, L_1)}{\pi_s(z_1, L_2)} > \frac{\pi_s(z_2, L_1)}{\pi_s(z_2, L_2)}.$$

In another word, if z_2 has higher profits in L_1 than in L_2, so does z_1. Necessarily,

$$L_s^*(z_1) > L_s^*(z_2).$$

Under technical assumptions, $L_s^*(z)$ is a strictly increasing function. Since the set of z is convex and $A(z, N, c_s)$ is such that the profit maximization problem is concave for all firms, the optimal set of city sizes is itself convex. It follows that $L_s^*(z)$ is invertible. It is also locally differentiable (using the fact that $A(z, N, c_s)$ is differentiable). The implicit function theorem applies, and I have

$$\frac{dL_s^*(z)}{dz} = -\frac{(\sigma_s - 1)\frac{\partial^2 \log A}{\partial z \partial N}\frac{\partial N}{\partial L}}{\frac{\partial^2 \log \pi_s}{\partial z^2}} > 0.$$

■

Proposition 5

Proof. Within a sector, $L_s(z)$ is strictly increasing in z (from Proposition 4). Therefore, if $L_s(z) = L_s(z')$, then we know $z > z'$. For simplicity, I denote $L_s(z)$ and $L_s(z')$ by L and L', respectively.

From Lemma 2, within a sector s, $N_s(z, L)$ is increasing in z and L. I get $N_s(z, L') > N_s(z', L')$. And since $L > L'$, I get $N_s(z) > N_s(z')$. For simplicity, I denote $N_s(z)$ and $N_s(z')$ by N and N', respectively.

Profit is proportional to $A(N_s(z, L), z)H(N_s(z), L)$. Under the assumption that $\frac{\partial A(N,z)}{\partial z} > 0$, i.e. firm profit is increasing in z, we have

$$\begin{aligned} A(N', z) &> A(N', z') \\ \implies A(N', z)H(N', L') &> A(N', z')H(N', L') \\ \implies \pi_s(z, N', L') &> \pi_s(z', N', L') \end{aligned}$$

where the last inequality comes from the fact that firms face the same wage in the same city.

Finally, $\pi_s(z, N, L) > \pi_s(z', N', L')$ as N and L are the profit maximizing choices for z. Therefore, I get $\pi_s(z) > \pi_s(z')$. Since revenue is proportional to profits, I obtain $r_s(z) > r_s(z')$.

Lastly, wage is proportional to size of the city. Hence $w_s(z) > w_s(z')$, if $L_s(z) > L_s(z')$. ■

Proposition 6

Proof. My argument follows the proof of Proposition 7 in Gaubert (Forthcoming). The proof covers both the case of the baseline assumption of the model (continuity and convexity of the support of z and L), and the case where the set of cities is exogenously given, and in particular discrete.

Denoted by $\mathcal{Z} : \mathcal{L} \times \mathcal{C} \to Z$ the correspondence that assigns to any $L \in \mathcal{L}$ and $c \in \mathcal{C}$ a set of z that chooses L at equilibrium (i.e. the matching function). Define $\bar{z}(L, c) = \max_z\{z \in \mathcal{Z}(L, c)\}$ as the maximum complexity level of a firm that chooses city size L in sector s characterized by c_s. To prove the results Proposition 6, I first prove the following three relevant lemmas.

Lemma 8 $\log \pi$ *is supermodular with respect to the triple* (z, L, c).

Recall $\log_s \pi$ is defined as:

$$\log \pi_s(z, L_c) = \log H(N, L) + \log A(N, z, c_s) + constant$$

Taking derivatives and applying the Envelope Theorem, I get,

$$\frac{\partial \log \pi}{\partial z} = \frac{\partial \log A}{\partial z}; \quad \frac{\partial \log \pi}{\partial L} = \frac{\partial \log H}{\partial L}; \quad \frac{\partial \log \pi}{\partial c_s} = \frac{\partial \log A}{\partial c_s}.$$

Taking cross-derivatives:

$$\frac{\partial^2 \log \pi}{\partial z \partial L} = \frac{\partial^2 \log A}{\partial z \partial N} \frac{\partial N}{\partial L} > 0$$

$$\frac{\partial^2 \log \pi}{\partial c_s \partial L} = \frac{\partial^2 \log A}{\partial c_s \partial N} \frac{\partial N}{\partial L} \geq 0$$

$$\frac{\partial^2 \log \pi}{\partial c_s \partial z} = \frac{\partial^2 \log A}{\partial c_s \partial z} + \frac{\partial^2 \log A}{\partial z \partial N} \frac{\partial N}{\partial c_s} + \frac{\partial^2 \log \mu}{\partial c_s \partial z} \geq 0$$

where the signs of inequalities are directly implied from Lemma 2, Assumption 1 and the assumption that firm-level complexity benefits firms in more complex sectors more, i.e. $\frac{\partial^2 \log A}{\partial c_s \partial z} \geq 0$. Lemma 8 follows directly.

Note that this result does not reply on an assumption on the convexity of $\mathcal{L}$. Checking the cross partials is sufficient to prove the supermodularity even if L is taken from a discrete set, as π can be extended straightforwardly to a convex domain, i.e. the convex hull of $\mathcal{L}$.

Lemma 9 $\log \pi(z, c_s)$ *is supermodular in* z *and* c_s.

The lemma can be obtained directly from the supermodularity of $\log \pi$ with respect to (z, L, c_s) and L is optimal city size that a firm z in sector s characterized by c_s chooses.

Lemma 10 $\bar{z}(L, c_s)$ *is increasing in* c_s.

The lemma can be obtained directly from the supermodularity of $\log \pi$ with respect to (z, L, c_s).

Using a classical theorem in monotone comparative statics, if $\log \pi(z, L, c_s)$ is supermodular in (z, L, c_s), as proven in Lemma 8, and $L^*(z, c_s) = \max_L \log \pi(z, L, c_s)$, then for all $c_H > c_L$,

$$(c_H, c_H) \geq (c_L, c_L) \implies L^*(z_H, c_H) \geq L^*(z_L, c_L)^{80}$$

[80] Note that everywhere, the sign $\geq$ denotes the lattice order on $\mathbb{R}^2$.

Define

$$\tilde{F}(L, c_s) = \Pr(\text{Firm from sector } s \text{ is in a city of size smaller than } L)$$
$$= F(\bar{z}(L, c_s))$$

where $F(\cdot)$ is the distribution of z of the firms in the sector. For any $z \in \mathcal{Z}$,

$$L^*(z, c_H) \geq L^*(z, c_L)$$

In particular, fix a given L,

$$L^*(\bar{z}(L, c_L), c_H) \geq L^*(\bar{z}(L, c_L), c_L) = L.$$

Because $L^*(z, c_H)$ is increasing in z, it follows that

$$z \in \mathcal{Z}(L, c_H) \implies z \leq \bar{z}(L, c_L) \implies \bar{z}(L, c_H) \leq \bar{z}(L, c_L).$$

I thus have

$$F(\bar{z}(L, c_H)) \leq F(\bar{z}(L, c_L))$$

and that $F(L, c)$ is increasing in c. This directly implies the first-order stochastic dominance of the geographic distribution of a high c sector versus that of a low c sector. ■

OA2.2 Instability of a homogeneous equilibrium

Proposition 11 *If agglomeration benefits are sufficiently strong relative to congestion costs, a homogeneous equilibrium cannot coexist in a locally stable equilibrium*

Proof. In a homogeneous equilibrium, all cities have the same size L and a symmetric distribution of firm types. Consider two cities, $L_1 = L_2$. Without loss of generality, consider perturbations of size $\epsilon > 0$ moving workers from city 1 to city 2. Since $\pi_s(z, L)$ is log-supermodular, the highest-z firms in city 1 have the most gain from a move and it is sufficient to consider perturbations of size ϵ in which all firms in the range $[z(\epsilon), \infty]$ move from city 1 to city 2. Since an interval of the highest-complexity firms, accompanied by the appropriate mass of workers in accordance to the firms' labor demand, moves from city 1 to city 2, $L_2' > L_1'$, with $L_2' = L_2 + \epsilon$ and $L_1' = L_1 - \epsilon$. The homogeneous equilibrium is only stable with respect to this perturbation only if

$$\begin{aligned}
&\log \pi_s(z(\epsilon), L_2) - \log \pi_s(z(\epsilon), L_1) \leq 0 \\
\implies &[\log A_s(N_s(z(\epsilon), L_2), z(\epsilon)) + \log H(N_s(z(\epsilon), L_2), L_2)] - \\
&[\log A_s(N_s(z(\epsilon), L_1), z(\epsilon)) + \log H(N_s(z(\epsilon), L_1), L_1)] \\
&\leq \frac{1-\eta}{\eta} L_2 - \frac{1-\eta}{\eta} L_1
\end{aligned}$$

This inequality is violated whenever z and the complementarity between N and z or between N and L is sufficiently high relative to η. ■

OA2.3 Properties of the heterogeneous equilibrium

In heterogeneous equilibria, Equation (19) characterizes the set of city sizes that necessarily exists in spatial equilibrium, i.e. no firms or workers would be better off by deviating from the optimal choices of city sizes. While the optimal city sizes are determined by the matching function, the density of different city sizes is obtained through the local labor market conditions, i.e. population living in a city of size L must equate to the total labor requirements from all firms that choose to locate in city L. Given that city-size is a continuous variable, it is easy to consider the cumulative distribution function for the city-size distribution $f_L(\cdot)$. Local labor market clearing condition dictates that, for all $L > L_0$ (where $L_0 = \inf(\mathcal{L})$, denoting the smallest city size in the equilibrium)

$$\int_{L_0}^{L} n f_L(n) dn = \sum_{s=1}^{S} M_s \int_{z_s(L_0)}^{z_s(L)} l_s(z) dF_s(z). \tag{43}$$

I can then obtain the city-size distribution $f_L(\cdot)$ by differentiating Equation (43) with respect to city size L and dividing by L on both sizes,

$$f_L(L) = \frac{1}{L} \left[\sum_{s=1}^{S} M_s \, \mathbf{1}_s(L) \, l_s(z_s(L)) \, f_s(z_s(L)) \, \frac{dz_s(L)}{dL} \right], \tag{44}$$

where $\mathbf{1}_s(L)$ is an indicator function, taking the value of 1 if sector s has firms in city L and 0 otherwise. Equation (44) gives an explicit expression for the distribution of city-sizes. Given the distribution of firm complexities, the equilibrium distribution of city size $f_L(\cdot)$, as shown in Equation (44), is unique. I get the following result:

Proposition 12 *The equilibrium city-size distribution $f_L(\cdot)$ is unique.*

Next, I discuss the stability of the heterogeneous equilibrium. Similar to the stability discussion for the homogeneous equilibrium, I prove the stability of the heterogeneous equilibrium through a perturbation exercise. Fix the set of equilibrium cities as well as the set of firms located in each cities. Consider a city. In equilibrium, its population is L and it has m firms of draw z. Labor demand for each firm is:

$$l = \frac{(\sigma_s - 1)^{\sigma_s}}{\sigma_s^{\sigma_s}} \frac{(A(N_s(z), z, c_s)H(N_s(z), L))^{\sigma_s - 1}}{w(L)^{\sigma_s}} R_s P_s^{\sigma_s - 1}.$$

From the local labor market condition,

$$m \frac{(\sigma_s - 1)^{\sigma_s}}{\sigma_s^{\sigma_s}} \frac{(A(N_s(z), z, c_s)H(N_s(z), L))^{\sigma_s - 1}}{w(L)^{\sigma_s}} R_s P_s^{\sigma_s - 1} = L,$$

I get wage $w(L)$ as a function of L. Recall that worker indirect utility is given by:

$$U(L) \propto w(L)^{\eta} L^{-(1-\eta)}$$

The equilibrium is stable if worker utility decreases if a small mass of individuals move away from the city. Note that I do not need to consider firms as firms are already maximizing their profits by locating in city L. I prove by contradiction, i.e. suppose $\frac{\partial \log U(L)}{\partial \log L} > 0$ instead.

$$\frac{\partial \log U(L)}{\partial \log L} = \eta \frac{w'(L)L}{w(L)} - (1-\eta) > 0$$

Differentiating local labor market clearing condition with respect with L, I get

$$m \frac{(\sigma_s - 1)^{\sigma_s}}{\sigma_s^{\sigma_s}} \frac{(A(N_s(z), z, c_s))^{\sigma_s - 1}}{w(L)^{\sigma_s}} R_s P_s^{\sigma_s - 1} \left[(\sigma_s - 1) \frac{\partial H}{\partial L} - \sigma_s \frac{w'(L)}{w(L)} \right] = 1. \tag{45}$$

From Equation (18), and the assumption that $\frac{w'(L)}{w(L)} L > \frac{1-\eta}{\eta}$, I get,

$$L \left[(\sigma_s - 1) \frac{\partial H}{\partial L} - \sigma_s \frac{w'(L)}{w(L)} \right] < -\frac{1-\eta}{\eta} < 0$$

A contradiction to Equation (45). I get the following result:

Proposition 13 *The heterogeneous equilibrium distribution of city size $f_L(\cdot)$ is stable.*

OA2.4 General equilibrium quantities

I now solve for the full set of general equilibrium quantities. The general equilibrium variables remaining to be determined are the aggregate revenues in the traded goods sector R, the mass of firms M_s and the sectoral price indexes P_s. To solve for the $2S + 1$ variables, I need $2S + 1$ equations, as specified below.

Using free entry condition for each sector $s = 1 \dots S$, I get

$$f_{E_s} P = \frac{(\sigma_s - 1)^{\sigma_s - 1}}{\sigma_s^{\sigma_s}} \xi_s R P_s^{\sigma_s - 1} \int_z \left(\frac{A(N_s(z), z, c_s) H(N_s(z, L), L)}{[(1-\eta) L_s(z)]^{\frac{1-\eta}{\eta}}} \right)^{\sigma_s - 1} dF_s(z). \tag{46}$$

where P is the aggregate price index for all tradable sectors. Given Cobb-Douglas preference,

$$P = \prod_{s=1}^{S} \left(\frac{P_s}{\xi_s} \right)^{\xi_s}.$$

Next, individual firms' production must sum up to aggregate production in each sector $s = 1 \dots S$,

$$1 = \frac{(\sigma_s - 1)^{\sigma_s - 1}}{\sigma_s^{\sigma_s - 1}} M_s P_s^{\sigma_s - 1} \int_z \left(\frac{A(N_s(z), z, c_s) H(N_s(z, L), L)}{[(1-\eta) L_s(z)]^{\frac{1-\eta}{\eta}}} \right)^{\sigma_s - 1} dF_s(z). \tag{47}$$

Lastly, using the national labor market clearing condition, I get

$$\bar{L} = \sum_{s=1}^{S} \frac{(\sigma_s - 1)^{\sigma_s}}{\sigma_s^{\sigma_s - 1}} M_s \xi_s R P_s^{\sigma_s - 1} \int_z \frac{[A(N_s(z), z, c_s) H(N_s(z), L)]^{\sigma_s - 1}}{((1-\eta) L_s^*(z))^{\frac{(1-\eta)}{\eta}}} dF_s(z) \tag{48}$$

Using Equations (47) and (48), I can solve for the aggregate revenue in the tradable sector:

$$\sum_{s=1}^{S} \frac{\sigma_s - 1}{\sigma_s} \xi_s \frac{\int_z \frac{[A(N_s(z), z, c_s) H(N_s(z), L)]^{\sigma_s - 1}}{((1-\eta) L_s^*(z))^{\frac{(1-\eta)}{\eta}}} dF_s(z)}{\int_z \left(\frac{A(N_s(z), z, c_s) H(N_s(z, L), L)}{[(1-\eta) L_s(z)]^{\frac{1-\eta}{\eta}}} \right)^{\sigma_s - 1} dF_s(z)} = \frac{\bar{L}}{R}. \tag{49}$$

Combining Equations (46) and (47), I get sectoral mass of firms:

$$M_s = \frac{\xi_s R}{\sigma_s f_{E_s} P} \tag{50}$$

Lastly, using Equations (46), I get the sectoral price indexes:

$$P_s^{\sigma - 1} = \frac{f_{E_s} P}{\frac{(\sigma_s - 1)^{\sigma_s - 1}}{\sigma_s^{\sigma_s}} \xi_s R \int_z \left(\frac{A(N_s(z), z, c_s) H(N_s(z, L), L)}{[(1-\eta) L_s(z)]^{\frac{1-\eta}{\eta}}} \right)^{\sigma_s - 1} dF_s(z)} \tag{51}$$

OA2.5 Descriptive evidence

In this part, I present descriptive evidence that is consistent with the theoretical results in Section 3.

OA2.51 Within-sector characteristics

The model predicts that in each sector, more complex firms sort into larger cities. This sorting of complexity generates sorting of other firm-level variables, including profits and revenue (Proposition 5). I first investigate how, within a sector, average firms' division of labor and labor payment change as city size increases.[81] In the model, the elasticities of firms' division of labor and firm revenue to city size are both positive within sectors. Empirically, I calculate the average establishment-level division of labor and labor payment within a sector-city cell and compute their elasticities with respect to city size.[82] Figure 10 plots the distribution of the two elasticities. For division of labor, it is positive for 91% of the observations, and is significantly negative for only two sectors, *growth of grains* and *sawmill*.[83] For labor payment, it is positive for 94% of the observations, and none of the negative estimates is significant. Results are therefore largely consistent with model predictions.

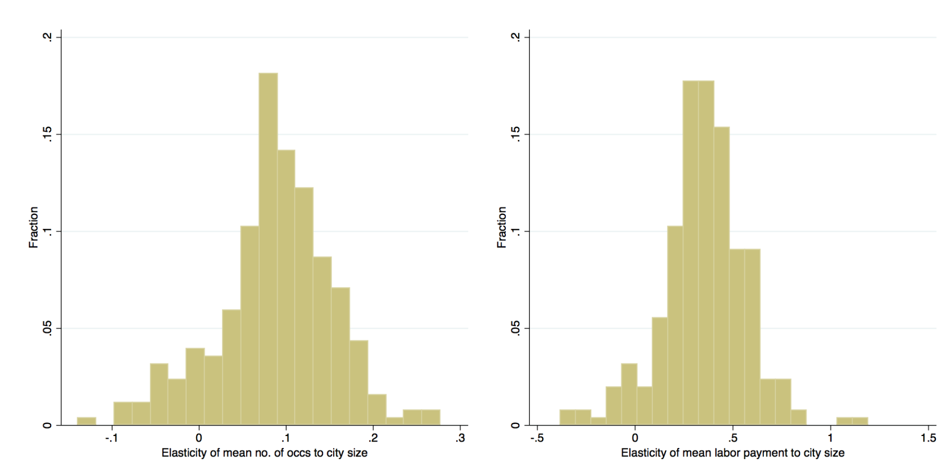


Figure 10: Elasticity of mean division of labor and labor payment to city size

Elasticity is generated by first running the regression: $\log\ \text{mean}\ N(L_j) = \alpha_s + \beta_s \log L_j + \epsilon_j$ (resp. $\log\ \text{mean labor payment}(L_j)$), sector-by-sector at the CNAE2.0 4-digit level.

OA2.52 Cross-sector characteristics

The model also predicts that the geographic distribution of firms in high-c_s sector first-order stochastically dominates that of firms in low-c_s sector, in Proposition 6. In other words, a larger share of firms is located in bigger cities for higher-c_s sectors. To test this prediction, I use two approaches. I first separate sectors into three broad categories based on sector-level complexities, and plot the distribution of firm size for each

[81] In the model, average labor payment is proportional to revenue. Labor payment is simply calculated as the total wage bill within an establishment.

[82] To have a meaningful number of establishments within each sector-city cell, I use the 4-digit CNAE2.0 code, which gives me 254 sectors.

[83] The results are not surprising since my model assumes that all locations are identical, whereas the location choices of these two sectors are driven by natural amenities, such as availability of arable land and forests.

group. As shown in Figure 11, high-c_s sectors display a clear first-order stochastic dominance relationship over medium- and low-c_s sectors.

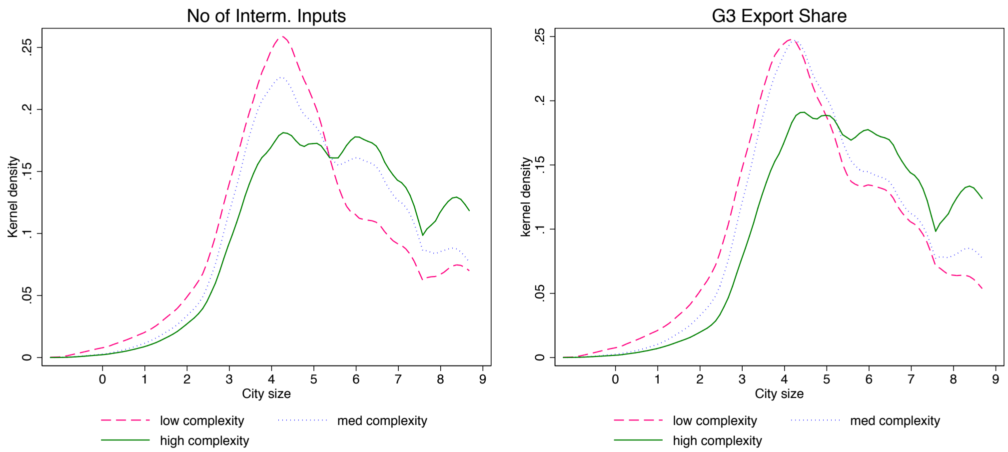


Figure 11: Distribution of firms across cities

I next use a more continuous measure of sector-level complexity. I calculate, for each sector, the share of firms located in *bigger* cities, where bigger cities are defined as the set of cities that host half of the population. I then estimate:

$$share_s = \alpha_0 + \alpha_1 c_s + \mathbf{X}_s + \varepsilon_s,$$

where an estimate of α_1 greater than 0 would be consistent with the model prediction. As shown in Table 28, α_1 is positive and precisely estimated for all specifications and using both complexity measures.

Dep var: Share of establishments in large cities						
	Intermediate inputs			G3 exp share		
	(1)	(2)	(3)	(4)	(5)	(6)
Log(Complexity)	.151***	.149***	.127***	.143***	.142***	.129***
	(.0258)	(.0258)	(.0262)	(.009)	(.009)	(.011)
No of Firms	No	Yes	Yes	No	Yes	Yes
Skill Intensity	No	No	Yes	No	No	Yes
Obs	269	269	269	269	269	269
R-sq	.15	.155	.215	.091	.102	.143

Robust standard errors in parentheses. Significance levels: * 10%, ** 5%, ***1%. Sectors are defined at 4-digit level using Brazillian CNAE system. All regressions include a sector fixed effect, defined at 2-digit CNAE level.

Table 28: Variation in the share of firms in big cities across sectors

OA2.6 Impacts of reduction in coordination costs

In the baseline model, I assume symmetric fundamentals. An exogenous improvement in ICT infrastructure, $\mathcal{I}$, changes the fundamentals in those cities. I denote the change by $\Delta\mathcal{I}$. For cities that undergo the infrastructure improvement, $\Delta\mathcal{I} > 0$; for the remaining cities, $\Delta\mathcal{I} = 0$. Additionally, I also separate the effects of city infrastructure on the costs of worker specialization from other mechanisms.[84] I re-write firms' original profit function as

$$\pi_s(z, L, N) = \kappa_s \left(\frac{A(N, z, c_s)\tilde{H}(N, \mathcal{I}(L), L)}{w(L)} \right)^{\sigma_s - 1},$$

where $\tilde{H}(N, \mathcal{I}(L), L) \equiv H(N, L)$ denotes the costs of division of labor, $\mathcal{I}(L)$ is an increasing function of L, and $\kappa_s = \frac{(\sigma_s - 1)^{\sigma_s - 1}}{\sigma_s^{\sigma_s}} R_s P_s^{\sigma_s - 1}$ is a sector-level constant.

In response to an improvement in city infrastructure, firms' profit function is now,

$$\pi_s(z, L, N) = \kappa_s \left(\frac{A(N, z, c_s)\tilde{H}(N, \mathcal{I}(L) + \Delta\mathcal{I}, L)}{w(L)} \right)^{\sigma_s - 1}.$$

Given Assumption 1, there is complementarity between city infrastructure and division of labor, i.e. the marginal cost of division of labor is decreasing in $\mathcal{I}$ or

$$\frac{\partial^2 H}{\partial N \partial \mathcal{I}} > 0.$$

I start by discussing the short-term partial equilibrium effect of the exogenous change. In the short term, I fix the current locations of firms and workers.

Proposition 7

Proof. Recall that in firm's problem, the first order condition with respect to N is:

$$\frac{\partial \log H}{\partial N} + \frac{\partial \log A}{\partial N} = 0 \tag{52}$$

Given my regularity conditions, the second order condition must also be met:

$$\frac{\partial^2 \log H}{\partial N^2} + \frac{\partial^2 \log A}{\partial N^2} < 0 \tag{53}$$

When there is an exogenous increase in $\mathcal{I}$, $\frac{\partial \log H}{\partial N}$ goes up, or the marginal cost of N, $-\frac{\partial \log H}{\partial N}$, decreases. Evaluated at the original level of N^*, the first order condition is now positive,

$$\left[\frac{\partial \log H}{\partial N} + \frac{\partial \log A}{\partial N} \right]_{|N = N^*} > 0. \tag{54}$$

To re-equalize, the first order condition must go down. Given the second order condition, the first order

[84] Recall that other mechanisms include, but are not restricted to, the learning advantages in larger cities as discussed in Appendix B.2.

condition decreases as N increases. Therefore an exogenous improvement in $\mathcal{I}$ increases the extent of worker specialization N.

Results in Lemma 2 state that the optimal level of $N_s(z, L)$ is increasing in (c, z, L). I can also rewrite the optimal division of labor as a function of the ICT infrastructure and city size,

$$\tilde{N}_s(z, \mathcal{I}(L), L) \equiv N_s(z, L).$$

There are two things to note: (i) $\tilde{N}_s(z, \mathcal{I}(L), L)$ is increasing in $\mathcal{I}$; (ii) $L_s(z)$ is in itself an increasing function of z and c_s from Proposition 4 and Lemma 8. Since $\mathcal{I}(L)$ is a function of city size, $\mathcal{I}(L)$ is also increasing in z and c_s. Therefore, when L is optimally chosen, $\tilde{N}_s(z)$ displays complementarity in $(z, \mathcal{I})$ and $(c_s, \mathcal{I})$. Intuitively, when z or c_s increases, it not only increases N directly, but also indirectly through L and $\mathcal{I}(L)$. When there is an improvement in city infrastructure $\mathcal{I}$, which reduces marginal cost of N at the city level, the change in N is higher for larger N, or $\frac{\partial^2 \tilde{N}_s(z)}{\partial \mathcal{I} \partial z} > 0$ and $\frac{\partial^2 \tilde{N}_s(z)}{\partial \mathcal{I} \partial c_s} > 0$. The intuition is again straightforward, given the complementarity between N and (c_s, z), the reduction in marginal cost benefits firms with higher c_s and/or z more, therefore the change in N is higher for high (c_s, z) firms.

Lastly, given Proposition 4, high-z firms locate in larger cities. Therefore, firms located in a larger city would increase N more if they experience an exogenous change in the level of city infrastructure. ■

In the longer term, firms can relocate across cities in response to the exogenous change in the city infrastructure. When firms' local labor demand changes, workers move across cities, resulting in changes in the equilibrium city-size distribution.

Corollary 14 *In equilibrium, an improvement in ICT infrastructure in a city increases the city size, and the level of division of labor, revenue and profit within firms in that city.*

Proof. In the long term, the exogenous change in ICT infrastructure would cause firms to re-optimize their location choices. Re-writing the optimal city size in Equation (18) by incorporating the change in city infrastructure, I get,

$$\frac{\tilde{H}_L}{\tilde{H}(N, \mathcal{I}(L) + \Delta\mathcal{I}, L)} = \frac{1-\eta}{\eta}\frac{1}{L}.$$

Given the complementarity between N and z, and between N and $\mathcal{I}$, there is positive assortative matching between z and $\mathcal{I}$. Denoted by $z_s^*(L, \mathcal{I}(L))$ the firm type z in city L. In the text, I show that within a sector, $z_s^*(L, \mathcal{I}(L))$ is strictly increasing in L. Consider two cities of the same size L, assume without loss that City 1 receives the new city infrastructure, given the complementarity, the equilibrium firm complexities in the two cities is,

$$z_s^{1*}(L, \mathcal{I}(L) + \Delta\mathcal{I}) > z_s^{2*}(L, \mathcal{I}(L)).$$

When there is an increase in $\mathcal{I}$, higher-z firms, with their higher willingness to pay for the local labor, would enter into the city. By law of demand, the increase in local labor demand would drive up the local wages. This has two consequences: (i) the more expensive labor cost would price out lower-z firms originally located in the city; and (ii) given the spatial indifference condition, the higher local wages would attract more workers into the city, according to the equilibrium wage function,

$$w(L) = \bar{w}[(1-\eta)L]^{\frac{1-\eta}{\eta}}.$$

In the long-run equilibrium, cities receive the new infrastructure would be occupied by firms with higher complexity draws. Since firms' division of labor, profit and revenue are all increasing in z, I get that the level of division of labor, revenue and profit within firms would all increase. The city size would increase, and the local workers would receive higher wages. ■

OA2.7 Model under costly trade

In this section, I prove that all theoretical results hold under costly trade assumption. My argument follows the proof in Appendix C1 in Gaubert (Forthcoming) and uses results from Allen and Arkolakis (2014).

Following the assumption in the base model, the economy consists of a continuum of locations $i \in \mathcal{S}$, where $\mathcal{S}$ is a compact subset of $\mathbf{R}^{\mathcal{N}}$. Trade is costly: trade costs are of the iceberg form and are described by the function $T : \mathcal{S} \times \mathcal{S} \to [1, \infty)$, where τ_{ij} is the quantity of a good needed to be shipped from city i in order for a unit of a good to arrive in city j. I discuss the single-sector results here, though all conclusions can be simply extended to the multi-sector case.

Price index for goods produced in city i is given by:

$$P_i = \left[\int_j \int_{z \in \mathcal{Z}(j)} \left(\frac{\tau_{ji} w_j}{\psi(z, L_j)} \right)^{1-\sigma} dF_j(z) dj \right]^{\frac{1}{1-\sigma}},$$

where $\mathcal{Z}(j)$ is the set of firms located in city i in equilibrium and $F_j(z)$ is the endogenous distribution of firms in city j.

We can define an average city-level productivity term:

$$\bar{\psi}_j = \left[\int_{z \in \mathcal{Z}(j)} \psi(z, L_j)^{\sigma-1} dF_j(z) \right]^{\frac{1}{\sigma-1}}. \tag{55}$$

Using Equation (55), price index can be re-written more compactly as:

$$P_i = \left[\int_j \left(\frac{\tau_{ji} w_j}{\bar{\psi}_j} \right)^{1-\sigma} dj \right]^{\frac{1}{1-\sigma}}. \tag{56}$$

Given CES preference, demand for good z produced in j from city j is $q_{ij}(z) = p_{ij}(z)^{-\sigma} w_j L_j P_j^{\sigma-1}$, whereas marginal cost for the good is $\frac{\tau_{ij} w_i}{\psi(z, L_i)}$. Combining, we get profit for firm z located in i,

$$\pi(z,i) = \frac{(\sigma-1)^{\sigma-1}}{\sigma^\sigma} \int_j \left(\frac{\tau_{ij} w_i}{\psi(z,L_i)}\right)^{1-\sigma} w_j L_j P_j^{\sigma-1} dj. \tag{57}$$

Note that the market access for firms in i is given by:

$$MA_i = \int_j \tau_{ij}^{1-\sigma} w_j L_j P_j^{\sigma-1} dj.$$

Substituting into Equation (57), we get,

$$\pi(z,i) = \frac{(\sigma-1)^{\sigma-1}}{\sigma^\sigma} w_i^{1-\sigma} \psi(z,L_i)^{\sigma-1} MA_i. \tag{58}$$

It is straightforward to see that Equation (62) displays log-supermodularity in (z, L_i). Therefore, in equilibrium, there is positive assortative matching in (z, L), i.e., more complex firms sort into larger cities, choosing greater extent of division of labor. As a results, firms in larger cities are also bigger and more productive. The final step in this proof involves in showing that L_i is the sufficient statistic for city i, i.e., distance between two cities plays no role in this economy. Note that since price index is a function of local wage and city size, we need to show that there is a one-to-one mapping between city size and wage.

Given free mobility of workers, utility, $U_i = \left[\frac{w_i}{P_i}\right]^\eta \left[\frac{L_i^{-1}}{1-\eta}\right]^{1-\eta}$, must be equalized across all cities. Using Equation (56) and some algebra, we can re-write the utility function as:

$$w_i^{1-\sigma} L_i^{(1-\sigma)\frac{\eta-1}{\eta}} = \bar{U}^{-1} \int_j \left(\frac{\tau_{ji} w_j}{\bar{\psi}_j}\right)^{1-\sigma} dj, \tag{59}$$

where $\bar{U} = \left[\frac{U}{\left(\frac{1}{1-\eta}\right)^{1-\eta}}\right]^{\frac{\sigma-1}{\eta}}$.

Next, using local goods market clearing condition:

$$w_i L_i = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \int_j \left(\frac{\tau_{ij} w_i}{\bar{\psi}_i}\right)^{1-\sigma} w_j L_j P_j^{\sigma-1} dj.$$

Multiplying both sides by $\left(\frac{w_i}{\bar{\psi}_i}\right)^{\sigma-1}$ and re-arranging, we get an expression for MA_i,

$$w_i^\sigma L_i \bar{\psi}_i^{1-\sigma} = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \int_j \tau_{ij}^{1-\sigma} w_j L_j P_j^{\sigma-1} dj = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} MA_i \tag{60}$$

Further re-arranging, we have,

$$w_i^\sigma L_i \bar{\psi}_i^{1-\sigma} = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \bar{U}^{-1} \int_j \tau_{ij}^{1-\sigma} w_j^\sigma L_j^{1-(\sigma-1)\frac{1-\eta}{\eta}} dj \tag{61}$$

The $2\mathcal{N}$ equations of (59) and (61) fit directly into the systems of equations in Allen and Arkolakis (2014), where the local congestion force is given by $L_i^{-\frac{1-\eta}{\eta}}$ and local productivity $\bar{\psi}_i$. Under further assumption that the trade cost is symmetric, i.e., $\tau_{ij} = \tau ji$, we can apply their results in Theorem 2, i.e. there exists unique vectors L_i and w_i in spatial equilibrium. Furthermore, we have

$$w_i^\sigma L_i \bar{\psi}_i^{1-\sigma} = \gamma w_i^{1-\sigma} L_i^{-(1-\sigma)\frac{1-\eta}{\eta}},$$

where γ is an endogenous constant in equilibrium.

Finally, from Equation (60), we get $MA_i = \gamma w_i^{1-\sigma} L_i^{-(1-\sigma)\frac{1-\eta}{\eta}}$. We can therefore re-write firm's profit function as

$$\pi(z,i) = \frac{(\sigma-1)^{\sigma-1}}{\sigma^\sigma}\gamma\left[\frac{\psi(z,L_i)L_i^{\frac{1-\eta}{\eta}}}{w_i^2}\right]^{\sigma-1} \tag{62}$$

In equilibrium, there must be a one-to-one mapping between city size and wage, i.e., no two cities with the same size can have different wages. Suppose the opposite is true, the firm will only choose the city that offers a lower wage. Furthermore, local wage has to be increasing in L, as firm profit is increasing in L and decreasing in w.

OA3 Empirics Appendix

OA3.1 Details of PNBL

Brazil enacted its National Broadband Plan (PNBL) in 2010, thought a presidential decree. The objective of the PNBL is to promote and disseminate the use of ICT to the lower-density and less-developed areas of Brazil. Until 2010, the distribution of broadband connections has been extremely uneven, closely reflecting the variation of population density across cities. The broadband was primarily provided by private telecommunication companies.[85] The private companies invested in the costly infrastructure only in highly developed where the population could afford the high service fees. This gap in broadband deployment raised concerns in the federal government. The government decided to take actions to stimulate broadband deployment adoption. In 2009, the first draft of the PNBL was released. The government proposed an investment amounting to US$41.9bil, of which US$27.2 billion from telecommunications operators and US$14.72 billion from government spending including tax cuts. After 6 months of discussion and deliberation, on May 12, 2010, President Luis Inàcio Lula da Silva signed Decree nr 7.175, which officially created PNBL.

A major initiative for the PNBL is the expansion of broadband backbone infrastructure. To implement this, Decree nr 7.175 addressed the recreation of the state-owned operator Telebras, which would build its own infrastructure or use other government-owned telecommunications infrastructure assets and other

[85] As of 2008, Brazil had 10 million fixed broadband lines in operation, out of which 63.7% were provided by its two biggest telecom companies, Oi and Telefonica.

infrastructure for example roads or power grid lines. The expansion of the backbone infrastructure was given a budget of $720mil USD.[86]

Telebras has been working with other companies and government organizations to expand the broadband backbone network in Brazil. As of 2014, the new broadband backbone extension, consisting mainly of optical fiber network, had reached 48,000km. The network now covers most of the country's states, and more importantly, improves the connectivity of regions which are otherwise too costly to receive broadband backbones. The fixed broadband connections in Brazil has increased from 15mil in 2010 to 22.5mil in 2014, as shown in Figure 12.

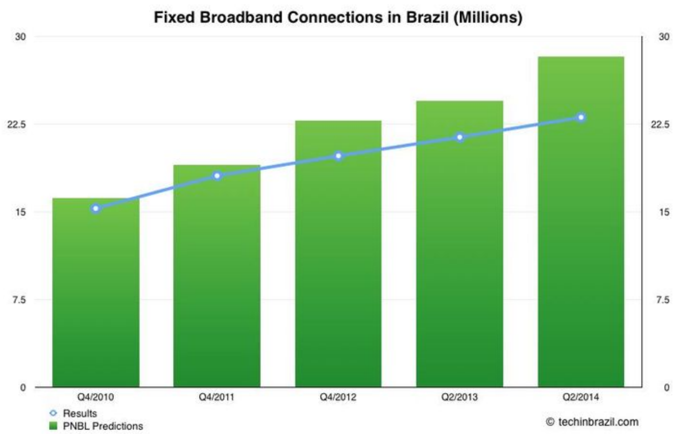


Figure 12: Growth of fixed broadband connections in Brazil

OA3.2 Broadband backbones and deployment technology

Figure 13 shows the supply chain of broadband internet in Brazil. The delivery of internet corresponds to four groups of major infrastructures. Listed in increasing order of "downstreamness" (and decreasing order of capacity), these four types of infrastructures are: submarine cables providing national / international connectivity, a national "backbone" of high-capacity (typically fiber optics) cables connecting submarine cables to the heartland of Brazil, smaller (usually radio or fiber) cables connecting national backbone to metropolitan base stations, and the "last-mile" infrastructure, consisting of fiber optic cables, wireless networks, coaxial cables and traditional telephone networks, to connect end users (Knight, 2016). In the analysis, I focus on the national backbone infrastructure. These are high-capacity fiber optic cables running from the coastal submarine cable landing points to the inland regions.

Since backbones use most exclusively fiber optic cables, there is a limit to its transmission range before losing all the data. The optimal length for each stretch is about 75 km (IGIC, 2004). The transmission distance is then extended by placing a device, called a *repeater*, at the end of the stretch to boost the signal.

[86]These government-owned telecommunications assets refer to fiber optic networks owned by government-owned companies Petrobras and Eletrobras, which cover many parts of the country and have a considerable amount of underused capacity.

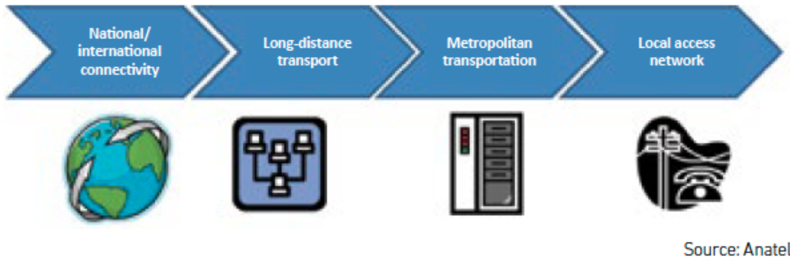


Figure 13: Broadband Supply Chain

Putting in the repeater is costly, and there is a limit to the number of repeaters that can be placed because it becomes no longer cost effective to do so. In general, up to four repeaters are implemented, making the maximum distance 400 km.

OA3.3 PNBL: Additional results

	(1)	(2)	(3) Interm. inputs	(4) G3 exp share
Dependent variable	Share of managers			
$Backbone_{jt}$	-.0114***	-.0087***	-.0072***	-.0085***
	(.0007)	(.0007)	(.001)	(.0008)
$Backbone_{jt} \times \log L_{ct_0}$		-.001***		
		(.0001)		
$Backbone_{jt} \times \log c_{st_0}$			-.0011***	-.0001
			(.0003)	(.0003)
Mean of outcome	.104	.104	.104	.104
Obs	777096	777096	777096	777096
R-sq	.731	.731	.731	.732

Robust standard errors clustered by municipality in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term, establishment and year FEs. High-skilled workers are defined as those with some college education and above.

Table 29: Impacts of broadband backbone on share of managers within establishment

	(1)	(2)	(3) Interm. inputs	(4) G3 exp share
Dependent variable	Skill intensity			
			Interm. inputs	G3 exp share
$Backbone_{jt}$	.0543***	.0667***	.0389***	.0621***
	(.0009)	(.001)	(.0009)	(.001)
$Backbone_{jt} \times \log L_{ct_0}$		.0081***		
		(.0002)		
$Backbone_{jt} \times \log c_{st_0}$			.0194***	.0061***
			(.0007)	(.0004)
Mean of outcome	.07	.07	.07	.07
Obs	777096	777096	777096	777096
R-sq	.628	.63	.629	.629

Robust standard errors clustered by municipality in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term, establishment and year FEs. High-skilled workers are defined as those with some college education and above.

Table 30: Impacts of broadband backbone on skill intensities within establishment

Dependent variable	Population (1)	Migration of workers (2)	No. of firms (3)	Relocation of firms (4)
$Backbone_{jt}$	.0258	.0711	.0148***	.04
	(.0287)	(.0566)	(.0024)	(.1018)
Obs	5022	3618	5022	1062
R-sq	.987	.716	.986	.225

Robust standard errors clustered by city in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term, city and year FEs.

Table 31: Impacts of broadband backbone on migration of workers and firms

OA4 Quantitative Appendix

OA4.1 Estimation procedure

To estimate $\chi_s = \{\alpha, \upsilon, c, \theta, \nu_z, \nu_L\}_s$, I use a method of simulated moment (Gourieroux, Monfort and Renault, 1993). The estimation is done for each sector separately. For each sector, I first construct a set of artificial Brazilian firms. Following Eaton, Kortum and Kramarz (2011), I draw a large sample of firms, 100,000 firms for each sector, to reduce the sampling variation in my simulation. Note that the number of simulated firms does not bear any relationship to the number of actual Brazilian firms. Firms operate as the model tells them, given some initial values of χ_s. In particular, I follow Gaubert (Forthcoming) and make firms choose optimal production location from 400 bins of normalized city sizes.[87] I then calculate the moments generated by the simulated economy. The steps are repeated until a find a set of moments that minimize the distance between the set of data moments and simulated moments, using the following criterion:

$$\hat{\chi}_s = \arg\min \, (m_{s,data} - m_{s,sim}(\chi_s))' J(m_{s,data} - m_{s,sim}(\chi_s))$$

The estimation follows the steps below:

1. I fix two set of random seeds from a uniform distribution on (0,1): one 100,000 for the firms; and one 100,000 $\times$ 400 for the firm-city-size-specific idiosyncratic shocks.

2. Given ν_z and ν_L, I use the random seeds to produce 100,000 realizations of firm complexities and 100,000 $\times$ 400 realizations for the idiosyncratic shocks.

3. For each city size, I use Equation (11) to calculate the optimal division of labor N^*.

4. For each city size, I plug N^* into Equation (33), to obtain the maximized firm productivity.

5. Based on the maximized firm productivity for each city bin, firms make a discrete choice of city size, according to Equation (34).

6. I then compute the 6 sets of moments described in Section 5.2.3.

7. I repeat Steps 1-6 to find parameters that minimize the objective function in Equation (35), using the particle swarm optimization (PSO) method (Kennedy and Eberhart, 1995).

[87] This restriction imposes 400 discrete choices of optimal city size for firms. Even though the choice set is exogenously given, the equilibrium city-size distribution is determined endogenously in general equilibrium.

Sector	Log wage bill			Log employment			Log no of occs			N
	mean	p25	p75	mean	p25	p75	mean	p25	p75	
Agriculture, and mining	10.94	9.94	11.7	2.05	1.39	2.71	1.42	.69	2.08	9239
Manufacture of food products, beverages and tobacco products	11	9.95	11.71	2.37	1.61	3	1.54	1.1	2.08	39831
Manufacture of textiles	11.15	10.06	11.95	2.4	1.61	3.14	1.55	.69	2.08	9742
Manufacture of wearing apparel	10.75	9.91	11.37	2.18	1.39	2.83	1.33	.69	1.79	46098
Manufacture of leather goods and footwear, leather tanning	11.15	10.11	11.96	2.52	1.61	3.3	1.54	1.1	2.08	11229
Manufacture and products of wood, except furniture	10.81	9.99	11.43	2.13	1.39	2.71	1.37	.69	1.95	14044
Manufacture of pulp, paper and paper products	11.45	10.36	12.28	2.58	1.61	3.37	1.78	1.1	2.4	4524
Publishing, printing and reproduction of recorded media	10.65	9.82	11.23	1.77	1.1	2.3	1.34	.69	1.79	11305
Manufacture of chemicals and chemical products	11.61	10.41	12.59	2.49	1.61	3.3	1.83	1.1	2.48	10017
Manufacture of pharmaceutical products	12.18	10.68	13.56	2.88	1.79	3.99	2.08	1.39	2.94	912
Manufacture of rubber and plastic products	11.5	10.44	12.34	2.58	1.79	3.37	1.73	1.1	2.3	15609
Manufacture of glass, ceramic, brick and cement products	10.95	10.08	11.62	2.28	1.61	2.89	1.42	.69	1.95	27008
Manufacture of basic metals	11.6	10.49	12.51	2.5	1.61	3.33	1.81	1.1	2.48	4523
Manufacture of fabricated metal products, except machinery	10.99	10.01	11.76	2.04	1.39	2.71	1.4	.69	1.95	34950
Manufacture of computer and electronic products	11.67	10.5	12.63	2.47	1.61	3.37	1.82	1.1	2.56	3751
Manufacture of electrical machines	11.68	10.58	12.59	2.58	1.61	3.4	1.86	1.1	2.48	4878
Manufacture of other equipments and machines	11.58	10.5	12.44	2.34	1.39	3.09	1.81	1.1	2.48	15287
Manufacture of automotive vehicles	11.6	10.36	12.51	2.54	1.61	3.3	1.81	1.1	2.48	6207
Manufacture of other transport equipment	11.66	10.43	12.6	2.57	1.61	3.4	1.86	1.1	2.48	1169
Manufacture of furniture	10.83	9.91	11.51	2.06	1.39	2.71	1.32	.69	1.95	18288
Manufacture of miscellaneous products, other mfg activities	10.89	9.92	11.62	2.1	1.39	2.71	1.39	.69	1.95	9801

Table 32: Summary statistics across sectors

OA4.2 Estimation results

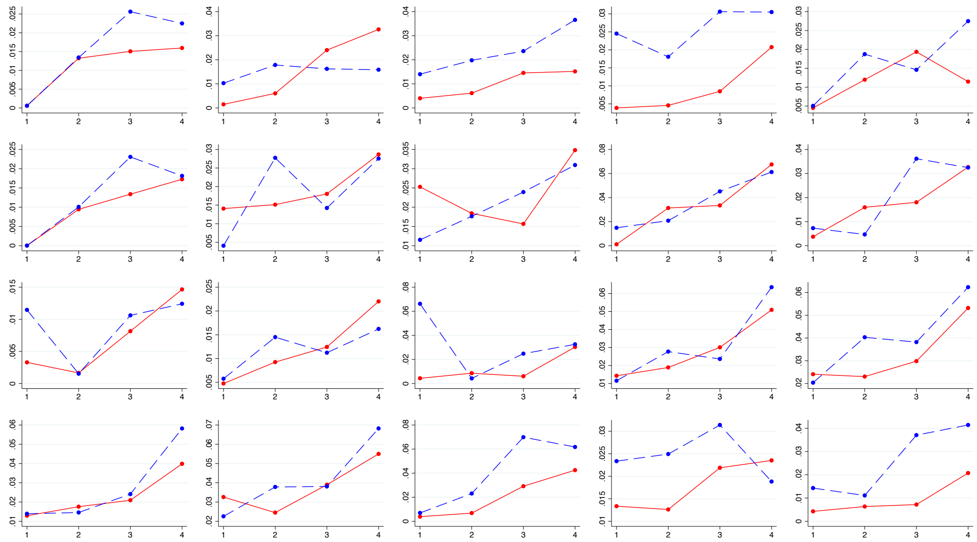


Figure 14: Average change in division of labor by city size
(Actual moments: solid red line; simulated moments: dashed blue line)

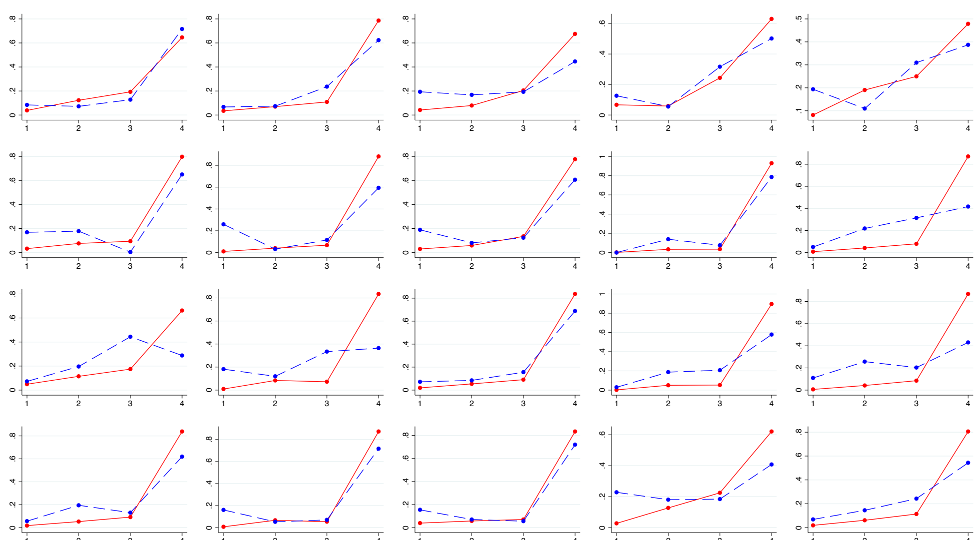


Figure 15: Distribution of employment across cities
(Actual moments: solid red line; simulated moments: dashed blue line)

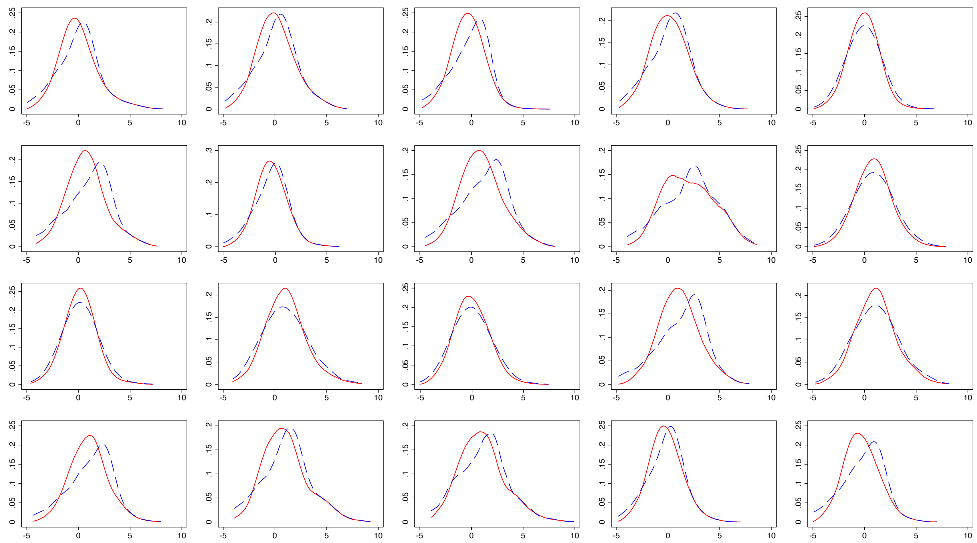


Figure 16: Distribution of firm labor payment

(Actual moments: solid red line; simulated moments: dashed blue line)

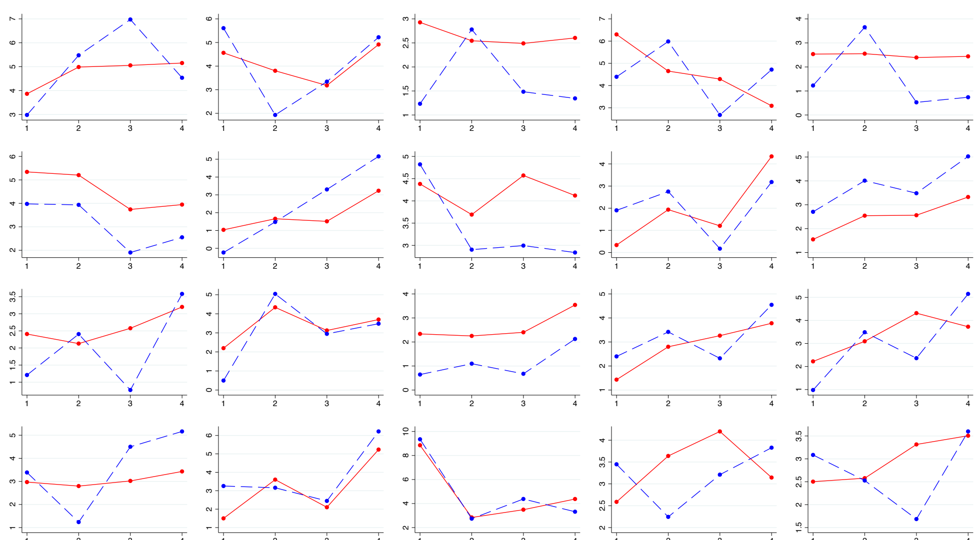


Figure 17: Average labor payment by city size

(Actual moments: solid red line; simulated moments: dashed blue line)

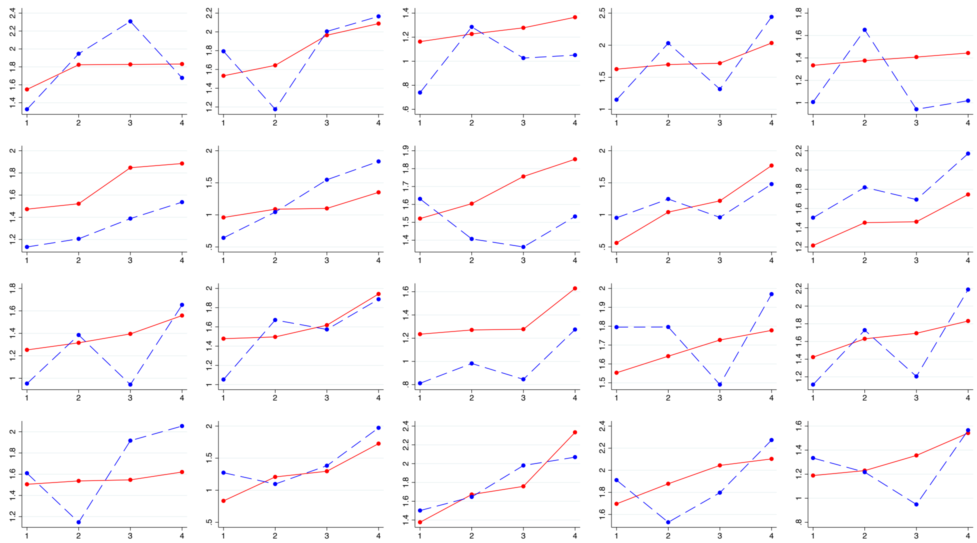


Figure 18: Average division of labor by city size
(Actual moments: solid red line; simulated moments: dashed blue line)

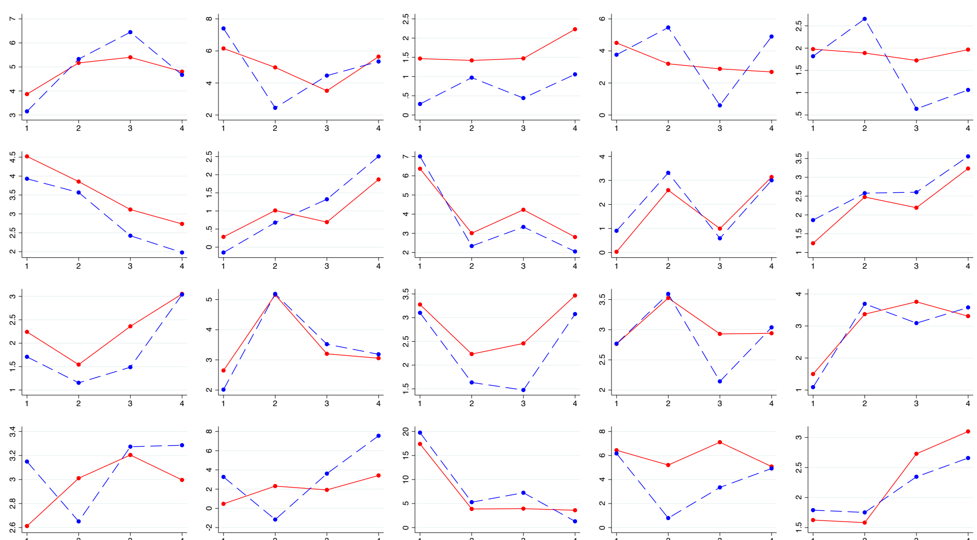


Figure 19: Variance of division of labor within city bins

OA5: Productivity impacts of the new ICT infrastructure

In this part, I illustrate how one could use the estimated model for policy evaluations. In particular, I use the estimated model to evaluate the impacts of the new broadband infrastructure on productivity and other outcomes.

Short-term impacts

In the model, an exogenous reduction in coordination costs would bring about general equilibrium effects, including the relocation of firms across cities and adjustment in city size when workers migrate internally in response to changes in local labor market demand. Most of these variables require a longer time horizon to be realized. Since my theory is static, the predictions can be seen as long-run general equilibrium effects. In Table 31, I show that an improved internet connection has no significant effect the migration of establishments or workers, within the observed period.[88] Since I find no significant migration of workers and firms in response to the new ICT infrastructure, I shut down firm sorting in this analysis to estimate the short-term productivity impact.

I find that, in the short-term, the average productivity in treated areas increase by 3.94 percentage points more than other areas. The productivity impact is generated through two channels: the direct impact of improved ICT infrastructure on productivity, and additional productivity increase due to firms' endogenous adjustment in the optimal division of labor. Using the estimated model, I shut down the second channel by fixing firms' division of labor at the level before the program. In doing so, the change in productivity reduces to 3.2% (or a 19% reduction), showing again that division of labor has substantial impact on firms' productivity.

Long-term impacts

Lastly, I use the estimated model to simulate the long-run general-equilibrium effects of improved ICT infrastructure by allowing firms and workers to move across space.

To evaluate the long-run general equilibrium impacts of the new policy, I adopt the following steps:

1. I fixed the aggregate number of workers, the set of firm-city-size specific idiosyncratic shocks, the distribution of firm complexities, and the number of cities in each city bin.

2. I first calibrate the local increase in ICT infrastructure using the estimated model, to match the reduced-form estimate in Section 4 on the impact of the new infrastructure on firms' division of labor.

3. From the spatial equilibrium estimated using the actual economy, I incorporate the infrastructure improvement to cities that receive the new infrastructure.

4. I recompute the optimal choices of city size by firms, taking into account the new infrastructure.

5. As the mix of firms within a city bin varies, the total labor demand for a given city size also changes. Since the number of cities in each city bin is fixed, the changes in that total local labor demand for a given city size would increase or reduce the size of each city bin.

[88] I only observe at most two years after the program, as the most recent RAIS data I have access to is for Yr 2014.

6. The change in city size feeds back to firms' production functions, affecting the local productivity and labor costs. I then recompute the optimal choices of city size by firms, taking into account the change in city size.

7. I iterate Steps 3-5, until I get a fixed point of this procedure in city sizes. The new city-size distribution defines the long-term economy.

Using the new long-term economy, I first estimate the local impacts using the following OLS regression:

$$\triangle_t \log y_m = \alpha + \beta Backbone_m + \varepsilon_m \tag{63}$$

where $\triangle_t \log y_m$ is the log change in the outcomes of interest y in city m before and after the treatment, and $Backbone_m$ is an indicator function taking the value of 1 if city m is connected to the new backbone and 0 otherwise. The variables I consider here are the number of establishments, city size, and average local productivity. Results from this specification are in Columns 1 to 3 in Table 33. In locations receiving the new infrastructure, the model predicts that the number of establishment grows by 7.7 percentage points relative to other locations. Correspondingly, the treated cities also experience a relative increase in the population of 7.8 percentage points.

Dependent var	Log change in no. of estb	Log change in city size	Log change in estb pdty
	(1)	(2)	(3)
Backbone	.0743***	.0751***	.0951***
	(.0011)	(.0033)	(.002)
Obs	558	558	558
R-sq	.923	.571	.432

Significance levels: * 10%, ** 5%, ***1%. All regressions include a constant term.

Table 33: Simulated long-term local impacts of PNBL

The new infrastructure also affects the average local productivity. The model predicts that relative to the control areas, the targeted cities would experience an increase in productivity by 9.98 percentage points. The productivity impact is higher than the short-term local impact of 3.94 percentage points because the long-run effects consist of both the effect of ICT infrastructure improvement, as well as productivity increase due to additional agglomeration externalities as firms and workers move into the targeted areas.[89]

In addition to evaluating the local impacts of PNBL, the calibrated model also allows me to compute the policy's long-term aggregate effects. As explained in Section 4, one of the key policy objectives of the program is to reduce spatial disparities. The literature (see, e.g., Kline and Moretti, 2014) points out that this

[89] These results are the same order of magnitude as the estimates by Hjort and Poulsen (2016), who find that access to broadband internet increases firm productivity by 15.7 percentage points in African countries. It is also intuitive that my estimates are lower, due to the model restriction that fast internet can only affect productivity by lowering the costs of division of labor, and ignores other potential productivity effects of the fast internet.

kind of spatially targeted policies may shift economic activities from one location to another. The aggregate impacts on productivity and welfare are therefore ambiguous. Using my estimates, I examine how the new infrastructure affects overall distribution of economic activities.

I compute the aggregate TFP and welfare effects of the policy, holding constant the treated areas.[90] The simulation shows that the expansion of broadband infrastructure has positive and small long-run effects on productivity and welfare. The PNBL increases the aggregate TFP by a mere 0.39 percentage point, and the aggregate welfare by 0.38 percentage point. Positive impacts to treated areas are largely offset by negative effects on other places, which is consistent with the qualitative results of Kline and Moretti (2014).

I last study the impact of the policy on the dispersion of spatial outcomes, by computing the Gini coefficients for the distributions of GDP per capita and city size in the economy. Despite low aggregate productivity and welfare effects, the policy achieves some success at reducing regional inequalities. Using the estimated model, I find that the expansion of broadband backbones reduces Gini indices by 0.7% and 1.4% for GDP per capita and city size, respectively.

[90] Aggregate TFP is constructed using the average sector-level productivity, $TFP = \prod_{s=1}^{S} TFP_s^{\xi_s}$, where $TFP_s = mean_s(\psi_{js})$. Welfare is measured by the worker's real income, which is constant across space. It is defined by $\bar{U} = \frac{w}{P^{\eta} p_H^{1-\eta}}$, where P is the aggregate price index, and p_H is the price of housing.