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

Lin Tian

INSEAD & CEPR

August 16 2021

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 positive 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, possibly by reducing costs associated with greater division of labor. Structural estimates derived from the model show that division of labor accounts for 17% of the productivity advantage 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. The theory also generates a set of auxiliary predictions of firms' responses to an exogenous shock to division of labor. Exploiting a quasi-experiment—the gradual roll-out of broadband internet infrastructure—I find causal empirical support for these predictions. Finally, the quasi-experiment also provides out-of-sample validation for the structural estimation: The model is successful in predicting the heterogeneous impacts of the new infrastructure across Brazilian cities.

*lin.tian@insead.edu. 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, Barcelona GSE Summer Forum, and other conferences, seminars and workshops for helpful discussions and comments. All errors are mine.

"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 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 quantitatively evaluating 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 in firms*. The idea that division of labor may contribute to spatial productivity difference was first discussed by Smith (1776), who proposes that factories 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 quantitative analyses, and show that division of labor within firms is an important source for the productivity advantage in larger cities.

To guide the empirical investigation into division of labor, I develop a theory of firm production and organization that links firm-level observables to the concept of division of labor. Production of goods in any firm requires combining a collection of tasks (Smith, 1776). The productivity benefits of division of labor comes from increasing returns of scale at the worker level in performing these tasks. The optimal contract is one in which firms subdivide the tasks into partitions with each worker specializing in a single partition, called an "occupation." The more partitions there are, the narrower the range of tasks that each worker specializes in, and the greater the division of labor. Guided by the theoretical insight, I construct a firm-level measure of division of labor using the number of partitions—or distinct occupations—within each firm.[2]

With this definition, I assemble a unique dataset on firm-level division of labor using a sample of matched employer-employee records of Brazilian firms.[3] The dataset allows me to document a new stylized fact on

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

[2]In a contemporaneous work, Becker et al. (2019) document, using German worker survey data that quantify the number of tasks performed within worker occupations, a negative relationship between a plant's count of occupations and the width of its average task range per occupation, i.e., the inverse of division of labor.

[3]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*

firms' division of labor: There is greater division of labor inside firms in larger cities.[4] The correlation remains largely unchanged when controlling for characteristics such as firm size and skill intensity, and when I use alternative definitions of division of labor.[5] I also provide tests to show that the observed correlation is not driven by multi-establishment firms or systematic differences in product varieties, share of informal workers, as well as number of tasks performed within the establishment boundaries across different cities.

Though robust, the observed correlation may potentially be driven by many different channels, and cannot be regarded as evidence showing the causal effect of city size on division of labor in and of itself. Separately identifying each channel is crucial in studying how division of labor contributes to the productivity advantage of cities. To unpack and quantitatively evaluate each channel, I develop a quantitative model in which the spatial distributions of firms' division of labor and productivity are determined jointly. The model generalizes the model of firm production organization in the previous section and embeds it in a standard spatial sorting model (Gaubert, 2018). Through the model, I propose potential mechanisms that generate the observed correlation between division of labor and city size. Firms, exogenously heterogeneous in their production *complexity*, choose division of labor and size of the city in which to locate to maximize profit. The model makes two key reduced-form assumptions, each of which is microfounded in the appendix: First, there is complementarity between division of labor and production complexity, e.g., more complex firms benefit relatively more from labor specialization; second, there is complementarity between division of labor and city size, e.g., larger cities reduce the costs of division of labor.[6,7] Since in equilibrium, more complex firms 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 through two channels: (i) a selection channel, i.e., more complex firms, which are firms that would choose greater division of labor in any given city, endogenously sort into larger cities in equilibrium; and (ii) a treatment channel, i.e., any given firm would choose a greater extent of division of labor in a larger city. Identifying the latter channel is important for the purpose of investigating how city size affects division of labor and, in turn, productivity. Without an experiment in which firms are randomly allocated across space, it is difficult to separate these two channels empirically. The quantitative model, on

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 the division of labor measure in Section 3 and Appendix A.

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

[5]For baseline analysis, I use the total number of 6-digit occupation codes to proxy the firm's division of labor. In robustness analyses, I use more aggregate occupation codes and a normalized measure for the heterogeneity of occupations within an establishment. See Section 3 for more details.

[6]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).

[7]I microfound the first assumption following closely the argument in Costinot (2009), as detailed in Appendix B.3. 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 (2009), 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 4 for more detailed discussions.

the other hand, allows me to do so by relying on the structure of the model.

I bring the model to data to recover estimates of the parameters for the quantitative analysis, using a method of simulated moments. I parameterize an extended version of the model, which incorporates the standard reduced-form agglomeration externalities in the urban literature (see, e.g., Allen and Arkolakis, 2014), the standard spatial sorting of firms (see, e.g., Gaubert, 2018), imperfect sorting of firms, 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 17% 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.[8] I further disentangle the roles played by spatial sorting of firms (i.e., the selection channel) and the direct effect of city size (i.e., the treatment channel) in another counterfactual experiment, in which I shut down the systematic sorting of firms due to division of labor. I estimate that each channel contributes approximately half of the 17% productivity advantage through division of labor.

In the final part of the paper, I provide external validations to the theoretical framework and the quantitative assessment through a quasi-experiment in Brazil. To do so, I hypothesize that one possible channel that generates the complementarity between division of labor and city size is through the provision of public infrastructure. In particular, larger cities provide better Information and Communications Technology (ICT) infrastructure, which potentially increases firms' division of labor.[9] If the hypothesis is true, an exogenous improvement in ICT infrastructure in certain areas will lead to an increase in division of labor for firms located in those areas. Importantly, the parameterized model generates two further predictions. First, in response to an exogenous improvement in ICT infrastructure in a given city, the increase in the extent of division of labor is larger for more complex firms, arising from the complementarity between division of labor and complexity. Second, in response to an exogenous improvement in ICT infrastructure in a set of cities, the increase in the extent of division of labor is larger for firms in larger cities because of the complementarity between division of labor and city size. I confront these model predictions for how variables respond to changes in ICT infrastructure with data, by exploiting the expansion of broadband infrastructure as part of the Brazilian National Broadband Plan (PNBL henceforth). 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

[8] Ellison and Glaeser (1999) find that natural advantage contributes to approximately 20% of productivity gains in larger cities. Serafinelli (2019) shows that firm-to-firm worker flows explain about 10% of agglomeration advantages in higher-density areas.

[9] 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 (see, e.g., Borghans and Weel, 2006; Varian, 2010; McElheran, 2014; and Bloom et al., 2014).

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.[10] I find evidence that validates my hypothesis that better ICT infrastructure is indeed one channel that generates the complementarity between division of labor and city size. More importantly, heterogeneities in the direct treatment effect are consistent with the model predictions specified above. Finally, the quasi-experiment also provides external validation for the structural estimates. Since the model is estimated using data before the implementation of the new infrastructure, I am able to compare the model-based predicted impact to the actual changes. Specifically, the estimated model predicts changes in the average division of labor within different cities in response to ICT infrastructure improvement, which I find are similar to the actual changes.

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, 2019; Behrens, Duranton and Robert-Nicoud, 2014; and Gaubert, 2018). My theoretical framework is most closely related to the one developed by Gaubert (2018), in which sorting of firms is generated by a reduced-form assumption that more productive firms benefit more 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 works by Caliendo and Rossi-Hansberg (2012) and Caliendo, Monte and Rossi-Hansberg (2015), which examine 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.[11] My paper is also related to Becker et al. (2019), which extend a Melitz (2003) model to relate within-plant occupation to worker task specialization and discuss the productivity and wage inequality implications of division of labor. I enrich these theoretical discussions by developing a spatial equilibrium framework that links a firm's decision on division of labor to its location

[10] I conduct an extensive set of robustness tests, including direct inspection of pre-trends, which supports a causal interpretation of my results.

[11] 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.

choice, to study the relationship between division of labor and city size and determine how firms' organization decisions contribute to spatial productivity differences.[12]

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. Incorporating heterogeneous firms is also essential to study how firm sorting affects division of labor across different cities. My paper is perhaps most related and complementary to the contemporaneous work by Becker et al. (2019), which uses German data to study empirically and quantitatively the impact of division of labor on within-plant wage dispersion and economy-wide wage inequality.

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.[13] 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 causally 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 outlines a stylized model that connects firm observables to division of labor. Section 3 describes the data and definitions, and documents the correlation between division of labor and city size. Section 4 develops a spatial equilibrium model with endogenous firm organization to unpack the channels driving the observed correlation. Section 5 summarizes the quantitative framework and the estimation process. Section 6 presents results from the counterfactual exercises. Section 7 details results from a quasi-experiment, which provide empirical support for the model. Section 8 concludes.

[12] 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.

[13] See Hjort and Tian (2021) and Draca, Sadun and Reenen (2009) for reviews of studies on developing and developed countries, respectively.

2 A stylized theory of division of labor

The first step in studying division of labor empirically is to find a sensible measure for division of labor at the firm level. To guide my empirical investigation, I develop a theory of firm production and organization, which links firm-level observables to the concept of division of labor directly. The broad logic of the model can be sketched as follows. Inspired by Smith (1776), I first observe that in any firm, production of a good requires combining a collection of tasks.[14] Following Costinot (2009), the production technology further generates increasing returns of scale at the worker level in performing individual tasks. I show that the optimal contract involves a firm subdividing these tasks into partitions (called "occupations") and each worker specializing in a single occupation. The more occupations there are, the narrower the range of tasks that each worker specializes in, and the greater the division of labor there is within the firm. Informed by the theoretical insight, I construct a firm-level measure of division of labor using the number of "partitions"—or occupation counts—within the firm.

2.1 Production technology

In each firm, a continuum of tasks in the unit interval $t \in [0, 1]$ must be performed to produce one unit of its output. Production follows the following technology:

$$Q = \int_0^\infty \left[\int_0^1 \mathbf{1}(t, u)^{\frac{\epsilon-1}{\epsilon}} dt \right]^{\frac{\epsilon}{\epsilon-1}} du; \quad 0 \leq \epsilon < 1, \tag{1}$$

where Q is the total units of output and $\mathbf{1}(t, u)$ is 1 if task t is performed on the u-th unit and 0 otherwise. The elasticity of substitution between tasks, ϵ, is less than 1, implying that tasks are complementary.[15]

Each worker is endowed with a fixed amount of labor supply, split between learning and production (Costinot, 2009). Prior to production, workers need to spend time to acquire competency in performing the tasks. The amount of labor required by worker i performing task t is given by:

$$l(i, t) = \int_0^\infty \mathbf{1}(i, t, u) du + z(t), \tag{2}$$

where $\mathbf{1}(i, t, u)$ is 1 if worker i performs task t on the u-th unit, and zero otherwise. $z(t)$ is the labor inputs required to learn how to perform task t. (2) implies that the production technology has increasing returns to scale at the worker-task level: Given the fixed cost of learning, the average cost of performing a task goes down when the task is performed on a greater number of units.

Finally, without loss, I normalize all tasks such that the worker's learning cost for each task is constant

[14] Smith (1776) notes that there are at least 18 distinct tasks associated in making a pin in a pin factory.

[15] In the extreme, when $\epsilon = 0$, tasks are perfect complements and the production follows a Leontief technology.

across the tasks,

$$z(t) \equiv z, \ \forall\, t \in [0,1].$$

Note that this normalization implies two things when we compare z across different firms: 1) Given the set of tasks, a higher z means that a task is more complex and thus requiring more time to learn; and 2) given the complexity of each task, a higher z also indicates that there are more tasks involved in producing the output.

2.2 Production organization

Firms organize the production by designing a set of occupations, $\mathcal{O}$, and writing a set of contracts, $\mathcal{C}$, that assign workers to these occupations. Formally, each firm has two control variables:

1. A partition $\mathcal{O} = \{\mathcal{O}_k\}_{k=1}^{N}$ of the sets of tasks in $[0,1]$; and

2. A mapping $\mathcal{C}(i) : [0,l] \times \mathbb{R}^+ \to \{\mathcal{O}, \emptyset\}$.

Lemma 1 *Suppose $\mathcal{O}^*$ and $\mathcal{C}^*$ are the optimal organization of a profit-maximizing firm, then*

1. *for all $i \in l$, there exists $k = 1 \dots N$ such that $\mathcal{C}^*(i,u) \in \{\mathcal{O}_k^*, \emptyset\}$ for all $u \in \mathbb{R}^+$; and*

2. *for all $k = 1 \dots N$, $\mathcal{O}_k^*$ is such that*
$$\int_{t \in \mathcal{O}_k^*} dt = \frac{1}{N}.$$

The formal proof can be found in Appendix B.5. Intuitively, the first part states that all workers specialize in one occupation. Since there is increasing returns to scale at the worker-task level, workers who know how to perform a given set of tasks involved in an occupation should perform it as many times as possible to minimize the learning costs. The second part states that all occupations include the same number of tasks. This can be explained by observing that worker productivity depends on the number of tasks included in an occupation. Specifically, a marginal decrease in the number of tasks in an occupation increases the time available for actual production. This increase is larger for occupations with more tasks. Since profit maximization requires that marginal changes in worker productivity be equalized across occupations, it also requires that each occupation includes the same number of tasks.[16]

2.3 Division of labor and occupations

Though stylized, this theory of production organization reveals an important link between the concept of division of labor and an observable variable at the firm level—the number of distinct occupations. Division

[16] It is useful to note that under Lemma 1, all workers with say 1 unit of labor supply endowment have $1 - \frac{z}{N}$ unit of labor available for production. Worker productivity is maximized when N is infinite and every worker learns an infinitesimal task. In other words, without costs associated with greater division of labor, profit maximization requires that each skill be used as intensively as possible. I defer discussion on the costs of division of labor to Section 4.

of labor is the extent of worker specialization within a firm. In the context of this model, a more specialized worker is one who performs fewer number of tasks. A firm organizes its production process by partitioning the tasks that have to be performed to produce its output into occupations and assign to its workers. The more occupations there are (higher N), the fewer the number of tasks that each worker specializes in (lower $\frac{1}{N}$), and the greater the division of labor. Admittedly, the model omits other features that may affect the relationship between division of labor and the observed number of occupations within the firm; however, it elucidates in a transparent way how the extent of division of labor can be correlated with the number of occupations. Within the firm boundary, a smaller count of occupations implies that the workers who do the jobs tend to carry out a wider range of tasks. Conversely, in firms with a larger count of occupations, each job only requires a narrower range of tasks to be performed. Notably, the approach to use occupation counts as a proxy for division of labor is validated empirically by Becker et al. (2019), which use German worker survey data to show that within-plant occupation counts are inversely correlated with the tasks performed by workers, and hence positively associated with the extent of division of labor.

In the ensuing parts, I build on this theoretical insight and measure division of labor empirically using the total number of distinct occupations within a firm. I also discuss the limitations of this measure and the various robustness checks in place to address them in the following section.

3 Data and stylized facts

In this section, I first describe the data sources and definitions used in the empirical analysis. Using the theoretical insights developed in the previous section and the dataset constructed, I then document a new stylized fact: There is greater division of labor within firms in larger cities.

3.1 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 the universe of establishments operating in the formal market. RAIS data contain linked employer-employee records. Both employers and employees have an incentive to accurately report relevant information: The former are liable for fines if they fail to 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 been used extensively in the literature (e.g., Dix-Carneiro and Kovak, 2017; Helpman et al., 2017). 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.[17]

These data have several 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), which studies division of labor at the broader industry level, 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, instead of industry level. 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.[18] 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 of the Brazilian CBO-02 codes, with a total of more than 2,500 occupation codes, and each accompanied by detailed task descriptions. 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 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 manufacturing sectors.[19]

3.2 Definitions

Guided by the model in Section 2, I construct a measure of within-firm division of labor, reflecting 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 allows me to more accurately measure the extent of task division in the actual production process.[20] 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 or nonsupervisory occupation

[17] RAIS reports earnings for December and average monthly earnings during employed months in the reference year. Following Dix-Carneiro and Kovak (2017), I use December earnings to avoid seasonal variation or month-to-month inflation.

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

[19] Manufacturing sector corresponds to Brazilian Industry Codes CNAE20 10000-32990.

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

codes within an establishment (henceforth referred to as the number of occupations within an establishment). I consider an alternative measure, called the "specialization index," to account for the difference in distribution of workers across occupations. This is defined as one minus the Herfindahl index across occupations within an establishment (Ciccone, 2002; Duranton and Jayet, 2011). 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:

$$N_j = 1 - \sum_{o=1}^{\mathcal{O}} \left(\frac{l_j(o)}{l_j} \right)^2, \tag{3}$$

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. Finally, for robustness tests, I use the more aggregate 4-digit CBO codes. A more detailed discussion on the construction of measures for division of labor is offered in Appendix A.

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 (Demográfico, 2000). For my analysis, I use all 558 microregions. To compare city sizes, I use a normalized measure based on the population density.[21]

3.3 Division of labor and city size

Using the dataset, I document a new stylized fact on division of labor: There is greater division of labor within firms in larger cities. Specifically, I use the following OLS regression to investigate the relationship between division of labor and city size:

$$\log N_{jt} = \alpha_0 + \alpha_1 \log L_{m(j)t} + \delta_{s(j)} + \delta_t + \mathbf{X}_{jt} + \varepsilon_{jt},$$

where N_{jt} is the division of labor within an establishment j in year t—measured either by the number of occupations or the specialization index defined in (3), $L_{m(j)t}$ is the size of city m in which establishment j is located in year t, $\delta_{s(j)}$ is the sector fixed effect, δ_t is the year fixed effect, and $\mathbf{X}_{jt}$ is a set of controls.[22]

Baseline results: Table 1 summarizes the relationship between division of labor and city size. Columns (1) and (2) show the unconditional and conditional correlations, respectively. In Column (2), we see that

[21] 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 and positively correlated, the choice of measure matters little for my analyses.

[22] The controls include establishment-level controls: establishment employment sizes, occupation categories (defined as 3-digit CBO occupation codes), and skill intensities within firms; and city-level controls: state fixed effects, sector diversity within the city (computed using the Herfindahl index of employment across sectors within a city), and the total employment of sector s in city m.

within the same sector, holding fixed establishment size and other controls, division of labor is strongly and positively correlated with city size. The estimated elasticity, at 0.025, implies that an establishment located in São Paolo has 19% more measured extent of division of labor compared to a similar establishment—in terms of its sector, size, skill mix, etc—located in Itaperuna, a medium sized city in Brazil.

Robustness: I next consider different subsets of firms to assess the robustness of the baseline results. First, firms with multiple establishments may allocate different organizational functions across its establishments located in different cities. To make sure that this potentially endogenous allocation is not driving the results, I study only mono-establishment firms. As shown in Column (3) of Table 1, the positive correlation between division of labor and city size remains strong.[23] Second, while I consider an empirical measure of division of labor using the heterogeneity of occupations within an establishment as guided by theory, it is possible that the number of occupation codes is correlated with other variables, e.g., the diversity of establishment outputs. That is, if establishments located in larger cities tend to produce a greater variety of products—thus pushing up the number of occupation codes—it will generate a spurious correlation between division of labor and city size. To investigate this possibility, I use only data from sectors that produce homogeneous products (Foster, Haltiwanger and Syverson, 2008), for which the potential for product diversification is limited. Results in Column (4) of Table 1 remain qualitatively similar to the baseline results, implying that the product diversity channel is unlikely driving the observed correlation. Finally, Columns (5) - (8) show analogous analysis using the alternative definition of specialization index. The positive correlation remains strong under this alternative measure.[24]

Dependent variable	Log no of occs				Specialization index			
	All		Mono-estb	Homog	All		Mono-estb	Homog
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (city size)	.0576***	.0253***	.0243***	.0303***	.0188***	.0137***	.0132***	.0143***
	(.002)	(.0011)	(.0011)	(.0103)	(.0007)	(.0005)	(.0005)	(.0048)
Other controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Obs	2960066	2960066	2776735	6111	2960066	2960066	2776735	6111
R-sq	.116	.86	.856	.912	.078	.526	.529	.554

Standard errors clustered by city in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include sector and year FEs. Specialization index is defined in (3). Establishment-level controls are establishment size, skill intensity, and occupation categories (defined as the number of 3 digit occupation codes) within the establishment. City-level controls are state dummy, Herfindahl index of employment across sectors within the city, 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. Mono-estb firms refers to firms with a single establishments. Homogeneous sectors include corrugated and solid fiber boxes, bread, carbon black, roasted coffee beans, ready-mixed concrete, wooden flooring, gasoline, ice, plywood, and sugar (Foster, Haltiwanger and Syverson, 2008).

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

[23] Interestingly, the results with only multi-establishment firms—with a firm fixed effect to control for cross-firm differences—also show a positive correlation between division of labor and city size. See more details in Appendix A.2.

[24] The results are also robust to using the more aggregate 4-digit and 3-digit CBO codes, as reported in Tables 3 and 4 in Appendix A.2.

In addition to the considerations above, another important dimension of firm heterogeneity that the theory in the previous section and this analysis abstract from is the variation in the set of tasks performed within establishment boundaries. If within a sector, there is systematic variation in the number of tasks performed in firms located across different cities, it may create bias in our correlation estimates. For example, if firms in larger cities tend to undertake a greater number of tasks, then the positive correlation reflects the variation of firm boundary, rather than the variation in extent of division of labor within the firm boundary. While I do not have data that allow me to measure firm boundaries, there is a literature arguing that firms tend to focus on a narrower range of fewer number of tasks in larger cities, since it is easier to outsource some peripheral functions (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 that establishments with the same "true" extent of division of labor could have fewer number of occupation codes in larger cities as some of the tasks necessary to produce the outputs are outsourced. To the extent that this effect is present, it would lead to a downward bias in my elasticity estimate. In other words, we can interpret the estimated correlation as a lower bound of the actual value. Additionally, though establishment boundaries are not directly observed in the data, I consider an additional check to account for potential biases introduced by systematic variation in boundaries across space. 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" documented in Fort (2017). I then estimate the correlation between division of labor and city size for these two groups separately. Reassuringly, correlation results remain both qualitatively and quantitatively close to the baseline estimates, with the correlations for establishments more likely to fragment not significantly different from those less likely to do so, as shown in Table 7 of Appendix A.2.

Finally, as an alternative way to control 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 will also partially address the problem of not observing informal workers within formal establishments. Based on the Brazilian Urban Informal Economy Survey, the share of informal workers is negatively correlated with firm size. As shown in Table 5 of Appendix A.2, the positive correlations across all deciles suggest that the result is unlikely driven by differences in informal employment across space.

Discussion: The positive correlations documented above, though robust, cannot be interpreted as causal relationships. Instead, these are general equilibrium observations since both division of labor and production location are endogenous to firms. Despite the extensive set of controls and robustness checks considered, two firms located in different cities may have unobservable differences that are correlated with both division of labor and the location choice. Therefore, to adequately answer my research question—how city size affects firm's division of labor and, in turn, productivity— I develop a model that captures various potential forces that drive the observed correlation in the next section.

4 A quantitative model

In this section, I lay out a model which incorporates various channels that can potentially generate the observed correlation between city size and firm's division of labor. The goal is to construct a parsimonious theoretical framework that has a quantitative bite, can be generalized in various ways, and ultimately can be used for policy analysis. The theory generalizes the model of firm production organization in Section 2 and embeds it in a standard spatial sorting model with heterogeneous firms (Gaubert, 2018). The general equilibrium framework allows us to study firm's organizational and locational decisions jointly.

The theory's basic logic can be sketched as follows. Firms are differentiated exogenously in their process (and/or product) *complexity*. Given city size, a firm determines its optimal division of labor based on its complexity. Larger cities have comparative advantages for firms with greater 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 relatively more from being in larger cities, there is, in equilibrium, positive assortative matching between firm complexity and city size. Through the lens of the model, the observed correlation between division of labor and city size is generated through two distinct channels: (i) a selection channel, i.e., more complex firms, which are firms that would choose greater division of labor in any given city, endogenously sort into larger cities in equilibrium; and (ii) a treatment channel, i.e., any given firm would choose a greater extent of division of labor in a larger city.

4.1 Set-up and agent's problem

The economy consists of a continuum of homogeneous individuals of mass $\bar{L}$. There is also a continuum of homogeneous sites that can potentially be developed into cities. The number of cities and their corresponding population sizes are endogenous. Following Gaubert (2018), 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. The economy has a continuum of heterogeneous firms producing in 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. Agents and firms are both perfectly mobile across cities.

Each individual in city L is endowed with 1 unit of labor supply, which they supply inelastically and earn a wage $w(L)$. Agents consume a bundle of traded good and land. For simplicity, they require one unit of land for accommodation and do not increase their utility by consuming more land. Following Behrens, Duranton and Robert-Nicoud (2014), the utility function is assumed to be:

$$U = Xu(L), \tag{4}$$

where X is the consumption of traded good and $u(L)$ is the local amenity.

Consumers choose varieties within the bundle of traded goods X according to a CES aggregator:

$$X = \left[\int x(z)^{\frac{\sigma-1}{\sigma}} dz \right]^{\frac{\sigma}{\sigma-1}}, \tag{5}$$

where $\sigma > 1$ is the elasticity of substitution across varieties z.[25]

The amenity level in city L, $u(L)$, on the other hand, reflects the possible congestion externalities imposed on urban amenities that are affected by the size of the city. Following Allen and Arkolakis (2014), I assume that the overall amenity in a city of size L can be written as:

$$u(L) = \kappa L^{-\eta}, \tag{6}$$

where $\kappa > 0$ and $\eta \geq 0$. $\kappa L^{-\eta}$ can be interpreted as the inverse of urban cost associated with residing in a larger city. In Appendix B.1, I show that this functional form assumption can be microfounded using a standard model of a monocentric city in which commuting costs increase with population size.

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 (6) into the utility function:

$$\bar{U} = \left[\frac{w(L)}{P} \right] \kappa L^{-\eta}, \tag{7}$$

where P is an aggregate price index for X. Since the final good is freely traded, P is same in all cities.

Given (7), I derive the equilibrium income of an agent in city L:

$$w(L) = \bar{w} \kappa^{-1} L^{\eta}, \tag{8}$$

where $\bar{w} = \bar{U}P$ is an endogenous variable to be pinned down in general equilibrium.

4.2 Firms and production

I turn now to the production side of the economy. Firms differ exogenously in their fundamentals, called the "complexity." Firms choose their division of labor, production scale, and production location to maximize profits. Firms engage in monopolistic competition, and outputs produced by firms are freely traded across space.[26]

[25] To keep the theoretical framework parsimonious, I consider a single-sector economy. The model can be easily extended to incorporate multiple sectors. Indeed, for the quantitative analysis in Section 5, I extend the baseline model to a multi-sector economy and allow for sector-specific parameters in the structural estimation to retain flexibility and incorporate systematic variations across different sectors.

[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 to derive my analytical results. In Appendix B.6, I provide proof to show that all theoretical results hold with costly trade.

4.2.1 Production Technology

In the model, firms are heterogenous in their *complexities*. One way to interpret firm complexity can be done through the lens of the model developed in Section 2: It denotes (i) the learning cost for each task required to produce the output; and (ii) the total number of tasks. A more complex product is therefore one that requires more tasks to be performed in its production process and/or one that involves more difficult tasks that require more learning time. Firm z produces its output using the following technology:

$$Q(z) = \psi(N, L; z)l, \tag{9}$$

where N denotes division of labor in firm z and l denotes the number of workers within the firm. The endogenously determined firm productivity, defined as output per worker, is given by $\psi(N, L; z)$, which depends on the key endogenous variables N, division of labor within the firm, and L, the city size, as well as the exogenous complexity parameter z.[27]

4.2.2 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 and 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.

4.2.3 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. Firm's problem can be formally expressed as follows:

$$\max_{N,l,p,L} \ pQ - w(L)l, \tag{10}$$

subject to:

$$Q = \psi(N, L; z)l. \tag{11}$$

Given the isoelastic preferences in (5), the demand schedule faced by firm z is:

$$p(z) = Q^{-\frac{1}{\sigma}} R^{\frac{1}{\sigma}} P^{\frac{\sigma-1}{\sigma}}, \tag{12}$$

where $Q = \bar{L}x(z)$, since quantity produced equals the product of the quantity demanded by each agent and the number of agents, R denotes the total revenue, and $P = \left(\int p(z)^{1-\sigma} dz\right)^{\frac{1}{1-\sigma}}$ denotes the price index.

[27] I assume that $\psi(N, L; z)$ is increasing in z, so that productivity is higher for a more complex firm and a firm will never choose to produce below its exogenously given level of complexity, i.e., $\frac{\partial \psi}{\partial z} > 0$.

Consider a firm of complexity draw z in a 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,

$$\max_{N,L} \pi(L, N; z) \equiv \max_{N,L} \frac{(\sigma-1)^{\sigma-1}}{\sigma^{\sigma}} \left(\frac{\psi(N, L; z)}{w(L)}\right)^{\sigma-1} RP^{\sigma-1}. \tag{13}$$

Given L, firms choose the optimal division of labor, $N(L; z)$, to maximize profits:

$$N(L; z) \equiv \arg\max_{N} \pi(L, N; z). \tag{14}$$

Substituting $N(L; z)$ into the profit function (13), I get the optimal profit of firm z in city L:

$$\pi^*(L; z) \equiv \frac{(\sigma-1)^{\sigma-1}}{\sigma^{\sigma}} \left(\frac{\psi(L; z)}{w(L)}\right)^{\sigma-1} RP^{\sigma-1}, \tag{15}$$

where $\psi(L; z) \equiv \psi(N(L; z), L; z)$. Lastly, firm employment, conditional on being in a city of size L, is given by

$$l(L; z) = (\sigma-1)\frac{\pi^*(L; z)}{w(L)}. \tag{16}$$

4.3 Spatial equilibrium

I characterize spatial equilibrium in this section. I show that under a simple assumption, 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.

In spatial equilibrium, homogeneous workers are indifferent across locations, while firms choose their locations optimally based on their complexity draws.[28] To fully analyze the characteristics of the equilibrium, I make the following assumptions:

Assumption 1 $\psi(N, L; z)$ *is twice-differentiable, and displays*

(a) strict log-supermodularity in firms' complexity z and division of labor N, i.e.,

$$\frac{\partial^2 \log \psi(N, L; z)}{\partial N \partial z} > 0;$$

and

(b) strict log-supermodularity in city size L and firms' division of labor, i.e.,

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

[28] See Appendix B.2 for a formal definition of the spatial equilibrium.

The first part of Assumption 1 states that there is complementarity between complexity and division of labor, e.g., a more complex production process benefits more from greater division of labor. The stylized model in Section 2 provides one microfoundation that generates such relationship. In that model, the gains from division of labor come from the savings on learning costs. Since more complex products require more learning time, the gains from worker specialization are higher for more complex firms.[29] The second part of Assumption 1 states that there is complementarity between city size and division of labor, e.g., 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 empirical support of this hypothesis in Section 7.[30] In what follows, I remain agnostic on the sources generating these relationships to highlight the generic features of an economy with such complementarities.

I highlight three noteworthy points before proceeding. First, it is important to note that while I could include all other cases of $\psi(N, L; z)$ 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, consistent with the empirical pattern documented in Section 3.

Second, the baseline model adopts a minimum set of assumptions necessary to obtain the general equilibrium outcome, in 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 additional terms that summarize other channels that might also increase firm productivity in larger cities. By separately identifying these channels, I can investigate and isolate the importance of division of labor in affecting productivity differences across cities. I discuss this in detail in Section 5.

Lastly, in Section 7, I present causal empirical evidence that is consistent with the log-supermodularity assumptions between N and z, and between N and L. I do so by focusing on one particular channel mentioned

[29] See Appendix B.3 for a formal discussion.

[30] 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.4, 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.4.

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.

4.3.1 Characteristics of the profit function

In my theoretical framework, there is complementarity between complexity and division of labor, and between city size and division of labor. Combining these assumptions generates the following results:

Lemma 2 *Suppose that Assumption 1 holds, firms' profit function, $\pi(L, N; z)$, displays log-supermodularity in (z, L, N).*

Lemma 3 *Suppose that Assumption 1 holds, the optimal division of labor given z and L, denoted by $N(L; z) = \arg\max_N \pi(L, N; z)$, increases in (z, L).*

Division of labor depends on the trade-off between gains and costs of specialization. Given the log-supermodularity between L and N, a larger city increases firms' division of labor, e.g., by lowering the costs at the margin. Similarly, the log-supermodularity between z and N implies that as complexity increases, division of labor goes up too, perhaps through increasing the benefits of worker specialization at the margin. Using a classic result in monotone comparative statics (Topkis, 1978), since the profit function $\pi(L, N; z)$ is log-supermodular in (z, L, N), once the firm solves for its optimal division of labor, $N(L; z)$, the profit function $\pi^*(L; z)$ displays log-supermodularity in (z, L).[31]

Lemma 4 *Denoted by $\pi^*(L; z) \equiv \max_N \pi(L, N; z)$, the optimal firm profit given z and L is log-supermodular in (z, L), if Assumption 1 holds.*

4.3.2 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

$$\pi^*(L; z) = \frac{(\sigma - 1)^{\sigma - 1}}{\sigma^\sigma} \left(\frac{\psi(L; z)}{w(L)} \right)^{\sigma - 1} R P^{\sigma - 1}.$$

[31] In addition to Assumption 1, I also assume that $\frac{\partial^2 \psi}{\partial z \partial L}$ is non-negative, i.e., more complex firms are not worse off in larger cities relative to less complex firms. In Section 5, I formalize this assumption and quantitatively assess the extent of direct interaction between firm complexity and city size.

Lemma 2 implies that the profit function shown in (15) is log-supermodular in (z, L), suggesting that more complex firms benefit more from being located in larger cities. However, given the 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 Appendix B.7 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 (15), the first order condition with respect to L is therefore:

$$\frac{\psi_L}{\psi} = \frac{\eta}{L}, \tag{17}$$

where $\psi_L = \frac{\partial \psi(L; z)}{\partial L}$.

In (17), $\frac{\psi_L}{\psi}$ corresponds to the marginal productivity benefit of being in a larger city. $\frac{\eta}{L}$ corresponds to the marginal cost of being in a larger city. It is equal to the extra costs due to more expensive labor costs. 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 with complexity z. Define the solution to (17) as

$$L^*(z) \equiv \arg\max_{L \geq 0} \pi^*(L; z). \tag{18}$$

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 $L^*(z)$ 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 $w(L^*(z))$. 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 optimal set of city sizes for firm distribution $f(z)$. Given $\mathcal{L}$, the optimal location choice for each firm z is defined by the following matching function:

$$L(z) = \arg\max_{L \in \mathcal{L}} \pi^*(L; z). \tag{19}$$

Using the definition in (19) and Lemma 4, I can invoke a classic theorem in monotone comparative statics (Topkis, 1978) and obtain the following key theoretical result.

Proposition 5 *Suppose that Assumption 1 holds. In the heterogeneous equilibrium, within a sector, high-z firms sort into larger cities. More formally, the matching function is increasing in z, or $L'(z) > 0$.*

The intuition for Proposition 5 is straightforward. Larger cities have higher urban costs due to congestion, so workers require higher wages in these locations. Larger cities attract firms because of the productivity advantage through division of labor. In particular, 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.[32]

4.4 Characterizing spatial equilibrium

In the heterogeneous spatial equilibrium, division of labor $N(z)$, profit $\pi(z)$, revenue $r(z)$, and firm size $l(z)$ are all determined by the matching function $L(z)$. The strict sorting of z generates the strict sorting of firm division of labor, profits and revenue. I denote the equilibrium variables using the following expressions:

$$N(z) = N(L(z); z), \tag{20}$$

$$\pi(z) = \frac{(\sigma-1)^{\sigma-1}}{\sigma^\sigma} \left(\frac{\psi(L(z); z)}{w(L(z))} \right)^{\sigma-1} P^{\sigma-1} R, \tag{21}$$

$$r(z) = \sigma \pi(z), \tag{22}$$

$$l(z) = \frac{\pi(z)}{(\sigma-1) w(L(z))}. \tag{23}$$

Given the results in Proposition 5, these firm-level observables also exhibit complementaries between firm complexity and city size, as stated in the following result:

Proposition 6 *In equilibrium, firms' division of labor, revenue, and profit all increase with city size. Formally, consider two firms z and z'. If $L(z) > L(z')$, then $N(z) > N(z')$, $\pi(z) > \pi(z')$, $r(z) > r(z')$, and $w(z) > w(z')$.*

In equilibrium, high-z firms sort into larger cities. This generates the motivating fact presented in Section 3: Firms' division of labor is greater in larger cities.[33] Through the lens of my model, I show how the correlation can be potentially achieved through two distinct channels: 1) *Selection channel:* more complex firms, which are firms that would choose greater division of labor in any given city, endogenously sort into larger cities in equilibrium; and 2) *Treatment channel:* given that larger cities facilitate greater division of labor for all firms, any given firm would endogenously choose greater division of labor if it is located in a larger city.

[32] In Appendix B.8, 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.

[33] In Appendix B.10, I provide further sector-specific evidence on the equilibrium results, i.e., positive correlations between division of labor and city size, as well as between firm revenue and city size.

Out of these two channels, only the *treatment channel* is relevant for the purpose of investigating how city size affects division of labor. Empirically, however, it is impossible to separate these two mechanisms short of running a natural experiment in which firms are randomly allocated across space. By developing a quantifiable model, I can tackle this question by relying on the structure of the model. Through a counterfactual experiment, I pin down the direct effect of city size by shutting down the selection channel. I describe the estimation strategies in more detail in the next section.

5 Quantitative analysis

In this section, I structurally estimate an extended version of the model following 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 remaining parameters. 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 extents of complementarities between division of labor and city size, and 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.

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 first remove establishments with annual labor payment below 10,000 Brazilian reals (approximately 2,000 USD) and trim the bottom and top 1% of the data. This leaves me with 192,286 establishments. For the estimation, I aggregate establishments into 20 sectors and allow parameters to vary by sectors. Summary statistics are reported in Table 8 of Appendix C.

5.1 Model specification

To carry out the structural estimation, I need to fully characterize the features of firm production function. I adopt the following functional form assumption for ψ:

$$\log \psi(N, L; z) \equiv (\log z)(1 + \log N)^c - (1 + \log N)(\log \tilde{L})^{-\theta}, \tag{24}$$

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

Under this functional form assumption, it is straightforward to see that the first term, $(\log z)(1+\log N)^c$, is strictly increasing in z and N. Importantly, the relationship between firm complexity and division of labor is determined by c. A positive value of c would confirm model assumption. When $c = 0$, I obtain a model in which the worker productivity is solely determined by firms' complexity draws. Analogously, the second term, $-(1 + \log N)(\log L)^{-\theta}$ is decreasing in L and N. θ governs the relationship between division of labor and city size. A positive value of θ is consistent with model assumption. If there is no relationship between N and L, then $\theta = 0$.

Additionally, following the conventional literature, I assume that $\log z$ is distributed according to a normal distribution with mean variance ν_z, truncated at its mean to prevent $\log z$ from being negative.

5.2 Model extensions

The parsimonious model presented in Section 4 constitutes a minimum set of elements to elucidate key mechanisms at work in order to focus on investigating the relationship between city size and division of labor, which in turn determines the relationship between city size and productivity *through this source.* To bring the model to data, however, I need to incorporate additional features to reflect other relevant forces at work that may also affect the observed relationship between city size and productivity.

Specifically, I adopt four extensions: (i) multiple sectors, (ii) other sources of agglomeration externalities, (iii) imperfect sorting of firms, and (iv) spatial equilibrium with a discrete set of cities. My extended model allows me to obtain results under less restrictive assumptions than Section 4, and to evaluate the contribution of division of labor to productivity differences across cities relative to other forces, on which my baseline model is silent.

Equation (25) shows the updated firm productivity after incorporating the first three extensions. First, I incorporate multiple sectors in the economy so that we can allow for sector-specific parameters to incorporate systematic differences across firms in different sectors. With multiple sectors, the bundle of traded goods X is a Cobb-Douglas combination of goods over $s = \{1, \ldots, S\}$ sectors.

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

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

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

where $\sigma_s > 1$ is the elasticity of substitution across varieties z within sector s. Lastly, the aggregate price index for the multi-sector composite good summarizes he price indexes P_s for all tradable sectors:

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

With this extension, all results in Section 4 hold for firms within a given sector.

Second, I include two terms in firms' productivity function that summarize other sources of agglomeration externalities, from which my model abstracts. The first term, $\alpha_s \log L$, incorporates productivity advantage of larger cities, which are not correlated with firm complexity or division of labor. This includes, but is not limited to, the sorting of skills among heterogeneous workers, knowledge spillover, and natural amenity differences. $\alpha_s > 0$ implies that a firm located in a larger city is more productive for reasons beyond division of labor. The second term, $\log z_j \log L^{\upsilon_s}$, incorporates potential direct interaction between firm complexity and city size. While the baseline 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, when $\upsilon_s > 0$, this term allows more complex firms to sort into larger cities for reasons beyond division of labor.

Next, 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 4.4 still hold.[34]

In summary, with these three extensions, productivity of a firm j with complexity draw z in sector s takes the following form:

$$\log \psi_j(N, L; z) = \alpha_s \log L + \log z_j \log L^{\upsilon_s} + (\log z_j)(1 + \log N)^{c_s} - (1 + \log N)(\log L)^{-\theta_s} + \epsilon_{jL}, \tag{25}$$

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, 2018). Additionally, when $\upsilon_s = 0$, I obtain a classic model of agglomeration

[34] 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 (2018) for an excellent discussion of this.

externalities without division of labor or firm sorting (see, e.g., Allen and Arkolakis, 2014).

Finally, I consider a discrete set of cities in estimating the model. 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. Note, however, 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. Cities, indexed by m, are ordered by their city size L_m. 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 $[\underline{z}_s(m), \bar{z}_s(m)]$. Spatial equilibria are determined by the following indifference condition:

$$\pi_s(\bar{z}, m) = \pi_s(\underline{z}, m + 1), \quad \forall L_m \in \mathcal{L}. \tag{26}$$

While the new spatial equilibria may no longer be unique, the equilibrium characteristics presented in Section 4.4 hold for both continuous and discrete cases.

5.3 Estimation procedure

5.3.1 Step one: Direct calibration

I begin by estimating the 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 elasticity of urban costs with respect to city size, 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, η corresponds to the elasticity of wages with respect to city size, from (8). To account for heterogeneity of workers across space, I calculate the elasticity using residuals from a Mincerian wage regression and obtain an elasticity of $\eta = 3.1\%$.[35]

5.3.2 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, the 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

[35] 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.

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.

5.4 MSM procedure and moments

Given the distributions of firm complexities and idiosyncratic firm-city-size shocks, parametric assumptions, and the parameters calibrated 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:

$$\log N_j^* \equiv \underset{N \in \mathbf{R}^+}{\arg\max} \ \log z_j (1 + \log N)^{c_s} - (1 + \log N)(\log L)^{-\theta_s}. \tag{27}$$

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

$$\log \psi_j(L; z) = \alpha_s \log L + \log z_j \log L^{\upsilon_s} + \log z_j (1 + \log N_j^*)^{c_s} - (1 + \log N_j^*)(\log L)^{-\theta_s} + \epsilon_{jL}. \tag{28}$$

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

$$\begin{aligned} \log L_s(z_j) &\equiv \underset{L \in \mathcal{L}}{\arg\max} \ \log \psi_j(L; z) - \log w(L) \\ &= \underset{L \in \mathcal{L}}{\arg\max} \ (\alpha_s - \eta) \log L + \log z_j \log L^{\upsilon_s} + \log z_j (1 + \log N_j^*)^{c_s} - \frac{1 + \log N_j^*}{(\log L)^{\theta_s}} + \epsilon_{jL}. \end{aligned} \tag{29}$$

To estimate the five parameters in χ_s, I match six sets of simulated and data moments for every sector: (i) geographic distribution of firms, (ii) firm-size distribution, (iii) cross-city variations in firm size., (iv) cross-city variations in division of labor, and (v) within-city variations in division of labor.[36]

The first three sets of moments jointly identify α, υ, ν_z and ν_L. The identification of the two complementarity parameters, c and θ, is possible because I observe firm-level division of labor. In equilibrium, the joint parameter $\frac{\theta}{1-c}$ captures the relationship between firms' division of labor and city size. By observing

[36] 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.

how the average division of labor increases across city sizes, I can identify $\frac{\theta}{1-c}$. Note that I purposely use the unconditional correlation between division of labor and city size because I want to retain the identifying variation arising from the general equilibrium effect of city size on firm's division of labor. Indeed, the model implies that the firms in larger cities are on average larger in size and more complex in their production processes. To separately identify c and θ, I consider within-city variations in firms' division of labor. Given a city size and within a sector, 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. See Appendix C for further discussions on moments and identification.

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

$$\hat{\chi}_s = \arg\min\,(m_{s,data} - m_{s,sim}(\chi_s))' J_s (m_{s,data} - m_{s,sim}(\chi_s)), \tag{30}$$

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).[37] 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 Appendix C.

5.5 Estimation results

In this section, I present results from the MSM estimation. Estimated parameters by sector are reported in Table 9 of Appendix C. I first examine model fit for the set of targeted moments (reported in Figures 7 to 11 in Appendix C). 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.

[37] 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.

I next move on to nontargeted moments. In particular, I consider two sets of nontargeted moments that combine the 20 sectoral estimation results. The first set considers the relative magnitude of c—the estimated complementarity parameter between complexity and division of labor—across different sectors. The estimation is made for each sector separately. I make no assumption on the relative size of c across sectors. However, if the sectoral complexity is higher, i.e., the average firm in a given sector is more relatively more complex, then one can expect that complementarity to be stronger too. To compare the estimated complementary parameters with sectoral complexities, I consider two data proxies. The first measure 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).[38] To relate the estimates of c to the two empirical proxies, I 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 12 of Appendix C.4 plots the rank of the estimates across sectors against the empirical measures, and shows that the two sets of values line up well.

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.[39] 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 Appendix C for detailed steps). As shown in Figure 13 of Appendix C.4, the estimated city-size distribution adheres to Zipf's law and follows the actual city-size distribution well.

6 Division of labor and productivity advantage of cities

Armed with the estimated model, I next conduct a counterfactual exercise to quantify the contribution of division of labor to the productivity gains in larger cities. 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 and firm sorting, the variance of firm complexity distribution, and firm-city-size idiosyncratic shocks, but also by firms' decisions regarding the extent of division of labor. Moreover, the productivity impacts through division of labor is driven by two forces—the direct effect of city size on

[38] 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.

[39] 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.

productivity by facilitating division of labor (i.e., the treatment channel) and the increase in productivity driven by spatial sorting of more complex firms (i.e., the selection channel). The counterfactual exercise detailed below unpacks these different forces.

I begin by studying how firm productivity varies city size in Brazil, using the following OLS regression on the simulated set of data:

$$\log \hat{\psi}_j = \beta_0 + \beta_1 \log L_j + \delta_{s(j)} + \epsilon_j \tag{31}$$

where $\hat{\psi}_j$ is the simulated firm productivity defined in (25) for firm j, L_j is the optimal city size chosen by the firm according to (29), and $\delta_{s(j)}$ is a sector fixed effect. Under this set up, β_1 is the elasticity of firm productivity with respect to city size. Running the OLS regression in (31), I get an OLS estimate of $\hat{\beta}_1 = 0.0881$.[40] This measure is 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 study how division of labor affects the productivity advantage of larger cities, I conduct the following counterfactual analysis, in which I shut down any productivity increase through the division of labor channel. This is achieved by making two changes in simulating the economy. First, I assume that firms choose their optimal city sizes according to the complexity draw and their firm-city-size specific shocks, instead of considerations related to division of labor, i.e.,

$$\log \tilde{L}_s(z_j) = \underset{L \in \mathcal{L}}{\arg\max}\ (\alpha_s - \eta) \log L + \log z_j (\log L)^{\upsilon_s} + \epsilon_{jL}. \tag{32}$$

Second, I do not allow firms' to choose their optimal division of labor, by fixing firm-level division of labor based on the average value within their respective 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 (31) using the new simulated data leads to an elasticity of firm productivity to city size of 0.0727. By this account, division of labor accounts for 17% of the productivity advantage in larger cities.[41] That is, absent of division of labor as a channel, the productivity advantage of larger cities would have been reduced by one-sixth. 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, 2019).[42]

[40] This implies productivity goes up by 8.81% when city size is doubled.

[41] Without the endogenous choice of division of labor, the elasticity estimate goes down by 0.0154 (0.0881 - 0.0727) , which is 17% (0.0154 / 0.0881) of the baseline elasticity.

[42] I also consider an alternative approach in which I re-estimate $\log \hat{\psi}_j$ by removing the standard agglomeration externalities,

A key challenge in the reduced-form exercise in Section 3 is that we cannot separately identify the relative contributions by firm's spatial sorting decisions (i.e., the selection channel) and the direct effect of city sizes (i.e., the treatment channel) in generating the correlation between division of labor and city size, which in turn determines the correlation between productivity and city size through division of labor. Using the estimated model, however, we can separate the two channels by relying on the structure of the model. Specifically, to examine the importance of *firm selection* to the 16% productivity contribution through division of labor, I conduct a second counterfactual exercise. In the model, firms sort into larger cities because larger cities foster greater division of labor, through the log-supermodularity assumption embedded in the firm productivity function between division of labor and city size. 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 according to (32), same as the first counterfactual exercise. However, once firms are allocated to a city size, they are allowed to choose the optimal division of labor based on their complexity draws and the city to which they are allocated. This counterfactual exercise effectively shuts down the firm selection channel, thereby allowing me to study the realized productivity with only the treatment channel that captures the direct effect of city size on division of labor. I find, by re-estimating (31), that the elasticity estimate drops to 0.0803, which is about half of the reduction in the first counterfactual experiment in which both channels are shut off. This implies that the firm sorting channel accounts for about half of the spatial productivity differences through division of labor.[43]

Relying on the structure of the model and the detailed Brazilian establishment-level data, I identify productivity advantage of cities through the channel of division of labor and further separate the relative contributions by the selection and treatment channels. Admittedly, as with most quantitative exercises, the question remains as to what extent the results are driven by the model assumptions. To lend further credibility to the key quantitative findings, I rely on a quasi-experiment in Brazil in the next section, in order to provide additional qualitative and quantitative support to my model.

7 External validation to proposed theory

In this section, I use Brazilian micro-level data to validate the qualitative and quantitative predictions from the model, through a quasi-experiment. The theoretical framework presented in Section 4 makes two key assumptions, i.e., the log-supermodularities between N and z, and between N and L. I first show that given

$\alpha \log L + \log z(1 + \log L)^{\upsilon}$, in (25). This assumes that the productivity advantage in larger cities only comes from my proposed channel of division of labor. Re-estimating (31) gives me similar results. I find that division of labor generates an elasticity estimate of 0.0151, which is 17% of the original value.

[43]Without systematic firm sorting, the elasticity estimate goes down by 0.0078 (0.0881 - 0.0803), which is 50% (0.0078 / 0.0154) of the contribution of division of labor to the spatial productivity difference. 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.

these assumptions, the model generates two predictions: In equilibrium, an exogenous shock to division of labor (i) affects more complex firms more, and (ii) affects firms in larger cities more. In practice, many factors potentially affect firms' decisions on division of labor. To establish a plausibly exogenous variation in division of labor, I rely on a quasi-experiment in Brazil in which a national infrastructure plan increases firms' division of labor. I use the quasi-experiment to examine the heterogeneities in the treatment effects in accordance with the theoretical predictions.

Additionally, the quasi-experiment also provides out-of-sample validation to the structural estimation, by examining quantitative predictions from the model. Since the model is estimated without using data after the implementation of the new infrastructure, the comparison between the model-based predicted impact and the actual changes provides an assessment on the fit of the estimated model.

7.1 Heterogeneous impact of division of labor shocks

Recall the firm productivity function in (25):

$$\log \psi_j(L; z) = \alpha \log L + \log z_j \log L^{\upsilon} + \log z_j (1 + \log N_j^*)^c - (1 + \log N_j^*) \log L^{-\theta} + \epsilon_{jL}.$$ [44]

Note that the optimal level of division of labor, under this specific parametric assumption, is

$$N_j^* = \exp[c^{\frac{1}{1-c}} (\log z)^{\frac{1}{1-c}} (\log L)^{\frac{\theta}{1-c}} - 1]. \tag{33}$$

Under the parametric assumption, the log-supermodularities in ψ between (N, z) and (N, L), under Assumption 1, are determined specifically by the signs of c and θ, for which positive values would imply the complementary relationships between these two pairs of variables. To provide further support for the assumptions, consider an infinitesimal exogenous shock to N, denoted by $\partial \mathcal{I}$, i.e.,

$$\partial \log N = \left(c^{\frac{1}{1-c}} (\log z)^{\frac{c}{1-c}} (\log L)^{\frac{\theta}{1-c}} - 1 \right) \partial \mathcal{I} > 0.$$ [45]

Given the complementary assumptions, the exogenous shock has heterogeneous effects for firms with different complexity parameters and located in different cities, as stated in the following proposition.

Proposition 7 *Consider an exogenous shock, $\partial \mathcal{I}$, to firm's division of labor, if $c > 0$ and $\theta > 0$, then:*

1. *Within a city, the increase in division of labor is higher for more complex firms, i.e.,*

$$\frac{\partial^2 \log N}{\partial \log z \, \partial \mathcal{I}} > 0.$$

[44] I omit the sector subscript on parameters for expositional simplicity.

[45] In this case, the envelop condition implies that the firms' and workers' location choices do not change.

2. Across cities, the increase in division of labor is higher for firms in larger cities, i.e.,

$$\frac{\partial^2 \log N}{\partial \log L \, \partial \mathcal{I}} > 0.$$

The formal proof can be found in Appendix B.5.[46] Intuitively, the shock reduces (increases) the costs (benefits) of division of labor at the margin. Given the complementarity between division of labor and complexity, more complex firms will benefit more from the shock, thus increasing their division of labor to a larger extent. Additionally, as stated in Proposition 5, more complex firms sort into larger cities given the complementarity between division of labor and city size. This implies that across cities, affected firms located in larger cities will also benefit more from the shock and increase their division of labor to a greater extent relative to those in smaller cities. These two predictions regarding the heterogeneity of the impacts from the local shock speak directly to the complementarity assumptions, which are at the heart of my theoretical framework. I next test the model predictions formally using a quasi-experiment in Brazil.

7.2 Quasi-experiment in Brazil

I rely on a quasi-experiment in Brazil to provide causal evidence in support of the model predictions in Proposition 8. The shock to division of labor arises from a national ICT infrastructure plan that affects various Brazilian regions differently. The quasi-experiment allows me to examine the heterogeneities in the treatment effects in the theoretical predictions. Moreover, by focusing on a specific hypothesis that generates the complementarity between division of labor and city size, I provide causal evidence on the existence of one potential microfoundation behind the assumed complementarity: Larger city provides better public infrastructure, through its larger tax bases—e.g., ICT infrastructure that reduces the communication frictions and coordination costs within a firm—thereby facilitating a greater extent of division of labor for firms in these larger cities. In Appendix B.4.2, I provide the formal description for the link of this microfoundation to the baseline model.

7.2.1 Background

The quasi-experiment is a large-scale ICT infrastructure project through the National Broadband Plan (*Programa Nacional de Banda Larga*, PNBL henceforth). In Brazil, the availability of broadband access closely reflects the country's wide variation in city size, as illustrated in Figure 1.[47]

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 broad-

[46] While the results focus on the parameterized model, in the appendix I provide proof for both the non-parameterized and parameterized models.

[47] 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.

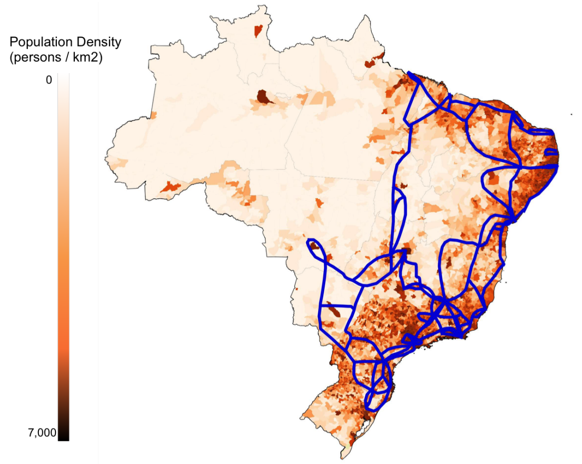


Figure 1: Broadband backbone and population density in 2010

band provision, leaving private operators to provide broadband infrastructure where they find it profitable to do so (Jensen, 2011; Knight, Feferman and Foditsch, 2016).[48] 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 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, Feferman and Foditsch, 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.[49] Between 2012 and 2014, PNBL added 48,000 km of new broadband backbone. Table 10 in Appendix D compares establishment characteristics between control and treatment groups. On average, treated establishments have greater division of labor and a higher share of managers, and are larger in size.[50]

7.2.2 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

[48]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, 2019).

[49]"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 Appendix D.2 for more discussion.

[50]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 Appendix D.1 for more extensive discussions on the background and details of the PNBL.

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

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.[51] 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, 2020), 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 sectoral complexity measures, I remove those establishments that relocate or change their sector classifications during the study period. This leaves 827,829 establishments over 9 years, or 91,981 establishment-year observations, for the empirical analysis.

[51]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.

7.3 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. This test ascertains that the new ICT infrastructure indeed affects firm's extent of division of labor. I run:

$$\log N_{jt} = \alpha + \beta Backbone_{jt} + \delta_j + \delta_t + \varepsilon_{jt}, \tag{34}$$

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 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.[52] 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 (2019), 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.[53] 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 makes predictions regarding heterogeneities in the treatment effects, as stated in Proposition 8. Specifically, the impacts of the new ICT infrastructure are larger for establishments located in larger cities relative to smaller cities, and for more complex establishments relative to less complex ones. While firm complexity is not directly observable, we can test the model prediction using proxies that measure the average firm complexities at the sector level. To this end, we revert to the two measures considered in Section 5.5: (1) number of intermediate inputs used by each sector in producing the sector-level outputs, and (2) the export share of goods by G3 economies. I test these predictions using (35) and (36). The model predicts that $\gamma > 0$ and $\omega > 0$.

$$\log N_{jt} = \alpha + \beta Backbone_{jt} + \gamma Backbone_{jt} \times \log L_{m(j),t_0} + \delta_j + \delta_t + \varepsilon_{jt}, \tag{35}$$

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

[53] 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.

$$\log N_{jt} = \alpha + \beta Backbone_{jt} + \omega Backbone_{jt} \times \log z_{s(j),t_0} + \delta_j + \delta_t + \varepsilon_{jt}, \tag{36}$$

where $\log L_{m(j),t_0}$ is the size of the city m in which establishment j is located and $\log z_{s(j),t_0}$ denotes the complexity of sector s that establishment j produces in. I use both measures of sector-level complexity for the regressions.[54]

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.[55] In Table 25 of Appendix D, 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.

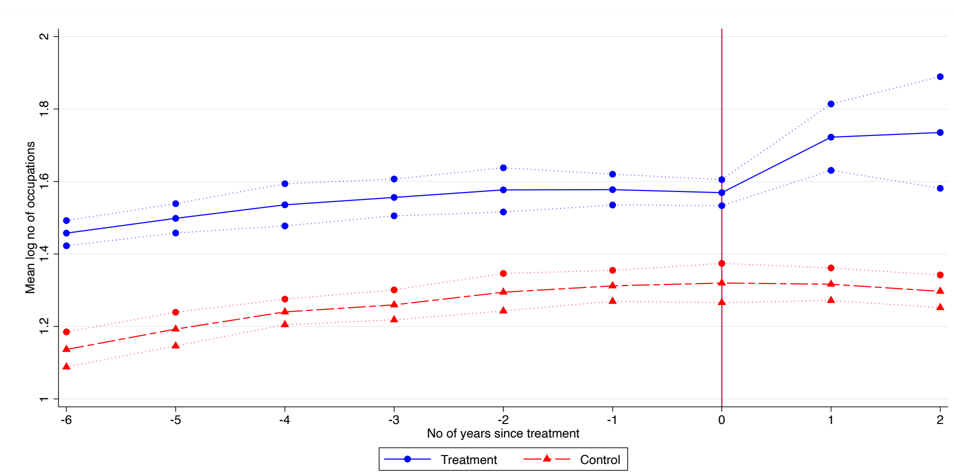


Figure 3: Log number of occs in treated versus control groups in Brazil
Dotted lines represent the 95% confidence bands.

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,

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

[55]Figure 16 in Appendix D shows the pre-trend graph for the specialization index. The two figures are very similar qualitatively.

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.

7.4 Results

Table 2 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 1.36 percentage point relative to other areas, whereas Column (3) shows that the specialization indices within these establishments increase by 0.09.

Next, I move on to the results that investigate the heterogeneous effects of the ICT shock. Columns (2) and (6) show the results for (35). 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. Finally, columns (3), (4), (7), and (8) illustrate results for (36). Again, consistent with the model prediction, the impacts of new ICT infrastructure are greater for establishments that produce in more complex sectors.

Before moving on, it is worth highlighting that the estimate of the treatment effect captures not just the direct effect of ICT on firm's division of labor, but also any potential indirect effect of this ICT shock on division of labor through other channels. When this happens, the total effect I estimate may differ from the theoretically defined partial effect of the model. However, as long as division of labor increases in response to the ICT shock, the theoretical predictions on the heterogeneities of the treatment effects still hold. Therefore, the discrepancy in the interpretations for the direct treatment effect does not undermine the objective of model validation. Furthermore, the quantitative evaluation of the model using the quasi-experiment targets the total effect in calibrating the magnitude of the shock. I discuss this in more detail in Section 7.5.

Alternative interpretations

I discuss two alternative interpretations of the results and describe the tests in place to ensure the validity of my interpretation. First, 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 (34), (35), and (36).[56] As shown in Panel B of Table 2, results are both qualitatively and quantitatively similar to the baseline results.

Second, 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 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.[57] 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 (34) to (36).[58] As shown in Panel C of Table 2, results are again similar to baseline results.

Robustness checks

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 D.3.

[56] 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.

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

[58] 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.

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								
$Backbone_{jt}$	.0136***	.0013	.0315	.0129	.0655***	.0116	.0728***	.0805***
	(.0029)	(.003)	(.0038)	(.018)	(.017)	(.0085)	(.014)	(.016)
$Backbone_{jt} \times \log L_{ct_0}$		.0092***				.0145***		
		(.0009)				(.0027)		
$Backbone_{jt} \times \log z_{st_0}$			.0153***	.004***			.02***	.0084***
			(.0031)	(.0031)			(.0054)	(.0012)
Mean of outcome	1.66	1.66	1.66	1.66	.57	.57	.57	.57
Obs	827829	827829	827829	827829	827829	827829	827829	827829
R-sq	.862	.862	.862	.854	.668	.668	.668	.668
Panel B: Excluding IT-related occupations								
$Backbone_{jt}$	.0133***	.0023	.0014	.012	.065***	-.011	.0629***	.0819***
	(.0034)	(.0025)	(.009)	(.0089)	(.0037)	(.0096)	(.0205)	(.0136)
$Backbone_{jt} \times \log L_{ct_0}$		.0089***				.0109***		
		(.0015)				(.0027)		
$Backbone_{jt} \times \log z_{st_0}$			.0154***	.0072***			.027***	.0038***
			(.0032)	(.0016)			(.0015)	(.0015)
Mean of outcome	1.65	1.65	1.65	1.65	.56	.56	.56	.56
Obs	827379	827379	827379	827379	827379	827379	827379	827379
R-sq	.862	.862	.862	.862	.667	.667	.667	.667
Panel C: Dropping occupation categories did not exist in the previous period								
$Backbone_{jt}$	.0096***	.009	.063	.009	.054***	.012	.0543***	.0708***
	(.0018)	(.0226)	(.081)	(.0321)	(.0037)	(.02)	(.0131)	(.0127)
$Backbone_{jt} \times \log L_{ct_0}$		.0132***				.0125***		
		(.0032)				(.0038)		
$Backbone_{jt} \times \log z_{st_0}$			.0123**	.0039***			.0134***	.0046***
			(.0053)	(.0015)			(.0053)	(.0011)
Mean of outcome	1.54	1.54	1.54	1.54	.42	.42	.42	.42
Obs	827829	827829	827829	827829	827829	827829	827829	827829
R-sq	.824	.824	.824	.824	.615	.615	.614	.615

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

Table 2: Impacts of fast internet on division of labor within establishments

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 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 Appendix D.4, I show that an improvement in internet connectivity reduces the share of managers within establishments, consistent with a reduction in coordination frictions.[59]

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 qualitative theoretical predictions. In the following section, I use the reduced-form estimates to evaluate the quantitative predictions of the model.

7.5 External validation to the structural estimates

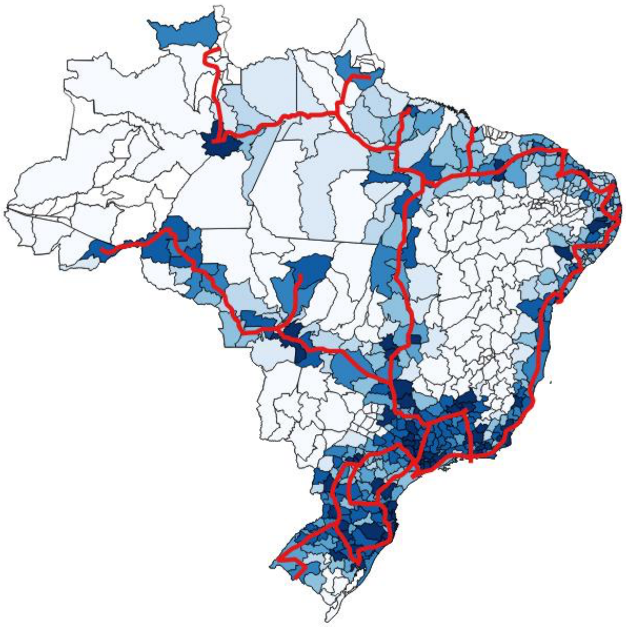
In addition to providing empirical support to the key model assumptions, I also use the quasi-experiment to provide out-of-sample validation to the estimated model. Since the model is estimated without using data after the implementation of new ICT infrastructure, I am able to compare the model-based predicted impact to the actual changes.[60]

I first compute the direct effect of the new infrastructure on division of labor, using the reduced-form analysis in Section 7.4. The new broadband infrastructure increases division of labor within firms in the treated areas by 1.36% more than the control areas. With the estimated model, I calibrate the magnitude of the productivity impact in treated cities in response to the broadband rollout to match the estimated average

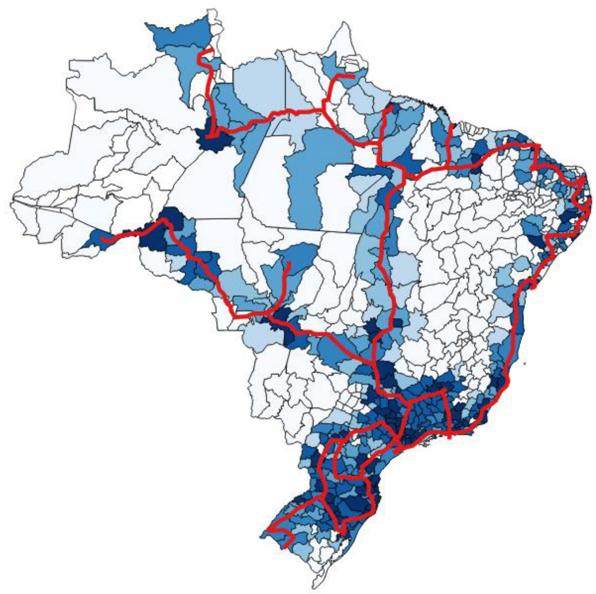
[59] Interestingly, as shown in Appendix D.4, the skill intensity within firms increased following the new broadband connection.

[60] While I do not use the quasi-experiment in the structural estimation directly, I rely on it to assess the reliability of the structural estimates. This is a commonly adopted approach in the literature, see, e.g., Todd and Wolpin (2006).

treatment effect on firms' division of labor, yielding a 3.9% productivity increase. I then feed the productivity shock to the simulated economy, assuming firms do not relocate spatially. Finally, I calculate the average city-level change in firms' division of labor based on the predicted distribution of firms and sectors within a city. In the model, cities populated by more high-z firms would undergo a higher average city-level increase in firms' division of labor due to the heterogeneity of the treatment across different complexities.



(a) Actual



(b) Predicted

Figure 4: Actual v.s. Simulated changes in firms' division of labor across cities

Even though I do not use post-program data in my estimation, 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 0.69.[61] Looking at Figure 4, 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.

In my judgment, the results above, together with the two non-targeted moments in Section 5.5, provide enough confidence in the model to use it to perform policy evaluations. In Appendix E, I illustrate how one could do so, by using the estimated model to evaluate the short-run and long-run impacts of the new broadband infrastructure on productivity and other outcomes in Brazil.

[61] The benchmark correlation is 0.28, which is obtained by assuming a uniform distribution of firms and sectors across all cities.

8 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. The structure of the model, combined with 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. Finally, through a quasi-experiment, I provide causal empirical evidence that supports a set of auxiliary theoretical predictions and validates the structural estimates.

This project is a step toward further unpacking the black box of agglomeration externalities. Identifying and quantitatively evaluating 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 efficiency 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

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. 2020. "On the road: Access to transportation infrastructure and economic growth in China." Journal of Development Economics, 145(C).

Banerji, Sourangsu, and Rahul Singha Chowdhury. 2013. "Wi-Fi & WiMAX: A Comparative Study." Indian journal of engineering, 2(5): 51–54.

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.

Becker, Sascha O., Hartmut Egger, Michael Koch, and Marc-Andreas Muendler. 2019. "Tasks, Occupations, and Wage Inequality in an Open Economy." Unviersity of California, San Diego, Working Paper.

Behrens, Kristian, Gilles Duranton, and Frèdèric 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.

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.

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. 2019. “A spatial knowledge economy.” American Economic Review, 109(1): 153–70.

Davis, Donald R, and Jonathan I Dingel. 2020. “The comparative advantage of cities.” Journal of International Economics, 123: 103291.

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

Demográfico, Censo. 2000. “documentação dos microdados da amostra.” Rio de Janeiro: IBGE.

Dix-Carneiro, Rafael, and Brian K Kovak. 2017. “Trade liberalization and regional dynamics.” American Economic Review, 107(10): 2908–46.

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-FranÁois Thisse. 2013. Economics of Agglomeration: Cities, Industrial Location, and Globalization. . 2 ed., Cambridge University Press.

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

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

Gaubert, Cecile. 2018. "Firm sorting and agglomeration." American Economic Review, 108(11): 3117–53.

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 J Redding. 2017. "Trade and inequality: From theory to estimation." The Review of Economic Studies, 84(1): 357–405.

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. 2019. "The Arrival of Fast Internet and Employment in Africa." American Economic Review, 109(3): 1032–79.

Hjort, Jonas, and Lin Tian. 2021. "The Economic Impact of Internet Connectivity in Developing Countries." INSEAD, Working Paper.

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

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

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

Kennedy, James, and Russell Eberhart. 1995. "Particle swarm optimization." Vol. 4, 1942–1948 vol.4, IEEE.

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, Flavio Feferman, and Nathalia Foditsch. 2016. Broadband In Brazil: past, present, and future. FGV Direito Rio.

Krugman, Paul. 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. 2009. Principles of economics: unabridged eighth edition. Cosimo, Inc.

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

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." Econometrica, 71(6): 1695–1725.

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.

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 . Vol. 4 of Handbook of Regional and Urban Economics, , ed. J. V. Henderson and J. F. Thisse, Chapter 49, 2119–2171. Elsevier.

Serafinelli, Michel. 2019. ""Good" Firms, Worker Flows, and Local Productivity." Journal of Labor Economics, 37(3): 747–792.

Smith, Adam. 1776. The Wealth of Nations. W. Strahan and T. Cadell, London.

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

Todd, Petra E., and Kenneth I. Wolpin. 2006. "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility." American Economic Review, 96(5): 1384–1417.

Topkis, Donald 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. University of Chicago Press.

A Data and stylized facts

A.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.[62] 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.[63] 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."[64] This leaves, in total, 1821 occupation codes in the dataset across all establishments.[65] 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.

[62] 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.

[63] 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.

[64] See Figure 5 for an illustration of the procedure.

[65] 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.

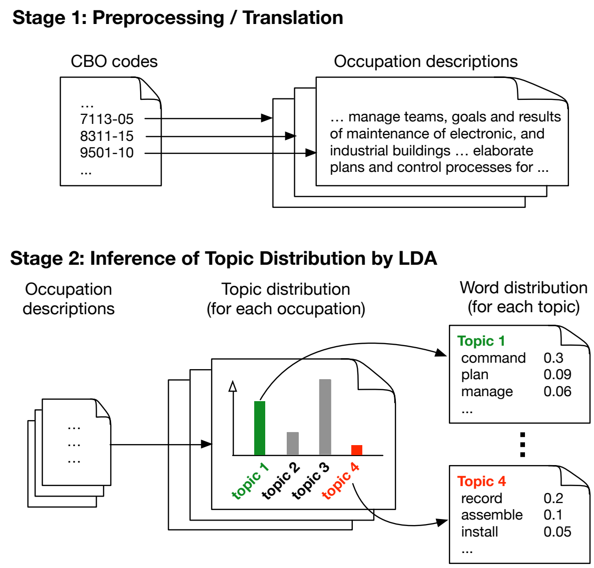


Figure 5: Removing managerial / supervisory occupations using the LDA technique

A.2 Additional results for stylized facts

First, 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 correlation between division of labor and city size, in which division of labor is measured at 4-digit occupation code level, instead of 6 digit. As shown in Tables 3 and 4, though lower in the values of the estimates, the positive correlations remain qualitatively consistent and quantitatively similar to the baseline results.

Dependent variable	Log no of occs				Specialization index			
	All		Mono-estb	Homog	All		Mono-estb	Homog
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (city size)	.055***	.024***	.0233***	.0281***	.0187***	.0137***	.0133***	.0144***
	(.0019)	(.001)	(.001)	(.0102)	(.0006)	(.0005)	(.0005)	(.0048)
Other controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Obs	2960066	2960066	2776735	6111	2960066	2960066	2776735	6111
R-sq	.119	.855	.852	.906	.081	.515	.518	.552

Standard errors clustered by city in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include sector and year FEs. Specialization index is defined in (3). Establishment-level controls are establishment size, skill intensity, and occupation categories (defined as the number of 3 digit occupation codes) within the establishment. City-level controls are state dummy, Herfindahl index of employment across sectors within the city, 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. Mono-estb firms refers to firms with a single establishments. Homogeneous sectors include corrugated and solid fiber boxes, bread, carbon black, roasted coffee beans, ready-mixed concrete, wooden flooring, gasoline, ice, plywood, and sugar (Foster, Haltiwanger and Syverson, 2008).

Table 3: Correlation of establishment's division of labor (measured at 4-digit level) and city size

I further divide establishments into deciles and study the correlation between firms' division of labor and

Dependent variable	Log no of occs				Specialization index			
	All		Mono-estb	Homog	All		Mono-estb	Homog
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (city size)	.0533***	.0236***	.0229***	.026**	.0192***	.0137***	.0131***	.0142***
	(.0018)	(.0009)	(.0009)	(.0102)	(.001)	(.0005)	(.0005)	(.0047)
Other controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Obs	2960045	2960045	2776714	6111	2960066	2960066	2776735	6111
R-sq	.132	.841	.837	.887	.111	.498	.499	.55

Standard errors clustered by city in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include sector and year FEs. Specialization index is defined in (3). Establishment-level controls are establishment size, skill intensity, and occupation categories (defined as the number of 3 digit occupation codes) within the establishment. City-level controls are state dummy, Herfindahl index of employment across sectors within the city, 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. Mono-estb firms refers to firms with a single establishments. Homogeneous sectors include corrugated and solid fiber boxes, bread, carbon black, roasted coffee beans, ready-mixed concrete, wooden flooring, gasoline, ice, plywood, and sugar (Foster, Haltiwanger and Syverson, 2008).

Table 4: Correlation of establishment's division of labor (measured at 3-digit level) and city size

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 5, the correlation remains positive for all deciles, suggesting that the result is unlikely driven by differences in informal employment across space.

Dependent variable							
Log (no of occupations)				Specialization index			
1st decile	.0011	6th decile	.0071***	1st decile	0	6th decile	.0047***
	(.0007)		(.0009)		(0)		(.0005)
2nd decile	.0022***	7th decile	.008***	2nd decile	.0012***	7th decile	.005***
	(.0005)		(.0011)		(.0003)		(.0006)
3rd decile	.0049***	8th decile	.0093***	3rd decile	.0033***	8th decile	.0055***
	(.0006)		(.0012)		(.0005)		(.0006)
4th decile	.006***	9th decile	.0096***	4th decile	.0043***	9th decile	.006***
	(.0007)		(.0012)		(.0005)		(.0007)
5th decile	.0056***	10th decile	.0177***	5th decile	.004***	10th decile	.011***
	(.0009)		(.0014)		(.0005)		(.0008)

Specialization index is defined in (3). All regressions include sector, state and year FEs. Establishment-level control is the skill intensity within the establishment. City-level controls are state dummy, Herfindahl index of employment across sectors within the city, 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.

Table 5: Correlation of the establishment's division of labor and city size, by decile

Next, Table 6 shows the results from multi-establishment firms, using the following specification:

$$\log N_{jt} = \alpha_0 + \alpha_1 \log L_{m(j)t} + \delta_{f(j)} + \delta_{s(j)} + \delta_t + \mathbf{X}_{jt} + \varepsilon_{jt},$$

where $\delta_{f(j)}$ is a firm fixed effect. Effectively, this specification allows me to study variation in extents of division of labor across establishments in different cities within the same firm. Again, the results are consistent with the baseline results in which there is strong positive correlation between division of labor and city size.

Dependent variable	Log no of occs		Specialization index	
	(1)	(2)	(3)	(4)
Log (city size)	.0148***	.0315***	.0094***	.0157***
	(.005)	(.0024)	(.0011)	(.001)
Other controls	No	Yes	No	Yes
Obs	172088	172088	172088	172088
R-sq	.562	.932	.461	.683

Standard errors clustered by city in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include firm, year and sector FEs. Specialization index is defined in (3). Establishment-level controls are establishment size, skill intensity, and occupation categories (defined as the number of 3 digit occupation codes) within the establishment. City-level controls are state dummy, Herfindahl index of employment across sectors within the city, 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.

Table 6: Correlation of division of labor and city size across establishments in multi-establishment firms

Dependent variable	Log no of occs		Specialization index	
	(1)	(2)	(3)	(4)
Log (city size)	.0488***	.0253***	.0171***	.014***
	(.0025)	(.0017)	(.001)	(.0007)
Log (city size) x Fragmentation Intensity	.0272	0	.0053	-.0008
	(.037)	(.002)	(.013)	(.0008)
Other controls	No	Yes	No	Yes
Obs	2960066	2960066	2960066	2960066
R-sq	.117	.86	.078	.526

Standard errors clustered by city in parentheses. Significance levels: * 10%, ** 5%, ***1%. All regressions include sector and year FEs. Specialization index is defined in (3). Establishment-level controls are establishment size, skill intensity, and occupation categories (defined as the number of 3 digit occupation codes) within the establishment. City-level controls are state dummy, Herfindahl index of employment across sectors within the city, 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. Fragmentation intensity is calculated using the share of fragmentation by 4-digit NAICS code from Fort (2017).

Table 7: Fragmentation intensity and correlation of division of labor and city size

Finally, I address concerns for variation in establishment boundaries through the following robustness check. Relying on the fragmentation index documented in Fort (2017), I categorize establishment into two groups based on their tendency to fragment their production process and/or outsource certain tasks to other firms (both domestically and internationally). I then run the following analysis:

$$\log N_{jt} = \alpha_0 + \alpha_1 \log L_{m(j)t} \mathbf{1}(Fragment) + \delta_{s(j)} + \delta_t + \mathbf{X}_{jt} + \varepsilon_{jt},$$

where $\mathbf{1}(Fragment)$ is dummy taking the value of 1 if the establishment falls in a 4-digit CNAE industry with a fragmentation index higher than the median measure of all industries. As shown in Table 7, establishments that are more likely to fragment do not have a significantly different correlation with city size relative to those that are less likely to fragment. This suggests that systematic variations in firm boundaries across cities are unlikely to drive the observed correlation.

B Theory Appendix

B.1 Microfounding urban costs

In this section, I provide a microfoundation for the reduced-form urban costs in (6) by modelling cities in a standard Alonso-Muth-Mills monocentric city framework.[66]

Production within each city takes place at a single point, called the central business district (CBD). Residential developments are aligned around the CBD, with length normalized to 1. Workers commute to the CBD at a cost, which are paid in worker's reference level of wages, $\bar{w}$.[67] The costly commute reflects opportunity cost of time. I further assume that the cost of a round trip from any location at a distance x from the CBD is:

$$c(x) = \tau x^{\eta}, \ \ \tau > 0 \text{ and } \eta \geq 0.$$

Within each city, workers choose their locations of residence to maximize utility given income and rent schedule. For simplicity, I assume that the rent is taxed and redistributed equally to all the residents within that city. Because each worker consumes a fixed amount of housing, worker's problem is effectively to choose a location x to minimize the combined cost of housing and commuting, i.e., $r(x)+c(x)$. Further, no arbitrage condition ensures that $r(x)+c(x)$ is the same for all residents in equilibrium.

With all these, the objective function can therefore be written as follows,

$$r(x) + c(x) = r(x) + \tau x^{\eta} = r\left(\frac{L}{2}\right) + \tau\left(\frac{L}{2}\right)^{\eta}, \tag{37}$$

where L is the total population in the city. Since the city is symmetric at CBD, the distance to the edge of the city is $L/2$. The second equality equates the combined urban cost for workers living at a distance x to the CBD to others living on the edge of the city.

Without loss, I normalize the rent at the edge of the city to 0. Substituting this to (37), I obtain the rent schedule within city L,

$$r(x) = \tau\left(\frac{L}{2}\right)^{\eta} - \tau x^{\eta}.$$

Integrating the rent schedule over the entire city yields total land rent:

$$R(L) = \int_0^L r(x)dx = 2\int_0^{L/2} r(x)\,dx = \frac{2\tau\eta}{\eta+1}\left(\frac{L}{2}\right)^{\eta+1},$$

[66]See Behrens, Duranton and Robert-Nicoud (2014) for a more recent discussion.

[67]Recall that $\bar{w}$ is treated as the numeraire in the model.

where the second equality takes advantage of the fact that rent is symmetric around the CBD.

For a worker living at distance x from the CBD, urban costs are the sum of her rent and commuting costs minus her share of the $R(L)$. The amenity values of living in a city of size L, defined as the inverse of the urban costs is therefore:

$$u(L) \equiv [c(x) + r(x) - \frac{R(L)}{L}]^{-1} = 2^{\eta}\tau^{-1}(\eta+1)L^{-\eta}. \tag{38}$$

Setting $\kappa \equiv 2^{\eta}\tau^{-1}(\eta+1)$, I obtain (6) in Section 4.

B.2 Definition of the spatial equilibrium

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 (8) as the numeraire. An equilibrium for a population $\bar{L}$ and firm with product distribution $f(z)$ in a set of locations $\mathcal{L}$ is characterized by a set of prices $\{w(L)\}$; a city-size distribution $f_L(\cdot)$; an optimal division of labor function $N(z)$; a location matching function $L(z)$; an employment function $l(z)$; a production function $Q(z)$; and a set of price index P and mass of firms M such that:

1. Workers maximize their utilities according to (4), given $w(L), p_H(L)$ and P.

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

3. Firms maximize profits according to (15), given $w(L)$ and P.

4. Aggregate production must be equal to the sum of individual firms' production:

$$1 = \frac{(\sigma-1)^{\sigma-1}}{\sigma^{\sigma}} MP^{\sigma-1} \int_z \left(\frac{\psi(N, L; z)}{[(1-\eta)L(z)]^{\frac{1-\eta}{\eta}}} \right)^{\sigma-1} dF(z). \tag{39}$$

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

$$f_E P = \frac{(\sigma-1)^{\sigma-1}}{\sigma^{\sigma}} RP^{\sigma-1} \int_z \left(\frac{\psi(N, L; z)}{[(1-\eta)L(z)]^{\frac{1-\eta}{\eta}}} \right)^{\sigma-1} dF(z). \tag{40}$$

6. The national labor market clears:

$$\bar{L} = \frac{(\sigma-1)^{\sigma}}{\sigma^{\sigma}} MRP^{\sigma-1} \int_z \frac{[\psi(N, L; z)]^{\sigma-1}}{[(1-\eta)L(z)]^{\frac{1-\eta}{\eta}\sigma}} dF(z). \tag{41}$$

7. The local labor markets clear:

$$\int_{L_0}^{L} n f_L(n) dn = M \int_0^{\infty} \mathbf{1}(L, z) l(z) dF(z) \quad \forall L > L_0, \tag{42}$$

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

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

B.3 Microfoundation for the complementarity between N and z

In Section 2, I follow Costinot (2009) and develop a stylized theory of firm production to guide the empirical measure for division of labor. In this section, I show that the model also provides a microfounded production function that generates log-supermodularities between division of labor N and firms' complexity draw z.

From Lemma 1, we know that the optimal contract assigns $\frac{1}{N}$ tasks to each worker, with N denoting the number of distinct occupations within the firm. Given that each task requires z units of training cost, the total training cost per worker is therefore $\frac{z}{N}$.. This further implies that each worker has $1 - \frac{z}{N}$ units of time available for production. The worker productivity is therefore given by,

$$\psi(N, z) \equiv 1 - \frac{z}{N}. \tag{43}$$

The result in (43) 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 cost of division of labor, efficiency requires that each skill be used as intensively as possible. Furthermore, it is straightforward to see that using this production process, $\psi(N, z)$ meets the condition specified in Assumption 1—$\psi(N, z)$ is log-supermodular in (z, N).

B.4 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.4.1 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. $R(L) = \kappa_1 \left(\frac{L}{2}\right)^{\eta+1}$, where κ_1 is a positive constant. Due to competition and free entry, land developers that invest less than $R(L)$ 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

$$\mathcal{I}(L) \equiv R(L) = \kappa_1 \left(\frac{L}{2}\right)^{\eta+1}, \tag{44}$$

where $\mathcal{I}(\cdot)$ denotes the optimal level of investment. Using (44), it can be readily seen that $\mathcal{I}(\cdot)$ is an increasing function of the city size, L.

Note that the result in (44) 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 (44) is the socially optimal level.[68]

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. In Section 7, I provide causal empirical evidence to support this assumption. Finally, 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.4.2 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 (2009), 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.

Recall that in the stylized model in Section 2, I assume that all tasks in $[0, 1]$ needs to be completed within a firm to produce any good. In this part, I further assume that 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 $K = \frac{1}{N}$.[69] 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

[68] 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.

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

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 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.[70]

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.[71] 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,[72]

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

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

$$Q = A(N, z)bl, \tag{45}$$

[70] 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.

[71] 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 (2020) for theoretical discussion and De la Roca and Puga (2017) for empirical evidence on this assumption.

[72] 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.

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

The firm's problem is therefore

$$\max_{N,b,L} \pi(z, L, N, b) = \max_{N,b,L} \kappa_2 \left(\frac{A(N, z)b}{(1 + \gamma(N, b, L))w(L)} \right)^{\sigma-1} RP^{\sigma-1}, \tag{46}$$

where κ_2 is a positive constant.

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 $\psi(\cdot)$ in (9) as:

$$\psi(N, L; z) \equiv \frac{A(N, z)b}{1 + \gamma(N, b, L)}.$$

When b is optimally chosen, $\psi(N, L; z)$ displays log-supermodularity in (N, L).

B.5 Proofs

This section presents the proofs to the propositions and lemmas discussed in the main text.

Lemma 1

Proof. I will prove the first part of the lemma by contradiction. Suppose $\mathcal{C}^*$ assigns a positive mass of workers $m \in l$ to multiple occupations.

Since the total number of occupations N is finite, there exists $l^{'} \subset l$ with positive mass such that all $m \in l^{'}$ perform the same set of occupations $\mathcal{O}' = \{\mathcal{O}^*_{k_h}\}_{h=1}^{H}$, where H is the total number of occupations for which these m workers perform.

Let l_{k_h} denote total amount of labor that workers $m \in l'$ allocate to performing $\mathcal{J}^*_{k_h} \in \mathcal{O}'$ under $\mathcal{C}^*$. Adding the workers' time constraints, we get:

$$l_{k_h} \leq \int_{m \in l'} \alpha(m, k_h) \left[1 - zc'\right] du \tag{47}$$

where $\alpha(m, k_h)$ is the share of after-training labor that worker m allocates to $\mathcal{O}^*_{k_h}$; and c' is the total number of tasks in $\mathcal{O}'$. By assumption, $c' < c^*_{k_h}$, the number of tasks in $\mathcal{O}^*_{k_h}$.

Now consider a contract $\tilde{\mathcal{C}}$ that reallocates workers $m \in l'$ across jobs according to the following rule:

1. For all $\mathcal{O}^*_{k_h} \in \mathcal{O}'$, $\tilde{\mathcal{C}}$ assigns a mass $\int_{m \in l'} \alpha(m, l_h) du - \epsilon$ of workers exclusively to $\mathcal{O}^*_{k_h}$.

2. The remaining $H\epsilon > 0$ of workers are not hired by the firm.

Each group of specialists can now allocate $\tilde{l}_{k_h}$ units of labor to $\mathcal{O}^*_{k_h}$, where $\tilde{l}_{k_h}$ is given by:

$$\tilde{l}_{k_h} = \int_{m \in H} \alpha(m, k_h)(1 - zc^*_k) du - \epsilon(1 - zc^*_k) \tag{48}$$

For ϵ small enough, (47), (48) and $c' > c^*_{k_h}$ imply $\tilde{l}_{k_h} > l_{k_h}$. As a result, any unit that can be produced under $\mathcal{C}^*$ can also be produced under $\tilde{\mathcal{C}}$, but that the total wage bill is smaller by $wH\epsilon > 0$, where w is the equilibrium wage. A contradiction that $\mathcal{C}^*$ being a solution of the profit maximization problem.

To prove the second part of the lemma, consider 2 occupations, $O_1 \in \mathcal{O}^*$ and $O_2 \in \mathcal{O}^*$. Assume that c_1 and c_2 are the number of tasks associated with O_1 and O_2, and M_1 and M_2 are mass of workers assigned to these two jobs. Total amount of labor available for performing each job is $M_1(1-zc_1)$ and $M_2(1-zc_2)$. Note that if the firm maximizes its profits, then all occupations must be performed on the same number of units Q. Otherwise, the firm could decrease the mass of workers performing one job, without decreasing output. This implies

$$M_1 \left(\frac{1 - zc_1}{c_1} \right) = M_2 \left(\frac{1 - zc_2}{c_2} \right) = Q$$

Now consider the following minimization problem:

$$\min_{c_1, c_2, M_1, M_2} M_1 + M_2$$

subject to

$$M_1(1 - zc_1) = c_1 Q$$

$$M_2(1 - zc_2) = c_2 Q$$

$$c_1 + c_2 = \bar{c}$$

where $\bar{c}$ is the total number of tasks in $O_1 \cup O_2$. After plugging the first 2 constraints into the objective function, wet

$$\min_{c_1, c_2} Q \cdot \left(\frac{c_1}{1 - zc_1} + \frac{c_2}{1 - zc_2} \right)$$

subject to

$$c_1 + c_2 = \bar{c}$$

The two necessary first-order conditions are given by:

$$\frac{Q}{(1 - zc_1)^2} = \theta$$

and

$$\frac{Q}{(1-zc_2)^2} = \theta$$

where θ is the Langrangian multiplier associated with $c_1 + c_2 = \bar{c}$. Combining these two FOCs with the last constraint, we get

$$c_1 = c_2 = \frac{\bar{c}}{2}.$$

This implies that, holding Q constant, the mass of workers necessary perform the $\bar{c}$ tasks in $O_1 \cup O_2$ is minimized when O_1 and O_2 include the same number of tasks.

In order to conclude the proof, we note that if profits are maximized, then holding the level of the output Q constant, the mass of workers must be minimized. So, if $\mathcal{O}^*$ is a solution of the profit maximization problem, then for any pair of occupations $\{O_1, O_2\} \subset \mathcal{O}^*$, O_1 and O_2 must include the same number of tasks. Since the total number of tasks that must be performed is c, we obtain,

$$\int_{t \in \mathcal{O}_k^*} dt = \frac{zc}{N}.$$

■

Lemma 2

Proof. Taking log of (13),

$$\log \pi(N, L; z) = constant + (\sigma - 1) \left[\log \psi(N, L; z) - \log w(L)\right]$$

Taking partial derivatives with respect to its arguments, I get

$$\frac{\partial \log \pi}{\partial z} = (\sigma - 1) \frac{\partial \log \psi}{\partial z};$$

$$\frac{\partial \log \pi}{\partial L} = (\sigma - 1) \left[\frac{\partial \log \psi}{\partial L} - \frac{\partial \log w(L)}{\partial L}\right];$$

$$\frac{\partial \log \pi}{\partial N} = (\sigma - 1) \frac{\partial \log \psi}{\partial N}.$$

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

$$\frac{\partial^2 \log \pi}{\partial z \partial L} \geq 0;$$

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

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

The last two inequalities come from Assumption 1. ■

Lemma 3

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

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

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

■

Lemma 4

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 $\pi(N, L; z)$ is log-supermodular in (z, L, N), then $\log \pi(L; z) \equiv \max_N \log \pi(N, L; z)$ is supermodular in (z, L). ■

Proposition 5

Proof. By Lemma 4, $\log \pi(z, L)$ is supermodular in (z, L).

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

$$\frac{\pi(L_1; z_1)}{\pi(L_2; z_1)} > \frac{\pi(L_1; z_2)}{\pi(L_2; z_2)}.$$

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

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

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

$$\frac{dL^*(z)}{dz} = -\frac{(\sigma - 1)\frac{\partial^2 \log \psi}{\partial z \partial N}\frac{\partial N}{\partial L}}{\frac{\partial^2 \log \pi}{\partial z^2}} > 0.$$

■

Proposition 6

Proof. $L(z)$ is strictly increasing in z (from Proposition 5). Therefore, if $L(z) = L(z')$, then we know $z > z'$. For simplicity, I denote $L(z)$ and $L(z')$ by L and L', respectively.

From Lemma 3, $N(z, L)$ is increasing in z and L. I get $N(z, L') > N(z', L')$. And since $L > L'$, I get $N(z) > N(z')$. For simplicity, I denote $N(z)$ and $N(z')$ by N and N', respectively.

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

$$\begin{aligned} \psi(N'; z) &> \psi(N'; z') \\ \implies \pi(N', L'; z) &> \pi(N', L'; z') \end{aligned}$$

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

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

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

Proposition 8

Proposition 8 *Consider an exogenous shock, $\partial \mathcal{I}$, to firm's division of labor, if $c > 0$ and $\theta > 0$, then:*

1. *Within a city, the increase in division of labor is higher for more complex firms, i.e.,*

$$\frac{\partial^2 \log N}{\partial \log z \, \partial \mathcal{I}} > 0.$$

2. *Across cities, the increase in division of labor is higher for firms in larger cities, i.e.,*

$$\frac{\partial^2 \log N}{\partial \log L \, \partial \mathcal{I}} > 0.$$

Proof. From (33), change in $\log N$ caused by an infinitesimal exogenous shock $\partial \mathcal{I}$ can be written as:

$$\frac{\partial \log N}{\partial \mathcal{I}} = c^{\frac{1}{1-c}} (\log z)^{\frac{1}{1-c}} (\log L)^{\frac{\theta}{1-c}} - 1 \tag{49}$$

Consider the cross partial derivative of (49) with respect to z and L, respectively:

$$\frac{\partial^2 \log N}{\partial \mathcal{I} \partial \log z} = \frac{1}{1-c} c^{\frac{1}{1-c}} (\log z)^{\frac{c}{1-c}} (\log L)^{\frac{\theta}{1-c}},$$

and

$$\frac{\partial^2 \log N}{\partial \mathcal{I} \partial \log L} = \frac{\theta}{1-c} c^{\frac{1}{1-c}} (\log z)^{\frac{1}{1-c}} (\log L)^{\frac{\theta-1+c}{1-c}}.$$

It can be readily seen that if $c > 0$ and $\theta > 0$, then $\frac{\partial^2 \log N}{\partial \mathcal{I} \partial \log z} > 0$ and $\frac{\partial^2 \log N}{\partial \mathcal{I} \partial \log L} > 0$.

The proof above shows that under the assumption in Section 5, there is heterogeneity in the impact of a shock to division of labor: 1) the increase in division of labor is higher for more complex firms and 2) the increase is also higher for firms in larger cities. ■

B.6 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 (2018) and uses results from Allen and Arkolakis (2014).

Following the assumption in the base model, the economy consists of a continuum of locations $m \in \mathcal{M}$, where $\mathcal{M}$ is a compact subset of $\mathbf{R}^{\mathcal{M}}$. Trade is costly: trade costs are of the iceberg form and are described by the function $T : \mathcal{M} \times \mathcal{M} \to [1, \infty)$, where τ_{mn} is the quantity of a good needed to be shipped from city m in order for a unit of a good to arrive in city n. 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 m is given by:

$$P_m = \left[\int_n \int_{z \in \mathcal{Z}(n)} \left(\frac{\tau_{nm} w_n}{\psi(L_n; z)} \right)^{1-\sigma} dF_n(z) dn \right]^{\frac{1}{1-\sigma}},$$

where $\mathcal{Z}(n)$ is the set of firms located in city m in equilibrium and $F_n(z)$ is the endogenous distribution of firms in city n.

We can define an average city-level productivity term:

$$\bar{\psi}_n = \left[\int_{z \in \mathcal{Z}(n)} \psi(L_n; z)^{\sigma - 1} dF_n(z) \right]^{\frac{1}{\sigma-1}}. \tag{50}$$

Using (50), price index can be re-written more compactly as:

$$P_m = \left[\int_n \left(\frac{\tau_{nm} w_n}{\bar{\psi}_n} \right)^{1-\sigma} dn \right]^{\frac{1}{1-\sigma}}. \tag{51}$$

Given CES preference, demand for good z produced in n from city m is $q_{mn}(z) = p_{mn}(z)^{-\sigma} w_n L_n P_n^{\sigma-1}$, whereas marginal cost for the good is $\frac{\tau_{mn} w_m}{\psi(L_m; z)}$. Combining, we get profit for firm z located in m,

$$\pi(z, m) = \frac{(\sigma - 1)^{\sigma - 1}}{\sigma^\sigma} \int_n \left(\frac{\tau_{mn} w_m}{\psi(L_m; z)} \right)^{1-\sigma} w_n L_n P_n^{\sigma-1} dn. \tag{52}$$

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

$$MA_m = \int_n \tau_{mn}^{1-\sigma} w_n L_n P_n^{\sigma-1} dn.$$

Substituting into (52), we get,

$$\pi(z,m) = \frac{(\sigma-1)^{\sigma-1}}{\sigma^{\sigma}} w_m^{1-\sigma} \psi(L_m; z)^{\sigma-1} MA_m. \tag{53}$$

It is straightforward to see that (53) displays log-supermodularity in (L_m, z). 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_m is the sufficient statistic for city m, 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, $U_m = \left[\frac{w_m}{P_m}\right] \kappa L_m^{-\eta}$, must be equalized across all cities. Using (51) and after some algebra, we can re-write the utility function as:

$$w_m^{1-\sigma} L_m^{(1-\sigma)\eta} = \tilde{U} \int_n \left(\frac{\tau_{nm} w_n}{\bar{\psi}_n}\right)^{1-\sigma} dn, \tag{54}$$

where $\tilde{U} = U^{1-\sigma}\kappa^{\sigma-1}$, where U is the utility level in equilibrium.

Next, using local goods market clearing condition:

$$w_m L_m = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \int_n \left(\frac{\tau_{mn} w_m}{\bar{\psi}_m}\right)^{1-\sigma} w_n L_n P_n^{\sigma-1} dn.$$

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

$$w_m^{\sigma} L_m \bar{\psi}_m^{1-\sigma} = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \int_n \tau_{mn}^{1-\sigma} w_n L_n P_n^{\sigma-1} dn = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} MA_m \tag{55}$$

Further re-arranging, we have,

$$w_m^{\sigma} L_m \bar{\psi}_m^{1-\sigma} = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \bar{U} \int_n \tau_{mn}^{1-\sigma} w_n^{\sigma} L_n^{1-(\sigma-1)\eta} dn \tag{56}$$

The $2M$ equations of (54) and (56) fit directly into the systems of equations in Allen and Arkolakis (2014), where the local congestion force is given by $L_m^{-\eta}$ and local productivity $\bar{\psi}_m$. Under further assumption that the trade cost is symmetric, i.e., $\tau_{mn} = \tau_{nm}$, we can apply their results in Theorem 2, i.e. there exists unique vectors L_m and w_m in spatial equilibrium. Furthermore, we have

$$w_m^{\sigma} L_m \bar{\psi}_m^{1-\sigma} = \gamma w_m^{1-\sigma} L_m^{(1-\sigma)\eta},$$

where γ is an endogenous constant in equilibrium.

Finally, from (55), we get $MA_m = \gamma w_m^{1-\sigma} L_m^{(1-\sigma)\eta}$. We can therefore re-write firm's profit function as

$$\pi(z,m) = \frac{(\sigma-1)^{\sigma-1}}{\sigma^\sigma}\gamma\left[\frac{\psi(L_m;z)L_m^{-\eta}}{w_m^2}\right]^{\sigma-1} \tag{57}$$

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.

B.7 Instability of a homogeneous equilibrium

Proposition 9 *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 $\pi(L;z)$ 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

$$\begin{aligned}
&\log\pi(L_2;z(\epsilon)) - \log\pi(L_1,z(\epsilon)) \leq 0\\
\implies &\log\psi(N(z(\epsilon),L_2);z(\epsilon)) - \log\psi(N(z(\epsilon),L_1);z(\epsilon))\\
&\leq \eta L_1 - \eta L_2
\end{aligned}$$

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

B.8 Properties of the heterogeneous equilibrium

In heterogeneous equilibria, (18) 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)

$$\int_{L_0}^{L} n f_L(n) dn = M \int_{z(L_0)}^{z(L)} l(z) dF(z). \tag{58}$$

I can then obtain the city-size distribution $f_L(\cdot)$ by differentiating (58) with respect to city size L and dividing by L on both sizes,

$$f_L(L) = \frac{1}{L}\left[M\,\mathbf{1}(L)\,l(z(L))\,f(z(L))\,\frac{dz(L)}{dL}\right], \tag{59}$$

where $\mathbf{1}(L)$ is an indicator function, taking the value of 1 if firm is in city L and 0 otherwise. Equation (59) 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 (59), is unique. I get the following result:

Proposition 10 *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:

$$l = \frac{(\sigma-1)^{\sigma}}{\sigma^{\sigma}} \frac{(\psi(N(z), L; z))^{\sigma-1}}{w(L)^{\sigma}} RP^{\sigma-1}.$$

From the local labor market condition,

$$M\frac{(\sigma-1)^{\sigma}}{\sigma^{\sigma}} \frac{(\psi(N(z), L; z))^{\sigma-1}}{w(L)^{\sigma}} RP^{\sigma-1} = L,$$

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

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

The equilibrium is stable if worker utility decreases if a small mass of individuals move into 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.

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

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

$$M\frac{(\sigma-1)^{\sigma}}{\sigma^{\sigma}} \frac{RP^{\sigma-1}}{w(L)^{\sigma}} \left[(\sigma-1)\frac{\partial \psi}{\partial L} - \sigma \frac{w'(L)}{w(L)}\right] = 1. \tag{60}$$

From Equation (17), and the assumption that $\frac{w'(L)}{w(L)}L > \eta$, I get,

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

A contradiction to Equation (60). I get the following result:

Proposition 11 *The heterogeneous equilibrium distribution of city size $f_L(\cdot)$ is stable.*

B.9 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 and the price index P. To solve for the 3 variables, I need 3 equations, as specified below.

Using free entry condition, I get

$$f_E P = \frac{(\sigma-1)^{\sigma-1}}{\sigma^\sigma}\xi R P^{\sigma-1}\int_z \left(\frac{\psi(N,L;z)}{\kappa^{-1}L(z)^\eta}\right)^{\sigma-1} dF(z). \tag{61}$$

Next, individual firms' production must sum up to aggregate production,

$$1 = \frac{(\sigma-1)^{\sigma-1}}{\sigma^{\sigma-1}} M P^{\sigma-1}\int_z \left(\frac{\psi(N,L;z)}{\kappa^{-1}L(z)^\eta}\right)^{\sigma-1} dF(z). \tag{62}$$

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

$$\bar{L} = \frac{(\sigma-1)^{\sigma}}{\sigma^{\sigma-1}} M\xi R P^{\sigma-1}\int_z \frac{\psi(N,L;z)}{\kappa^{-1}L(z)^\eta} dF(z) \tag{63}$$

Using Equations (62) and (63), I can solve for the aggregate revenue:

$$\frac{\sigma-1}{\sigma}\xi\frac{\int_z \frac{\psi(N,L;z)^{\sigma-1}}{\kappa^{-1}L(z)^\eta} dF(z)}{\int_z \left(\frac{\psi(N,L;z)}{\kappa^{-1}L(z)^\eta}\right)^{\sigma-1} dF(z)} = \frac{\bar{L}}{R}. \tag{64}$$

Combining Equations (61) and (62), I get sectoral mass of firms:

$$M = \frac{\xi R}{\sigma f_E P} \tag{65}$$

Lastly, using Equations (61), I get the sectoral price indexes:

$$P^{\sigma-1} = \frac{f_E}{\frac{(\sigma-1)^{\sigma-1}}{\sigma^\sigma}\xi R \int_z \left(\frac{\psi(N,L;z)}{\kappa^{-1}L(z)^\eta}\right)^{\sigma-1} dF(z)} \tag{66}$$

B.10 Descriptive evidence

In this part, I present descriptive evidence that is consistent with the theoretical results in Section 4. The model predicts that, more complex firms sort into larger cities. This sorting of complexity generates sorting of other firm-level variables, including profits and revenue (Proposition 6). I first investigate how, average firms' division of labor and labor payment change as city size increases.[73] The model features a single sector economy, in which the elasticities of firms' division of labor and firm revenue to city size are both positive. Empirically, I first divide firms into different sectors to control for cross-sector heterogeneities. I then 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.[74] Figure 6 plots the distribution of the two elasticities. For division of labor, it is positive for 93% of the observations, and significantly negative for only two sectors, *sawmill* and *ferroalloy production*.[75] For labor payment, it is positive for 95% of the observations, and none of the negative estimates is significant. Results are therefore largely consistent with model predictions.

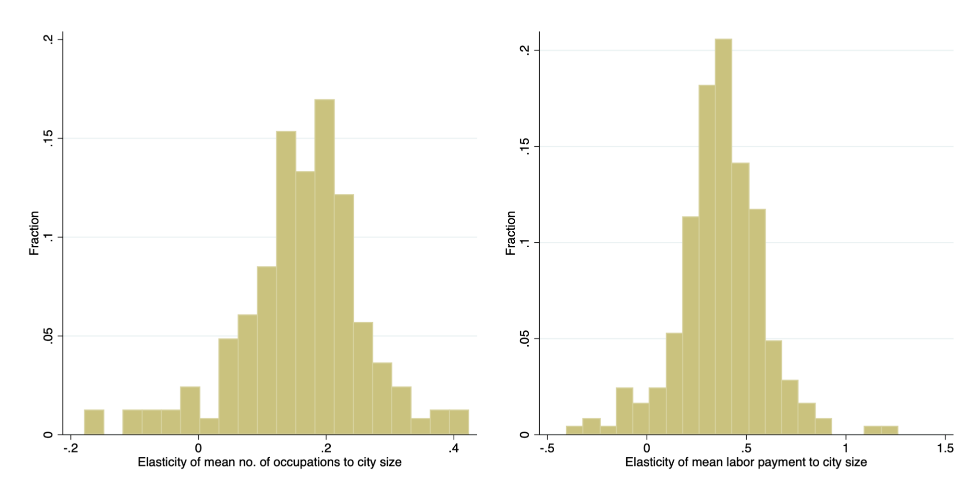


Figure 6: Elasticity of mean division of labor and labor payment to city size

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

C Quantitative Appendix

This section provides further details to the quantitative exercise in the paper. The summary statistics of the data used in the quantitative analysis are shown in Table 8. I first describe the estimation procedure,

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

[74] To have a meaningful number of establishments within each sector-city cell, I use the 4-digit CNAE2.0 code, which gives me 248 sectors.

[75] The results are not surprising since my model assumes that all locations are identical, whereas the location choices of these this sector are driven by natural amenities, such as availability of forests and iron ores.

then explains the identification strategy and the corresponding moments, and finally presents the estimation results by presenting the goodness of fit to both the targeted and non-targeted moments.

C.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 (2018) and make firms choose optimal production location from 400 bins of normalized city sizes.[76] 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 (14) to calculate the optimal division of labor N^*.

4. For each city size, I plug N^* into Equation (25), 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 (29).

6. I then compute the 6 sets of moments described in Section 5.4.

7. I repeat Steps 1-6 to find parameters that minimize the objective function in Equation (30), using the particle swarm optimization (PSO) method (Kennedy and Eberhart, 1995).

[76] 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	
Manufacture and products of wood, except furniture	10.45	9.55	11.19	1.98	1.1	2.71	1.55	1.1	1.95	11966
Manufacture of automotive vehicles	11.29	10.01	12.34	2.47	1.39	3.33	1.93	1.39	2.48	4507
Manufacture of basic metals	11.27	10.12	12.25	2.4	1.39	3.26	1.92	1.39	2.48	3687
Manufacture of chemicals and chemical products	11.26	9.97	12.32	2.33	1.39	3.22	1.91	1.39	2.48	7729
Manufacture of computer and electronic products	11.36	10.13	12.45	2.39	1.39	3.33	1.93	1.39	2.48	2941
Manufacture of electrical machines	11.34	10.14	12.39	2.46	1.39	3.37	1.94	1.39	2.48	3655
Manufacture of fabricated metal products, except machinery	10.66	9.59	11.53	1.94	1.1	2.64	1.6	1.1	2.08	26025
Manufacture of food products, beverages and tobacco products	10.57	9.48	11.32	2.13	1.1	2.83	1.63	1.1	2.08	30770
Manufacture of furniture	10.47	9.49	11.24	1.93	1.1	2.64	1.53	1.1	1.95	13310
Manufacture of glass, ceramic, brick and cement products	10.61	9.69	11.36	2.17	1.39	2.83	1.57	1.1	1.95	19184
Manufacture of leather goods and footwear, leather tanning	10.8	9.68	11.73	2.41	1.39	3.3	1.66	1.1	2.08	10132
Manufacture of mischellaneous products, other mfg activities	10.54	9.48	11.39	1.98	1.1	2.71	1.58	1.1	1.95	6823
Manufacture of other equipments and machines	11.27	10.16	12.24	2.26	1.39	3.04	1.94	1.39	2.48	11452
Manufacture of other transport equipment	11.39	10.09	12.46	2.54	1.61	3.47	2.01	1.39	2.56	833
Manufacture of pharmaceutical products	11.88	10.41	13.3	2.73	1.61	3.77	2.14	1.39	2.83	748
Manufacture of pulp, paper and paper products	11.17	10.02	12.11	2.45	1.39	3.33	1.88	1.39	2.4	3610
Manufacture of rubber and plastic products	11.16	10.06	12.09	2.44	1.61	3.26	1.83	1.1	2.3	12172
Manufacture of textiles	10.78	9.63	11.68	2.24	1.39	3.04	1.68	1.1	2.2	7935
Manufacture of wearing apparel	10.36	9.45	11.08	2.02	1.1	2.71	1.51	1.1	1.95	38981
Publishing, printing and reproduction of recorded media	10.31	9.39	11.02	1.61	.69	2.2	1.5	1.1	1.95	9622

Table 8: Summary statistics across sectors

C.2 Moments and identification

Geographic distribution of firms

The first 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 second 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 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.

C.3 Estimation results

The parameter estimates are shown in Table 9.

Sector	$\hat{\alpha}$	$\hat{\upsilon}$	$\hat{c}$	$\hat{\theta}$	$\hat{\nu_z}$	$\hat{\nu_L}$
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 9: Estimated parameters

C.3.1 Targeted Moments

Model fit for the set of targeted moments are shown in Figures 7 to 11.

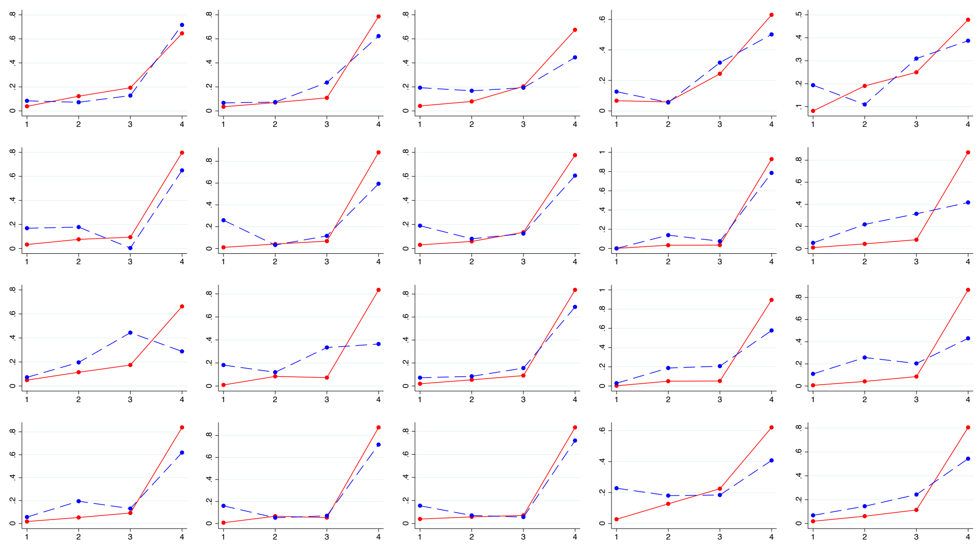


Figure 7: Distribution of employment across cities
(Actual moments: solid red line; simulated moments: dashed blue line)

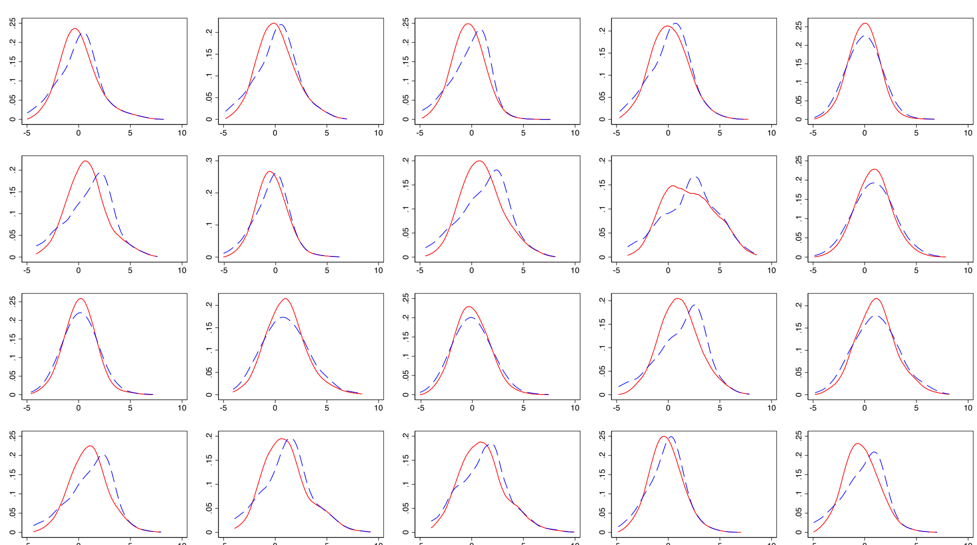


Figure 8: Distribution of firm labor payment
(Actual moments: solid red line; simulated moments: dashed blue line)

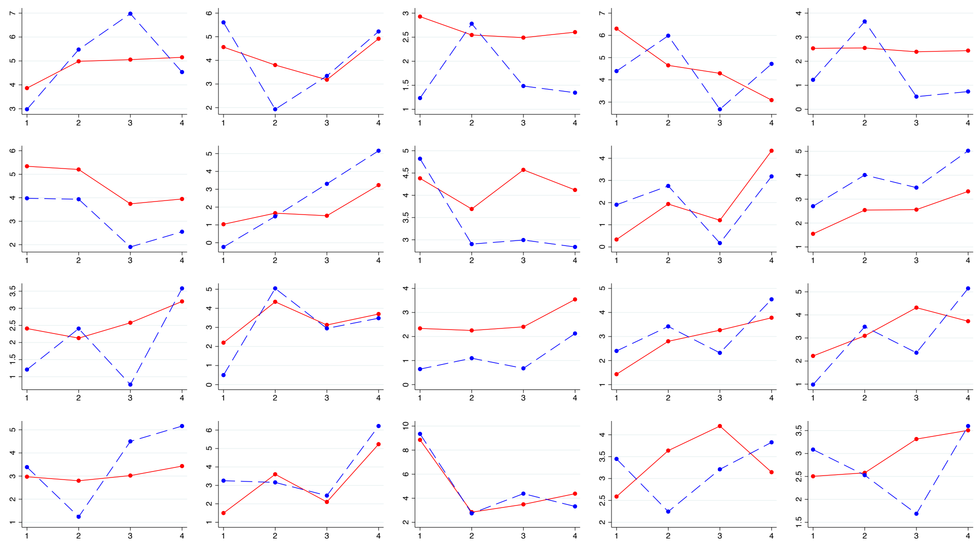


Figure 9: Average labor payment by city size
(Actual moments: solid red line; simulated moments: dashed blue line)

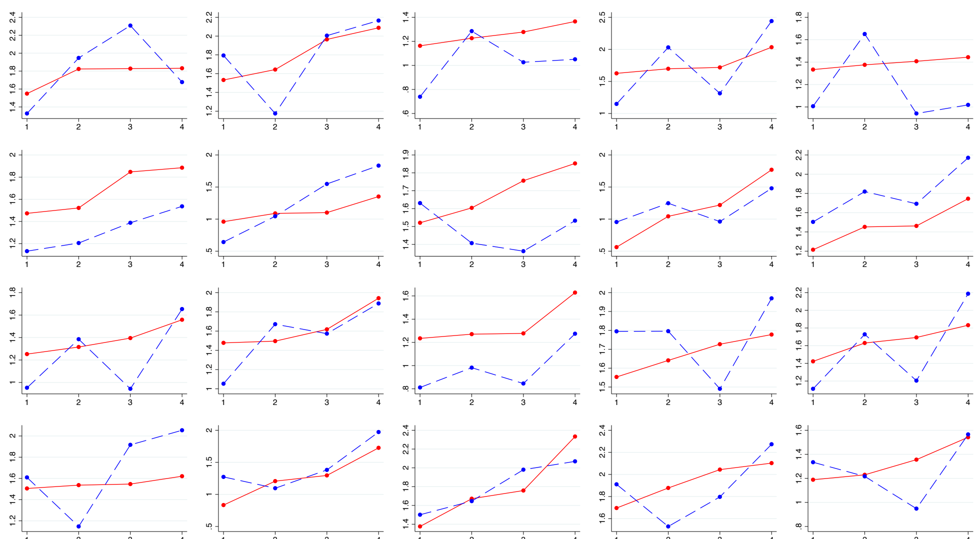


Figure 10: Average division of labor by city size
(Actual moments: solid red line; simulated moments: dashed blue line)

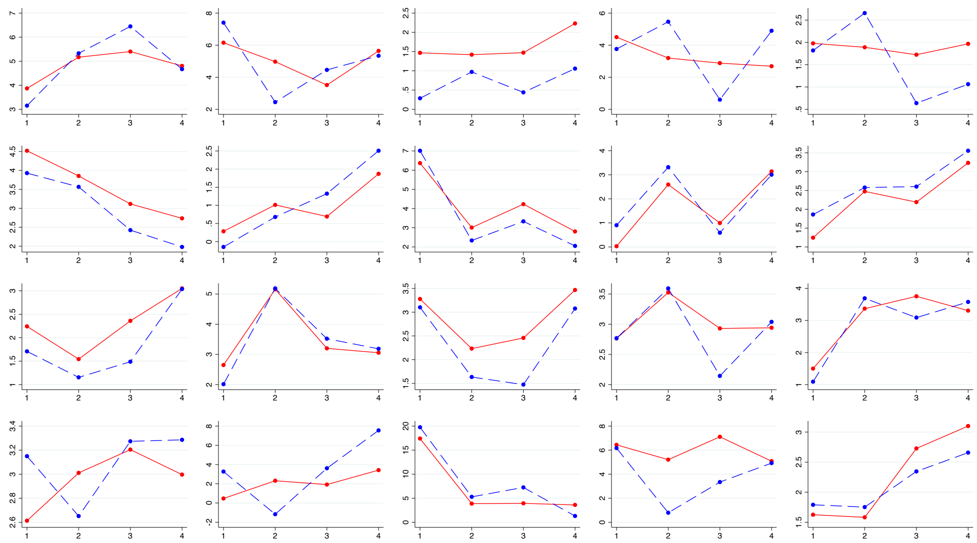


Figure 11: Variance of division of labor within city bins

C.4 Non-targetted Moments

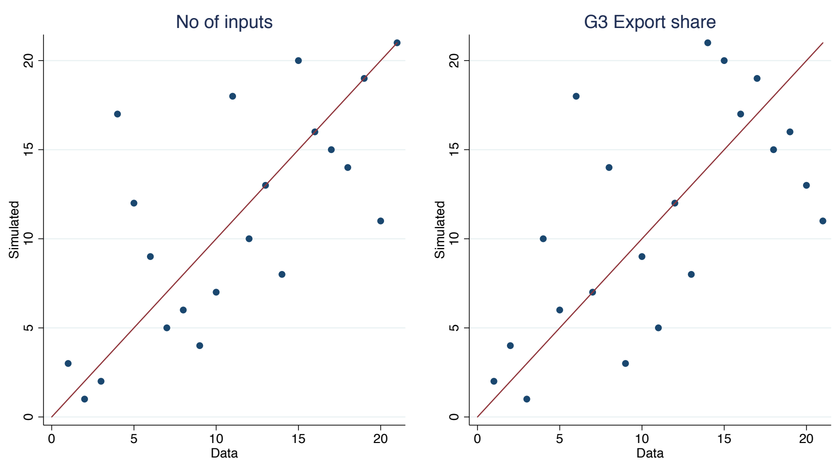


Figure 12: Rank correlations of complexity measures

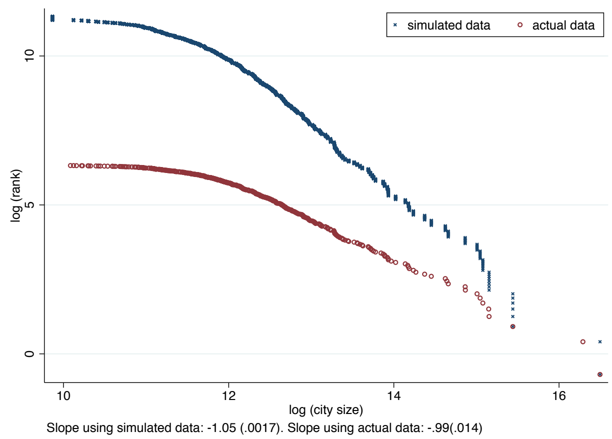


Figure 13: City size distribution

D Empirics Appendix

I provide further details to the quasi-experiment in this part. I start by providing more background information to the policy experiment and describing technological details that I leverage in my identification strategy. Next, I provide results for the robustness tests. Finally, I present additional results form the quasi-experiment that hints at the potential mechanisms at work to facilitate greater worker specialization following the ICT improvement. Table 10 compares characteristics of establishments in the treated and control areas before the quasi-experiment.

	Treatment	Control	All
Log (no of occupations)	1.54 (.98)	1.28 (.978)	1.47 (.982)
Specialization index	.6 (.263)	.557 (.275)	.588 (.267)
Share of managers	.347 (.398)	.393 (.393)	.334 (.397)
Estb size	40.29 (123)	37.03 (112)	39.36 (120)
Log (city size)	5.52 (1.76)	4.52 (1.05)	5.18 (1.79)
Total	.71	.29	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 10: Establishment-level characteristics before the treatment

D.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.[77] 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 infrastructure for example roads or power grid lines. The expansion of the backbone infrastructure was given a budget of $720mil USD.[78]

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 14.

D.2 Broadband backbones and deployment technology

Figure 15 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, Feferman and Foditsch, 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.

[77] 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.

[78] 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.

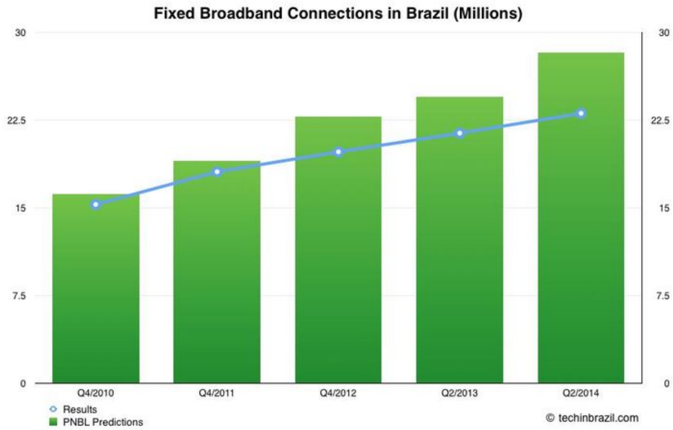


Figure 14: Growth of fixed broadband connections in Brazil

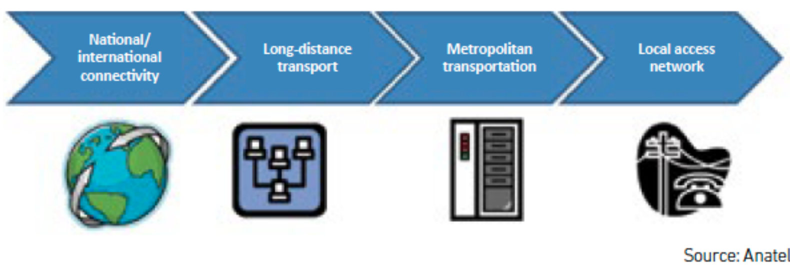


Figure 15: Broadband Supply Chain

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. 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.

D.3 Robustness tests

In this section, I detail the robustness tests I run for the regressions specified by Equations (34), (35) and (36) in Section 7. I start by showing the pre-trends using the specialization index as the alternative definition for division of labor. As shown in Figure 16, 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 12, 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 2. 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 12 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.[79] 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.[80] As shown in Table 29, 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 14, The results are essentially unchanged.

While I argue in Section 7 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 15, the results are not sensitive to excluding establishments in locations 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 16, 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

[79] See Acemoglu and Autor (2011) for a review of the literature on skill-biased technological changes, and Hjort and Poulsen (2019) for direct evidence that impacts from improvement in ICT infrastructure is skill biased.

[80] 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.

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 17, 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 18, 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 18, 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 19, I exclude establishments that are either very near (<10th percentile) or very far (>90th percentile) from the backbone network. In Table 20, I only consider establishments in urban areas by dropping establishments located in microregions with a density lower than 400 persons/km^2.[81] In Table 21, I drop establishments in very large cities.[82] The results are robust to all three tests.

In Table 22, 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 23, I drop the observations in 2010 and 2011. The estimates remain essentially the same as the baseline results.

Results in Table 24 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 on establishment-level variables of interest. In Table 25, 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 26, 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-

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

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

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 (34). 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 17 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 2.

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

$$\log N_{jt} = \alpha + \beta Backbone_{jt} + \gamma Backbone_{jt} \times \log L_{c(j),t_0} + \upsilon Backbone_{jt} \times \log c_{s(j),t_0} + \delta_j + \delta_t + \theta_{m(j)} \times t + \varepsilon_{jt}. \quad (67)$$

As shown in Table 27, 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.

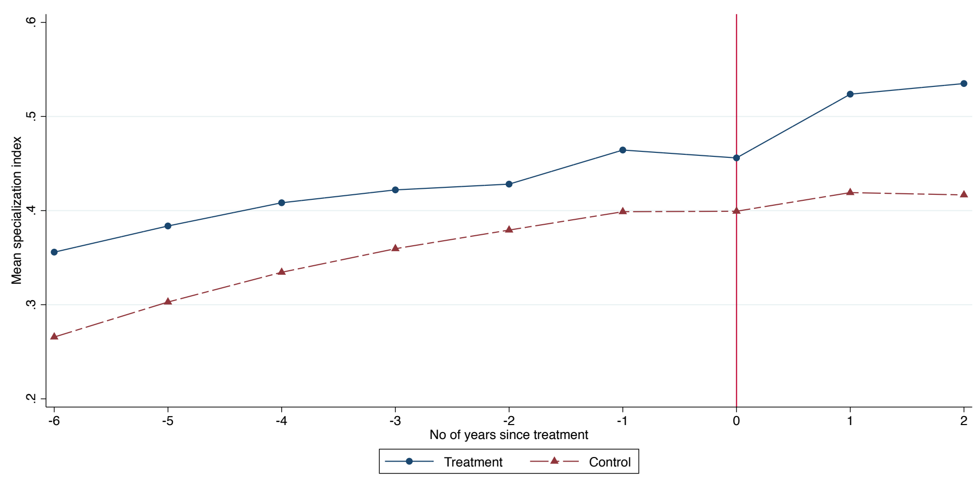


Figure 16: Specialization index in treated versus control groups in Brazil

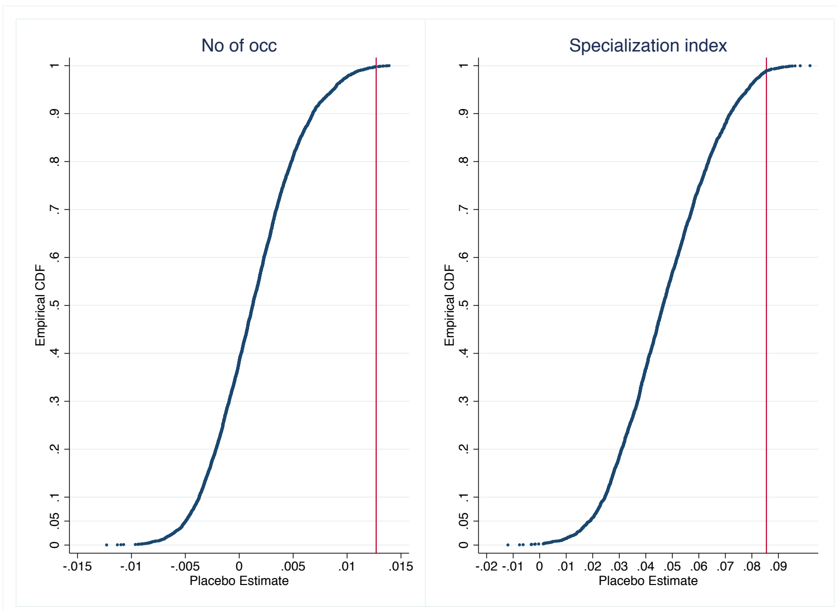


Figure 17: Distribution of placebo estimates

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 (34) 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 2.

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 11: 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 12: 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 13: 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)	(4)	(5)	(6)	(7)	(8)
			Interm. inputs	G3 exp share			Interm. inputs	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 14: Broadband connection and division of labor, only mono-establishment firms

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
$Backbone_{jt}$	.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 15: 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 16: 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 17: 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 18: 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 19: 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) Interm. inputs	(4) G3 exp share	(5)	(6)	(7) Interm. inputs	(8) 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 20: Broadband connection and division of labor, excluding establishments located in rural areas

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)
$Backbone_{jt}$	.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 21: 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 22: Broadband connection and division of labor, separating firms based on export intensity

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
$Backbone_{jt}$	.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 23: Broadband connection and division of labor, excluding observations in Year 2010 and 2011

Dependent variable	Log (No of occs)			Specialization index		
	(1)	(2) Interm. inputs	(3) G3 exp share	(4)	(5) Interm. inputs	(6) 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 24: 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 25: Broadband connection and division of labor, with lead controls

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)
$Backbone_{jt}$	.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 26: 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 27: Broadband connection and division of labor, combining two interactions

D.4 Additional results

In this section, I present additional results on the impact of the improved ICT infrastructure on other establishment-level variables.

	(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 28: 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 29: 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 30: Impacts of broadband backbone on migration of workers and firms

E Policy evaluation

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.

E.1 Short-term impacts

In the model, an exogenous shock to firm's division of labor 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

30, I show that an improved internet connection has no significant effect the migration of establishments or workers, within the observed period.[83] 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.

E.2 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 7 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.

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.

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

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

$$\triangle_t \log y_m = \alpha + \beta Backbone_m + \varepsilon_m \tag{68}$$

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 31. 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 31: 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.[84]

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 7, 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 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.[85] The

[84]These results are the same order of magnitude as the estimates by Hjort and Poulsen (2019), 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.

[85]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 $\bar{U} = \frac{w\kappa}{PL^{\eta}}$,

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 0.28 percentage point, and the aggregate welfare by 0.29 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 0.65% and 1.5% for GDP per capita and city size, respectively.

where P is the aggregate price index.