Wage Inequality and the Spatial Expansion of Firms*

Benny Kleinman

Princeton University†

November 26, 2022

Click here for latest version.

Abstract

Multi-region firms increasingly dominate the U.S. economy. I study the implications of this trend for labor market inequalities. I document that multi-region service firms account for most of the rise in wage inequality since the 1980s, and provide evidence on the uneven nature of their spatial expansion: larger firms operate establishments in more locations, while hiring more skilled labor and paying higher wages in spatially-concentrated headquarters. I integrate this structure into a general equilibrium model, in which (a) firms open branches to serve local markets; (b) the output of headquarters workers is non-rival across branches; (c) firms have wage-setting power. The resulting wage distribution depends on the full network of firm spatial activity, and inequality rises with firms' geographical scope. The model admits tractable aggregation despite its complex micro-structure. I estimate it for 391 U.S. labor market areas and infer frictions to spatial expansion from the universe of HQ-branch linkages. Quantitatively, the decline in these frictions since the 1980s can account for multiple trends in U.S. labor markets, including rising inequality across establishments – between and within firms – and higher inequality and segregation across space.

*I am grateful to my advisors Gene Grossman, Ezra Oberfield, Stephen Redding, and Richard Rogerson for their continued guidance and support. I am particularly grateful to Esteban Rossi-Hansberg for insightful discussions, and to Ernest Liu and Jose P. Vasquez for detailed comments. This paper has also benefited from discussions with Mark Aguiar, Elena Aguilar, Juanma Castro-Vincenzi, Oren Danieli, Greg Dobbels, Pablo D. Fajgelbaum, Manuel García-Santana, Cecile Gaubert, John R. Grigsby, Gregor Jarosch, Hugo Lhuillier, Isabela Manelici, Joan Monras, Eduardo Morales, Hannah Rubinton, Gianluca Violante, and Conor Walsh, as well as numerous seminar participants. I thank Shai Bernstein, Shirley Liu, and Hannah Rubinton for help with the data. This research benefited from financial support from the International Economics Section (IES), the Simpson Center for the Study of Macroeconomics, the Industrial Relations Section (IRS), and the Washington Center for Equitable Growth.
Any views expressed are those of the authors and not those of the U.S. Census Bureau. The Census Bureau's Disclosure Review Board and Disclosure Avoidance Officers have reviewed this information product for unauthorized disclosure of confidential information and have approved the disclosure avoidance practices applied to this release. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 2542.

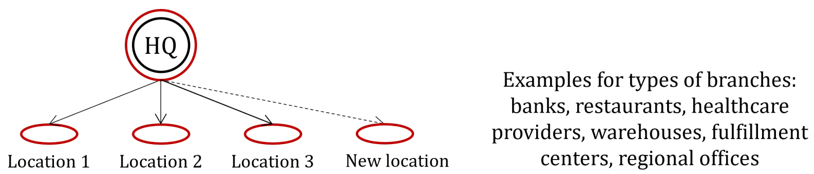
†Princeton University, binyamin@princeton.edu.

1 Introduction

The U.S. economy is increasingly dominated by multi-region firms that operate establishments across multiple local markets. This trend is evident across a wide range of service sectors, from retail and food services to finance and healthcare. In this paper, I study the labor market implications of this trend, and link it to the evolution of the wage distribution in the economy. I argue that the spatial expansion of firms has played an important role in the rise of wage inequality in recent decades, and that it can account for multiple dimensions of the rise in inequality, speaking simultaneously to growing wage dispersion across establishments, across firms and across space.

I conceptualize multi-region firms as networks of local service providers linked by national headquarters – see Figure 1. This structure emerges when output is partially non-tradable, such that a portion of the firm's value added must be generated near its consumers by opening a local establishment. Examples include banks, retail chains, healthcare networks, real-estate management companies and telecommunications firms. I investigate the distributional implications from the expansion of these firms in five steps. First, I provide evidence for the increased prevalence of this class of firms in the economy and for their central role in the rise of inequality. Second, I develop a general equilibrium model that incorporates this form of organization. Third, I provide an analytical characterization of the relationship between firm structure and inequality in the model. Fourth, I estimate the model and recover costs to spatial firm expansion. Fifth, through a series of model counterfactuals, I show that the decline in these costs can account for multiple trends in U.S. labor markets since the 1980s.

Figure 1: Illustration - the structure of multi-region service firms



I begin with an empirical investigation of multi-region firms. I document four stylized facts that serve to both highlight their centrality in the rise of wage inequality and guide the construction of the model. I use confidential Census Bureau data on the structure and payroll of U.S. firms, and complement it with commercial job-postings data from Burning Glass Technologies that allows me to analyze the spatial distribution of jobs within firms. First, I show that between 1980-2017, firms have experienced substantial spatial expansion across a wide range of service sectors. Second, I show that multi-region service firms tend to organize in accordance with the headquarters-branch structure described above, and that firm labor demand is uneven across space: firms hire more skilled labor and pay higher wages in their spatially-concentrated headquarters. Third, I show that multi-region service firms account for most of the rise in wage inequality between 1980-2017.[1] Finally, I show that rising inequality across different establishments within firms accounts for almost half of the role of multi-region firms, highlighting the importance of their internal structure for the study of wage inequality.

To rationalize these trends and analyze the labor market implications of firm expansion, I develop a spatial general equilibrium model with the above form of organization at its core. The model introduces three key features to the production structure of firms. First, firm expansion is spatial: firms can become larger by opening branches in more local markets. Second, firms can hire headquarters workers to improve productivity across all branches, subject to internal communication frictions. Third, factor intensities differ between headquarters and branches, with higher skill-intensity at the headquarters. Consequentially, a large firm in the model is a combination of multiple spatially-dispersed branches and a large skill-intensive headquarters. Firms decide where to locate their headquarters and branches and face firm-specific upward-sloping labor supply curves in each location. Despite this complex micro-structure, the model admits tractable aggregation, and its equilibrium can be represented as a set of non-linear equations only in terms of region-level aggregates.

In the model's equilibrium, the distribution of wages depends on the structure of multi-region firms and their geographical scope. Larger firms serve more markets and pay higher wages across all their locations, both due to exogenous differences in productivity, and due to endogenous differences in the extent of headquarters hiring. At the same time, larger firms are also characterized by higher wage inequality between their headquarters and branches: HQ-level workers affect firm revenues across all locations, while branch-level workers affect only local revenues, resulting in growing differences in compensation as the firm serves more markets. In this environment, a shock that facilitates firm expansion – such as a reduction in within-firm communication frictions[2] – is able to raise inequality both across firms and within them, consistent with the data.

The model also generates implications for regional inequality. Firms demand more skilled labor and pay higher wages in their headquarters locations. Local markets can specialize in hosting headquarters and supplying services to other regions, resulting in spatial differences in income and skill intensity. The scope for such specialization depends on the ability of firms to expand in space. The model can thus generate an increase in regional inequality and segregation following a reduction in frictions to firm expansion, consistent with the data.

[1]Throughout the analysis I focus mostly on within-industry wage inequality. Note, however, that the role of multi-region firms is even higher when considering rising overall inequality. See Section 2 for additional details.

[2]For example, consider improved transportation and communication infrastructure such as air travel, phone-calls, emails and video-calls, allowing firms to open branches in more locations. I later describe in detail how I measure these frictions and their change over time.

To quantify these effects, I estimate the model for 391 labor market areas of the contiguous U.S. in 1980. I employ a Simulated Method of Moments (henceforth SMM) approach, targeting moments from a variance decomposition of wages that utilizes the structure of multi-region firms. In particular, I infer key parameters of the production function from the importance of firm linkages and headquarters locations for the decomposition of establishment-level wages in the data. Region-level fundamental parameters are recovered through inversion of the model's equilibrium conditions, nested in the SMM loop. The estimated model is able to capture key features of the spatial distribution of economic activity and of the wage structure of multi-region firms.

In the main counterfactual analysis, I consider three shocks to the baseline 1980 equilibrium. The first shock is a decline in within-firm communication frictions, recovered from the universe of HQ-branch linkages in the U.S. over 1980-2017. I find that these frictions have declined significantly over time, capturing improvements in transportation and communication technologies, and changes in the regulatory environment that made it easier for firms to operate establishments in more locations. The second shock is a change in demand, motivated by the significant increase in the aggregate expenditure on services in the data ("structural transformation"). The third shock is a homogenous increase in the skill-intensity of production, in line with the literature on skill-biased technical-change.

Quantitatively, I find that lower frictions to firm expansion can account for multiple secular changes in the U.S. labor markets since the 1980s. Inequality across establishments – within and between firms – rises by around twenty percent of the empirical change over 1980-2017, including higher wage dispersion between headquarters and branches. Inequality across local labor markets rises by a similar magnitude. Markets with ex-ante high concentration of service-sector headquarters such as New York City experience an increase in the relative price and quantity of skilled labor, raising the urban wage premium and concentration of skilled labor in large cities. When lower frictions to expansion are combined with higher demand for services or with the homogenous increase in the skill intensity of production, the rise in inequality accounts for around 40%-50% of the change in the data. This increase results from the amplified effect of lower frictions to firm expansion, and not from the individual effect of the other shocks.

I use the model to compute the welfare implications of these changes. Welfare is captured by a measure of real income that reflects individual earnings and local prices of housing and non-tradable services. Holding preferences constant, the measured reduction in spatial frictions to firm expansion is associated with average employment-weighted welfare gains of around 4%, resulting from higher economy-wide productivity and variety of jobs and services. However, these gains exhibit substantial heterogeneity. For example, real income in Fayetteville, AR (the HQ location of Walmart) rises by as much as 18%, while in Pittsburgh, PA it declines by 6%. Real income increases the most for households that reside in locations with ex-ante specialization in hosting headquarters, and particularly so for skilled labor in these locations. Regional income growth is associated with an increase in housing prices, but also with an increase in the variety of local jobs and services.

I conclude by comparing the above mechanism on the spatial expansion of firms to other mechanisms in the literature on rising wage inequality. I highlight that the model can simultaneously speak to several dimensions of wage inequality that have been studied in the literature, as I further detail below. At the same time, it generates distinctive predictions that existing studies do not speak to, such as rising wage dispersion across different establishments within firms – conditional on skill group and location. I provide further evidence in support of these predictions and show that they are quantitatively important for the overall rise in inequality. Other benefits of my approach include the alignment of the model with observed changes in the spatial organization of production; and the focus on services sectors, which drive the empirical changes in firm organization and wage dispersion.

This paper connects to several strands of recent literature. First, I relate to the recent literature that documents empirically the spatial expansion of firms, such as Cao et al. (2017), Aghion et al. (2019), Jiang (2021), Rossi-Hansberg et al. (2021), and Hsieh and Rossi-Hansberg (2022). Building on this literature, I turn to study the labor market implications of firm expansion, with a focus on the relationship between the spatial organization of firms and the distribution of wages in the economy.

Second, I relate to the literature that documents rising wage dispersion across firms and establishments in the U.S., including Barth et al. (2016), Song et al. (2019) and Haltiwanger et al. (2022). I complement the results in these studies by providing new evidence on the central role of multi-region firms and their organization in generating these trends. I also rationalize these facts by modelling the expansion of multi-region firms in a spatial general equilibrium setting.

Third, this paper relates to the literature on "the great divergence" in the spatial economics literature which studies the increased skill differentials in allocations and wages across space, e.g. Berry and Glaeser (2005), Moretti (2012), Diamond (2016), Giannone (2017), Eckert (2019), Rubinton (2020), Card et al. (2021), and Eckert et al. (2022). I contribute a new mechanism to explain these patterns that focuses on within-firm trade across regions and its increased importance due to firm geographical expansion.

From a theoretical perspective, I relate to studies that model the expansion of firms through space, including Jia (2008) and Holmes (2011) in the industrial organization literature; Argente et al. (2020), Oberfield et al. (2020), and Giroud et al. (2021) in the macroeconomics literature; and Helpman (1984), Ramondo and Rodríguez-Clare (2013), Tintelnot (2017), and Arkolakis et al. (2018) in the literature on trade and multinational firms. My two main innovations are to model endogenous headquarters-level decisions that shape branch-level productivity and wages; and to study spatial firm expansion in an economic geography setting that allows for location decisions by both firms and workers, with firms determining locations of both headquarters and branches.

Fifth, I relate to the large literature that documents and models the flows of intangible knowledge and know-how within firms, including Atalay et al. (2014), Fort (2017), Alviarez et al. (2020) and Ding et al. (2022). I emphasize the role of skilled-labor at the firm's headquarters in the production of such knowledge. In addition, I show that this structure has important implications for the rise of wage inequality, in light of firms' growing ability to spread intangible knowledge to their branches through space. I use the structure of the model to overcome the challenge in measuring these flows, and recover frictions to within-firm communication using the empirical network of HQ-branch linkages.

Finally, I relate to the empirical literature on firm-wage setting in multi-location firms, e.g. Derenoncourt et al. (2021) and Hazell et al. (2021). I contribute to this literature by providing micro-foundations for why branch-level wages across space are more similar within firms than across firms, while also highlighting the importance of wage differentials between firm headquarters and branches.

The remainder of the paper is structured as follows. In Section 2, I provide evidence for the spatial expansion of firms in recent decades and for the centrality of multi-region service firms in the growth of wage inequality. In Section 3, I lay out a model of multi-region service firms in spatial general equilibrium. I characterize the equilibrium in Section 4, highlighting how this type of firm structure shapes labor market inequalities. In Section 5, I estimate the model, matching key aspects of the United States in 1980, and evaluate the model's fit to the data. Section 6 analyzes the labor market implications of firm expansion through the lens of the model using a series of counterfactuals, with a focus on declining within-firm communication frictions and rising expenditure on services. In Section 7, I discuss how the mechanism put forward in this paper relates to existing mechanisms in the literature on wage inequality. Section 8 concludes.

2 Descriptive evidence

In this Section, I provide evidence on the spatial expansion of firms in recent decades, on the structure of multi-region firms, and on their role in the rise of wage inequality. I begin with a description of the two main data sources that I use.

2.1 Data and definitions

The main data source is the U.S. Census Bureau Longitudinal Business Database (henceforth LBD), which provides establishment-level data on employment, payroll, location, NAICS industry code and the identity of the firm that owns the establishment.

Additional details on the data and the sample selection are provided in Appendix Section B. I define establishment level wages as the ratio of annual payroll to establishment employment. My definition of a firm follows the standard Census Bureau firm identifiers that link different establishments together based on IRS employer identification numbers (EINs) and ownership data from enterprise-level surveys.

Firm-level industry and sector. In the LBD, establishments are classified into industries, but multi-establishment firms do not have a unique industry identifier. I define firm-level industry and firm-level sector according to the 4-digits and 2-digits NAICS industry codes that account for the largest share of the firm's payroll, respectively. See additional details in Appendix B. I classify firms as "Service firms" if establishments in services-producing sectors – as defined by the Bureau of Economic Analysis (BEA) – account for at least half of the firm's total payroll. Otherwise, firms are defined as "Goods-producing firms".

Geography. Throughout the paper, I define a region or a local labor market by 1990 commuting zones (CZs) in Tolbert and Sizer (1996). A multi-region firm as a firm that has establishments in at least two commuting zones. In the model, I consider a slight aggregation of these commuting zones that groups together very small neighboring commuting zones into labor market areas (henceforth LMAs). I focus on the contiguous U.S., excluding Alaska, Hawaii and the American territories, yielding a total of 722 commuting zones that aggregate into 391 LMAs.

Firm headquarters and main location. The Census Bureau lacks a clear identifier for headquarters establishments and headquarters locations of multi-establishment firms. To address this problem, I apply two alternative approaches. The first approach uses establishments classified under the NAICS-55 industry code ("Management of Companies and Enterprises") to detect headquarters activity. Under this approach, I define the *headquarters location* of a firm as the commuting zone with the highest share of payroll out of its NAICS-55 establishments. I restrict this definition to firms in which at least half of their NAICS-55 payroll is in a single commuting zone. This restriction turns out to exclude only a small percentage of all firms with any NAICS-55 establishments, since payroll in this sector tends to be highly spatially-concentrated within firms. The main problem with this approach is that most firms do not have any NAICS-55 establishments, including a high percentage of the multi-establishment firms.[3] Therefore, as an alternative approach, I define the headquarters market of a firm based on the commuting zone that includes the firm-level mailing address from the Business Register.[4]

[3]This is not due to lack of headquarters activity in these firms. The absence of NAICS-55 establishments in the Census data can indicate that the headquarters establishments are engaged in additional activities and are classified accordingly. This seems to be particularly common in the information and finance sectors, in which firms seems to classify headquarters establishments in accordance to their core activity. Alternatively, it could just indicate measurement errors in the industry classification.

[4]For the subset of firms with both a firm-level mailing address and NAICS-55 establishments, there is a very high correlation in the identity of the headquarters commuting zone across these two approaches.

Burning Glass Technologies (BGT) and Dun & Bradstreet Data. A limitation of the LBD data is the lack of job-level or worker-level information. Therefore, I complement the analysis using online job-posting data from Burning Glass Technologies,[5] which provides information about the spatial distribution of jobs within firms. In these data, a firm is defined based on the set of job postings that share the same codified employer name.[6] The BGT data lacks establishment identifiers, but the geographic location of the posting is known, allowing me to compare firms' postings across different commuting zones and LMAs. These data are only available since 2010, but it is nevertheless useful for understanding how multi-region service firms are structured.

To identify the headquarters market of firms in BGT, I match data from the Dun & Bradstreet individual business files, which include establishment-level data with firm linkages and classification of establishments according to their role within the firm (e.g. headquarters and branches). This matching process results in around 75,000 multi-region firms, out of which around 64,000 are in service sectors. The Dun & Bradstreet data also prove useful for some aspects of the model's calibration, as I explain in Section 5.[7]

I now provide a set of stylized facts on the structure of multi-region service firms and their centrality in the rise of inequality. These facts will guide features of the model that I develop below.

2.2 The spatial expansion of firms

I begin by providing basic statistics on the expansion of firms in recent decades, highlighting the growing importance of multi-region service firms in the economy.[8] Table 1 shows that between 1980 and 2017, the employment share of multi-region service firms has increased from 24% of the total workforce to 41% of the total workforce. This reallocation reflects both the general transition from goods to services ("structural transformation") and the evolution of services from single-region to multi-region firms. Looking only within the services sector, the share of multi-region firms has increased from 35% to 48%.

In the service sector, spatial expansion accounts for much of the growth in average firm size. Employment in single establishment service firms has increased by 3% from an average of 12.7 employees to 13.1 employees, whereas the average multi-region firm has expanded its employment by over 70%, from 432 to 748 employees. Moreover, the

[5] See https://www.economicmodeling.com/.

[6] Burning Glass Technologies claim to invest much effort into name codification to ensure that they capture the same entities.

[7] When using the Dun & Bradstreet data, I am mostly utilizing the information on firm linkages and the spatial distribution of economic activity. These are two dimensions with high-quality coverage in this dataset. See for example Barnatchez et al. (2017) for evidence on the high spatial correlation for the geographic distribution of establishments between Dun & Bradstreet and administrative datasets.

[8] Other papers that have highlighted the importance of the extensive margin for firm expansion include Hsieh and Rossi-Hansberg (2022) and Cao et al. (2017).

Table 1: **Statistics on firm spatial expansion**

Type of firm	Year	Firms	% of workforce	Emp. per firm	Estabs. per firm	CZs per firm
Goods-producing firms	1980	639800	31%	38.0	1.2	1.1
Goods-producing firms	2017	778400	14%	23.3	1.1	1.1
Services, single estab.	1980	2407000	40%	12.7	1.0	1.0
Services, single estab.	2017	3769000	38%	13.1	1.0	1.0
Services, multi-estab. and single-CZ	1980	69000	5%	53.2	2.9	1.0
Services, multi-estab. and single-CZ	2017	72500	7%	125.3	3.5	1.0
Services, multi-estab. and multi-CZ	1980	43500	24%	432.6	12.2	5.0
Services, multi-estab. and multi-CZ	2017	72500	41%	748.2	20.4	6.4

Note: Data from the Census Bureau Longitudinal Business Database. Goods- and services-producing firms stand for firms with at least half of their wage bill across all establishments in goods- and services-producing sectors, respectively. Multi-estab. stands for a firm with at least two establishments, and multi-CZ stands for a firm with establishments in at least two 1990 commuting zones. See Section 2.1 for additional details on sample selection and definitions.

expansion in employment in multi-region firms reflects mostly an increase in the average number of establishments per firm, not an increase in employment per establishment. Much of this expansion took place across local markets, with the average number of commuting-zones per firm increasing by almost 30%. Finally, note that in contrast to services, average firm size and average number of establishments per firm in the goods sector have both declined.[9] For the sake of brevity, Table 1 combines all types of firms in the goods sector together, but additional breakdown by firm type are reported in Appendix Table 10.

The pattern of spatial firm expansion is common to a wide range of services sectors. Figure 2 reports the log change in the average number of establishments per firm across selected 2-digit NAICS sectors relative to 1980, taken from the Census Bureau Business Dynamics Statistics dataset, for firms with at least five employees. While this ratio remained roughly constant in manufacturing, it rose in all service sectors, from 0.13 in Accommodation and Food Services, to 1.19 in Real Estate, Rental and Leasing.

I summarize these findings in the following:

Fact 1. ***Firm spatial expansion:*** *Between 1980-2017, the average number of markets per firm and the share of employment in multi-region firms have both increased in services-producing sectors.*

[9] Holmes and Stevens (2014) provide one potential explanation for this trend, based on the idea that rising international trade has shifted the production of large-scale standardized goods to other countries, leaving the domestic production of manufacturing goods more concentrated on custom or specialty goods. In any case, in this paper I do not analyze the decline in the size and scope of manufacturing firms, and focus mostly on the expansion of firms in services.

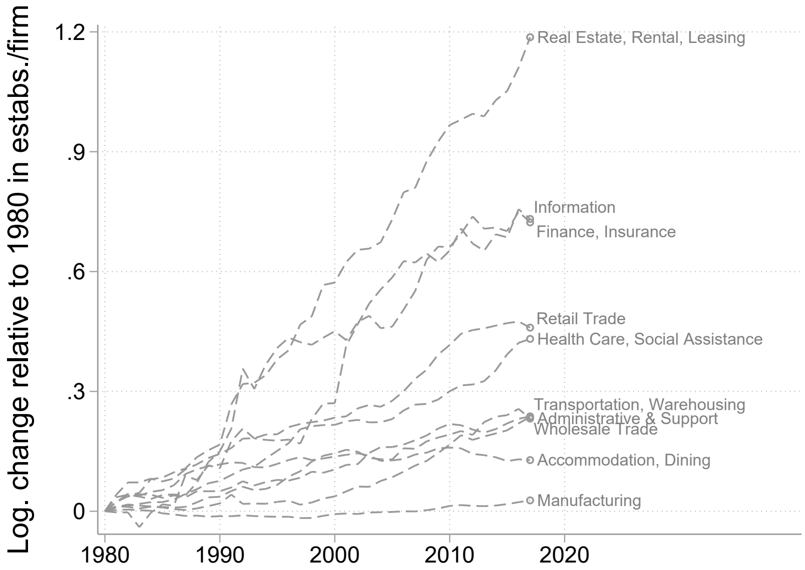


Figure 2: Firm expansion by sector: This figure shows the log change in the average number of establishments per firm relative to 1980 by economic sector, for firms with at least 5 employees. Data from the Business Dynamics Statistics dataset.

2.3 The structure of multi-region service firms

I now provide evidence on the spatial organization of multi-region service firms. I show that the composition of their labor demand is uneven across space, motivating the link between spatial expansion and heterogenous labor market effects. In particular, I highlight the prevalence of a structure comprised of spatially-dispersed branches and spatially-concentrated, skill-intensive, headquarters.

2.3.1 Firm spatial structure - industry composition

I first characterize the spatial organization of firms through the lens of their industry composition in the LBD, using establishment-level NAICS industry identifiers. For each firm, I divide its establishments into three categories. First, I classify establishments as belonging to the firm's core sector if they have the same 2-digit industry classification as that of their firm. Second, I classify establishments as performing headquarters activity if they do not belong to the firm's core sector, and are classified in one of the business-services sectors (NAICS sectors 51, 52, 53, 54, and 55). Most of this category – 80% of the total payroll in this category – is accounted for by explicit headquarters activity, as captured by establishments in the NAICS 55 sector ("Management of Companies and Enterprises"). The remaining establishments are classified as "other establishments". The last group would typically include wholesale and retail establishments for firms outside of the wholesale and retail sectors, as well as some manufacturing establishments.

Table 2 shows how total payroll in multi-region service firms is distributed across these three categories. I focus on the group of firms with a clear headquarters location based on NAICS-55 activity as defined in Section 2.1.[10] The typical firm spends around 85% of its payroll on establishments in its core sector, 11% on headquarters activity, and 4% on other types of establishments. Columns 2 and 3 in the same table reveal that the distribution of headquarters activity within the firm is highly spatially concentrated in its headquarters market: 30% of the payroll in the firm's HQ commuting zone is accounted by establishments that perform headquarters activity, relative to 3% in all other commuting zones. Evidently, headquarters activity is characterized by high spatial concentration within firms, and almost all of the firm's operations outside of its HQ commuting zone is accounted for by establishments in its main sector of activity.[11]

Table 2: The industry composition of multi-region service firms

Firm activity	(1) % of firm payroll	(2) % of HQ-CZ payroll	(3) % of non-HQ-CZ payroll
Establishments in the firm's core sector	84.4%	67.7%	92.1%
Headquarters activity	11.2%	29.2%	3.0%
Other establishments	4.4%	3.1%	4.9%
Total	100%	100%	100%

Note: Data from the Census Bureau Longitudinal Business Database. The table shows how the total payroll of multi-region service firms is distributed across different activities, in firms with clearly identifiable headquarters location, as discussed in Section 2.1. Column (1) shows the distribution of firm payroll across these activities for the firm as a whole; Column (2) shows it for establishments in the firm's headquarters commuting zone; and Column (3) shows it for all other commuting zones. Establishments in the firm's core sector are those with an industry classification identical to the firm-level 2-digit main NAICS code, as discussed in Section 2.1. Headquarters activity is defined by establishments that do not belong to the firm's core sector and are classified in one of the business services sectors (2-digit NAICS codes 51, 52, 53, 54, 55). Most of this category (80%) is accounted for by establishments in the NAICS-55 sector ("Management of Companies and Enterprises").

2.3.2 Firm spatial structure - occupational composition

Next, I investigate the distinction between headquarters activity and branch activity through the spatial composition of jobs within firms, using data from Burning Glass Technologies. First, I show that hiring at the headquarters market of firms tends to include spatially-concentrated occupations that are less common in firms' non-HQ locations. I then show that these occupations tend to be more skill intensive.

[10]This group of firms has the advantages of: (a) providing informative industry classifications; and (b) have a clear headquarters location. Multi-unit service firms outside of this group tend to be smaller and report very homogenous industry classifications across all of their establishments, which prevents us from using industry classification to analyze the firm's structure. In particular, it is not clear if in these firms all establishments perform the same activity, or if they report a generic industry code for all their establishments regardless of the activity performed in the establishment.

[11]Table 12 in the appendix reports a similar decomposition at a more granular level, including a breakdown of the firm's core sector into the core 4-digit industry and other establishments in the core sector; and a breakdown of headquarters activity into NAICS-55 establishments and a residual. The main takeaways remain similar to the analysis in this section.

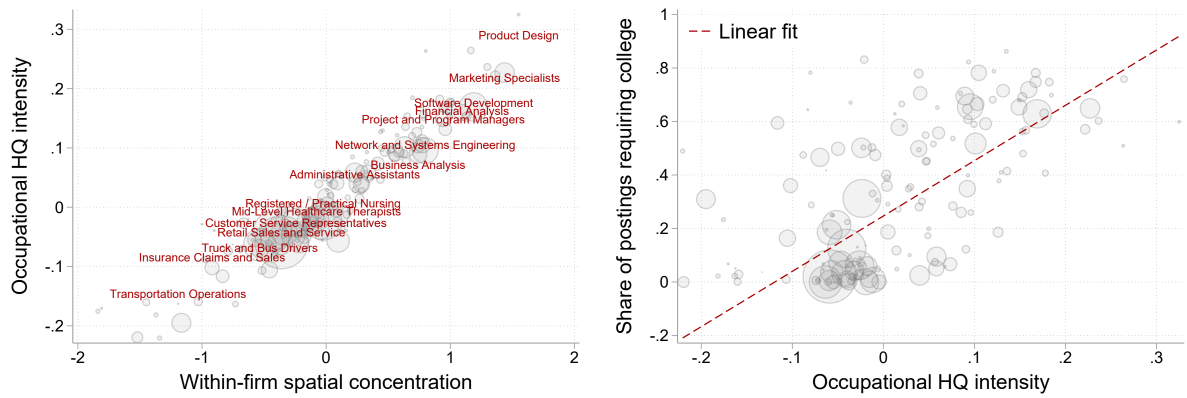
To measure the geographic concentration of occupations within firms, I calculate a Herfindahl-Hirshmann index using data on firm job postings in different markets. High concentration means that an occupation tends to be hired only in a small subset of the firm's locations, while low concentration means that an occupation tends to be hired in most of the firm's locations. Specifically, I define this measure as

$$HHI_{fo} \equiv \frac{\left(\sum_j \left(\frac{n_{foj}}{n_{fo}}\right)^2\right) - \frac{1}{J_{fo}}}{1 - \frac{1}{J_{fo}}},$$

where n_{foj} is the number of job-postings for firm f, occupation o in commuting zone j over the years 2010-2019, n_{fo} is the sum of n_{foj} across the firm's markets, and J_{fo} is the total number of markets with fo postings. The adjustment to J_{fo} yields the normalized HHI, which controls for variation in the number of markets in which firms are active. I then compute the postings-weighted average of HHI_{fo} across all firms to obtain an occupation-level measure HHI_o.[12] The resulting index measures how spatially concentrated each occupation is within firms relative to other occupations in the same sector. For example, Retail Managers and Registered Nurses are very spatially-dispersed within firms, while Product Designers and Financial Analysts are very spatially-concentrated.

Figure 3: The spatial distribution of occupations within firms

(a) Occupational spatial concentration within-firms **(b)** HQ-intensity and skill requirements



Note: This figure shows how different occupations are spatially distributed within the firm. Panel (a) shows that spatially-concentrated occupations within firms are likely to be hired in the headquarters market of firms. The y-axis shows the share of job-postings that are located in the headquarters market (commuting zone) of firms. The x-axis shows the average across firms in the log of within-firm normalized Herfindahl–Hirschman index (HHI). Both measures are shown after demeaning NAICS-2 sector-level effects. The HHI is computed for each firm and occupation separately across all commuting zones in which firms are active. Panel (b) shows the share of job-postings that require a college degree for each occupation against the above measure of occupational headquarters-intensity. In both panels, the size of the circle represents total job-postings for an occupation.

[12] I demean sectoral fixed effects to account for the fact that some occupations tend to be hired by more spatially-concentrated industries.

In the left panel of Figure 3, I show how these measures vary with the probability that an occupation is hired at the headquarters market of a firm, which I label "HQ-intensity". While high HQ-intensity implies high concentration, the converse is not true: occupations can in principle be concentrated in markets that do not host the firm's headquarters. The figure shows a very strong relation between the two measures, implying that firms tend to hire the spatially-concentrated occupations in their headquarters market, and that headquarters hiring differs notably from the hiring of other establishments. The right panel of Figure 3 reveals that there is a strong positive correlation between the HQ-intensity measure and the share of postings in each occupation that demand a college degree. Evidently, tasks that are performed at firm headquarters are more skill-intensive than tasks performed in other establishments.

Table 3 further reports how total demand for skill varies across different firm locations, comparing the headquarters commuting zone to other locations. Demand for skill is again measured by the percent of job-postings that explicitly require a college degree. While sectors vary greatly in their skill-intensity, all service sectors share the common pattern of hiring more skilled labor in their headquarters market than elsewhere. The differences between the headquarters market of firms and their other markets are meaningful, with differences of between 8% (in Health-Care) and 38% (in Wholesale-Trade).

Table 3: The distribution of demand for skill within firms

NAICS code	Sector	% Postings in HQ-CZ	% College		Δ
			HQ CZ	Other CZs	
42	Wholesale Trade	15%	54%	16%	38%
44-45	Retail Trade	9%	42%	7%	35%
48-49	Transportation and Warehousing	10%	36%	7%	30%
51	Information	30%	63%	43%	20%
52	Finance and Insurance	22%	52%	39%	12%
53	Real Estate and Rental and Leasing	12%	30%	11%	19%
54	Professional, Scientific, and Technical Services	25%	65%	53%	11%
56	Administrative and Support, Waste Management	15%	33%	13%	20%
62	Health Care and Social Assistance	37%	32%	25%	8%
71	Arts, Entertainment, and Recreation	26%	26%	14%	13%
72	Accommodation and Food Services	12%	15%	5%	10%
81	Other Services	24%	38%	19%	19%

Note: Data from Burning Glass Technologies online job-postings. The table is based on online job postings by firms in service sectors that are active in at least two commuting zones and have a matched headquarters location from Dun & Bradstreet data, according to the procedure described in Section 2.1. The sector of a firm is determined by the main 2-digit NAICS code in its job postings, and the table shows sectoral aggregates across all firms in a given sector. The third column shows the percent of job postings in the headquarters commuting zones of firms out of their total postings. The fourth column shows the percent of postings in the headquarters location of firms that require a college degree. The fourth column shows the corresponding figure in all other non-headquarters commuting zones in which firms are active.

Headquarters activity is also associated with higher wages, beyond what can be explained by the above differences in skill-intensity. In appendix Section D, I provide a series of such conditional correlations to demonstrate this result. I regress posted wages in the Burning Glass data on my measure of HQ-intensity, controlling for a wide range of job characteristics to demonstrate the robustness of this pattern (education, location, firm, industry, year). I further show that this headquarters premium is higher in firms that operate in more markets. These patterns are consistent with the model developed below, and I will provide additional details on them in Section 5 that confronts the model with the data.

I conclude these findings with the following fact:

Fact 2. *Headquarters activity is spatially-concentrated within firms, and it is characterized by greater skill-intensity and higher wages than branch-level activity.*

2.4 Multi-region firms and wage inequality

I now turn to show that multi-region service firms played a key role in the rise of wage inequality. I focus on the variance of log average payroll across establishments in the U.S.,[13] which has been shown in previous studies to mirror the increase in the variance of log individual earnings.[14] This measure increased by 0.17 points between 1980 and 2017. Naturally, much of this increase comes from sectoral differences, with evidently different trends between high-skilled sectors such as finance and information and low-skilled sectors such as retail trade. In the model, I will consider a generic services sector and will not explicitly address sectoral differences other than the broad distinction between goods and services. Therefore, from this point on, I focus on the within-industry rise in inequality.[15] The left panel in Figure 4 shows the decomposition of the total trend in the variance of log wages into the part due to industry differences and the part due to within-industry inequality. Between 1980 and 2017, slightly over 50% of the rise in inequality occurs within narrowly-defined industries,[16] with an overall increase of 0.084 in the within-industry variance.

The right panel of Figure 4 separates the universe of establishments into two groups, multi-establishment firms and single-establishment firms, and shows the change in within-

[13] Throughout the analysis and in all references to variances of log wages, I weight different establishments by their employment.

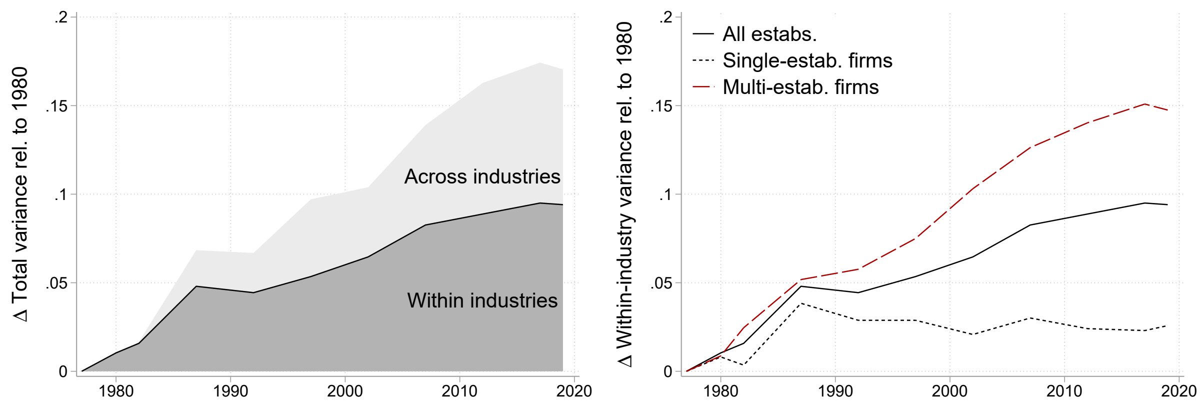
[14] See Barth et al. (2016). The authors also use the Longitudinal Employer-Household Dynamics (LEHD) data to show that the rise in inequality across U.S. establishments accounts for most of the rise in inequality across individuals.

[15] I control for 4-digit firm-level NAICS industry code as defined in Subsection 2.1. Note that I do not absorb establishment-level industry, since within-firm inequality between different kinds of establishments - particularly headquarters and branches - plays an important role in the suggested mechanism.

[16] Haltiwanger et al. (2022) find that a slightly higher share – around 60% – of the rise in total inequality in the LEHD is due to differences between industries starting from the late 90s. Indeed, as can be seen in Figure 4, the importance of differences between industries becomes larger in the latter part of the sample, reconciling the figures reported here with their findings.

Figure 4: The role of multi-establishment firms in the rise of inequality

(a) ΔVariance within and between industries **(b)** ΔVariance within industries by firm type



Note: Panel (a) shows changes in the employment-weighted variance of log average payroll across establishments in the Longitudinal Business Dataset (LBD) in selected years relative to 1978, separating the total increase into the part due to rising variance between industries (upper shaded area) and within industries (lower shaded area), where a firm's industry is defined by the NAICS 4-digit code with the maximal share of its total payroll. Panel (b) reports the change in the within-industry variance for all establishments (solid-black line), for multi-establishment firms (dashed-red line) and for single-establishment firms (dotted-gray line).

industry variance of log wages for each group separately. Note that the variance for multi-establishment firms includes both changes across firms and changes between different establishments within firms. The figure shows a sharp increase in the variance for establishments that belong to multi-establishment firms. In contrast, for single-establishment firms, there is only a mild increase in the 1980s, and if anything, a declining trend since then.

I quantify the importance of multi-establishment firms for the rise in inequality using a variance decomposition of log wages. Consider a partition of the universe of establishments into G distinct groups. The total change in variance equals to the sum of several components:

$$\begin{aligned}\underbrace{\Delta\sigma^2_t}_{\text{Change in variance}} &= \underbrace{s_{g',0}\left(\Delta\sigma^2_{g',t}\right)}_{\text{Variance within group } g'} + \underbrace{\sum_{g\in G}\left(\Delta s_{g,t}\right)\sigma^2_{g,0}}_{\text{Emp. reallocation across groups}} \\ &+ \underbrace{\sum_{g\in G}\left(\Delta s_{g,t}\right)\left(\Delta\sigma^2_{g,t}\right)}_{\text{Comovement of variance and emp.}} + \underbrace{\underbrace{\sum_{g\in G/g'} s_{g,0}\left(\Delta\sigma^2_{g,t}\right)}_{\text{Variance within other groups}} + \underbrace{\sum_{g\in G}\Delta s_{g,t}\left(\mu^2_{g,t}-\mu^2_t\right)}_{\text{Variance between groups}}}_{\text{Residual}} \end{aligned} \tag{1}$$

where $s_{g,t}$ is the employment-share of group g in period t; $\mu_{g,t}$ is the employment-weighted mean of log-wages in group g, period t; $\sigma^2_{g,t}$ is the employment-weighted variance in group g, period t; μ_t is the aggregate mean; and 0 and 1 mark the initial and the final period.

In this equation, the first term captures the rise in variance for a particular group $g' \in G$, e.g. multi-region service firms, multiplied by its initial share of total employment. The second term captures reallocation of employment across groups, keeping constant the variance of each group at its base level. The third term captures cross-changes: it adds a positive value to total variance when a group with rising variance also sees an increase in its employment share. Finally, I include all other terms in a residual, which comprised of the rising variance within all other groups $g \in G/g'$ and rising variance between groups.

Table 4 shows this decomposition for the total rise in within-industry wage inequality. Column (1) considers this decomposition when singling out multi-establishment firms as the group g', and shows that over 80% of the rise in overall (within-industry) inequality is due to rising variance for this group, consistent with the pattern in Figure 4 above. Column (2) repeats this decomposition when singling out the group of service-sector firms, highlighting the key role of services in rising inequality. Note that in this case, the reallocation and cross-changes terms play a bigger role, due to rising role of services in the economy. Finally, Column (3) repeats this decomposition when narrowing down the focus group to only multi-market service firms. In this case, 40% of the overall increase in inequality is due to rising inequality for this group, and 28% is due to the reallocation to this group and the cross effect of reallocation and rising variance. The first term is smaller relative to Column (1) since as we have seen in Table 1, multi-market service firms accounted for only 24% of total employment in 1980. Still, even holding this share constant, the rising variance in the group of multi-market service firms explains 40% of the overall change. Together with the reallocation of employment to this group, it explains 68% of the total rise in inequality. I summarize this finding in:

Fact 3. *Multi-region firms and inequality:* *Multi-region service firms account for most of the rise in within-industry wage inequality between U.S. establishments over 1980-2017. This reflects both an employment reallocation to these firms and rising inequality for this group of firms.*

We can further decompose the increase in the variance for multi-region service firms, $\Delta\sigma^2_{\text{MR-serv},t}$, to changes within firms and changes between firms:

$$\Delta\sigma^2_{\text{MR-serv},t} = \underbrace{\Delta_t \mathbb{V}\left[\bar{\ln w_f}\right]}_{\text{Between-firms - 55\%}} + \underbrace{\Delta_t \mathbb{V}\left[\ln w_{fj} - \bar{\ln w_f}\right]}_{\text{Within-firms, between establishments - 45\%}}$$

where w_{fj} is the average wage in establishment j in firm f and $\bar{\ln w_f}$ is the employment-weighted average of log wages across all establishments in firm f. Rising differences within firms – between their different establishments – explain around 45% of the overall change $\Delta\sigma^2_{\text{MR-serv},t}$, highlighting the importance of internal firm structure for the rise in inequality. I have thus established:

Table 4: Decomposition of the rise in wage inequality

% of the overall increase in variance (1980-2017)	(1) **Multi-estab. firms**	(2) **Service firms**	(3) **Multi-CZ services**
Rising variance within the group of firms	83%	70%	40%
Changes in employment shares b/w groups (reallocation)	-4%	6%	8%
Comovement of variance and employment shares	9%	10%	20%
Residual	12%	13%	32%
Total change across all firms in the economy	**100%**	**100%**	**100%**

Note: Data from the Census Bureau Longitudinal Business Database. Each column is a separate decomposition of the total increase in within-industry wage inequality across all establishments between 1980 and 2017. The first row shows the share of total increase in variance due to rising variance in the group of firms that is mentioned at the top of each column, matching the first RHS term in Equation (1). The second row shows the share due to changes in employment between that group and the other firms in the sample (employment reallocation), holding constant the change in variance in each group, matching the second RHS term in Equation (1). The third row shows the share that is due to the cross-product of rising variance and rising employment share, matching the third RHS term in Equation (1), and the fourth row is a residual so that the sum for each column is 100%. See Section 2.4 for additional details.

Fact 4. *Close to half of the rise in inequality for multi-region service firms is between different establishments within these firms.*

To recap, I have shown that multi-region service firms: (a) have experienced substantial spatial expansion between 1980-2017, and account for a growing share of the total workforce; (b) tend to organize as spatially-concentrated headquarters and spatially-dispersed branches, with greater skill-intensity and higher wages at their headquarters; (c) account for most of the rise of within-industry wage inequality between 1980-2017; and (d) experienced an increase in wage inequality across their different establishments, accounting for much of their role in the overall rise in inequality.

3 A spatial model of firm expansion

In this section I develop a model to analyze how changes in the spatial scope of firms shape labor market outcomes. The model introduces two novel features into an otherwise standard quantitative spatial general equilibrium framework. First, I introduce a production function for multi-region service firms, according to which firms need to open a branch to serve any given market, and the output of their headquarters labor is a non-rival input across branches. Second, I assume that workers have idiosyncratic preferences for different employers,[17] generating monopsony power for the firm in each of

[17] See e.g. Card et al. (2018).

its markets.[18]

3.1 Setup

The economy consists of N regions, indexed by i and j. There are S types of workers who differ in skill and indexed by s, with an aggregate supply of $\bar{L}_s$ workers of type s. All households consume services and tradable-goods, and can choose their region i and employer ν.[19] For simplicity, I assume that tradable-goods are homogenous, freely traded and produced using constant returns to scale (henceforth CRS) technology. Since the price of goods is identical in all regions, I choose it as the numeraire. In the services sector, households consume a continuum of varieties in a setting of monopolistic competition. These varieties are partially non-tradable in the sense that part of the value added is generated near the consumer, by hiring local labor; while another part is tradable and generated by hiring workers at a potentially distant headquarters.

3.2 Households

Households choose their region i, employer ν, consumption of tradable-goods Q_i^g and the bundle of local services Q_i to maximize

$$U_{is\nu} = b_{i\nu} \left(Q_i\right)^{\beta} \left(Q_i^g\right)^{1-\beta}, \quad Q_i = \left[\int_{\omega \in \Omega_i} \left(q_i\left(\omega\right)\right)^{\frac{\sigma-1}{\sigma}} d\omega\right]^{\frac{\sigma}{\sigma-1}}, \tag{2}$$

where $b_{i\nu}$ is an household-specific idiosyncratic preference shock for region i and employer ν; $\beta \in (0,1)$ is the expenditure share on services; Q_i^g is a homogenous quantity of goods; and the bundle Q_i aggregates all service varieties ω available in market i, using a constant elasticity of substitution $\sigma > 1$. I denote the price index for local services by P_i, and the price index that aggregates goods and services by $\bar{P}_i$.[20]

Labor supply. Households draw the set of idiosyncratic shocks $b_{i\nu}$ from a nested Fréchet distribution,[21] which guides their sorting decisions across space and across employers. The upper nest reflects preferences across locations with dispersion given by ξ, capturing a regional labor supply elasticity. The lower nest reflects preferences across

[18] In appendix H, I recast the model with a different mechanism to generate firm-specific wages that allows for unobserved worker heterogeneity. While this implies an alternative interpretation for why workers are paid different wages conditional on their skill group and location (focusing on workers' ability and not on preferences), it maintains the wage structure of multi-region firms and their role in the rise of wage inequality over time.

[19] An employer can be the local branch of some firm or its headquarters.

[20] The services price index is given by $P_i = \left(\int_{\omega \in \Omega_i} P_i\left(\omega\right)^{1-\sigma} d\omega\right)^{\frac{1}{1-\sigma}}$. Since the price of goods is set to unity, $\bar{P}_i = P_i^{\beta}$. In the quantification of the model in Section 5 I add local housing, which also enters into the formula for the regional final price index $\bar{P}_i$.

[21] Specifically, I assume that $F\left(\boldsymbol{b}\right) = \exp\left(-\sum_{i=1}^{N}\left(\int_{\nu \in \mathcal{V}_i} b_{i\nu}^{-\epsilon} d\nu\right)^{\frac{\xi}{\epsilon}}\right)$.

employers[22] with dispersion given by ϵ, capturing an employer-level labor supply elasticity. As a result, an employer ν in region i that wishes to hire $l_{is}(\nu)$ workers of skill s is required to pay a wage

$$w_{is}(\nu) = L_{is}^{-\frac{1}{\epsilon}} W_{is} l_{is}(\nu)^{\frac{1}{\epsilon}}, \quad W_{is} = \left(\int_{\nu \in \mathcal{V}_i} w_{is}(\nu)^{\epsilon} d\nu \right)^{\frac{1}{\epsilon}}, \tag{3}$$

where L_{is} is the measure of type-s workers that choose location i[23] and W_{is} is the regional wage index for type-s workers.[24] When the employer-specific labor supply elasticity is infinite ($\epsilon \to \infty$), all local employers pay a common wage equal to the index W_{is}. Otherwise, firms need to pay higher wages to attract more workers.

3.3 The services production function

The production structure of multi-region firms is the key novel element of the model. I define a firm in the services sector by its technology to produce a unique variety. Consider a firm headquartered in market i with the capability to produce a variety ω. The firm can choose to open a branch in any market j by paying a fixed cost f in units of the local final good. A firm that hires l_{ijs} workers of type s produces $q_{ij}(\omega)$ units of output in that market according to

$$q_{ij}(\omega) = \varphi_{ij}(\omega) \prod_{s=1}^{S} l_{ijs}^{\alpha_s}, \tag{4}$$

where $\varphi_{ij}(\omega)$ is the productivity of the firm in market j conditional on having its headquarters in market i; and α_s is the output-elasticity of type-s labor in branch-level production. For example, consider the case of Starbucks branches in San Francisco. In this case, ω stands for Starbucks; i for Seattle – the headquarters location of Starbucks; j for San Francisco; and l_{ijs} for Starbucks worker of skill s in its San Francisco branches.[25]

[22] An employer in the model is a combination of a firm (i.e. a variety ω) and a local market. For example, agents draw different shocks $b_{i\nu}$ for a Starbucks branch in New-York and a Starbucks branch in Seattle. In addition, I assume that headquarters jobs and branch-level jobs for the same firm are distinct jobs: working for the Starbucks HQ in Seattle is not the same as working for one of its branches there. Therefore, firms face multiple labor supply curves: one for each market in which they have a branch, and another one for their headquarters.

[23] The probability that workers of type s choose location i is given by $\frac{L_{is}}{\sum_{j=1}^{N} L_{js}} = \frac{\left(W_{is}/\bar{P}_i\right)^{\xi}}{\sum_{j=1}^{N}\left(W_{js}/\bar{P}_j\right)^{\xi}}$. It increases with real income in location i, given by the ratio of the local s-specific wage index $W_{i,s}$ to the local price index $\bar{P}_i$.

[24] This index aggregates wages across all available employers in market i, $\mathcal{V}_i$. It takes into account both the level of wages and the variety of jobs.

[25] When I quantify the model in Section 5, I define a region in the model as a U.S. commuting zone. Accordingly, a branch in the model aggregates all establishments of a given firm in a commuting zone. One of the advantages of this approach is that the model can remain generic with respect to the specific industry and types of firms that it considers. For example, Walmart might choose fewer establishments to serve a local market relative to Starbucks, but the model encompasses both by aggregating all activity within a local market.

The productivity term $\varphi_{ij}(\omega)$ captures the importance of multi-region firms and comprises four parts:

$$\varphi_{ij}(\omega) = A_{ij} z_{\omega,i} \phi_{\omega,j} \left(\prod_{s=1}^{S} h_{is}^{\rho_s} \right)^{\gamma}. \tag{5}$$

First, all firms that are headquartered in market i and serve market j share a common productivity shifter A_{ij}. This shifter captures all exogenous factors that affect the ability to operate a HQ-branch relationship between i and j, such as distance, availability of transportation and communication infrastructure, and the regulatory environment.[26,27] In the quantification of the model below, I will assume that $A_{ij} \equiv A_i/\tau_{ij}$, where A_i is a unilateral productivity component, and τ_{ij} is a bilateral and symmetric term that captures spatial frictions for operating HQ-branch linkages. The second term, $z_{\omega,i}$, is firm-level productivity, affecting production across all of firm ω branches. The subscript i captures the dependency of this firm-level component on the firm's headquarters location.[28] The third term, $\phi_{\omega,j}$, is a branch specific shock, reflecting all factors that make location j a particularly successful branch for the firm. In the Starbucks example, A_{ij} captures the ability of any Seattle-based firm to operate branches in San Francisco; $z_{\omega,i}$ captures the productivity of Starbucks across all its branches; and $\phi_{\omega,j}$ is a shock unique to Starbucks in San Francisco.

The final component in the productivity composite (5) reflects the firm's choice of inputs in the headquarters. Firms choose the measure of workers h_{is} to hire in their HQ, where ρ_s is the elasticity of the HQ composite to type-s labor. The output of these workers is non-rival across branches, with the parameter γ governing the elasticity of branch productivity to HQ inputs. In the Starbucks example, h_{is} represents workers in its Seattle HQ, such as Starbucks designers, food-scientists, and programers.

For future reference, the four productivity components are summarized in Table 5 below.

[26]Inter alia, it also includes all unilateral exogenous features that make location i particularly attractive for firms to place their headquarters there. For example, the availability of an airport might make it easier for businesses to manage their branches from i throughout all destinations j. Note that the supply of labor (including the relative supply of skilled labor) and the size of the market are endogenously determined in the model, so they are excluded from this term.

[27]The U.S. banking sector provides an example in which regulations were responsible for particularly low A_{ij}, since in the early 1980s banks in many states were still forbidden from operating branches in other states. In the model, this would be captured as $A_{ij} \to 0$ when i and j are in different states. The removal of these regulations would be interpreted as a transition to positive A_{ij}.

[28]The firm has a single productivity $z_{\omega,i}$ once its headquarters location is set. However, as we will see later, it faces multiple $z_{\omega,i}$ – one for each market – when making its headquarters location decision.

Table 5: Firm productivity components

Firm productivity term	Interpretation
A_{ij}	Exogenous bilateral shifter, common to all firms with HQ at i and a branch at j
$z_{\omega,i}$	Firm-specific draw, common to all the firm's branches if it is HQed at i
$\phi_{\omega,j}$	Firm-branch-specific draw in location j - no direct effect on other branches
$\left(\prod_{s=1}^{S} h_{is}^{\rho_s}\right)^{\gamma}$	Endogenous choice of HQ labor, affecting all the firm's branches

Note: Productivity components for a firm ω wih headquarters at market i and a branch at market j.

3.4 Firm's production problem

From this point on, I omit the reference to the firm's variety ω, since all firms with the same headquarters location i and productivities z_i and $\boldsymbol{\phi} \equiv \{\phi_j\}_{j=1}^{N}$ make the same decisions. The firm's problem is to choose branch-level labor $\mathbf{l} \equiv \{l_{js}\}$ across all markets j and HQ-level labor $\mathbf{h} = \{h_{is}\}$ to maximize profits $\pi_i(z_i, \boldsymbol{\phi})$. Profits are given by the sum of sales across all markets, after deducting variable and fixed costs across all active branches and payments to headquarters labor. The firm internalizes that it faces a downward-sloping demand curve in each market, with demand elasticity σ; and an upward-sloping labor supply curve for each branch and for the HQ, with labor supply elasticity ϵ. Profits for a i-headquartered firm are thus given by[29]

$$\begin{aligned}\pi_i(z_i, \boldsymbol{\phi}) = \max_{\boldsymbol{h},\boldsymbol{l}} \sum_j & \underbrace{E_j^{\frac{1}{\sigma}} P_j^{\frac{\sigma-1}{\sigma}} q_{ij}(z_i, \boldsymbol{\phi})^{\frac{\sigma-1}{\sigma}}}_{\text{Sales in market } j} \\ - \sum_j \mathbb{I}_j & \underbrace{\left(\sum_s W_{js} L_{js}^{-\frac{1}{\epsilon}} l_{js}^{\frac{\epsilon+1}{\epsilon}} - \bar{P}_j f\right)}_{\text{Costs in market } j} - \underbrace{\sum_s W_{is} L_{is}^{-\frac{1}{\epsilon}} h_{is}^{\frac{\epsilon+1}{\epsilon}}}_{\text{HQ costs in market } i},\end{aligned} \tag{6}$$

where $\mathbb{I}_j$ is an indicator that equals to 1 if the firm is active in market j.

I impose two additional assumptions to facilitate tractable aggregation of the model. First, I assume that firms choose their HQ inputs $\boldsymbol{h}$ before the branch-level shocks $\boldsymbol{\phi}$ are observed, and that they make branch-level decisions $\boldsymbol{l}$ after $\boldsymbol{\phi}$ are observed. Therefore, firms decide on $\boldsymbol{h}$ by forming expectations about which markets they will serve, based on their knowledge of the HQ location i and their firm-level productivity z_i. Second, I assume that the branch-level shocks $\boldsymbol{\phi}$ are drawn independently from a common Pareto distribution with shape parameter θ.[30]

[29] Note that under this specification, every firm has headquarters and branches. A single-establishment firm in the data is therefore captured in the model by a combination of small headquarters and a single branch in the same location. This can be thought of as combination of workers and management in a single establishment. The model remains agnostic on whether these activities occur in the same establishment or different establishments, conditional on occurring in the same labor market.

[30] In practice, I assume that the shape parameter is scaled by θ and given by $\left(\frac{\sigma}{\sigma-1} - \frac{\epsilon}{1+\epsilon}\right)\theta$ for

3.5 Entry and headquarters location decisions

Firms can enter freely in any initial market n by paying f_e units of the local final good. Upon entry, firms observe their firm-level productivities $\boldsymbol{z}$ across all potential markets and choose where to locate their headquarters to maximize profits $\pi_i(z_i, \boldsymbol{\phi})$,

$$\max_i \mathbb{E}_{\boldsymbol{\phi}}\left[\pi_i\left(z_i, \boldsymbol{\phi}\right)\right]. \tag{7}$$

Firms thus sort in space based on the knowledge of their potential productivity in each location. However, I introduce a simple friction in this process that allows for a lasting effect of the initial entry location.[31] Specifically, I assume that firms retain only a share $c \in (0,1)$ of their productivity if they chose a HQ location i that is different from their initial entry location n. The productivities $\boldsymbol{z}$ are thus a product of some baseline productivity $z_{0,i}$ that is drawn upon entry, and the parameter c if $i \neq n$.[32] This friction creates a link between the entry location and the headquarters location, while still allowing for sorting: As $c \to 1$, the entry cost is sunk, and firms sort independently of their initial entry location; as $c \to 0$, ex post sorting is impossible, and the headquarters location is determined by the initial entry location, before firms observe their HQ-level productivities. The loss of productivity expressed in c captures all kinds of advantages that entrepreneurs face when deciding on keeping their main location of operations in their home market, including familiarity with local economic conditions and regulations, as well as home-bias in preference that might affect the productivity of the firm's management. This structure allows to encompass a range of assumptions on entry and headquarters location by varying the parameter c.[33] Finally, I assume that the baseline productivities $z_{0,i}$ are drawn from an i.i.d Fréchet distribution with shape parameter $\eta > 1$.

To recap, the entry and headquarters-location block of the model introduces three parameters: entry cost f_e, the dispersion of HQ-level shocks η, and the productivity loss from moving headquarters away from the firm's entry location, c.

3.6 The tradable-goods sector

Tradable goods are produced under conditions of perfect competition and constant returns to scale, and are freely traded across locations. Firms produce using a Cobb-Douglas production function that combines different types of labor with skill-intensities

analytical convenience, and that this term is greater than one. Note that since $\sigma > 1$ and $\epsilon > 1$, it is always the case that $\frac{\sigma}{\sigma-1} - \frac{\epsilon}{1+\epsilon} > 0$. Also note that σ and ϵ remain constant throughout the counterfactual analysis.

[31] As I discuss later, this allows more variation in firm size across space, which is a key feature of the data.

[32] Specifically, $z_i = cz_{0,i}$ for market i if $i \neq n$, and $z_i = z_{0,i}$ if $i = n$.

[33] In Section 5 below I estimate c and find it to be closer to 0, implying a strong connection between headquarters locations and the initial entry locations.

$\alpha_{g,s}$, such that $\sum_{s=1}^{S} \alpha_{g,s} = 1$. I allow productivity to differ across locations, such that goods-producing firms in region i have a total factor productivity of $A_{g,i}$. Similarly to the services sector, firms can enter freely after paying f_e units of the local final good.[34]

This completes the set-up of the model. I now turn to characterize key properties of the equilibrium, focusing on the structure of wages and inequality.

4 Equilibrium characterization

4.1 Aggregation

A key property of the model is that it admits tractable aggregation despite the complex firm micro-structure. Specifically, as I show in detail in Appendix Section F, equilibrium can be expressed in terms of region-level variables, without keeping track of firm-level decisions. Firm-level decisions are then obtained as power functions of these regional variables and the exogenous firm productivity terms z and $\boldsymbol{\phi}$. The set of equilibrium regional variables is regional employment and wage indices – L_{is} and W_{is} – for all markets and skill-groups; the mass of service-sector entrants, service-sector headquarters and goods-producing firms in each market – $M_{e,i}$, M_i and $M_{g,i}$, respectively; aggregate revenues for i-headquartered service firms and goods-producing firms, R_i and $R_{g,i}$; the services price index P_i; and average productivity indices Z_i, capturing the average HQ-level productivity z_i of firms that place their headquarters in i. I summarize this property of the model in the following Lemma:

Lemma 1. *The model's equilibrium can be expressed as a set of non-linear equations that include as variables only region-level aggregates.*

Proof: see Appendix Section F.

4.2 Firm wages and inequality

I now turn to characterize wages and inequality. I begin with the structure of branch-level wages, as summarized in the following proposition:

Proposition 2. *Consider a firm headquartered in market i with a branch in market j, characterized by firm-level productivity z and branch-level productivity ϕ. The wages that*

[34] Note that since firms in this sector operate in the same labor markets as service firms, they also face the same firm-specific upward-sloping labor supply curves and earn rents due to labor market power. The free entry condition ensures that net profits are zero in equilibrium.

this firm pays for workers of type s admit a log-linear structure given by

$$\log w_{ijs}(z,\phi) = const_s + \underbrace{\frac{1}{\epsilon+1}\log\left(W_{js}^{\epsilon}L_{js}^{-1}\bar{P}_j^{\frac{\theta-1}{\theta}}\Upsilon_j^{\frac{1}{\theta}}\right)}_{\textit{Local market effect}} \quad \underbrace{+\frac{\chi}{\epsilon}\log\phi}_{\textit{Branch-specific shock}}$$

$$\underbrace{+\frac{\chi}{\epsilon}\frac{1}{1-\gamma\chi\theta}\log z}_{\textit{Fundamental firm productivity}} \quad \underbrace{+\frac{\gamma\chi}{\epsilon+1}\log\frac{\Lambda_i}{\Gamma_i}}_{\textit{HQ market effect}} \quad \underbrace{+\frac{\chi}{\epsilon}\log A_{ij}}_{\textit{HQ-branch frictions}},$$

with $\chi \equiv \frac{\frac{\epsilon}{\epsilon+1}\frac{\sigma-1}{\sigma}}{1-\frac{\epsilon}{\epsilon+1}\frac{\sigma-1}{\sigma}}$ *and where* Υ_j*,* Λ_i *and* Γ_i *are regional aggregates that capture the market-potential of operating a branch in* j*, the attractiveness of operating headquarters in* i*, and the cost of headquarters labor in* i*, respectively.*[35]

Proof: see Appendix Section F.

Proposition 2 highlights how wages in this model differ from more standard settings without multi-region firms. I discuss the five terms included in this wage decomposition in turn. The first term is the local market effect that exists in every spatial model - both competitive and monopsonistic - due to the segregation of labor markets across space. When the labor supply elasticity ϵ is infinite, this is the only term that remains, and the wage $w_{ijs}(z,\phi)$ collapses to the market wage index W_{js}.[36] The second term is the branch-specific effect, that exists in every model with wage differentiation across (single-establishment) firms, and depends on the branch productivity draw ϕ.

The other three components are unique to the current setting of multi-region firms. First, firms with higher fundamental productivity z pay higher wages across all of their branch locations. When the HQ-branch elasticity γ is zero, the effect of higher z is the same as higher branch productivity ϕ.[37] With a positive γ, the effect of higher z is amplified: the greater market size of more productive firms leads them to hire more HQ workers, raising productivity across all branches by a magnitude determined by γ. Second, conditional on z, being headquartered in a more attractive location i – either due to larger market size or cheaper HQ labor – also leads to more HQ hiring and higher branch productivity. This is the headquarters market effect, which disappears when $\gamma \to 0$.

[35]See Appendix Section F for a discussion of these terms. The market potential term is given by $\Upsilon_j \equiv \left(E_j^{\frac{1}{\sigma}}P_j^{\frac{\sigma-1}{\sigma}}C_j^{-(1-\psi)}\bar{P}_j^{-\psi\frac{\theta-1}{\theta}}\right)^{\frac{\theta}{\psi}}$ with $\psi \equiv 1-\frac{\sigma-1}{\sigma}\frac{\epsilon}{1+\epsilon}$; the cost shifter of branch-level labor is given by $C_j \equiv \prod_s\left(\alpha_s W_{js}^{-1}L_{js}^{\frac{1}{\epsilon}}\right)^{-\alpha_s}$; the cost shifter of headquarters labor is given by $\Gamma_i \equiv \prod_s\left(\rho_s W_{is}^{-1}L_{is}^{\frac{1}{\epsilon}}\right)^{-\rho_s}$; and regional headquarters attractiveness is given by $\Lambda_i \equiv \tilde{\iota}\left(\sum_{j=1}^{N}A_{ij}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}}\Upsilon_j\right)^{\frac{1}{\iota}}\Gamma_i^{-\frac{1-\iota}{\iota}}$, where $\iota \equiv 1-\gamma\frac{\epsilon}{\epsilon+1}\frac{\sigma-1}{\sigma}\frac{\theta}{\psi}$ and $\tilde{\iota}$ is a constant that subsumes various parameters.

[36]With a finite ϵ, it decreases in labor supply L_{js} since there is less labor market competition for workers; increases in the market potential from opening branches Υ_j; and decreases in the local price index $\bar{P}_j$, which prices the fixed cost of opening branches.

[37]This can be seen by the same coefficient χ/ϵ when $\gamma = 0$.

Finally, the bilateral frictions captured by A_{ij} also affect the branch productivity and therefore its wages.

Proposition 2 also rationalizes the empirical findings in Section 2 and related findings in the labor literature on wage setting in multi-region firms. Through the firm-level productivity z and HQ-location effect i, it provides a rationale for why establishments of the same firm pay similar wages across all of the local labor markets in which they operate, as documented in Hazell et al. (2021). This is also consistent with the large role of inequality *across firms* in accounting for inequality *across establishments*.[38] The proposition also claims that part of the firm effect that is common to all of its branches derives from the location of its headquarters. This result is consistent with findings in the literature on wage setting in multinational firms, in which affiliate wages are increasing with the average wage in the headquarters country of the firm, as in Setzler and Tintelnot (2021). As part of the model estimation in Section 5, I provide evidence for the quantitative importance of this channel, and show that headquarters locations account for between 30%-50% of the wage variance across multi-region firms.

Another important aspect of Proposition 2 is that the wage distribution depends on the spatial equilibrium through two channels: the establishment location j and the headquarters location i. Therefore, changes in the spatial distribution of economic activity can affect not only inequality across regions, but also inequality within any region. In addition, the wage distribution is shaped by the full HQ-branch network, highlighting the importance of solving for firm headquarters and branch locations when studying changes in inequality.

Proposition 2 focuses on branch-level wages. I now turn to characterizing the relationship between headquarters wages and branch wages, as stated in the following:

Proposition 3. *Firm size and inequality between headquarters and branches:*

a) Across firm types: consider two firms with headquarters in location i and HQ-level productivities z_i' and z_i such that $z_i' > z_i$. Assume that both are active in market j and have the same branch-level productivity shock ϕ_j. Then the ratio of headquarters wage to the wage in branch j for each skill type s is higher in firm z_i' than in firm z_i.

b) For a given firm type: Let $r_{ij}(z_i, \phi_j)$ be firm revenues in some market j. For each skill type s, the ratio of headquarters wages $w_{is}^{HQ}(z_i)$ to branch wages in the firm's headquarters market i, $w_{iis}(z_i, \phi_i)$, is increasing in the ratio of total expected firm

[38] This is true even after controlling for the local labor market effects j that account for the fact that firms open establishments in different markets. In fact, the rise in variance across commuting zones accounts for only a very small part of the overall increase in inequality. Accordingly, the role of inequality between firms in Section 2 cannot be explained purely by differences in the location choices of firms for their establishments.

revenues across all locations, $r_i(z_i) \equiv \mathbb{E}_\phi\left[\sum_j r_{ij}(z,\phi_j)\right]$*, to local revenues in the headquarters market* $r_{ii}(z_i,\phi_i)$*.*

Proof: see Appendix Section F.

The first part of Proposition 3 states that other things equal, firms with higher z_i have larger wage differentials between headquarters and branches. These firms serve more locations, and accordingly have a greater incentive to hire headquarters-level labor to improve the firm's technology across all these branches. Therefore, the marginal product of headquarters labor rises relative to that of workers employed at the individual branch. Using the same logic, the second part of Proposition 3 states that for a given firm productivity z_i, HQ-branch differentials rise as the firm expands into more markets, captured by the greater ratio between total firm sales and local firm sales at its headquarters market.

4.3 Inequality across worker types and spatial segregation

The model also yields predictions for inequality across worker types and how it varies over space, as summarized in the following:

Proposition 4. *Suppose that* s *is the group of high-skilled workers.*

a) The economy-wide payroll-share of high-skilled workers is independent of the spatial distribution of economic activity, and given by

$$\frac{\alpha_s + \rho_s\gamma + \alpha_{g,s}\frac{\sigma}{\sigma-1}\frac{1-\beta}{\beta}}{1+\gamma+\frac{\sigma}{\sigma-1}\frac{1-\beta}{\beta}}.$$

It increases in the share of services β *when services as a whole are more skilled-intensive than production of tradable-goods:* $\frac{\gamma}{\gamma+1}\rho_s + \frac{1}{\gamma+1}\alpha_s > \alpha_{g,s}$*; and it increases in the HQ-branch elasticity* γ *when headquarters production is more skill-intensive than branch production:* $\rho_s > \alpha_s$*.*

b) The payroll-share of high-skilled workers in some region i *depends on its specialization in providing headquarters services, as summarized by the ratio of total sales by locally-headquartered firms,* R_i*, to domestic expenditure on services,* E_i*.*
A sufficient condition for it to rise with R_i/E_i *is that headquarters are sufficiently skill intensive:* $\rho_s > \alpha_{g,s} + \frac{1}{\gamma}(\alpha_{g,s}-\alpha_s)$*; and that the productivity loss from reallocation of headquarters is small (*$c \to 0$*).*

c) In the limit economy with no multi-region service firms, i.e. $A_{ij} \to 0\ \forall i \neq j$*, there are no differences in the payroll-share of high-skilled workers across space.*

Proof: see Appendix Section F.

Proposition 4 highlights the importance of multi-region firms for the distribution of wage income across skill groups. The ability to operate branches in other markets allows certain local markets (like Seattle or New York City) to specialize in providing headquarters services. When a local market i hosts many or particularly large headquarters, the revenues by all i-headquartered firms, R_i, are large relative to local spending on services, E_i. If headquarters activity is very skill-intensive (high ρ_s) – as in the data – it yields a high payroll share for skilled labor in market i. This result gives a natural explanation for the empirical co-location of headquarters activity and skilled labor, and for the rise in the skill premium and the relative supply of skilled labor in local labor markets that specialize in providing headquarters services.[39] In the limit economy with no multi-region firms, the high-skilled payroll share is equalized across all labor markets. More generally, a reduction in barriers to firm expansion (higher A_{ij} for $i \neq j$) increases the scope for specialization and the resulting spatial differences in compensation for skilled workers.

4.4 Additional features of the equilibrium

I now briefly mention several additional, distinctive features of the equilibrium that arise from the structure of multi-region firms in the model.

Gravity for provision of services. Total revenues of i-headquartered firms from branches in market j admit a log-linear gravity structure. They can be represented as a sum of a headquarters-location fixed effect, a branch-location fixed effect, and the bilateral frictions A_{ij}.[40] As I discuss in Section 5, this gravity structure provides a good fit for bilateral HQ-branch flows across local labor markets, when A_{ij} is parameterized as a function of distance. This pattern is reminiscent of the widely-studied gravity structure for multinational firms in the international context.

Trade imbalances. Inter-regional trade in goods is often characterized by large imbalances. The model rationalizes this pattern through cross-region flows in (within-firm) headquarters services. These flows are typically unobserved, but can be large in an economy with meaningful activity of multi-region firms, and rise as firms expand in space. The balance of traded goods – regional expenditure minus revenues – is given in the model by

$$\underbrace{\frac{1-\beta}{\beta}E_i - R_{g,i}}_{\text{Traded goods balance}} = \underbrace{\frac{\psi}{\theta}\left(R_i - E_i\right)}_{\text{HQ payments balance}} + \underbrace{\left(f_e \bar{P}_i M_{e,i} - \frac{\psi\iota}{\theta}R_i\right)}_{\text{Firms' profit transfers}},$$

[39]Note that for this effect to take place, the model does not require for services as a whole to be more skill-intensive than the goods sector, but only the tradable part of value-added that crosses regional borders.

[40]Specifically, they are given by $R_{ij} = \frac{A_{ij}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}}\Upsilon_j}{\sum_{j'=1}^{N} A_{ij'}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}}\Upsilon_{j'}}R_i$, where recall that Υ_j captures market potential for location j, and R_i captures total revenues by i-headquartered firms.

such that net-importers of tradable-goods are net-exporters of headquarters services, and vice versa.[41]

Headquarters sorting. Recall from Equation (7) that firms choose their headquarters location to maximize expected profits. As I detail further in Appendix F, the probability that a firm chooses some location i is summarized by the regional HQ-attractiveness shifter Λ_i.[42] This shifter reflects large market size (high market potential term Υ_i); high connectivity to other large markets (high A_{ij}s vis-à-vis other markets with high Υ_j); and high supply of labor that is intensive in HQ-level production. This pattern aligns with results from empirical studies of HQ location decision, e.g., Bel and Fageda (2008).

4.5 Extension: unobserved heterogeneity and worker screening

In the baseline version of the model, workers get different wages conditional on location and skill group due to their idiosyncratic preferences for employers. In Appendix Section H, I develop the model with alternative micro-foundations for such differences in wages, focusing on unobserved worker productivity. Following Helpman et al. (2010), I assume that workers differ in their productivity, and that it is imperfectly observed by firms. Firms can screen higher productivity workers in a frictional labor market subject to a convex screening cost, and wages at any given branch or at the headquarters are determined by the marginal surplus of the firm from employing another worker there.[43] Production in each branch and in the firm's headquarters depends on the number of hired workers and on their average productivity, reflecting complementarities across workers with different abilities. In Appendix Proposition A.1, I show that the main results about the wage structure of multi region firms and the connection between firm scope and within-firm inequality (Propositions 2 and 3) continue to hold. Furthermore, I show that in such an environment, headquarters are endogenously more skill-intensive than branches, and that this gap grows as firms expand in space (See Proposition A.2). I choose the version with worker idiosyncratic preferences as the baseline model for the quantitative analysis due to its greater tractability.

[41] The constants ψ and ι subsume various model parameters. The third source of bilateral flows in this expression – $\left(f_e \bar{P}_i M_{e,i} - \frac{\psi\iota}{\theta} R_i\right)$ – is due to shifting of firm profits across space due to differences between entry locations and headquarters locations. However, in the quantified model, the net-flows that arise from this part are small.

[42] Specifically, it is given by $\frac{\left(c^{\zeta \mathbb{I}_{n\neq i}}\Lambda_i\right)^{\frac{\eta}{\zeta}}}{\sum_{m=1}^{N}\left(c^{\zeta \mathbb{I}_{n\neq m}}\Lambda_m\right)^{\frac{\eta}{\zeta}}}$, where ζ is a constant that subsumes various parameters of the model, reflecting the elasticity of firm profits to changes in Λ_i; and recall that c is the loss of firm productivity when choosing HQ location i that is different from the entry location n.

[43] This assumption on wage determination is obtained as a solution to the "Rolodex" game in Brügemann et al. (2019), which corrects the Stole–Zwiebel bargaining protocol from Stole and Zwiebel (1996). See Appendix H, Helpman et al. (2010) and Brügemann et al. (2019) for additional details.

5 Model Quantification

I now turn to estimate the model and use it to quantify the importance of firm expansion for the overall rise in wage inequality. To this end, I match key features of the U.S. economy in 1980. In Section 6, I will subject this baseline economy to a series of shocks that reflect changes in the economic environment between 1980-2017. I define a region in the model as a 1990 labor market area (LMA) of the contiguous United States, yielding $N = 391$ regions.

5.1 Adding regional amenities and housing

I first enrich the model with two additional components that help to discipline the baseline equilibrium and are common in spatial equilibrium models. First, agents choosing location i now also enjoy an exogenous amenity component B_i that enters multiplicatively in their utility function.[44] This allows me to match exactly the spatial distribution of wages and employment in 1980.

Second, I add local housing, which introduces a second dispersion force on top of the idiosyncratic preference shocks. Agents spend a share δ of their income on housing and a share $1 - \delta$ on the composite of goods and services. I assume a constant housing supply elasticity ϱ and that the rights to housing rents are owned by immobile absentee landlords who have a similar structure of preferences for consumption as workers.[45]

5.2 External calibration of the households block

I divide the model's parameters into three groups. The first group includes parameters that govern household preferences and the housing market. This group consists of commonly-used building blocks from the literature and I rely on existing estimates to calibrate it. The second group includes regional productivity and amenity fundamentals, which I invert from the model's equilibrium conditions. The final group includes parameters from the production block of the model, which I estimate using Simulated Method of Moments (henceforth SMM).

[44] In this case, the probability that an agent of type s chooses location i is $\frac{L_{is}}{\sum_{j=1}^{N} L_{js}} = \frac{\left(B_i W_{is}/\bar{P}_i\right)^{\xi}}{\sum_{j=1}^{N}\left(B_j W_{js}/\bar{P}_j\right)^{\xi}}$, where $\bar{P}_i$ is the price index of the local final good.

[45] The only change in the set of equilibrium conditions following the introduction of housing is that the aggregate price index, previously given by P_i^{β}, is now equal to

$$\bar{P}_i = \left(\frac{\delta}{\beta\left(1-\delta\right)}\right)^{\frac{\delta}{(1+\varrho)-\delta\varrho}} P_i^{\beta\left(1-\frac{\delta}{(1+\varrho)-\delta\varrho}\right)} E_i^{\frac{\delta}{(1+\varrho)-\delta\varrho}}.$$

I.e., conditional on the price of services P_i, local cost of living is increasing in expenditure E_i due to higher cost of housing, where recall that E_i is the local expenditure on services. The exponent on local expenditure is higher when agents spend more on housing (higher δ) or when housing supply is relatively inelastic (lower ϱ).

I begin with the external calibration of the household block. I calibrate the expenditure share on services (β) to match the share of services value-added in national accounts. I set the dispersion of idiosyncratic shocks across regions (ξ) to 2.8, in line with the range of values in the trade and spatial literature, e.g. Galle et al. (2017). I set the dispersion of idiosyncratic shocks across employers (ϵ) to 5.0, matching recent estimates of the average wage markdown in Lamadon et al. (2022), Berger et al. (2022) and Azar et al. (2019), which is given in the model by $\epsilon/(1+\epsilon)$. The elasticity of substitution across varieties σ is set to 5.0, to match a price markup of 25% over marginal cost. This is also in line with the range of existing estimates on substitution between goods in the trade literature, e.g. Costinot and Rodríguez-Clare (2014). The expenditure share on housing is set to 0.24, following Davis and Ortalo-Magné (2011). Finally, I set the local housing supply elasticity to the population-weighted estimate of 1.75 in Saiz (2010). Table 6 summarizes the calibration of this part of the model.

Table 6: Calibration of the households block of the model

Parameter	Interpretation	Source	Value
ξ	Dispersion of location preference shocks	Galle et al. (2017)	2.8
ϵ	Dispersion of employer preference shocks	Lamadon et al. (2022), Berger et al. (2022)	5.0
σ	EoS between varieties	Costinot and Rodríguez-Clare (2014)	5.0
β_k	Sectoral expenditure shares	Direct computation - BEA NIPA (1980)	[0.64,0.36]
δ	Housing expenditure shares	Davis and Ortalo-Magne (2011)	0.24
ϱ	Housing supply elasticity	Saiz (2010)	1.75

I also compute the skill elasticities in production directly from the data. I set $S = 2$ and choose the skill groups to represent workers with and without a college degree. I compute the skill elasticities $\bar{\alpha} \equiv \{\alpha_s, \rho_s, \alpha_{g,s}\}_{s=1}^{S}$ according to the respective shares of total compensation in the data. To this end, I use the BEA KLEMS Integrated Industry-Level Production Account (KLEMS) statistics[46] that provide sectoral accounts of the distribution of labor compensation by college attainment over 1987-2019. I extrapolate the time series backward to get values for 1980. The resulting compensation shares for college graduates out of total compensation in the headquarters bundle (ρ_s), branch-level bundle (α_s) and goods-production ($\alpha_{g,s}$) are 0.36, 0.26 and 0.2, respectively.[47]

5.3 Model inversion

The set of location fundamentals $\mathcal{A} \equiv \left\{ \{A_{ij}\}_{j=1}^{N}, B_i, A_{g,i} \right\}_{i=1}^{N}$ includes bilateral productivity shifters in the services sector, regional amenities, and regional productivity in

[46] https://www.bea.gov/data/special-topics/integrated-industry-level-production-account-klems.

[47] These are the values for 1980. In one of the counterfactuals, I consider a change in these parameters and update them to their 2017 level, given by 0.8, 0.58, and 0.44.

tradable-goods. I obtain $\mathcal{A}$ by inverting the model's equilibrium conditions, conditional on all other parameters.[48]

To recover A_{ij}, I first assume that it is a product of a unilateral productivity term A_i and symmetric bilateral frictions τ_{ij},

$$A_{ij} = A_i / \tau_{ij}, \tag{8}$$

such that $\tau_{ij} > 1$ for $i \neq j$ and $\tau_{ii} = 1$. The set of τ_{ij} captures frictions for operating branches away from the headquarters market. These frictions can be recovered in a similar procedure to the method proposed by Head and Ries (2001) to recover trade costs in the international trade literature. The main difference with respect to trade flows is that the flows of value added within firms - between their headquarters and branches - are typically unobserved. However, bilateral activity in the form of HQ-branch relationships is observed, since many datasets report firm linkages and ownership structures. Let M_{ij} be the measure of firms with headquarters in location i and a branch in location j. τ_{ij} can be computed by double-differencing this expression and applying the symmetry assumption:[49]

$$\tau_{ij} = \left(\frac{M_{ij} M_{ji}}{M_{ii} M_{jj}} \right)^{-\frac{1}{\frac{2\theta}{\psi} \frac{\sigma - 1}{\sigma}}}. \tag{9}$$

I implement this procedure using the Dun & Bradstreet data, which has two main advantages relative to the LBD in this context. First, in Dun & Bradstreet, every multi-establishment firm has a clear headquarters location, while in the LBD the headquarters location cannot be inferred for a large share of firms, especially in earlier periods.[50] Second, since the number of bilateral frictions increases exponentially with the number of locations, using the LBD data would require to restrict the number of regions in the model or the output from the model to avoid disclosure risk. The Dun & Bradstreet data is well-suited for this purpose, since it puts great emphasis on documenting firm linkages, and since the spatial distribution of firms and establishments is one of the dimensions in

[48] Therefore, I implement this inversion for every guess of the model's parameters when estimating the production block below.

[49] In the model, M_{ij} can be shown to equal

$$M_{ij} = \frac{\theta - 1}{\theta} \frac{\psi}{\bar{P}_j f} \frac{A_{ij}^{\frac{\theta}{\psi} \frac{\sigma-1}{\sigma}} \left(\frac{\Lambda_i}{\Gamma_i} \right)^{1-\iota} Z_i M_i}{\sum_{n=1}^{N} A_{nj}^{\frac{\theta}{\psi} \frac{\sigma-1}{\sigma}} \left(\frac{\Lambda_n}{\Gamma_n} \right)^{1-\iota} Z_n M_n} E_j.$$

Multiplying by M_{ji} and dividing by $M_{ii} M_{jj}$ yields an expression that depends only on τ_{ij} and model parameters.

[50] In 1980, inferring headquarters locations using the NAICS-55 establishments as described in Section 2.1 is possible for a smaller share of the overall sample relative to 2017. In addition, inferring it using the firm-level mailing address of firms is impossible since this data doesn't exist for the earlier decades.

which it compares well to administrative datasets.[51]

I implement this procedure for 1980 and 2017, in order to both recover the 1980 equilibrium, and to later consider a shock to the set bilateral frictions according to how they changed over this time period. One remaining concern is that the quality of the data has improved over time, which might obscure the inference of changes in τ_{ij} between 1980 and 2017. Therefore, I use a smoothed version of M_{ij} to recover τ_{ij}, by taking the fitted values $\hat{M}_{ij}$ from a Poisson regression of M_{ij} on headquarters-market i fixed effects, branch-market j fixed effects, and a flexible polynomial of distance, using the Poisson Pseudo-Maximum Likelihood estimator from Silva and Tenreyro (2006). This specification accounts for most of the variation in the raw M_{ij}, with a pseudo R-squared of 0.79.[52]

Figure 5 displays key features of the recovered frictions and how they changed over time. The left panel of the figure shows that the frictions are increasing in distance, and that this dependence wanes for very large distances. Second, there is a clear decline in the level of bilateral frictions across all distance levels, capturing improvements in transportation and communication technologies in recent decades. The right panel of the figure shows one potential source of such improvements: higher prevalence of bilateral air-travel, which increased travel opportunities between headquarters and branches. Conditional on distance, bilateral pairs of LMAs with a greater increase in bilateral air travel have also witnessed a greater decline in the recovered frictions τ_{ij}.[53]

I recover unilateral fundamentals $A_i, A_{g,i}, B_i$ by solving for the model's equilibrium with inversion of specific equilibrium conditions. I briefly discuss here the moments in the data that I match as part of this procedure, and leave additional details on the inversion for Appendix G. I obtain the set of unilateral productivity shifters in services A_i by exactly matching the aggregate revenues of all i-headquartered firms, R_i. These measures are obtained by attributing firm-level revenues for all service-sector firms in the Dun & Bradstreet data to their headquarters location, as implied by the model.[54] I recover the regional productivity in production of goods $A_{g,i}$ as a residual to match exactly the aggregate wage bill in each LMA from the BLS Quarterly Census of Employment and Wages (QCEW). Finally, I recover the set of amenities B_i by exactly matching the distribution of 1980 employment across regions.

[51] See Barnatchez et al. (2017) for additional details.

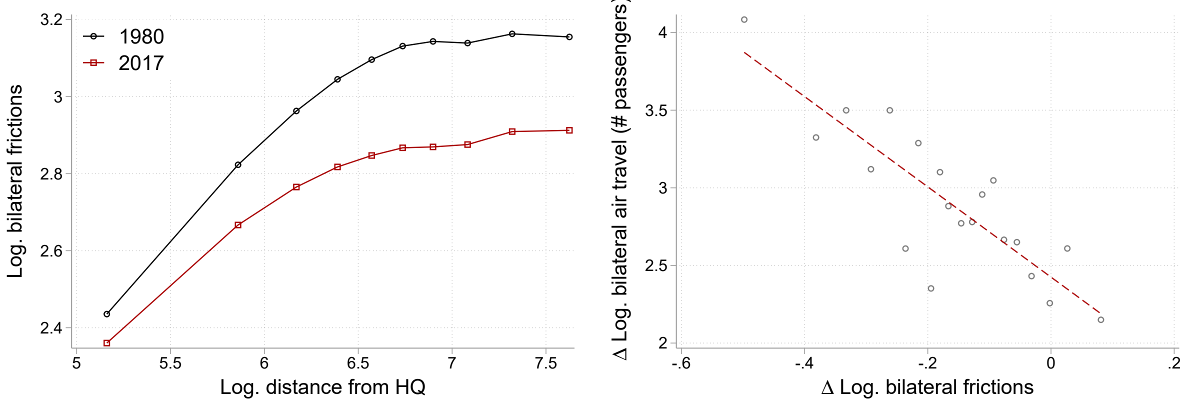
[52] Figure 14 in the appendix provides additional characterization of this procedure. First, it shows that conditional on headquarters-location and branch-location fixed effects, the number of bilateral linkages declines with distance. Secondly, it plots $\hat{M}_{ij}$ against M_{ij} to demonstrate the that the procedure of smoothing M_{ij} yields a tight fit.

[53] This pattern aligns with previous findings that emphasize the positive effect of air-travel on within-firm spillovers and productivity, e.g. Giroud (2013).

[54] On top of the advantage of having a clear headquarters location for every firm as discussed above, the Dun & Bradstreet data has an additional advantage for this exercise of providing revenue data for all firms in 1980. As shown in Appendix G, the aggregated R_i measures exhibit sensible behavior and strongly correlate with other regional aggregates from administrative data.

Figure 5: **The reduction in HQ-branch frictions (1980-2017)**

(a) Firm communication frictions against distance **(b)** Δ Communication frictions and air-travel



Note: Panel (a) shows a binscatter plot for log of the model-recovered bilateral within-firm communication frictions τ_{ij} against the log of distance between markets i and j for 1980 and 2017. Panel (b) shows a binscatter plot for the 1980-2017 changes in the log of bilateral air travel – growth in the log number of passengers in direct flights between each two markets – against the log-change in the bilateral frictions τ_{ij}, controlling for bilateral distance.

5.4 Simulated method of moments for the production block

The vector of production parameters $\Theta \equiv (\gamma, \theta, \eta, c, f, f_e)$ includes the elasticity of branch productivity to headquarters inputs, γ; dispersion of branch-level and firm-level productivity shocks, θ and η; productivity loss from reallocation of headquarters, c; and fixed production and entry costs, f and f_e. I estimate these parameters by matching key moments on firm organization and wage structure from Section 2. Below I provide brief intuition for which empirical moment helps to identify each of these parameters, though recall that in practice all moments are jointly determined in equilibrium by all of these parameters.

To facilitate the identification of f and f_e, I target the *average number of markets* per multi-region firm and the *average size of firms.* To facilitate the identification of γ, θ and η, I use the model-based decomposition of wages from Proposition 2. First, since η governs the dispersion of z_i, it relates directly to the role of firm-level wage effects, and therefore to the importance of *between-firm variance* of log wages. Second, γ affects the magnitude of market size effects: as $\gamma \to 0$, the HQ-market effect from Proposition 2 disappears. I therefore target the part of between-firm variance that is due to *variation between headquarters markets* (as opposed to variation across firms with headquarters in the same market). Third, θ affects the dispersion of branch-level wages within the firm, so I target the *overall variance of log wages.* Finally, the productivity loss from setting the headquarters away from the entry market, c, affects the allocation of headquarters across space and the responsiveness of the number of headquarters to the regional HQ-attractiveness shifter Λ_i, so I inform this parameter by the covariance of the mass of headquarters and average wages across space.

To estimate Θ, I minimize the loss function $\mathcal{L}(\Theta) \equiv \left(m(\Theta) - \tilde{m}_{1980}\right)' \boldsymbol{W} \left(m(\Theta) - \tilde{m}_{1980}\right)$ where $m(\Theta)$ is the vector of simulated moments from the model; $\tilde{m}_{1980}$ are the equivalent moments for 1980 in the data; and $\boldsymbol{W}$ is a weighting-matrix, which I set to be diagonal and inversely proportional to the squared values of $\tilde{m}_{1980}$, expressing the moments in percentage-deviation terms.[55]

Results from this estimation procedure are provided in Table 7. I estimate a value of 0.13 for the HQ-branch elasticity parameter γ. This value reflects a payroll-share of approximately 12% for headquarters-level labor, and as such aligns with the expenditure share implied by the distribution of firms payroll from Section 2.4. I also estimate a high value for the productivity loss from moving headquarters away from the initial entry location (captured by a low value for c), reflecting limited HQ sorting and important role for the initial firm entry location.

Table 7: Parameters for the production block of the model

Parameter	Interpretation	Estimate	Main targeted moment	Model	Data (1980)
θ	Shape parameter of branch-level shocks	1.3	Var. of log. wages, MR firms - overall	0.22	0.22
η	Shape parameter of firm-level shocks	5.53	Var. of log. wages, MR firms - between firms	0.12	0.11
γ	Elasticity of branch productivity to HQ inputs	0.13	Var. of log. wages, MR firms - between HQ markets	0.05	0.05
c	Productivity loss from moving HQ	0.18	Regional covariance of headquarters and wages	0.16	0.2
f	Fixed cost of operating a branch	8.31	Average number of markets per multi-region firm	5.04	5.02
f_e	Fixed entry cost	301	Average employment per multi-region firm	462	484

5.5 Characterization of the baseline equilibrium

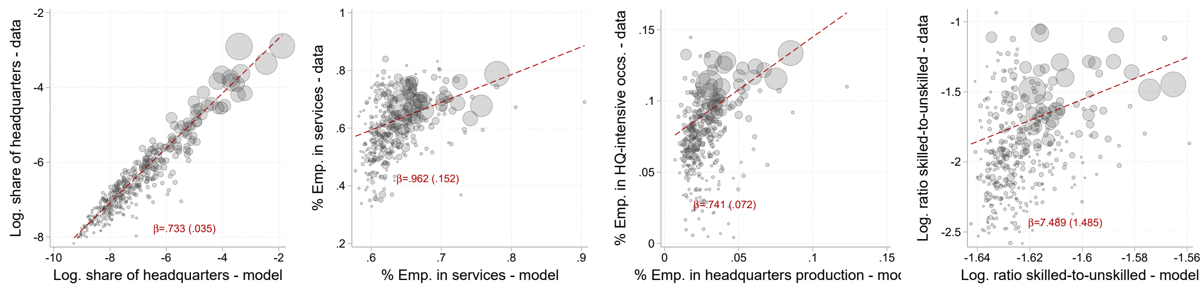
In this section, I report features of the baseline 1980 equilibrium, and show how the model compares to the U.S. economy in 1980 when addressing moments that were not targeted in the estimation procedure. I begin with the spatial distribution of economic activity. By construction, the model matches exactly the distribution of employment and average wages across space. Figure 6 shows the variation across space of other variables predicted by the model. Each subplot shows a single moment across all LMAs, in the model (x-axis) and in the data (y-axis). The size of each circle reflects total LMA employment. Panel (a) focuses on the distribution of headquarters across locations and reveals a tight fit between the model and the data. Panel (b) shows the regional share of employment in services sectors (out of total regional workforce). The corresponding measure in the data is computed from the Census Business Dynamics Statistics (BDS). The model yields too little specialization in tradable-goods for the small LMAs, but still the relationship is strong and positive. In particular, the model captures well the positive association between local market size and specialization in services.

[55]In practice, I employ the TikTak algorithm for global optimization from Arnoud et al. (2019) with 500 starting points, setting the Nelder–Mead method as the local minimizer.

Panel (c) shows the share of employment in headquarters tasks as a share of the total regional workforce. I choose as the corresponding measure in the data the employment share in the most HQ-intensive occupations, as computed in Section 2.3.2. In line with panel (b), larger markets tend to specialize in providing headquarters services, with greater employment share in the associated occupations. Finally, panel (d) shows the (log) ratio of skilled to unskilled labor across markets. In the model, the only reason for differences in this measure across space is specialization in providing headquarters services, as summarized in Proposition 4. All other reasons such as within-sector specialization, skill-specific amenities, and regional differences in the factor-bias of technology are excluded. Nevertheless, there is a strong and positive correlation for spatial skill-intensity between the model and the data. Quantitatively, this single mechanism can account for slightly over 20% of the differences in skill-intensity across space.

Figure 6: Regional characteristics in the baseline equilibrium

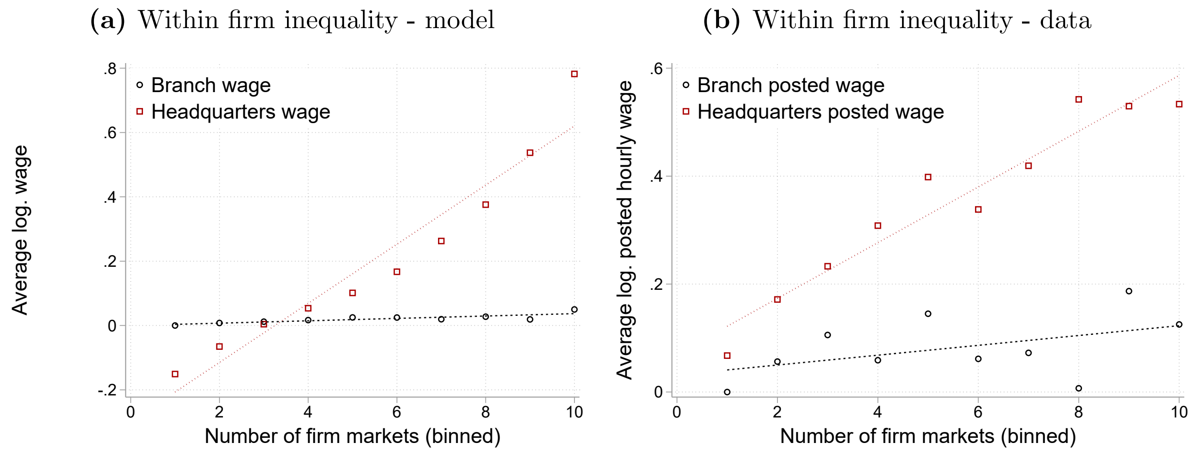
(a) Share of HQs **(b)** % Emp. in services **(c)** % Emp. in HQ prod. **(d)** Log. skill ratio



Note: Each subplot shows a single labor market characteristic across labor market areas (LMAs), in the model (x-axis) and in the data (y-axis). Circle size captures total employment in each labor market. Panel (a) shows the distribution of headquarters across locations. Panel (b) shows regional share of employment in services sectors. Panel (c) shows the share of employment in headquarters tasks. The corresponding measure in the data is the share of employment in the most headquarters-intensive occupations, as defined in Section 2.3.2, with the threshold chosen such that the total share of employment in headquarters (across all locations) is equal to its share in the model. Panel (d) shows the (log) ratio of skilled to unskilled labor across markets.

Another key feature of the model is its predictions for the distribution of wages within firms. Figure 7 shows how branch-level wages and headquarters-level wages are related to a firm's geographical scope. I cluster firms into 10 equal-sized bins that capture geographical scope (1 reflects the smallest number of markets; 10 reflects the largest number of markets). The left panel plots separately wages in branches and wages in headquarters of simulated firms from the model. The right panel plots an empirical equivalent using posted wages of multi-region service firms from Burning Glass Technologies. In both cases, wages rise with firm size, but this increase is much steeper for headquarters jobs, and the HQ-branch wage gap is higher for large firms - in line with Proposition 3. Quantitatively, this gap is somewhat larger in the model for the highest bin (around 70% difference in the model, relative to around 40% in the data), but this might also reflect the fact that job postings in the data capture only part of expected worker compensation, especially in HQ jobs of very large firms. Overall, the model captures well the structure of wages in multi-region firms and the relation of HQ-branch wage dispersion to firm size.

Figure 7: Within-firm inequality in the model and in the data



Note: Panel (a) plots average simulated log. wages from the model against 10 equally-sized bins that capture the geographical scope of firms, with bin number 1 representing firms that operate in the smallest number of markets and bin number 10 representing firms that operate in the largest number of markets. The red line captures average headquarters wages and the black line captures average branch wages. Panel (b) repeats this exercise with posted wages of multi-region service firms from Burning Glass, where a job posting is classified under "headquarters" if: (a) it is hired in the headquarters market of a multi-region firm; and (b) its occupational headquarters intensity as defined in Section 2.3.2 is above a threshold that guarantees the same overall share of HQ jobs in the data and in the model.

6 Counterfactual analysis of firm spatial expansion

I now turn to evaluate how different shocks to the baseline 1980 equilibrium affect the evolution of firm structure and inequality. I focus on three main shocks that capture key changes in the economic environment between 1980-2017. As described below, these shocks can be measured directly from the data, without a full re-estimation of the model for 2017. In particular, I measure them without using data on the evolution of the wage distribution.

The first shock that I consider is a change in the ability of firms to expand in space, as captured by the bilateral frictions τ_{ij} which were recovered in Section 5.3.[56] Recall that a reduction in these frictions captures improvements in transportation and telecommunication infrastructure, as well as the removal of regulations that impede firms from operating branches across multiple markets. Rather than feeding in the full matrix of changes in

[56]Note that while I used the market-clearing block of the model to recover these shocks, it did not require estimation of the full model or its parameters. I used no data on wage inequality in this procedure, and these shocks reflect only changes in the geographical structure of firms over time, as measured by changes in the number of branches per firm across markets in Equation 9.

τ_{ij} that reflects both a common trend and idiosyncratic changes, I lower every τ_{ij} by the average recovered change across all location pairs.[57]

Next, I consider a change in households preferences in the form of higher expenditure share on services (β), motivated by the large reallocation of economic activity from goods-producing to services-producing sectors over this time period ("structural transformation"). I measure an increase in β from 64% to 79% in national accounts.[58] On top of the realistic nature of this change, it has direct implications for the expansion of firms, since the payoff of opening an additional branch is higher relative to the cost of doing so in an environment with more demand for services.[59] Intuitively, when demand for services is particularly low (β close to 0), firms have no incentive to pay the fixed cost f to operate branches.

Finally, I consider a shock that reflects homogenous skill-biased technical change (henceforth SBTC). Recall that the skill-intensities $\alpha \equiv \{\alpha_s, \rho_s, \alpha_{s,g}\}_{s=1}^{S}$ are constant by construction in the model. However, in the data, they rise over time for the group of college graduates. This change has been studied extensively as a driver of wage inequality in the SBTC literature, and therefore is a natural shock to investigate.[60]

I now turn to study the effect of these shocks on key moments in the model, with a focus on inequality across firms and across local labor markets. I summarize the results in Table 8. The upper panel of the table displays changes in levels following each shock – or combination of shocks – relative to the 1980 equilibrium, and the lower panel displays changes in terms of percentages of the respective empirical trends over 1980-2017. I start by analyzing each shock separately, as well as the effect of interactions between pairs of shocks (Columns 1-6 in Table 8). This allows me to provide intuition for the sign and magnitude of the effects of each shock and asses their relative importance, and to connect to the analytical results from Section 4. I then conclude with the joint effect of all three shocks combined (Column 7).

The effects of lower frictions to expansion (τ). I begin with the effect of a decline in τ_{ij} in Column (1). As a result of this change, firms expand in space: the number of markets served by the average firm rises by 0.3 log points (equivalent to over

[57]Note that since τ_{ii} equals to unity for all i, this shock captures a common reduction in the cost of operating branches away from the headquarters location, and it does not constitute a positive productivity shock across all firm's markets.

[58]In the growth and development literature on structural transformation, the change in β is endogenized, e.g. using non-homothetic preferences or non-unitary elasticity of substitution between consumption of goods and services. Since this is not the main subject of the paper, I retain the Cobb-Douglas preferences structure and consider a simple exogenous change in β. An interesting extension of the current study would be to study the interaction between richer preference structure and spatial firm expansion.

[59]Recall that the fixed cost of opening a branch is denominated in terms of the local final good, both of which do not scale proportionally with β as the demand for services does.

[60]In this exercise, I study how firm spatial expansion interacts with SBTC using the structure of the model: first by evaluating the effect of higher α in the current setting, and secondly by comparing it to the effects of lower τ. In Section 7 below, I provide additional reduced-form analysis of how the mechanism in this paper differs from SBTC.

100% of the empirical change). Inequality also rises: the variance of log wages across establishments[61] rises by 0.025 points (18% of the change in the data). Roughly half of this change is within firms, in line with results from Proposition 3. Inequality also rises between regions, capturing 21% of the increase in variance of log wages across local labor markets in the data. These spatial disparities take a particular form in the data and in the model: wage growth is larger in more populated labor markets, especially skilled labor, captured by the regression slopes of (log) average wages and college premium on (log) regional employment.

Table 8: Model counterfactuals - shocks to the 1980 equilibrium

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Changes in levels						
	Data: Δ1980-2017	**$\Delta\tau$**	**$\Delta\beta$**	**$\Delta(\tau,\beta)$**	**$\Delta\alpha$**	**$\Delta(\tau,\alpha)$**	**$\Delta(\tau,\beta,\alpha)$**
	Cross-section of firms						
Log number of markets per firm	0.246	0.300	0.038	0.182	0.029	0.298	0.181
Variance of log wages: multi-region service firms	0.140	0.025	0.008	0.055	0.023	0.046	0.074
Variance of log wages: within-firms	0.070	0.015	0.021	0.045	0.031	0.035	0.055
	Cross-section of labor market areas						
Variance of log regional wages	0.034	0.007	0.003	0.022	0.002	0.012	0.038
Regression slope: log wage on log size	0.060	0.005	-0.016	0.011	0.005	0.017	0.025
Regression slope: log skill-premium on log size	0.052	0.011	0.009	0.012	0.003	0.023	0.025
Regression slope: log skill-ratio on log size	0.051	0.014	0.010	0.013	0.005	0.028	0.029
	% of changes in data (1980-2017)						
	Cross-section of firms						
Log number of markets per firm	100%	122%	16%	74%	12%	121%	74%
Variance of log wages: multi-region service firms	100%	18%	6%	39%	17%	33%	53%
Variance of log wages: within-firms	100%	21%	30%	64%	44%	50%	79%
	Cross-section of labor market areas						
Variance of log regional wages	100%	21%	10%	65%	7%	35%	112%
Regression slope: log wage on log size	100%	8%	-26%	19%	8%	29%	41%
Regression slope: log skill-premium on log size	100%	22%	17%	23%	7%	44%	48%
Regression slope: log skill-ratio on log size	100%	28%	20%	26%	10%	56%	57%

Note: this table shows changes in key moments from the model following a shock to the baseline 1980 equilibrium. The upper panel includes changes in levels, and the lower panel includes percentage changes relative to the 1980-2017 empirical trends. Column (1) shows changes in the data. Column (2) considers changes in the model following a homogenous decline in the HQ-branch frictions τ according to the average decline recovered in Section 5.3. Column (3) considers an increase in the aggregate household spending on services β. Column (4) considers the joint effect of a change in τ and a change in β. Column (5) considers the effect of changes in the skill-intensities in production α, and Column (6) considers the joint effect of changes in τ and α. Finally, Column (7) considers the joint effect of all three shocks.

To gain more intuition into the effects of lower τ, I investigate how the spatial equilibrium responds to small changes that reveal the first order effects of such a shock. Figure

[61] Recall that an establishment in the model is either the combined headquarters of the firm or the representative branch in a single labor market area.

8 plots the resulting changes in key regional aggregates against ex-ante specialization in supplying headquarters services (log of R_i/E_i in the model). This measure is an important sufficient statistic for the regional impact of changes in τ. Locations with high specialization in services gain the most from this change: firms in these regions experience a market size effect and expand into new markets, yielding greater demand for headquarters services in their home location. Each circle represents a labor market area (weighted by 1980 employment), with the rightmost circle capturing the New-York City LMA (henceforth NYC), which is revealed to be particularly specialized in services. The left panel of the figure shows the impact on local wages and prices. Locations like NYC experience higher wage growth, greater decline in the price index of services (due to increased variety), and an increase in housing prices. The middle panel decomposes average wage growth into its different components in the model. In line with the model's logic, headquarters wages rise the most in locations like NYC, following the elevated demand from firm expansion; and wages in the tradable-goods sector experience a relative decline. Finally, the ratio of skilled to unskilled wages rises in these locations since headquarters activity is relatively skill-intensive, as can be seen in the right panel of that figure. Overall, these changes are consistent with key patterns in the rise of spatial disparities in recent decades, and provide intuition for the spatial effects of lower τ in Table 8.

The effects of higher expenditure on services (β). Higher β yields quite similar effects to a reduction in the spatial frictions τ (Column 3 in Table 8). To see this, recall from Proposition 4 that it raises the aggregate skill-premium.[62] Therefore, the price of headquarters labor rises relative to branch-level labor, leading to higher inequality within firms. In addition, inequality rises across space, with wages again going up the most in locations like NYC that specialize in services.[63] One key difference relative to the reduction in τ is that the urban wage premium (the regression slope of log regional wage on log size) goes down. This difference comes from the fact that higher β makes locations with greater variety of services – typically large markets – relatively more attractive,[64] resulting in greater supply of labor following an influx of workers from other markets.

Higher expenditure on services also amplifies the impact of lower spatial frictions τ when both shocks are considered together (Column 4 in Table 8). There are two reasons for this amplification. First, in this scenario, HQ wages rise both due to the market-size effect of firm expansion (since HQ labor is non-rival across branches) and due to the higher price of skill (since HQ labor is skill-intensive). Second, the market size effect of firm expansion is greater when there is more demand for services; at the extreme case of

[62]This follows from the fact that services as a whole are more skill-intensive than tradable-goods in the data.

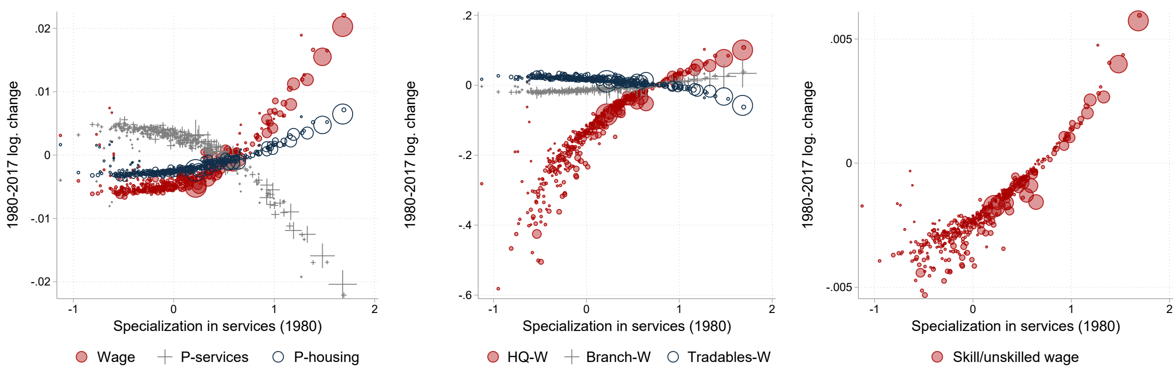
[63]See Appendix Figure 19 for a replication of Figure 8 with a small change in β.

[64]Recall that without housing, the regional price index is given by P_i^β. Therefore, the cost of living changes across space even before general equilibrium effects take place. This change can be thought of as an increase in amenities in locations with ex-ante lower P_i.

$\beta \rightarrow 0$, a decline in τ has no effect. Jointly, these shocks generate a greater increase in wage inequality than their individual effects – across firms and space.

The effects of a homogeneous SBTC shock (α). The main effect of the SBTC shock (Column 5 in Table 8) is to increase inequality within firms, since it boosts the price of the skill-intensive HQ labor. However, the effects on other measures of inequality are milder relative to the previous two shocks. In addition, similar to higher β, it also amplifies the effect of declining τ (Column 6 in Table 8). The same logic applies in this case: the price of HQ labor rises due to high skill-intensity and due to the market size effect from greater expansion. As before, this effect is stronger in markets with ex-ante specialization in services like NYC. Overall, the joint effect of higher α and lower τ accounts for much of the rise in inequality across firms and across space. Combining all three shocks generates even larger effects (Column 7 in Table 8), accounting for most (74%) of the spatial expansion of firms; over half of the rise in inequality across and within firms; all of the rise in spatial inequality; and around half of the increase in the association between skill-intensity and market size.

Figure 8: The spatial impact of a small reduction in τ



Note: This figure shows regional changes in the model following a small reduction in communication frictions τ. The x-axis in all three subplots is regional specialization in services in the 1980 equilibrium, given by the log of E_i/R_i in the model. Each circle represents a labor market area and the size of the circles capture total LMA employment in the baseline equilibrium. The left panel shows log changes in the regional average wage index (solid red circles), regional price of housing (empty blue circles), and regional price index of services (shaded gray circles). The center figure shows log changes in the three components of the regional wage index: headquarters wage index (solid red circles), branches wage index (shaded gray circles), and tradable-goods wage index (empty blue circles). The right panel shows changes in the log ratio of skilled to unskilled wages.

To recap, through these counterfactuals, the model is able to jointly account for multiple trends in U.S. labor markets since the 1980s. It highlights a quantitatively significant role for the reduction in spatial frictions that impede firm expansion (τ). The role of these frictions is amplified when combined with greater demand for services or homogenous skill biased technical change.

7 Relationship to existing theories of rising inequality

In this section, I briefly discuss how the above model relates to other mechanisms discussed in the literature on wage inequality.

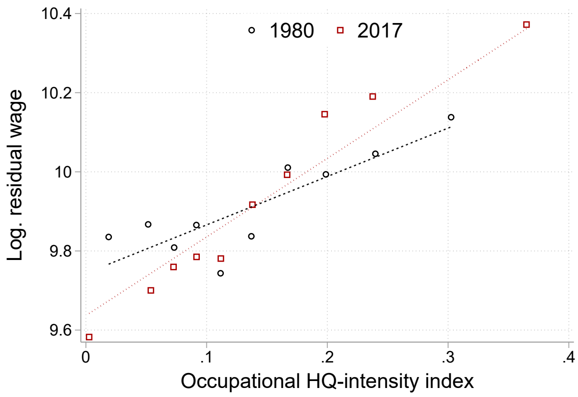
Skill-Biased Technical Change (SBTC). Tracing back to Katz and Murphy (1992), SBTC has been seen as a key driver of wage inequality. My model complements this literature in two senses. First, it provides a particular form of SBTC that aligns with recent changes in the spatial organization of firms. Therefore, it can speak to changes in the distribution of income across skill groups as summarized in Proposition 4. Second, the model provides a set of predictions that go *beyond changes in skill-premium*, and are quantitatively important for the overall rise in inequality. Inter alia, the model predicts that (a) *conditional* on worker skill and location, workers at headquarters-level jobs earn higher wages; (b) this relationship has a systematic connection to firm size, with greater HQ-branch gaps in larger firms; (c) this conditional headquarters wage premium has risen over time, given the increase in the geographical scope of firms.

I now provide further evidence in support of these additional predictions. I have already demonstrated that HQ-branch wage inequality is systematically related to firm size (see Figure 7), and that this pattern holds when conditioning on educational attainment (see the discussion in Section 2.3.2 and Appendix D). However, the data in Section 2 is not well-suited to demonstrate the importance of these predictions for rising inequality within skill groups over time, due to the lack of worker characteristics in the LBD, and the limited time span of the BGT data. I turn instead to evidence from individual-level Census data, focusing on dispersion in individual earnings from the 1980 Decennial Census and the 2015-2019 American Community Survey (henceforth ACS).[65] Figure 9 plots the average earnings of workers against the measure of occupational headquarter-intensity from Section 2.3.2, after *controlling for educational attainment*, interacted with worker commuting zone. The slope from a pooled regression of this residual wage on the HQ-intensity measure in 1980 is 1.26 (standard error of 0.027), and the slope has increased between 1980 and 2017 by a magnitude of 0.75 (standard error of 0.037). Therefore, in line with the model's predictions, we find that there is a clear positive relationship between wages and the likelihood that a job is performed in firm headquarters, even after controlling for location and educational attainment. In addition, this relationship has strengthened over the past decades, as would be expected in response to firms' spatial expansion.

To demonstrate that the rise in HQ-branch differentials could potentially play a meaningful role in the overall increase in inequality – in particular when comparing to the rise in the college premium – I perform a variance decomposition of individual earnings over

[65] I obtain these data from Ruggles et al. (2021).

Figure 9: The conditional headquarters premium



Note: This figure shows a binscatter plot of individual wages against the headquarters-intensity of their occupation, after controlling for individual location and educational attainment.

time. I estimate the wage regression

$$\log w_i = \underbrace{\beta^{HQ}\text{HQ}_{occ(i)}}_{\text{Occupational HQ-intensity}} + \underbrace{\text{College}_i + CZ_i + \text{College}_i \times CZ_i}_{\text{College attainment and commuting zone}} + \epsilon_i,$$

where $\log w_i$ is the log of individual i wage earnings; $\text{HQ}_{occ(i)}$ is the occupational HQ-intensity from Section 2.3.2 for individual i; College_i is an indicator variable for college attainment; CZ_i is a commuting zone fixed effect; and ϵ_i is a residual. We have already seen in Figure 9 that β^{HQ} is positive and has increased over time. Table 9 further shows that the increase in β^{HQ} can account for a large part of the increase in the variance of log individual earnings between 1980 and 2017, even when controlling for the joint effect of the growing college premium and increasing inequality across commuting zones. The variance of log individual earnings has increased by 0.17 points between 1980 and 2017, in line with the rise in inequality across establishments in Section 2.[66] The interacted indicators for college attainment and commuting zone account for 18% of this overall increase, while the component that captures the occupational HQ-intensity accounts for 46%. In addition, 20% of the overall increase is due to the covariance of these two terms, since high-skilled individuals in high-wage locations tend to work in headquarters-level jobs. To sum up, the distinction between HQ and branch-level jobs is quantitatively important for the rise in inequality between 1980-2017, even after accounting for differences in location and educational attainment.

Outsourcing. The literature on outsourcing has documented a decline in wages for outsourced workers relative to peers with similar characteristics who remain within the boundary of the firm – see e.g. Goldschmidt and Schmieder (2017). Therefore, the growth

[66] The similarity between the rise of inequality in the cross-section of establishments and in the cross-section of individuals has been noted in Barth et al. (2016).

Table 9: Variance decomposition of individual earnings

Component of log individual earnings	Δ Variance 1980-2017	% of total Δ
Commuting zone and college attainment	0.03	18%
HQ-intensity of the occupation	0.08	46%
Covariance	0.03	20%
Residual	0.03	16%
Total	**0.17**	**100%**

Note: this table shows a variance decomposition of individual wage earnings from the 1980 Decennial Census and 2015-2019 American Community Survey. Individual earnings are regressed on the interaction of commuting-zone and college attainment fixed effects, and on the continuous measure of occupational headquarters intensity from Section 2.3.2 of the paper. The table shows changes over time in the variance for the predicted part of earnings based on each of these components, and the residual change in variance.

of outsourcing has been proposed as a potential explanation for the rise in inequality. While outsourcing can raise wage dispersion across firms, it does not account for rising dispersion within firms. If anything, simple models of outsourcing would imply reduced within-firm inequality, following outsourcing of workers that do not take part in the firm's core activity. However, in the data, inequality rises within firms in parallel to its increase between firms. A mechanism that relies on the spatial expansion of firms can speak to these patterns.

Functional specialization. A closely related literature on functional specialization studies the separation of production into headquarters and plants across space,[67] providing a rationale for the observed increase in spatial segregation by skill. I emphasize the increase in the geographical scope of firms – conditional on the ability to separate branches from headquarters – which arises naturally in the context of non-tradable sectors. Such expansion can also generate a rise in segregation, along with other dimensions of inequality that are important in the data.

8 Conclusion

This paper argues that the spatial expansion of service firms in recent decades can explain multiple secular trends in the U.S. labor markets since the 1980s. I document that multi-region service firms have experienced substantial spatial expansion and account for most of the rise in wage inequality over 1980-2017. I link these trends by providing evidence on the uneven nature of firm expansion: firms become larger by opening branches in more markets, while hiring more skilled workers and paying higher wages in spatially-concentrated headquarters. I integrate this structure into a spatial general equilibrium with firm wage setting, in which firms hire branch-level workers across multiple local markets, and the output of headquarters workers is skill-intensive and non-rival across branches. The model generates distinctive labor market implications, including rising

[67] See e.g. Duranton and Puga (2005).

within-firm inequality as firms expand in space, and a link between firm expansion and the rise of spatial inequality and segregation. The estimated model can account for a substantial share of the observed rise in inequality along these dimensions. It also highlights an important role for the interaction between declining within-firm communication frictions and rising demand for services in generating the rise in inequality.

I highlight four main benefits of the focus on firm expansion in explaining the rise in inequality. First, my model generates distinctive predictions that are quantitatively important in the data and are not covered by existing theories in the literature on wage inequality, such as the growth of wage dispersion across establishments within firms. Second, it is able to simultaneously account for multiple trends in the U.S. labor market. Third, it aligns with the observed trends in the organization of production and the expansion of multi-establishment firms. Finally, it provides a natural explanation for the concentration of these trends in services, due to the non-tradable nature of output in these sectors.

More broadly, the model demonstrates the importance of multi-region firms and the network of HQ-branch linkages for many economic questions, including inequality, the propagation of shocks across space, and the effects of structural transformation. Another relevant topic is the interaction of this firm structure with globalization, since many of the major U.S.-based service firms have parallel operations overseas. Finally, the model offers a framework to investigate policies that shape the location decisions of these firms, including regional business and tax incentives to attract firm headquarters (e.g. as in the case of Amazon's second headquarters), and policies to mitigate the social costs of inter-region competition for these companies.

References

Aghion, Philippe, Antonin Bergeaud, Timo Boppart, Peter J Klenow, and Huiyu Li, "A theory of falling growth and rising rents," Technical Report, National Bureau of Economic Research 2019.

Alviarez, Vanessa I, Javier Cravino, and Natalia Ramondo, "Firm-embedded productivity and cross-country income differences," Technical Report, National Bureau of Economic Research 2020.

Argente, David, Sara Moreira, Ezra Oberfield, and Venky Venkateswaran, "Scalable Expertise," 2020.

Arkolakis, Costas, Natalia Ramondo, Andrés Rodríguez-Clare, and Stephen Yeaple, "Innovation and production in the global economy," *American Economic Review*, 2018, *108* (8), 2128–73.

Arnoud, Antoine, Fatih Guvenen, and Tatjana Kleineberg, "Benchmarking global optimizers," Technical Report, National Bureau of Economic Research 2019.

Atalay, Enghin, Ali Hortaçsu, and Chad Syverson, "Vertical integration and input flows," *American Economic Review*, 2014, *104* (4), 1120–48.

Azar, José, Ioana Marinescu, Marshall Steinbaum, and Bledi Taska, "Concentration in US labor markets: Evidence from online vacancy data," *Labour Economics*, 2020, *66*, 101886.

_ , Steven Berry, and Ioana Elena Marinescu, "Estimating labor market power," *Available at SSRN 3456277*, 2019.

Barnatchez, Keith, Leland Dod Crane, and Ryan Decker, "An assessment of the national establishment time series (nets) database," *Finance and Economics Discussion Series (FEDS)*, 2017.

Barth, Erling, Alex Bryson, James C Davis, and Richard Freeman, "It's where you work: Increases in the dispersion of earnings across establishments and individuals in the United States," *Journal of Labor Economics*, 2016, *34* (S2), S67–S97.

Bel, Germa and Xavier Fageda, "Getting there fast: globalization, intercontinental flights and location of headquarters," *Journal of Economic Geography*, 2008, *8* (4), 471–495.

Berger, David, Kyle Herkenhoff, and Simon Mongey, "Labor market power," *American Economic Review*, 2022, *112* (4), 1147–93.

Berry, Christopher R and Edward L Glaeser, "The divergence of human capital levels across cities," *Papers in regional science*, 2005, *84* (3), 407–444.

Brügemann, Björn, Pieter Gautier, and Guido Menzio, "Intra firm bargaining and Shapley values," *The Review of Economic Studies*, 2019, *86* (2), 564–592.

Cao, Dan, Henry R Hyatt, Toshihiko Mukoyama, and Erick Sager, "Firm growth through new establishments," *Available at SSRN 3361451*, 2017.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline, "Firms and labor market inequality: Evidence and some theory," *Journal of Labor Economics*, 2018, *36* (S1), S13–S70.

_ , Jesse Rothstein, and Moises Yi, "Location, location, location," *UC Berkeley*, 2021.

Costinot, Arnaud and Andrés Rodríguez-Clare, "Trade theory with numbers: Quantifying the consequences of globalization," in "Handbook of international economics," Vol. 4, Elsevier, 2014, pp. 197–261.

Davis, Morris A and François Ortalo-Magné, "Household expenditures, wages, rents," *Review of Economic Dynamics*, 2011, *14* (2), 248–261.

Derenoncourt, Ellora, Clemens Noelke, David Weil, and Bledi Taska, "Spillover effects from voluntary employer minimum wages," Technical Report, National Bureau of Economic Research 2021.

Diamond, Rebecca, "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000," *American Economic Review*, 2016, *106* (3), 479–524.

Ding, Xiang, Teresa C Fort, Stephen J Redding, and Peter K Schott, "Structural Change Within Versus Across Firms: Evidence from the United States," 2022.

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

Eckert, Fabian, "Growing apart: Tradable services and the fragmentation of the us economy," *mimeograph, Yale University*, 2019.

_ , Sharat Ganapati, and Conor Walsh, "Urban-Biased Growth: A Macroeconomic Analysis," Technical Report, National Bureau of Economic Research 2022.

Fort, Teresa C, "Technology and production fragmentation: Domestic versus foreign sourcing," *The Review of Economic Studies*, 2017, *84* (2), 650–687.

_ **and Shawn Klimek**, "The effects of industry classification changes on US employment composition," *CES Working Paper 18-28, US Census Bureau*, 2018.

Galle, Simon, Andrés Rodríguez-Clare, and Moises Yi, "Slicing the pie: Quantifying the aggregate and distributional effects of trade," Technical Report, National Bureau of Economic Research 2017.

Giannone, Elisa, "Skilled-biased technical change and regional convergence," *University of Chicago. Unpublished manuscript*, 2017.

Giroud, Xavier, "Proximity and investment: Evidence from plant-level data," *The Quarterly Journal of Economics*, 2013, *128* (2), 861–915.

_ **, Simone Lenzu, Quinn Maingi, and Holger Mueller**, "Propagation and amplification of local productivity spillovers," Technical Report, National Bureau of Economic Research 2021.

Goldschmidt, Deborah and Johannes F Schmieder, "The rise of domestic outsourcing and the evolution of the German wage structure," *The Quarterly Journal of Economics*, 2017, *132* (3), 1165–1217.

Haltiwanger, John C, Henry R Hyatt, and James Spletzer, "Industries, mega firms, and increasing inequality," Technical Report, National Bureau of Economic Research 2022.

Hazell, Jonathon, Christina Patterson, H Sarsons, and B Taska, "National wage setting," Technical Report, Working Paper 2021.

Head, Keith and John Ries, "Increasing returns versus national product differentiation as an explanation for the pattern of US-Canada trade," *American Economic Review*, 2001, *91* (4), 858–876.

Helpman, Elhanan, "A simple theory of international trade with multinational corporations," *Journal of political economy*, 1984, *92* (3), 451–471.

_ **, Oleg Itskhoki, and Stephen Redding**, "Inequality and unemployment in a global economy," *Econometrica*, 2010, *78* (4), 1239–1283.

Holmes, Thomas J, "The diffusion of Wal-Mart and economies of density," *Econometrica*, 2011, *79* (1), 253–302.

_ **and John J Stevens**, "An alternative theory of the plant size distribution, with geography and intra-and international trade," *Journal of Political Economy*, 2014, *122* (2), 369–421.

Hsieh, Chang-Tai and Esteban Rossi-Hansberg, "The Industrial Revolution in Services," 2022.

Jia, Panle, "What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry," *Econometrica*, 2008, *76* (6), 1263–1316.

Jiang, Xian, "Information and Communication Technology and Firm Geographic Expansion," Technical Report, working paper 2021.

Katz, Lawrence F and Kevin M Murphy, "Changes in relative wages, 1963–1987: supply and demand factors," *The quarterly journal of economics*, 1992, *107* (1), 35–78.

Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler, "Imperfect competition, compensating differentials, and rent sharing in the US labor market," *American Economic Review*, 2022, *112* (1), 169–212.

Moretti, Enrico, *The new geography of jobs*, Houghton Mifflin Harcourt, 2012.

Oberfield, Ezra, Esteban Rossi-Hansberg, Pierre-Daniel Sarte, and Nicholas Trachter, "Plants in space," Technical Report, National Bureau of Economic Research 2020.

Ramondo, Natalia and Andrés Rodríguez-Clare, "Trade, multinational production, and the gains from openness," *Journal of Political Economy*, 2013, *121* (2), 273–322.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Nicholas Trachter, "Diverging trends in national and local concentration," *NBER Macroeconomics Annual*, 2021, *35* (1), 115–150.

Rubinton, Hannah, "The geography of business dynamism and skill biased technical change," *FRB St. Louis Working Paper*, 2020.

Ruggles, Steven, Sarah Flood, Sophia Foster, Ronald Goeken, Jose Pacas, Megan Schouweiler, and Matthew Sobek, "Ipums usa: Version 11.0 [dataset]," *Minneapolis, MN: IPUMS*, 2021, *10*, D010.

Saiz, Albert, "The geographic determinants of housing supply," *The Quarterly Journal of Economics*, 2010, *125* (3), 1253–1296.

Setzler, Bradley and Felix Tintelnot, "The effects of foreign multinationals on workers and firms in the United States," *The Quarterly Journal of Economics*, 2021, *136* (3), 1943–1991.

Silva, JMC Santos and Silvana Tenreyro, "The log of gravity," *The Review of Economics and statistics*, 2006, *88* (4), 641–658.

Song, Jae, David J Price, Fatih Guvenen, Nicholas Bloom, and Till Von Wachter, "Firming up inequality," *The Quarterly journal of economics*, 2019, *134* (1), 1–50.

Stole, Lars A and Jeffrey Zwiebel, "Intra-firm bargaining under non-binding contracts," *The Review of Economic Studies*, 1996, *63* (3), 375–410.

Tintelnot, Felix, "Global production with export platforms," *The Quarterly Journal of Economics*, 2017, *132* (1), 157–209.

Tolbert, Charles M and Molly Sizer, "US commuting zones and labor market areas: A 1990 update," Technical Report 1996.

Appendices

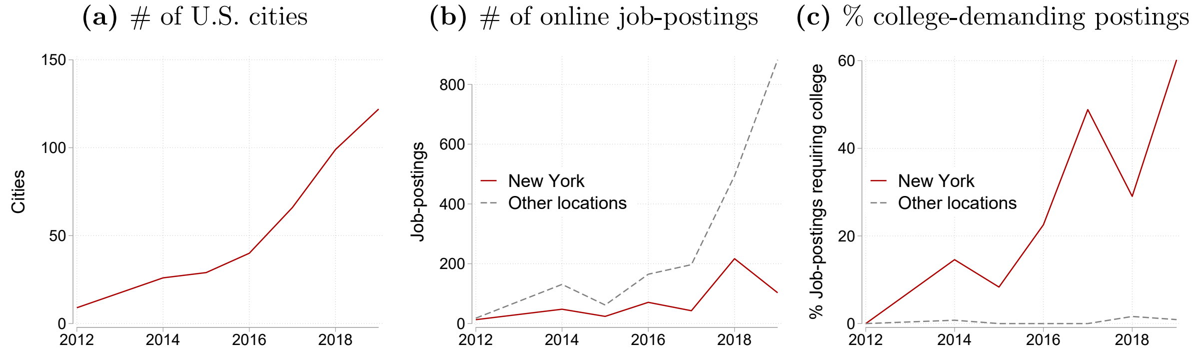
A Case studies for firm spatial expansion

In this section, I provide two case studies that demonstrate the heterogenous labor market implications of the spatial expansion of firms. Note that in both examples, I use only publicly available data and make no use of the confidential Census Bureau data.

A.1 The expansion of Shake-Shack

The first case study is the expansion of Shake Shack, an American fast casual restaurant chain based in New York City, which opened its first restaurant outside of New York in 2010. I explore its demand for labor through the lens of its online job postings, as collected by Burning Glass Technologies (BGT). Panel (a) in Figure 10 below shows the number of cities with postings by Shake Shack in the BGT data over time, which has risen in parallel to its expansion into more cities across the U.S.. Panel (b) shows that most of these postings were in Shake Shack's new locations, with a significant share of them still posted in the original market of New York City. Panel (c) shows the skill intensity of these jobs, as measured by the share of total postings that explicitly require a college degree. While most of the posted jobs are low-skilled according to this measure, the jobs that Shake Shack opens in New York City are increasingly high-skilled, in line with the expansion of its headquarters in that location. By 2019, 60% of the New York City job postings required a college degree. A closer look reveals that these jobs include traditional headquarters-level occupations, such as management, design, marketing and information-technology specialists.

Figure 10: Example: the expansion of Shake Shack (online job-postings data)



This example demonstrates the heterogenous effects of firm expansion across the two main dimensions that are considered in the model: first, in the cross-section of establishments, between firm headquarters and its branches; secondly, across space, between

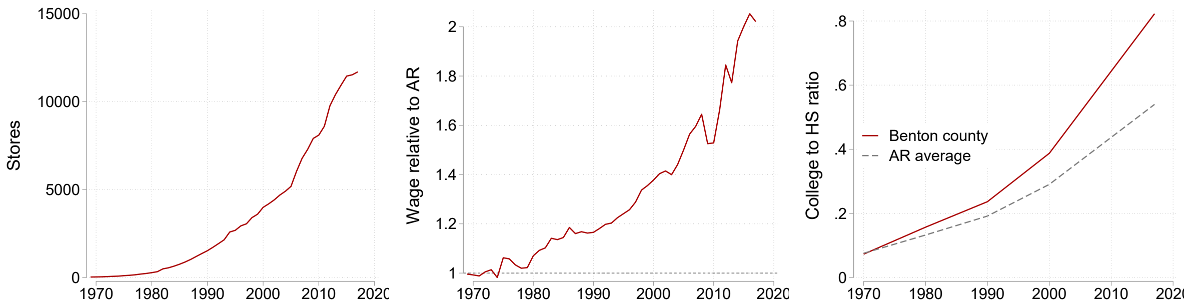
New York City and other locations in the U.S.. This example suggests potentially meaningful distributional implications at the economy as a whole, given that the *average firm* in the economy has exhibited this kind of spatial expansion in recent decades, as was demonstrated in Figure 2 of the paper.

A.2 The expansion of Walmart

The second example is the expansion of the retail corporation Walmart, headquartered in Benton county (Bentonville City), Arkansas. Walmart is an outlier in the sense that it is a particularly large firm with headquarters in (what used to be) a relatively remote location. However, precisely because of this property, it provides a good example for the regional effects of firm spatial expansion at its headquarters market. Panel (a) of Figure 11 below shows the spatial expansion of Walmart as measured by its total number of stores (data from Walmart Inc.). The chain's expansion outside of Arkansas started in the late 1960s and has been growing exponentially since then. Panel (b) and (c) show regional outcomes for Benton county, Arkansas, relative to the Arkansas average. In the early 1970s, Benton county used to be an average Arkansas county, with a similar wage to the rest of the state and with similar skill-intensity, as measured by the ratio of workers with a college degree to workers with only high-school diploma. Since then, in parallel to the expansion of Walmart's headquarters in that county, the average wage has diverged from the rest of the state, such that by now it is more than double the average Arkansas wage. At the same time, it has experienced an influx of college graduates, leading to stronger skill-deepening than the rest of the state. These predictions are all in line with the central mechanism of the model, in which the aggregate spatial expansion of firms leads to greater growth of income and demand for skilled labor in locations that specialize in providing headquarters services.

Figure 11: Example: the expansion of Walmart (Walmart inc. data)

(a) # of Walmart stores overall **(b)** Benton to AR avg. wage ratio **(c)** Benton & AR college/HS ratio



This example demonstrates the heterogenous effects of firm spatial expansion once aggregated to the regional level, with divergence in income and demand for skill across different areas within a single state. As in the case of the previous example of Shake Shack's expansion, it is suggestive of potentially important heterogenous effects for the economy as a whole given the observed spatial expansion of the average firm in the data.

B Additional details on the data

B.1 LBD

Sample selection. I follow a similar sample selection procedure to Barth et al. (2016). I drop observations with non-positive employment or payroll, as well as establishments with over 100,000 employees which are likely to capture miscoded records. I compute average wages as the ratio of total annual payroll to total establishment employment. I convert wages to 1982 dollars using the Consumer Price Index and exclude establishments that have an average wage less than half the yearly equivalent of the 1982 minimum wage of $3.35 an hour for a 40-hour week. I also omit firms in the utilities sector.

Firm-level industry code. In the LBD, establishments are classified into industries, but multi-establishment firms do not have a unique industry identifier. I define the firm's industry according to the 4-digit NAICS code that accounts for the largest share of the firm's payroll.[68] For example, a firm's industry is classified as "Restaurants and Other Eating Places" if establishments with a NAICS code of 7225 constitute most of the firm's total wage bill.[69] Similarly, I define firm-level sector as the 2-digit NAICS sector that accounts for the largest part of its payroll. In the example above, the firm's sector would be "Accommodation and Food Services" (NAICS 72). To classify firms into goods and services, I follow the standard Bureau of Economic Analysis (BEA) definitions for "goods-producing industries" and "services-producing industries".[70] I define a firm as a "service firm" if establishments in services-producing sectors account for at least half of its total payroll.

B.2 Dun & Bradstreet

I use the Dun & Bradstreet Historical Records which provide data on U.S. private and public companies going back to 1969, with the exception of 1981 and 1984. Each establishment in a multi-establishment firm is linked to its headquarters establishment, and therefore it is possible to infer the headquarters location for all establishments in the data. I combine establishments from different headquarters into the same firm identifier if they share a common parent company and company name. I omit observations with missing firm linkages. I omit establishments in Public Administration and other selected industries that are beyond the scope of the LBD dataset such as Membership Organiza-

[68] Throughout the paper, I employ the longitudinally consistent industry codes from Fort and Klimek (2018) that address changes in U.S. industry classification schemes over time.

[69] I construct a separate category for firms without a clear industry classification when no single industry covers at least 40% of the firm's payroll. This group of firms accounts for only a small share of total employment and wage bill, so most firms in the data have a clear firm-level industry identifier.

[70] "Goods-producing industries" include agriculture, mining, construction and manufacturing. "Services-producing industries" include all other sectors with the exception of utilities which are excluded from the analysis.

tions. I employ similar geographic definitions as described for the LBD data in Section 2.1.

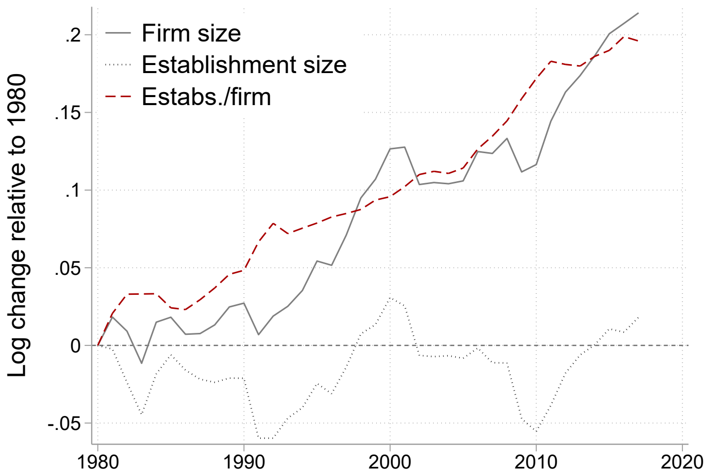
B.3 Burning Glass Technologies

Data on online job postings is obtained from Burning Glass Technologies, a business analytics company, which extracts information from the near-universe of online job postings from a variety of online sources such as job boards and company websites. BGT employs a designated algorithm to avoid double counting of postings across multiple sources. The data covers 2010-2019, and includes extracted information on employer, job location, occupation, education-requirement, and for a small subset (approximately a fifth of total observations) also posted wages. See Azar et al. (2020) for additional information on this data.

To analyze the spatial structure of firms with this data, I merge information on firm organization from the above Dun & Bradstreet dataset. I do so using name and location matching for firms that operate in at least two commuting zones in both datasets. I can then deduce for each job posting where is the headquarters of its firm located, and whether the job is located at the firm's headquarters market or not.

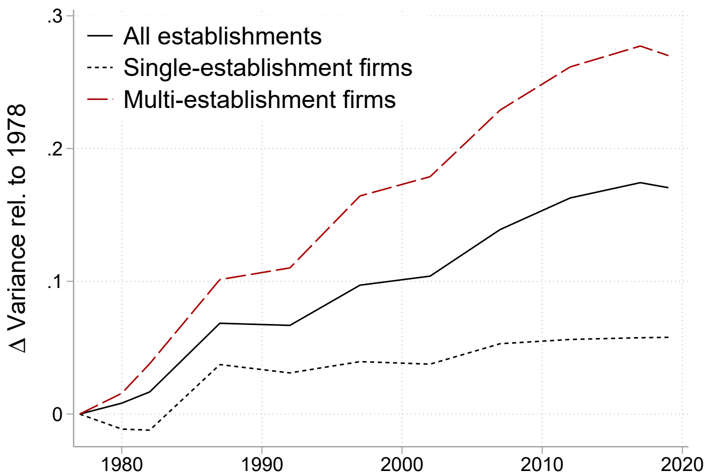
C Additional figures and tables

Figure 12: Firm expansion through the extensive margin

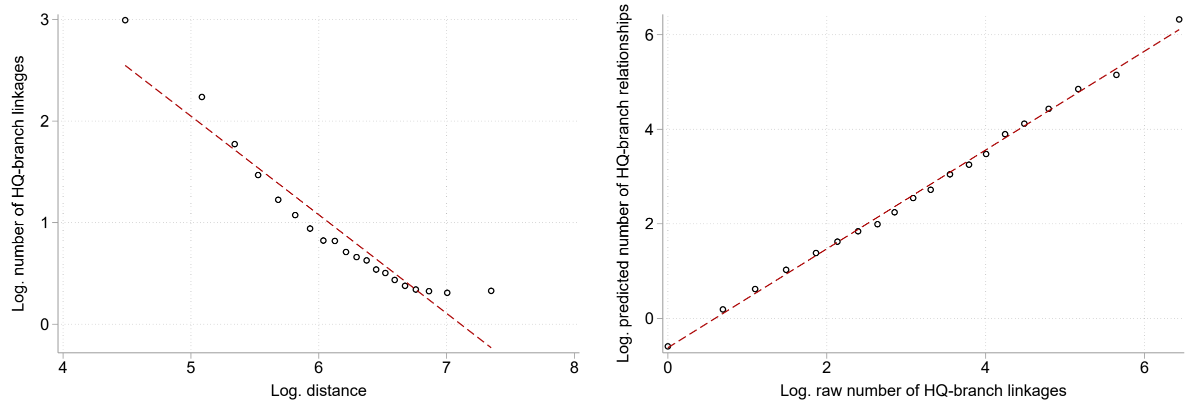


Note: This figure shows that changes in the average number of establishments per firm account for all the increase in average firm size (defined as employment per firm) since 1980, while average establishment size (employment per establishment) remained constant. Data from the U.S. Census Business Dynamics Dataset for firms with at least five employees.

Figure 13: The role of multi-establishment firms in the rise of inequality - raw wages



Note: This figure shows changes in the employment-weighted variance of log average payroll across establishments in the Longitudinal Business Dataset (LBD) in selected years relative to 1978: for all establishments (solid-black line), for multi-establishment firms (dashed-red line) and for single-establishment firms (dotted-gray line). Relative to Figure 4 in the paper, this figure shows this decomposition without first demeaning industry fixed effects.



(a) HQ-branch linkages against distance **(b)** Predicted HQ-branch linkages

Figure 14: HQ-branch linkages: Panel (a) shows the log of bilateral number of headquarters-branch linkages across 1990 labor market areas against the log of distance after controlling for headquarters-location and branch-location fixed effects. Panel (b) shows these bilateral linkages on the x-axis against the predicted part from a Poisson regression of the bilateral linkages on headquarters-location fixed effect, branch-location fixed effect and a polynomial of distance on the y-axis. The data comes from the 1980 venue of the Dun & Bradstreet establishments dataset. Additional details are provided in Section 5.3.

Figure 15: The spatial distribution of headquarters in the data

(a) All HQs and main offices **(b)** HQs serving at least 5 states

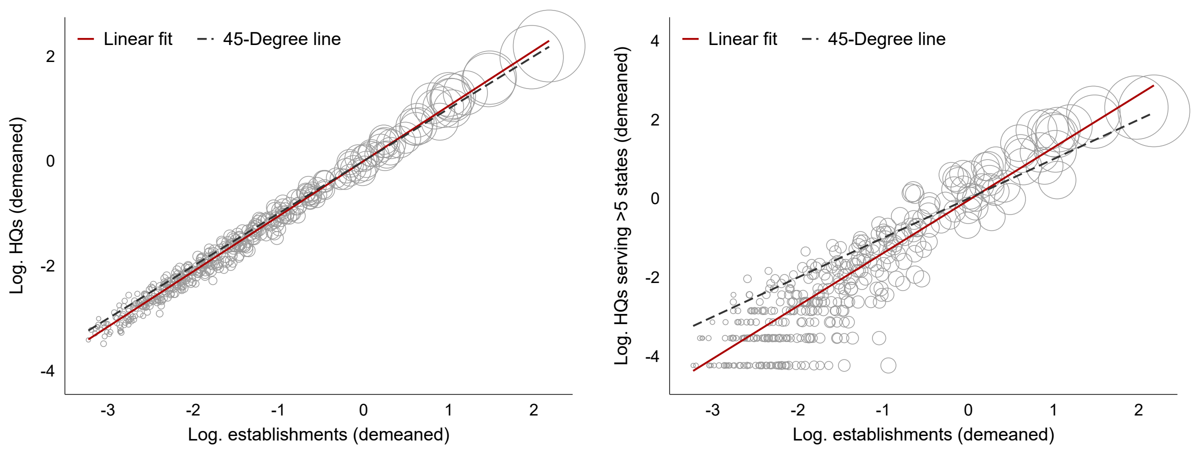


Figure 16: Note: This figure shows the log number of headquarters and main offices from the Dun and Bradstreet data against the log number of total establishments from the Business Dynamics Statistics dataset across commuting zones. Both measures are relative to the national average. Panel (a) shows this relationship for all headquarters in the D&B dataset, and panel (b) shows this relationship only for "large" firms that serve at least five states.

Figure 17: HQ-branch gravity for firm job-postings across services sectors

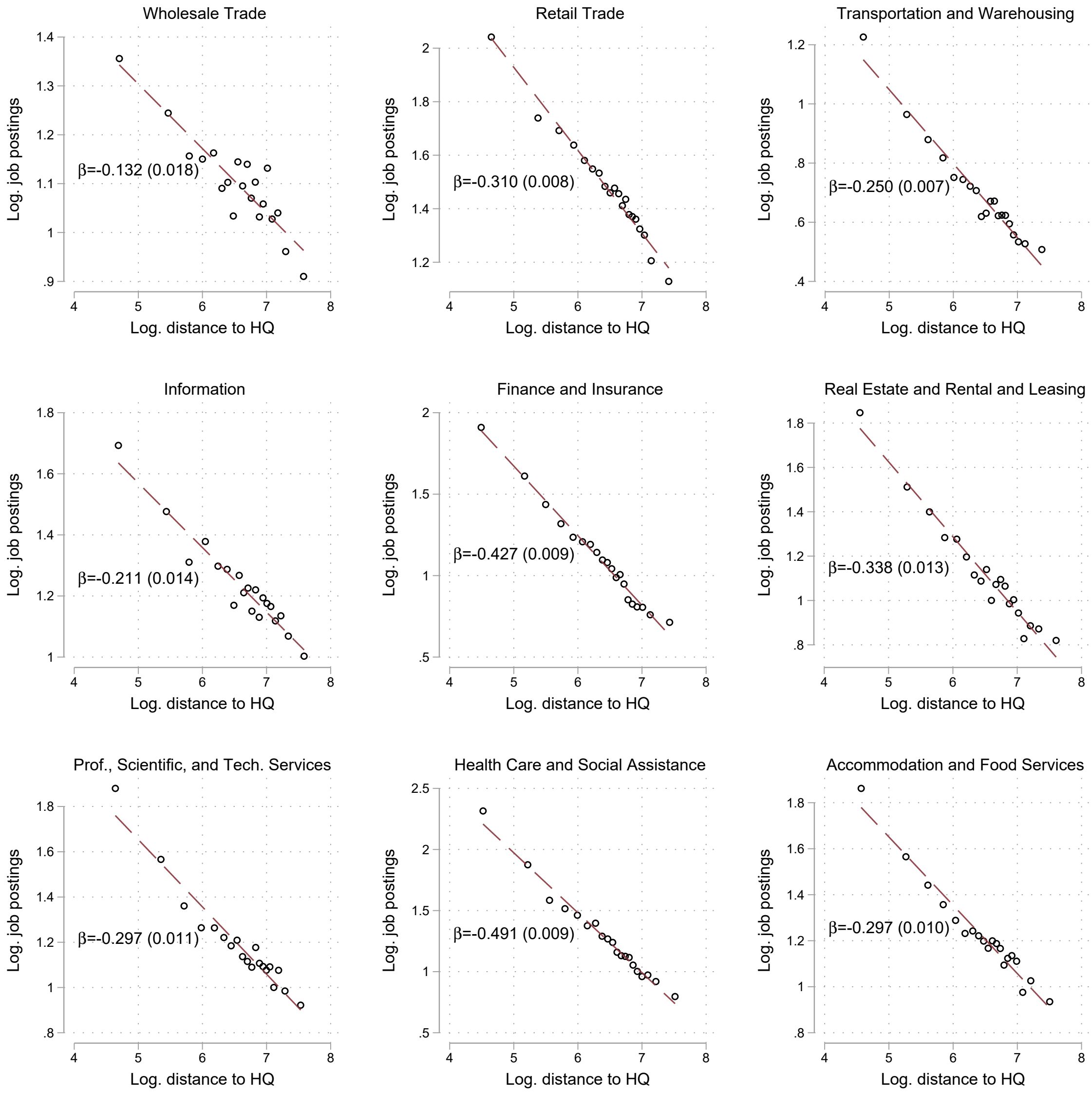
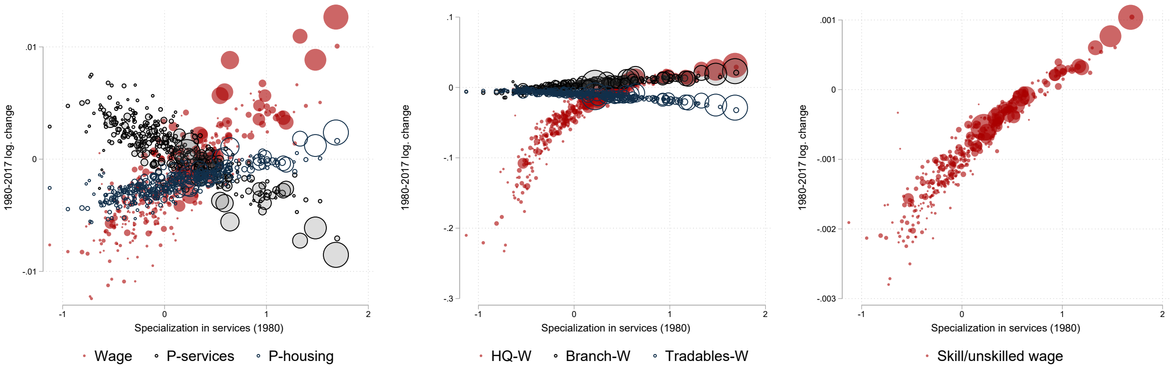


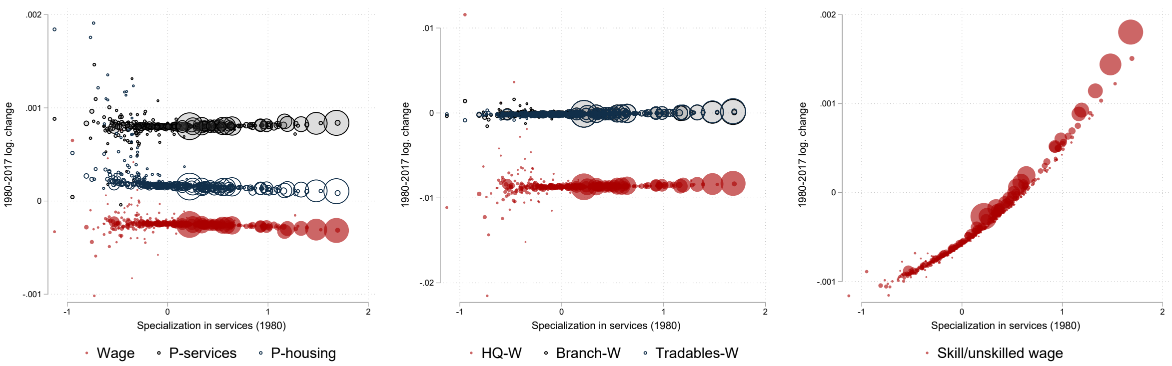
Figure 18: Note: This figure reports binscatter plots for the volume of online job postings against the distance between the posting's location and its firm headquarters location. A location is defined as a 1990 commuting zone, and the figures control for headquarters-location and job-location fixed effects. The job postings data is from BGT, after merging it with firm headquarters locations from the D&B dataset. Each subplot represents a different NAICS sector (2-digits code).

Figure 19: The spatial impact of a small increase in β



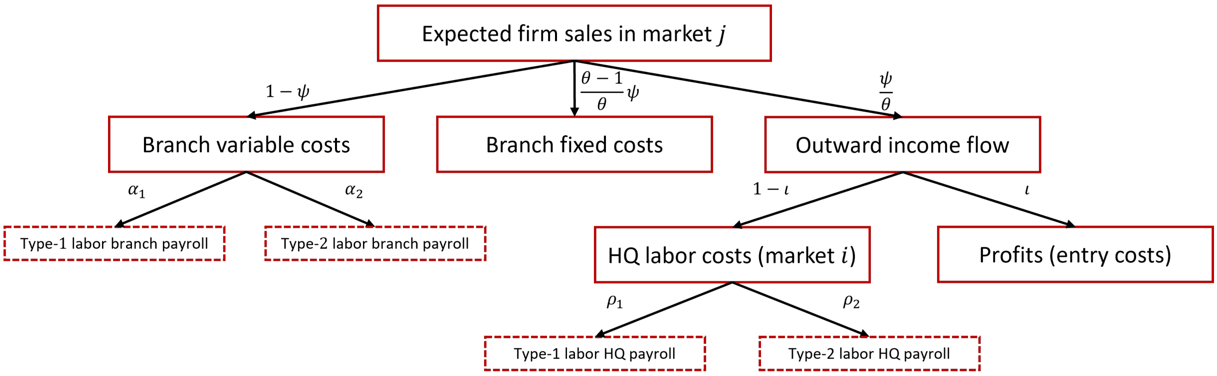
Note: This figure shows regional changes in the model following a small increase in the expenditure share on services β. The x-axis in all three subplots is regional specialization in services in the 1980 equilibrium, given by the log of E_i/R_i in the model. Each circle represents a labor market area and the size of the circles capture total LMA employment in the baseline equilibrium. The left panel shows log changes in the regional average wage index (solid red circles), regional price of housing (empty blue circles), and regional price index of services (shaded gray circles). The center figure shows log changes in the three components of the regional wage index: headquarters wage index (solid red circles), branches wage index (shaded gray circles), and tradable-goods wage index (empty blue circles). The right panel shows changes in the log ratio of skilled to unskilled wages.

Figure 20: The spatial impact of a small increase in ρ_s



Note: This figure shows regional changes in the model following a small increase in the intensity of college graduates in headquarters production ρ_s. The x-axis in all three subplots is regional specialization in services in the 1980 equilibrium, given by the log of E_i/R_i in the model. Each circle represents a labor market area and the size of the circles capture total LMA employment in the baseline equilibrium. The left panel shows log changes in the regional average wage index (solid red circles), regional price of housing (empty blue circles), and regional price index of services (shaded gray circles). The center figure shows log changes in the three components of the regional wage index: headquarters wage index (solid red circles), branches wage index (shaded gray circles), and tradable-goods wage index (empty blue circles). The right panel shows changes in the log ratio of skilled to unskilled wages.

Figure 21: The composition of firm revenues in the model



Note: This figure shows how the expected sales of a i-headquartered firm in some market j are distributed across locations and different types of costs. "Outward income flow" stands for the residual of local revenues after paying local variable and fixed costs. The figure considers two types of worker skills, such that $S = \{1, 2\}$. The composite parameters ψ and ι are given by $\psi \equiv 1 - \frac{\sigma-1}{\sigma}\frac{\epsilon}{1+\epsilon}$ and $\iota \equiv 1 - \gamma\frac{\epsilon}{\epsilon+1}\frac{\sigma-1}{\sigma}\frac{\theta}{\psi}$.

Table 10: Additional statistics on firm structure in goods-producing sectors

Type of firm	Year	Firms	% of workforce	Emp. per firm	Estabs. per firm	CZs per firm
Goods, single estab.	1980	623000	11%	49	1.00	1.00
Goods, single estab.	2017	762000	7%	38	1.00	1.00
Goods, multi-estab. and single-CZ	1980	6900	1%	210	1.78	1.00
Goods, multi-estab. and single-CZ	2017	5400	1%	230	1.03	1.00
Goods, multi-estab. and multi-CZ	1980	9900	19%	8942	69.66	18.73
Goods, multi-estab. and multi-CZ	2017	11000	6%	3535	48.22	11.50

Note: Data from the Census Bureau Longitudinal Business Database. The table includes statistics for firms with at least half of their wage bill in establishments in goods-producing sectors. Multi-estab. stands for a firm with at least two establishments, and multi-CZ stands for a firm with establishments in at least two 1990 commuting zone. See Section 2.1 for additional details on sample selection and definitions.

Table 11: Examples for job postings by selected multi-region service firms

Firm	Type of branch	HQ CZ	CZs	Top occupations in HQ-CZ	Top occupations in other CZs
Equity Residential	Apartments building	Chicago city, IL	61	Accounting Professionals Human Resources Specialists Network and Systems Engineering	Property and Facilities Managers Non-Technical Sales Maintenance and Repair
Bank of America	Bank branch	Charlotte city, NC	453	Software Development Business Analysis Network and Systems Engineering	Financial Sales Investment Specialists Banking and Lending
Delta Airlines	Airport hub/office	Atlanta city, GA	189	Marketing Specialists Project and Program Managers Software Development	Maintenance and Repair Hospitality and Travel Customer Service Representatives
Major League Soccer	Stadium / sports center	New York city, NY	21	Software Development Communications and Public Relations Marketing Specialists	Marketing Specialists Sports Non-Technical Sales
T-Mobile	Store / Repair center	Seattle city, WA	407	Software Development Marketing Specialists Network and Systems Engineering	Non-Technical Sales Retail Managers Retail Sales and Service
Union Pacific Railroad	Rail / repair station	Omaha city, NE	287	Engineering Managers Network and Systems Engineering Software Development	Skilled Construction and Building Trades Maintenance and Repair Rail Transportation
Unitedhealth Group	Local office / clinic	Minneapolis, MN	602	Network and Systems Engineering Business Analysis Software Development	Healthcare Administrators and Managers Advanced Nursing and Physician Assistants Registered / Practical Nursing
Walmart	Store	Fayetteville city, AR	630	Human Resources Specialists Software Development Retail Managers	Retail Sales and Service Logistics and Operational Support Retail Managers

Note: This table shows examples for job postings by firms from different service sectors from the Burning Glass Data. The HQ-CZ is the headquarters commuting zone of the firm. CZs stands for the number of commuting zones with postings by this firm in the data. The fifth column shows the top three occupations in the postings of the firm at its headquarters market, and the last column to the right shows the top three occupations in all other locations.

Table 12: The industry composition of multi-region service firms

Firm activity	(1) % of firm payroll	(2) % of HQ-CZ payroll	(3) % of non-HQ-CZ payroll
Core industry (main 4-digit NAICS)	34.8%	25.4%	39.2%
Core sector (main 2-digit NAICS)	49.6%	42.4%	52.9%
Explicit headquarters activity (NAICS-55)	9.0%	26.8%	0.9%
Non-core business services (NAICS-5*)	2.2%	2.4%	2.1%
Other non-core activity	4.4%	3.1%	4.9%
Total	100%	100%	100%

Note: This table extends Table 2 from Section 2.3.1 with more granular categories of firm activity. Data from the Census Bureau Longitudinal Business Database. The table shows how the total payroll of multi-region service firms is distributed across different activities, in firms with clearly identifiable headquarters location, as discussed in Section 2.1. The categories of activities are: Core industry - establishments in the same 4-digits NAICS code as the firm-level classification; Core sector - establishments in the same 2-digits NAICS code as the firm; Explicit headquarters activity - establishments in the NAICS-55 sector ("Management of Companies and Enterprises"); Non-core business services - establishments outside of the core sector with a 2-digit NAICS-code of 51, 52, 53 or 54; Other non-core activity - all other establishments. The HQ-market of a firm is the commuting zone with the highest share of NAICS-55 activity.

D The headquarters wage premium

In this section, I provide further evidence for the wage differentials between HQ-level tasks and branch-level tasks. To this end, Table 13 reports regressions of log posted wages in the Burning Glass data against the occupational HQ-intensity measure developed in Section 2.3.2 of the main text. Column (1) reports this relationship when controlling for whether the job posting requires a college degree, its commuting zone and its industry (NAICS-3 code). Comparing the least HQ-intensive occupation to the most HQ-intensive occupation is associated with a wage differential of around 25% that is not explained by educational attainment, location, or industry. Column (2) shows that this relationship is even stronger when comparing only occupations within the same firm by controlling for firm fixed effects. Column (3) further shows that the difference between headquarters wages and branch-level wages is larger in firms that serve more markets (more commuting zones), in line with Proposition 3 from Section 4. Columns (4)-(6) repeat Columns (1)-(3) in a multi-year setting with year fixed effects. In particular, Column (6) reports also the coefficient on the Log average number of markets served by the firm, which is no longer constant at the firm level when different years are considered. Firms post higher wages as they expand in space, but particularly so for headquarters-level tasks.

Table 13: The headquarters wage premium

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome: log posted wage	Single-year			Multi-year		
Occupational HQ-intensity	0.249***	0.437***	0.228***	0.362***	0.455***	0.315***
	(0.0753)	(0.0380)	(0.0350)	(0.0713)	(0.0357)	(0.0323)
College requirement	0.406***	0.348***	0.347***	0.387***	0.335***	0.335***
	(0.0105)	(0.00768)	(0.00774)	(0.0108)	(0.00783)	(0.00782)
HQ-intensity × Log CZs			0.0730***			0.0129*
			(0.0239)			(0.00743)
Log CZs						0.0503**
						(0.0216)
Commuting zone FE	Y	Y	Y	Y	Y	Y
Industry FE	Y	N	N	Y	N	N
Firm FE	N	Y	Y	N	Y	Y
Year FE	N	N	N	Y	Y	Y
Observations	3281863	3281863	3281863	7253996	7253996	7253996

Note: This table shows the results from a regression of a log posted wages in the Burning Glass data on various job characteristics. Columns (1)-(3) consider only 2019 - the year with the largest number of postings that include wages. Columns (4)-(6) consider a multi-year setting covering 2015-2019. Earlier years (2010-2014) are omitted due to limited information on posted wages. "Occupational HQ-intensity" captures the measure developed in Section 2.3.2. "College requirement" is an indicator that equals to one if a posting explicitly requires a college degree. "Log CZs" is the average number of commuting zones in which the firm is active. Standard errors in parentheses, clustered at the firm level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

E A shift-share analysis of firm expansion

In this appendix, I provide additional evidence for the mechanism in the model by exploring the differential exposure of different regions to national industry-level expansion trends. In the model, locations with ex-ante specialization in providing within-firm headquarters services see the greatest increase in income and demand for skill following a decline in frictions to firm expansion. I extend this logic to a multi-industry setting, in which locations that ex-ante specialize in providing headquarters-services in industries with greater decline in such frictions should see the greatest increase in income and demand for skilled labor. A simple example to this logic can be seen in the case of Bentonville, AR from Appendix Section A. In 1970, this local market already specialized in providing headquarters services to retail branches. Since then, the Retail Trade sector witnessed significant technological changes that led to substantial spatial firm expansion, as can be seen in the rise of the average establishments per firm for this sector as a whole in Figure 2 of the main text. Accordingly, this local market experienced greater income growth and skill deepening than the rest of Arkansas.

Specifically, for each commuting zone i and period t, I define the following regional exposure to firm spatial expansion EE_{it}:

$$EE_{it} \equiv \sum_{k \in \text{Industries}} \left(\frac{N^{HQ}_{ik,1980}}{\sum_{k'} N^{HQ}_{ik',1980}} \right) \times \log \left(\frac{\overline{CZ}_{k,t}}{\overline{CZ}_{k,1980}} \right),$$

where $N^{HQ}_{ik,1980}$ is the number of firm headquarters located in region i for industry k in the Dun and Bradstreet dataset; and $\bar{CZ}_{k,t}$ is the *national* average number of commuting-zones per firm in industry k for period t. This shift-share variable yields a high value for region i, period t, when the headquarters that region i hosted in 1980 tend to be from industries that have witnessed greater national spatial expansion between 1980 and t.

Table 14 below shows the results from a regression of a series of regional outcomes on these expansion-exposure measures. I omit small commuting zones with less than 5000 workers, and I include the years 1980, 1990, 2000, and 2017. The first outcome is log average wages from the BLS Quarterly Census of Employment and Wages (QCEW). The second outcome is the regional log college premium (average wage for college graduates relative to workers without a college degree), computed from the decennial census files for 1980-2000, and from the 2015-2019 American Community Survey for 2017.[71] Due to the noisier nature of this data, this outcome has less observations, and smaller commuting zones are omitted. The third outcome is the log regional ratio of college-graduates to workers without a college degree.[72] Columns (1)-(3) report the coefficients

[71]Data taken from Ruggles et al. (2021).

[72]This data is obtained from the USDA county-level data: https://www.ers.usda.gov/data-products/county-level-data-sets/.

from regressions of these three outcomes on the expansion-exposure measure, controlling for commuting-zone and year fixed effects, and weighting by commuting zone 1980 employment. Columns (4)-(6) add controls for total (log) commuting zone employment and establishments; and for the manufacturing employment share in a commuting zone, to control for differential structural transformation, though adding these controls does not affect much the estimated coefficients.

In line with the rationale of the mechanism in the model, commuting zones that were more exposed to national firm expansion trends by hosting headquarters of relevant industries in 1980 have witnessed greater income growth. In addition, they have experienced greater increase of the relative price and the relative quantity of skilled-labor, indicating of an upward shift in the demand for skilled labor in these markets relative to less-exposed markets. Note that all initial commuting-zone characteristics such as its size or skill intensity in 1980 are absorbed in the commuting zone fixed effect.

Table 14: Expansion Exposure

	(1) Log. Wage	(2) Log. SP	(3) Log. HL	(4) Log. Wage	(5) Log. SP	(6) Log. HL
Expansion Exposure	1.150***	0.668**	0.670***	0.867***	0.674**	0.616***
	(0.188)	(0.275)	(0.224)	(0.208)	(0.284)	(0.208)
Log. Employment				0.121	-0.287***	0.0393
				(0.103)	(0.0900)	(0.174)
Log. Establishments				0.0850	0.236**	0.0287
				(0.0882)	(0.0933)	(0.201)
Manufacturing Share				-0.332	0.582**	-0.904***
				(0.300)	(0.232)	(0.191)
N	2332	1650	2332	2332	1650	2332
CZ and Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table shows the results from a regression of a series of regional outcomes on the expansion-exposure measures described in the text. Small commuting zones with less than 5000 workers are omitted. The included years are 1980, 1990, 2000, and 2017. SP is the regional college premium (average wage for college graduates relative to workers without a college degree), computed from the decennial census files for 1980-2000, and from the 2015-2019 American Community Survey for 2017. Due to the noisier nature of this data, this outcome has less observations. HL is the ratio of college-graduates to workers without a college degree. Columns (1)-(3) report the coefficients from regressions of these three outcomes on the expansion-exposure measure, controlling for commuting-zone and year fixed effects, and weighting by commuting zone 1980 employment. Columns (4)-(6) add controls for total (log) commuting zone employment and establishments, and for the manufacturing employment share. Standard errors in parentheses, clustered at the commuting zone level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

F Additional model derivations

F.1 Additional derivations for the households block

Recall that households maximize

$$U_{is\nu} = b_{i\nu} \left(Q_i\right)^{\beta} \left(Q_i^g\right)^{1-\beta}, \quad Q_i = \left[\int_{\omega\in\Omega_i} \left(q_i\left(\omega\right)\right)^{\frac{\sigma-1}{\sigma}} d\omega\right]^{\frac{\sigma}{\sigma-1}},$$

where $b_{i\nu}$ is an household-specific idiosyncratic preference shock for region i and employer ν; $\beta \in (0,1)$ is the expenditure share on services; Q_i^g is a homogenous quantity of goods; and the bundle of services Q_i aggregates all service varieties ω available in market i, Ω_i, using a constant elasticity of substitution $\sigma > 1$. The ideal price index dual to the composite of services is given by

$$P_i = \left(\int_{\omega\in\Omega_i} P_i\left(\omega\right)^{1-\sigma} d\omega\right)^{\frac{1}{1-\sigma}},$$

and I denote the total expenditure on services in market j by E_j. Note that E_j includes expenditure by households, i.e. a share β of total household income, as well as additional expenditure by firms that will be specified later. The aggregate regional price index faced by each consumer is given by

$$\bar{P}_i = P_i^{\beta}.$$

The idiosyncratic shocks $\boldsymbol{b}$ are assumed to be independently draw from

$$F\left(\boldsymbol{b}\right) = \exp\left(-\sum_{i=1}^{N}\left(\int_{\nu\in\mathcal{V}_i} b_{i\nu}^{-\epsilon} d\nu\right)^{\frac{\xi}{\epsilon}}\right), \tag{A.1}$$

where the parameter ξ governs the dispersion of shocks across different markets and thus captures the region-level labor supply elasticity; ϵ governs the dispersion of shocks across different employers within each location and thus captures the employer-specific labor supply elasticity; and $\mathcal{V}_i$ is the set of available employers in location i.

The implied regional labor supply, given by the probability that an agent of type s chooses region i, equals to:

$$\frac{L_{is}}{\sum_{j=1}^{N} L_{js}} = \frac{\left(W_{is}/\bar{P}_i\right)^{\xi}}{\sum_{j=1}^{N}\left(W_{js}/\bar{P}_j\right)^{\xi}} \tag{A.2}$$

where L_{is} is the measure of households of type s that choose location i, and W_{is} is the regional skill-specific ideal wage index, aggregating the employer-specific wages $w_{is}(\nu)$:

$$W_{is} = \left(\int_{\nu\in\mathcal{V}_i} w_{is}\left(\nu\right)^{\epsilon} d\nu\right)^{\frac{1}{\epsilon}}. \tag{A.3}$$

W_{is} is the welfare-relevant wage-index, accounting for the level of wages and for the variety of available employers.

Conditional on choosing location i, the mass of type-s workers that choose employer ν is given by

$$l_{is}\left(\nu\right) = L_{is}\left(\frac{w_{is}\left(\nu\right)}{W_{is}}\right)^{\epsilon}. \tag{A.4}$$

F.2 Solving the branch problem

I start by deriving closed-form solutions for the firm's branch-level decisions. First, conditional on the composite of headquarters labor $h_i \equiv \prod_{s=1}^{S} h_{is}^{\rho_s}$ and on activity in market j, we solve for the optimal production decisions in that market:

$$\pi_{0,ij}\left(z\right) = \max_{\{l_s\}_{s=1}^{S}} E_j^{\frac{1}{\sigma}} P_j^{\frac{\sigma-1}{\sigma}} \left(q_{ij}\left(z\right)\right)^{\frac{\sigma-1}{\sigma}} - \sum_{s=1}^{S} W_{js} L_{js}^{-\frac{1}{\epsilon}} l_{ijs}^{\frac{\epsilon+1}{\epsilon}}. \tag{A.5}$$

The implied j-market revenues $r_{ij}\left(z_i,\phi_j\right)$, type-$s$ employment $l_{ijs}\left(z_i,\phi_j\right)$ and type-s wages $w_{ijs}\left(z_i,\phi_j\right)$ are given by

$$\begin{aligned} r_{ij}\left(z_i,\phi_j\right) &= \left(1-\psi\right)^{\frac{1}{\psi}\left(1-\psi\right)} \left(E_j^{\frac{1}{\sigma}} P_j^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\psi}} C_j^{-\frac{1-\psi}{\psi}} \varphi_j\left(z_i,\phi_j\right)^{\frac{1}{\psi}\frac{\sigma-1}{\sigma}}, \\ l_{ijs}\left(z_i,\phi_j\right) &= W_{js}^{-\frac{\epsilon}{\epsilon+1}} L_{js}^{\frac{1}{\epsilon+1}} \left(\left(1-\psi\right)\alpha_s r_{ij}\left(z_i,\phi_j\right)\right)^{\frac{\epsilon}{\epsilon+1}}, \\ w_{ijs}\left(z_i,\phi_j\right) &= W_{js}^{\frac{\epsilon}{\epsilon+1}} L_{js}^{-\frac{1}{\epsilon+1}} \left(\left(1-\psi\right)\alpha_s r_{ij}\left(z_i,\phi_j\right)\right)^{\frac{1}{\epsilon+1}}, \end{aligned} \tag{A.6}$$

where ψ captures the share of operating profits from sales in market j[73]

$$\psi \equiv 1 - \frac{\sigma-1}{\sigma}\frac{\epsilon}{1+\epsilon},$$

C_j is a regional cost-shifter that summarizes the price and availability of labor for branch-level production in that market

$$C_j \equiv \prod_{s}\left(\alpha_s W_{js}^{-1} L_{js}^{\frac{1}{\epsilon}}\right)^{-\alpha_s},$$

and ϕ_j is the firm's productivity at market j. Intuitively, the firm has greater revenues in market j when the market size is larger (higher E_j), there is less competition (higher P_j), branch-labor is cheaper (lower C_j), and the firm's productivity is higher. Branch-level

[73] I define operating profits in this context as revenues minus local labor costs, excluding fixed costs and headquarters labor costs.

employment and wages are both increasing in the firm's local revenues, and branch-level wages are decreasing in the local supply of labor L_{js} and increasing in the local wage index W_{js}. Note that in the limit of perfectly competitive labor markets, $\epsilon \to \infty$, branch-level wages are given by the market j skill-specific wage index W_{js}, which itself becomes the average wage for group s in that market.

F.3 Solving the branch activity decision

The firm serves market j if these operating profits are greater than the fixed cost $\bar{P}_j f$, where $\bar{P}_j$ is the aggregate price index in market j which we use to price the fixed costs f. This decision takes place when the realization of the market-specific idiosyncratic shock ϕ_j is large enough, or $\psi r_{ij}(z) > \bar{P}_j f$. I.e. we require

$$\phi_j^{\frac{1}{\psi}\frac{\sigma-1}{\sigma}} > \frac{\bar{P}_j f}{\psi (1-\psi)^{\frac{1}{\psi}(1-\psi)} \left(E_j^{\frac{1}{\sigma}} P_j^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\psi}} C_j^{-\frac{1-\psi}{\psi}} \left(A_{ij}\varphi_i(z)\right)^{\frac{1}{\psi}\frac{\sigma-1}{\sigma}}}. \tag{A.7}$$

For convenience, denote

$$r_{0,ij}(z) \equiv (1-\psi)^{\frac{1}{\psi}(1-\psi)} \left(E_j^{\frac{1}{\sigma}} P_j^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\psi}} C_j^{-\frac{1-\psi}{\psi}} \left(A_{ij}\varphi_i(z)\right)^{\frac{1}{\psi}\frac{\sigma-1}{\sigma}}. \tag{A.8}$$

The threshold productivity for activity is thus equal to

$$\tilde{\phi}_j^{\frac{1}{\psi}\frac{\sigma-1}{\sigma}} = \frac{\bar{P}_j f}{\psi r_{0,ij}(z)}, \tag{A.9}$$

and the probability for activity is

$$Pr\left(\phi_j^{\frac{1}{\psi}\frac{\sigma-1}{\sigma}} > \frac{\bar{P}_j f}{\psi r_{0,ij}(z)}\right) = \left(\frac{\bar{P}_j f}{\psi r_{0,ij}(z)}\right)^{-\theta}. \tag{A.10}$$

The resulting unconditional sales are

$$\bar{r}_{ij}(z) = \frac{\theta}{\theta-1}\left(\frac{\bar{P}_j f}{\psi}\right)^{1-\theta} \left(r_{0,ij}(z)\right)^{\theta}, \tag{A.11}$$

which can also be written as

$$\bar{r}_{ij}(z) = \tilde{\psi} f^{1-\theta} \Upsilon_j \left(A_{ij} z_i h_i^{\gamma}\right)^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}}, \tag{A.12}$$

where

$$\tilde{\psi} \equiv \left(\frac{\theta}{\theta-1}\psi^{\theta-1}(1-\psi)^{\frac{\theta}{\psi}(1-\psi)}\right),$$

and the regional aggregate Υ_j captures *market potential* in location j

$$\Upsilon_j \equiv \left(E_j^{\frac{1}{\sigma}} P_j^{\frac{\sigma-1}{\sigma}} \bar{P}_j^{-\psi \frac{\theta-1}{\theta}} C_j^{-(1-\psi)} \right)^{\frac{\theta}{\psi}}. \tag{A.13}$$

The market potential term Υ_j conveniently summarizes all terms that make some market j more attractive for operating branches, including greater market size E_j, the degree of competition (captured by P_j), the cost of hiring branch-level labor in that market C_j, and the price of the fixed branch cost $\bar{P}_j$.

The resulting unconditional revenues $\bar{r}_j(z_i)$ are a power function of j's market potential, the bilateral friction between i and j, A_{ij}, and the firm-level productivity components affecting all branches: fundamental firm productivity z_i and endogenous headquarters-level bundle h_i. We thus get a standard gravity structure for firm-level expected sales, and they are log linear in a firm-level component $(z_i h_i^{\gamma})^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}}$, the destination j market potential Υ_j, and bilateral frictions between the firm's headquarters location i and market j. Similarly to unconditional sales, expected unconditional employment and wage levels are given by

$$\bar{l}_{ijs}(z) = W_{js}^{-\frac{\epsilon}{\epsilon+1}} L_{js}^{\frac{1}{\epsilon+1}} \left((1-\psi)\alpha_s\right)^{\frac{\epsilon}{\epsilon+1}} \frac{\theta-1}{\theta-\frac{\epsilon}{\epsilon+1}} \left(\frac{\bar{P}_j f}{\psi}\right)^{\frac{\epsilon}{\epsilon+1}-1} \bar{r}_{ij}(z), \tag{A.14}$$

$$\bar{w}_{ijs}(z) = W_{js}^{\frac{\epsilon}{\epsilon+1}} L_{js}^{-\frac{1}{\epsilon+1}} \left((1-\psi)\alpha_s\right)^{\frac{1}{\epsilon+1}} \frac{\theta-1}{\theta-\frac{1}{\epsilon+1}} \left(\frac{\bar{P}_j f}{\psi}\right)^{\frac{1}{\epsilon+1}-1} \bar{r}_{ij}(z). \tag{A.15}$$

F.4 Headquarters and firm-wide solutions

With the above branch-level solutions, I now continue to derive closed-form solutions for headquarters-level decisions. Based on its expected revenues across all markets, the firm solves for optimal headquarters-level hiring $\{h_{is}\}_{s=1}^{S}$ and the implied headquarters wages $\{w_{is}\}_{s=1}^{S}$:

$$\max_{\{h_s\}_{s=1}^{S}} \tilde{\psi} f^{1-\theta} \left(\sum_{j=1}^{N} A_{ij}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \Upsilon_j\right) \left(z\left(\prod_s h_{is}^{\rho_s}\right)^{\gamma}\right)^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} - \sum_{s=1}^{S} W_{is} L_{is}^{-\frac{1}{\epsilon}} h_{is}^{\frac{\epsilon+1}{\epsilon}}. \tag{A.16}$$

The first order conditions yield

$$\rho_s \underbrace{\gamma \frac{\epsilon}{\epsilon+1} \frac{\sigma-1}{\sigma} \frac{\theta}{\psi}}_{1-\iota} \frac{\psi}{\theta} \bar{r}_i(z) = W_{is} L_{is}^{-\frac{1}{\epsilon}} h_{is}^{\frac{\epsilon+1}{\epsilon}}, \tag{A.17}$$

and the resulting firm choices are

$$h_{is}(z) = \left(\rho_s (1-\iota) \frac{\psi}{\theta} W_{is}^{-1} L_{is}^{\frac{1}{\epsilon}} \bar{r}_i(z)\right)^{\frac{\epsilon}{\epsilon+1}}, \tag{A.18}$$

$$w_{is}(z) = W_{is}^{\frac{\epsilon}{\epsilon+1}} L_{is}^{-\frac{1}{\epsilon+1}} \left(\rho_s (1-\iota) \frac{\psi}{\theta} \bar{r}_i(z)\right)^{\frac{1}{\epsilon+1}}. \tag{A.19}$$

where $\bar{r}_i(z_i)$ captures expected firm revenues across all markets, internalizing optimal branch-level and headquarters-level decisions, and on which I elaborate further below. According to the above expressions, the firm hires more headquarters workers and pays them higher wages when firm-level revenues across all markets are higher. This dependency captures the non-rival nature of the HQ output. Accordingly, firms with better expansion opportunities (i.e., higher $\bar{r}_i(z_i)$ conditional on z_i) have larger headquarters and higher branch-level productivity. As the elasticity of branch-productivity to HQ-hiring γ approaches zero, the firm has no incentive to hire HQ labor. Finally, the firm hires more s-type workers and pays them lower wages when the market-level supply L_{is} is higher and the ideal wage index W_{is} is lower.

Internalizing these optimal decisions, the above expected total revenues $\bar{r}_i(z_i)$ take a simple form

$$\bar{r}_i(z_i) = \Lambda_i z_i^{\frac{\sigma-1}{\sigma}\frac{\theta}{\iota\psi}}. \tag{A.20}$$

where ι is a constant capturing the share of operating profits that ends up going to the firm's owners.

The regional aggregate Λ_i conveniently captures all forces that make location i more attractive for operating headquarters, and is given by

$$\Lambda_i \equiv \tilde{\iota} \underbrace{\left(\sum_{j=1}^{N} A_{ij}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \Upsilon_j\right)^{\frac{1}{\iota}}}_{\text{HQ market access}} \underbrace{\left(\prod_s \left(\rho_s W_{is}^{-1} L_{is}^{\frac{1}{\epsilon}}\right)^{-\rho_s}\right)^{-\frac{1-\iota}{\iota}}}_{\text{Cost and supply of HQ labor}}, \tag{A.21}$$

where $\tilde{\iota}$ is a constant that subsumes various parameters, and we define the headquarters-labor cost shifter

$$\Gamma_i \equiv \prod_s \left(\rho_s W_{is}^{-1} L_{is}^{\frac{1}{\epsilon}}\right)^{-\rho_s}. \tag{A.22}$$

Expected firm sales to specific market j are thus given by

$$\bar{r}_{ij}(z) = \frac{A_{ij}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \Upsilon_j}{\sum_{j'=1}^{N} A_{ij'}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \Upsilon_{j'}} \Lambda_i z^{\frac{\sigma-1}{\sigma}\frac{\theta}{\iota\psi}}. \tag{A.23}$$

Figure 21 summarizes the breakdown of expected revenues in the model to different types of costs and to profits. A share $1-\psi$ of the firm's expected sales in each market j are paid as compensation to local branch labor, divided across the different skill groups in accordance to the branch-level production function. A share $\psi(1-1/\theta)$ of sales is expected to be spent on fixed branch costs in each market. The residual ψ/θ of revenues is split between payments to HQ labor ($(1-\iota)$ percentage of this residual), in the headquarters location i, and between firm profits (ι percentage of this residual).

F.5 Entry and headquarters locations

The equilibrium allocation of headquarters across space admits a simple solution. Recall that the individual firm chooses headquarters location i to maximize its expected profits from Equation (7).

We now obtain a simple expression for these profits, given by

$$\mathbb{E}_{\boldsymbol{\phi}}\left[\pi_i\left(c^{\mathbb{I}_{i\neq n}} z_{0,i}, \boldsymbol{\phi}\right)\right] = \frac{\psi\iota}{\theta}\Lambda_i z_i^{\frac{\sigma-1}{\sigma}\frac{\theta}{\iota\psi}},$$

which is just a constant share $\psi\iota/\theta$ of expected revenues from Equation (A.20). This share captures the market power of the firm in output and labor markets, captured by the elasticities σ and ϵ, both of which appear in the composite constants ψ and ι. In addition, it captures the dispersion of branch-level shocks θ that governs the probability of serving individual markets and the HQ-branch elasticity γ that appears in ι and governs the ability of the firm to raise revenues across all markets through the non-rival HQ inputs.

The log-linearity of expected profits in the headquarters-productivity term z_i, together with the assumption that the baseline HQ productivity draws $z_{0,i}$ are Fréchet-distributed, yield a simple expression for the probability that a firm that entered in location n chooses i as its headquarters base, denoted by λ_{ni}:

$$\lambda_{ni} = \frac{c^{\eta\mathbb{I}_{n\neq i}}\Lambda_i^{\eta/\left(\frac{\sigma-1}{\sigma}\frac{\theta}{\iota\psi}\right)}}{\sum_{m=1}^{N} c^{\eta\mathbb{I}_{n\neq m}}\Lambda_m^{\eta/\left(\frac{\sigma-1}{\sigma}\frac{\theta}{\iota\psi}\right)}}. \tag{A.24}$$

According to this expression, locations with higher Λ_i attract a greater share of entrants from each market n. Note that in the extreme case that the productivity lose from moving the HQ away from the entry market is particularly high, $c \to 0$, we obtain that all firms stay in their entry market, and there is no sorting based on the set of productivity draws $\{z_{0,i}\}_{i=1}^N$. In this case, $\lambda_{ni} = 1$ for $n = i$ and is zero otherwise.

Finally, ex-ante expected profits for market n are given by

$$\pi_{e,n} = \tilde{\pi}\left[\left(1-c^{\eta}\right)\Lambda_n^{\eta/\left(\frac{\sigma-1}{\sigma}\frac{\theta}{\iota\psi}\right)} + \sum_{i=1}^{N} c^{\eta}\Lambda_i^{\eta/\left(\frac{\sigma-1}{\sigma}\frac{\theta}{\iota\psi}\right)}\right]^{\left(\frac{\sigma-1}{\sigma}\frac{\theta}{\iota\psi}\right)/\eta}, \tag{A.25}$$

with $\tilde{\pi} \equiv \frac{\psi\iota}{\theta}\Gamma\left(1-\left(\frac{\sigma-1}{\sigma}\frac{\theta}{\iota\psi}\right)/\eta\right)$, where $\Gamma(\cdot)$ is the Gamma-function. This expression shows that ex-ante profits depend on the HQ-attractiveness shifters Λ_i across all locations given the ability to place headquarters away from the firm's entry market n. In the extreme case when such reallocation is associated with a great productivity loss, i.e. $c \to 0$, expected profits simply scale with the value of Λ_n in the entry location n. When $c \to 1$, expected profits become a standard power mean of Λ_i across all locations, and the entry location does not matter for the firm's headquarters choice (i.e., the entry cost is a sunk cost).

F.6 Characterization of firm wages - Propositions 2 and 3

Armed with solutions to the firm's problem, we can now obtain results for Propositions 2 and 3. Consider a firm headquartered in location i with a branch in location j, firm-level productivity z and branch-level productivity ϕ. Combining the expressions for branch wages and revenues from Equation A.6, the wages that it pays for workers of type s in market j are given by

$$w_{ijs}(z,\phi) = W_{js}^{\frac{\epsilon}{\epsilon+1}} L_{js}^{-\frac{1}{\epsilon+1}} \left((1-\psi)\,\alpha_s\,(1-\psi)^{\frac{1}{\psi}(1-\psi)}\left(E_j^{\frac{1}{\sigma}} P_j^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\psi}} C_j^{-\frac{1-\psi}{\psi}} \varphi_{ij}(z,\phi)^{\frac{1}{\psi}\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\epsilon+1}}.$$

Based on the solution for the firm headquarters workers from Equation A.18, the productivity term $\varphi_{ij}(z,\phi)$ equals to

$$\varphi_{ij}(z,\phi) = A_{ij} z \phi \left(\prod_{s=1}^{S} h_{is}^{\rho_s}\right)^{\gamma} = A_{ij}\phi\left((1-\iota)\,\frac{\psi}{\theta}\frac{\Lambda_i}{\Gamma_i}\right)^{\frac{\epsilon}{\epsilon+1}\gamma} z^{\frac{1}{\iota}}.$$

Plugging it into the expression for branch wages yields an equation for wages in terms of regional aggregates and firm productivities

$$w_{ijs}(z,\phi) = \text{const}_s W_{js}^{\frac{\epsilon}{\epsilon+1}} L_{js}^{-\frac{1}{\epsilon+1}} \left(\left(E_j^{\frac{1}{\sigma}} P_j^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\psi}} C_j^{-\frac{1-\psi}{\psi}}\right)^{\frac{1}{\epsilon+1}} \tag{A.26}$$

$$\times\left(\left(A_{ij}\phi\left(\frac{\Lambda_i}{\Gamma_i}\right)^{\frac{\epsilon}{\epsilon+1}\gamma} z^{\frac{1}{\iota}}\right)^{\frac{1}{\psi}\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\epsilon+1}}. \tag{A.27}$$

Substituting $\left(E_j^{\frac{1}{\sigma}} P_j^{\frac{\sigma-1}{\sigma}} C_j^{-(1-\psi)}\right)^{\frac{1}{\psi}} = \Upsilon_j^{\frac{1}{\theta}} \bar{P}_j^{\frac{1}{\theta}-1}$ from the definition of the market potential term Υ_j, taking logs and rearranging yields the expression from Proposition 2. Recall in this step that for the sake of expositional clarity, the proposition defines the constant $\chi \equiv \frac{\frac{\sigma-1}{\sigma}\frac{\epsilon}{\epsilon+1}}{1-\frac{\sigma-1}{\sigma}\frac{\epsilon}{\epsilon+1}} = \frac{1-\psi}{\psi}$.

To get the first result in Proposition 3, recall first the headquarters-level wages are given by

$$w_{is}(z) = W_{is}^{\frac{\epsilon}{\epsilon+1}} L_{is}^{-\frac{1}{\epsilon+1}} \left(\rho_s (1-\iota) \frac{\psi}{\theta} \Lambda_i z^{\frac{\sigma-1}{\sigma}\frac{\theta}{\iota\psi}}\right)^{\frac{1}{\epsilon+1}}.$$

Using this expression and the expression from Equation A.26, I compute the ratio of headquarters wages to branch wages $\frac{w_{is}(z)}{w_{ijs}(z,\phi)}$. Now consider two i-headquartered firms with firm productivities z and z' that are active in market j and have the same branch level productivity ϕ. Dividing the expression for $\frac{w_{is}(z)}{w_{ijs}(z,\phi)}$ by the one for $\frac{w_{is}(z')}{w_{ijs}(z',\phi)}$ yields

$$\left(\frac{w_{is}(z)}{w_{ijs}(z,\phi)} / \frac{w_{is}(z')}{w_{ijs}(z',\phi)}\right) = \left(\frac{z}{z'}\right)^{(\theta-1)\frac{\sigma-1}{\sigma}\frac{1}{\iota\psi}\frac{1}{\epsilon+1}}.$$

Since $\theta > 1$, this ratio is increasing in z/z'. Therefore, the HQ-branch wage differential is higher in the more productive firm, yielding the first part of Proposition 3.

For the second part of Proposition 3, simply divide Equation A.19 by the expression for branch wages from Equation A.18, setting $i = j$. Then we get that the ratio of headquarters wages to branch-level wages in the firm's headquarters market is increasing in the ratio of total expected firm sales across all markets to local branch sales in the headquarters market.

F.7 Aggregation of firm revenues

First, we express aggregate bilateral sales R_{ij}, which capture the total revenues of i-headquartered firms from establishments in market j, by aggregating Equation (A.23) above:

$$R_{ij} = \frac{A_{ij}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \Upsilon_j}{\sum_{j'=1}^{N} A_{ij'}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \Upsilon_{j'}} \Lambda_i Z_i M_i, \tag{A.28}$$

where M_i is the mass of firms with headquarters in market i, and Z_i is an aggregate term capturing the average productivity of firms with headquarters in location i:

$$Z_i \equiv \int_z z^{\frac{\sigma-1}{\sigma}\frac{\theta}{\iota\psi}} dG_i(z), \tag{A.29}$$

where $G_i(z)$ is the distribution of firm productivity z for firms with headquarters in location i. Total sales by firms headquartered in location i are thus given by

$$R_i = \sum_{j=1}^{N} R_{ij}. \tag{A.30}$$

We can also express bilateral sales in terms of the expenditure shares of location j on varieties headquartered in location i, χ_{ji}, given by

$$\chi_{ji} = \frac{A_{ij}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \left(\frac{\Lambda_i}{\Gamma_i}\right)^{1-\iota} Z_i M_i}{\sum_{n=1}^{N} A_{nj}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \left(\frac{\Lambda_n}{\Gamma_n}\right)^{1-\iota} Z_n M_n}, \tag{A.31}$$

so that

$$R_{ij} = \chi_{ji} E_j. \tag{A.32}$$

F.8 Services price index

To get the services price index, first note that

$$E_j = \sum_i R_{ij} = \Upsilon_j \sum_i \tilde{\psi} f^{1-\theta} A_{ij}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \left((1-\iota)\frac{\psi}{\theta}\frac{\Lambda_i}{\mathcal{C}_i}\right)^{1-\iota} Z_i M_i. \tag{A.33}$$

Then from the definition of Υ_i

$$\begin{aligned}
P_i^{-\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} &= \bar{P}_i^{1-\theta} E_i^{\frac{\theta}{\psi}\frac{1}{\sigma}} C_i^{-\frac{\theta}{\psi}(1-\psi)} \Upsilon_i^{-1} \\
&= \bar{P}_i^{1-\theta} E_i^{\frac{\theta}{\psi}\frac{1}{\sigma}-1} C_i^{-\frac{\theta}{\psi}(1-\psi)} \frac{E_i}{\Upsilon_i} \\
&= \tilde{\psi}\left((1-\iota)\frac{\psi}{\theta}\right)^{1-\iota} \bar{P}_i^{1-\theta} E_i^{\frac{\theta}{\psi\sigma}-1} C_i^{-\frac{\theta}{\psi}(1-\psi)} \sum_n f^{1-\theta} A_{ni}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \left(\frac{\Lambda_n}{\Gamma_n}\right)^{1-\iota} Z_n M_n \\
&= \tilde{\psi}\left((1-\iota)\frac{\psi}{\theta}\right)^{1-\iota} \bar{P}_i^{1-\theta} E_i^{\frac{\theta}{\psi\sigma}-1} C_i^{-\frac{\theta}{\psi}(1-\psi)} \sum_n f^{1-\theta} A_{ni}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \left(\frac{R_n}{\Gamma_n}\right)^{1-\iota} (M_n Z_n)^{\iota},
\end{aligned}$$

Where the last line uses the fact that

$$R_n = \Lambda_n M_n Z_n,$$

So that

$$P_i = \left[\tilde{\iota} \bar{P}_i^{1-\theta} E_i^{\frac{\theta}{\psi\sigma}-1} C_i^{-\frac{\theta}{\psi}(1-\psi)} \sum_n A_{ni}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \left(\frac{R_n}{\Gamma_n}\right)^{1-\iota} (M_n Z_n)^{\iota}\right]^{\frac{1}{-\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}}}, \tag{A.34}$$

where recall that

$$\tilde{\iota} \equiv \left(\tilde{\psi} f^{1-\theta}\left((1-\iota)\frac{\psi}{\theta}\right)^{1-\iota}\right)^{\frac{1}{\iota}}.$$

F.9 The tradable-goods sector

Firms in the tradable-goods sector in location i solve the following profit-maximization problem

$$\max_{\{l_{is}\}_{s=1}^{S}} A_{g,i} \prod_{s=1}^{S} l_{is}^{\alpha_{g,s}} - \sum_{s=1}^{S} W_{\text{is}} L_{is}^{-\frac{1}{\epsilon}} l_{is}^{\frac{\epsilon+1}{\epsilon}}. \tag{A.35}$$

The first order condition yields

$$A_{g,i} l_{is} = \frac{\epsilon+1}{\epsilon} W_{\text{is}} L_{is}^{-\frac{1}{\epsilon}} l_{is}^{\frac{\epsilon+1}{\epsilon}}. \tag{A.36}$$

The aggregate revenues of the tradable-goods sector in market i are thus given by

$$R_{g,i} = M_{g,i} \left(\left(\frac{\epsilon}{1+\epsilon} \frac{1}{C_{g,i}} \right)^{\frac{\epsilon}{1+\epsilon}} A_{g,i} \right)^{1+\epsilon}, \tag{A.37}$$

where $M_{g,i}$ is the mass of goods-producing firms that are active in location i, and $C_{g,i}$ is the tradable-goods cost shifter:

$$C_{g,i} \equiv \prod_{s} \left(\alpha_{g,s} W_{is}^{-1} L_{is}^{\frac{1}{\epsilon}} \right)^{-\alpha_{g,s}}. \tag{A.38}$$

Free entry in the goods sector equates the average local entry cost $\bar{P}_i f_e$ to the average sectoral profit, and is given by the following equation

$$M_{g,i} \bar{P}_i f_{g,i} = \frac{1}{1+\epsilon} R_{g,i}. \tag{A.39}$$

F.10 Aggregate wage index

Recall that the aggregate wage index for labor of type s in region i is given by

$$W_{is} = \left(\int_{\nu \in \mathcal{V}_i} w_{is} (\nu)^{\epsilon} \, d\nu \right)^{\frac{1}{\epsilon}},$$

where $\mathcal{V}_i$ is the set of local employers, comprised of local branches, headquarters and goods-producing firms. I compute W_{is}^{ϵ} for each of these groups separately, and then combine them.

I begin with the part of the wage index due to branch-level wages. Consider a firm with headquarters in location n that has a branch in location i, characterized by firm-level productivity z and branch-level productivity ϕ. From the solution of the branch problem, it pays the following wages to workers of type s

$$w_{nis}(z,\phi) = W_{is}^{\frac{\epsilon}{\epsilon+1}} L_{is}^{-\frac{1}{\epsilon+1}} \left((1-\psi)\,\alpha_s r_{ni}(z,\phi)\right)^{\frac{1}{\epsilon+1}},$$

where the branch revenues are given by

$$\begin{aligned} r_{ni}(z,\phi) = {} & (1-\psi)^{\frac{1}{\psi}(1-\psi)} \left((1-\iota)\frac{\psi}{\theta}\right)^{\frac{\epsilon}{\epsilon+1}\frac{1}{\psi}\frac{\sigma-1}{\sigma}\gamma} \\ & \times \left(E_i^{\frac{1}{\sigma}} P_i^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\psi}} C_i^{-\frac{1-\psi}{\psi}} A_{ni}^{\frac{1}{\psi}\frac{\sigma-1}{\sigma}} \left(\frac{\Lambda_n}{\Gamma_n}\right)^{\frac{1}{\psi}\frac{\sigma-1}{\sigma}\frac{\epsilon}{\epsilon+1}\gamma} z^{\frac{1}{\iota}\frac{1}{\psi}\frac{\sigma-1}{\sigma}} \phi^{\frac{1}{\psi}\frac{\sigma-1}{\sigma}}. \end{aligned}$$

The part of the regional wage index due to branch-level wages is obtained by aggregating the expected value of $w_{nis}(z,\phi)^{\epsilon}$ across all firms in the economy

$$\left(W_{is}^{\text{Branches}}\right)^{\epsilon} \equiv \sum_{n=1}^{N} M_n \mathbb{E}_{z,\phi}\left[\mathbb{I}_{ni}(z,\phi)\left(w_{nis}(z,\phi)\right)^{\epsilon}\right],$$

where $\mathbb{I}_{ni}(z,\phi)$ equals to one if a firm with productivities (z,ϕ) serves market i, thus taking care of the fact that only a subset of n-headquartered firms serve market i for each n. This expression in turn is given by

$$\begin{aligned} \left(W_{is}^{\text{Branches}}\right)^{\epsilon} = {} & \left((1-\psi)\,\alpha_s\right)^{\frac{\epsilon}{\epsilon+1}} \frac{\theta-1}{\theta-\frac{\epsilon}{\epsilon+1}} \left(\frac{\psi}{f}\right)^{\frac{1}{\epsilon+1}} \\ & \times \sum_{n=1}^{N} W_{is}^{\epsilon\frac{\epsilon}{\epsilon+1}} L_{is}^{-\frac{\epsilon}{\epsilon+1}} \bar{P}_i^{-\frac{1}{\epsilon+1}} \frac{A_{ni}^{-\frac{1}{\psi}\frac{\sigma-1}{\sigma}} \Upsilon_i}{\sum_{i'=1}^{N} A_{ni'}^{-\frac{1}{\psi}\frac{\sigma-1}{\sigma}} \Upsilon_{i'}} R_n. \end{aligned}$$

Next, recall from the solution to the headquarters problem that a firm with productivity z with headquarters in i pays the following wages for type-s HQ workers:

$$w_{is}(z) = W_{is}^{\frac{\epsilon}{\epsilon+1}} L_{is}^{-\frac{1}{\epsilon+1}} \left(\rho_s(1-\iota)\frac{\psi}{\theta}\Lambda_i z^{\frac{\sigma-1}{\sigma}\frac{\theta}{\iota\psi}}\right)^{\frac{1}{\epsilon+1}}.$$

Aggregating w_{is}^{ϵ} across all i-headquartered firms yields the HQ component of the wage index

$$\left(W_{is}^{\text{HQ}}\right)^{\epsilon} \equiv M_n \mathbb{E}_z\left[\left(w_{is}(z)\right)^{\epsilon}\right].$$

Performing the integration yields

$$\left(W_{is}^{\text{HQ}}\right)^{\epsilon} = M_i Z_{w,i} W_{is}^{\epsilon\frac{\epsilon}{\epsilon+1}} L_{is}^{-\frac{\epsilon}{\epsilon+1}} \left(\rho_s(1-\iota)\frac{\psi}{\theta}\Lambda_i\right)^{\frac{\epsilon}{\epsilon+1}},$$

where $Z_{w,i}$ is an aggregate productivity measure that accounts for labor market power and given by

$$Z_{w,i} \equiv \int_i z^{\frac{\theta}{\iota}\frac{1-\psi}{\psi}} dG_i(z) \, .$$

Note that when $\epsilon \to \infty$, $Z_{w,i} \to Z_i$. Finally, the component of the wage index that accounts for goods-producing firms can be show to equal

$$\left(W_{is}^{\text{Goods}}\right)^{\epsilon} = W_{is}^{\epsilon\frac{\epsilon}{\epsilon+1}} L_{is}^{-\frac{\epsilon}{\epsilon+1}} M_{g,i} \left(\alpha_{g,s} \frac{\epsilon}{1+\epsilon} \frac{R_{g,i}}{M_{g,i}}\right)^{\frac{\epsilon}{\epsilon+1}} .$$

Combining all three terms yields the regional wage index

$$W_{\text{is}} = \left[\left(W_{is}^{\text{Branches}}\right)^{\epsilon} + \left(W_{is}^{\text{HQ}}\right)^{\epsilon} + \left(W_{is}^{\text{Goods}}\right)^{\epsilon}\right]^{\frac{1}{\epsilon}} . \tag{A.40}$$

F.11 Equilibrium in regional aggregates (Lemma 1)

I now present the full set of non-linear equations that determine the model's equilibrium. I separate this set of equations into four blocks.

Labor market clearing. First, the labor market clearing block determines the wage levels W_{is} and employment levels L_{is}. This block includes the wage-index Equation (A.40) above and the sorting of household across space:

$$\frac{L_{is}}{\sum_{j=1}^{N} L_{js}} = \frac{\left(W_{is}/\bar{P}_i\right)^{\xi}}{\sum_{j=1}^{N} \left(W_{js}/\bar{P}_j\right)^{\xi}}$$

Services market clearing and regional balance of payments. Secondly, services price indices P_i, expenditure levels E_i, and revenues R_i, are determined by market clearing in the services market,

$$R_i = \sum_{j=1}^{N} \frac{A_{ij}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \left(\frac{\Lambda_i}{\Gamma_i}\right)^{1-\iota} Z_i M_i}{\sum_{n=1}^{N} A_{nj}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \left(\frac{\Lambda_n}{\Gamma_n}\right)^{1-\iota} Z_n M_n} E_j$$

the expression for the services price index

$$P_i = \left[\tilde{\iota} \bar{P}_i^{1-\theta} E_i^{\frac{\theta}{\psi\sigma}-1} C_i^{-\frac{\theta}{\psi}(1-\psi)} \sum_n A_{ni}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \left(\frac{R_n}{\Gamma_n}\right)^{1-\iota} \left(M_n Z_n\right)^{\iota}\right]^{\frac{1}{-\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}}} ,$$

and the regional balance of payments,

$$\left(\frac{1-\beta}{\beta} + \frac{\psi}{\theta}\right) E_i = R_{g,i} + (1-\iota)\frac{\psi}{\theta} R_i + M_{e,i}\bar{P}_i f_e,$$

where recall that

$$\Lambda_i \equiv \tilde{\iota} \sum_{j=1}^{N} A_{ij}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \Upsilon_j^{\frac{1}{\iota}} \Gamma_i^{-\frac{1-\iota}{\iota}}, \quad \Gamma_i \equiv \prod_s \left(\rho_s W_{is}^{-1} L_{is}^{\frac{1}{\epsilon}}\right)^{-\rho_s}$$

and

$$\Upsilon_j \equiv \left(E_j^{\frac{1}{\sigma}} P_j^{\frac{\sigma-1}{\sigma}-\beta\psi\frac{\theta-1}{\theta}} C_j^{-(1-\psi)}\right)^{\frac{\theta}{\psi}}, \quad C_j \equiv \prod_s \left(\alpha_s W_{js}^{-1} L_{js}^{\frac{1}{\epsilon}}\right)^{-\alpha_s},$$

and that the aggregate price index $\bar{P}_i$ in the equilibrium with housing is given by

$$\bar{P}_i = \left(\frac{\delta}{\beta(1-\delta)}\right)^{\frac{\delta}{(1+\varrho)-\delta\varrho}} P_i^{\beta\left(1-\frac{\delta}{(1+\varrho)-\delta\varrho}\right)} E_i^{\frac{\delta}{(1+\varrho)-\delta\varrho}}.$$

Tradable-goods production and entry. Third, the regional mass of tradable-goods firms $M_{g,i}$ and revenues $R_{g,i}$ are determined by the optimality and free entry conditions for this sector:

$$R_{g,i} = M_{g,i}\left(\left(\frac{\epsilon}{1+\epsilon}\frac{1}{C_{gi,}}\right)^{\frac{\epsilon}{1+\epsilon}} A_{g,i}\right)^{1+\epsilon}, \quad C_{g,i} \equiv \prod_s \left(\alpha_{g,s} W_{is}^{-1} L_{is}^{\frac{1}{\epsilon}}\right)^{-\alpha_{g,s}}$$

and

$$M_{g,i}\bar{P}_i f_{g,i} = \frac{1}{1+\epsilon} R_{g,i}$$

.

Entry and optimal headquarters-location decisions in the services sector. Finally, all remaining aggregates M_i, $M_{e,i}$, Z_i and $Z_{w,i}$ are determined in the entry and HQ-location block for the services sector. The aggregate productivity index Z_i is given by

$$Z_i = \Gamma\left(1 - 1/\left(\frac{\eta}{\zeta}\right)\right) \sum_o \frac{M_{e,o}}{M_i} c^{\zeta \mathbb{I}_{o\neq i}} \lambda_{oi}^{1-\frac{\zeta}{\eta}}, \tag{A.41}$$

and the labor market productivity term $Z_{w,i}$ is given by

$$Z_{w,i} = \Gamma\left(1 - 1/\left(\frac{\eta}{\zeta_w}\right)\right) \sum_o \frac{M_{e,o}}{M_i} \lambda_{oi}^{1-\frac{\zeta}{\eta}} c^{\zeta_w \mathbb{I}_{o\neq i}}, \tag{A.42}$$

where $\Gamma(\cdot)$ is the Gamma-function, ζ is a constant of parameters given by $\zeta \equiv \frac{\sigma-1}{\sigma}\frac{\theta}{\iota\psi}$, and the headquarters probability shares λ_{ni} are given by

$$\lambda_{ni} = \frac{\left(c^{\zeta \mathbb{I}_{n\neq i}} \Lambda_i\right)^{\frac{\eta}{\zeta}}}{\sum_{m=1}^{N} \left(c^{\zeta \mathbb{I}_{n\neq m}} \Lambda_m\right)^{\frac{\eta}{\zeta}}}.$$

The mass of entry $M_{e,i}$ for each market can be recovered from the condition that balances the flow of firms across locations

$$M_i = \sum_n \lambda_{ni} M_{e,n},$$

and the mass of headquarters M_i is given by $M_i = R_i/(\Lambda_i Z_i)$.
This completes the set of non-linear equations that are required to characterize all regional aggregates in the model.

F.12 Aggregate and regional skill-intensity (Proposition 4)

First note that the total compensation for labor of type s in location i is given by

$$L_{is}W_{a,is} = \underbrace{\alpha_s \frac{\sigma-1}{\sigma}\frac{\epsilon}{\epsilon+1}E_i}_{\text{Branch-level compensation}} + \underbrace{\rho_s\gamma\frac{\epsilon}{\epsilon+1}\frac{\sigma-1}{\sigma}R_i}_{\text{HQ compensation}} + \underbrace{\alpha_{g,s}\frac{\epsilon}{1+\epsilon}R_{g,i}}_{\text{Tradable-goods compensation}}, \tag{A.43}$$

where note that $W_{a,is}$ is the *arithmetic* mean regional wage for group s, which is different from W_{is} when $\epsilon \in (1, \infty)$. To get the share of group s in total aggregate labor compensation, we sum the above across all regions, and divide by the sum across all regions and skill groups, to obtain

$$\frac{W_{a,s}L_s}{W_a L} = \frac{\alpha_s + \rho_s\gamma + \alpha_{g,s}\frac{\sigma}{\sigma-1}\frac{1-\beta}{\beta}}{1+\gamma+\frac{\sigma}{\sigma-1}\frac{1-\beta}{\beta}}, \tag{A.44}$$

where we have used the market clearing condition $\sum_i R_i = \sum_i E_i$ and the fact that $\sum_i R_{g,i} = ((1-\beta)/\beta)\sum_i E_i$.

Now suppose that $c \to 0$, i.e. firms are headquartered in their entry market. The regional balance of payments becomes

$$\frac{1-\beta}{\beta}E_i - R_{g,i} = \frac{\psi}{\theta}\left(R_i - E_i\right).$$

Plugging it into Expression (A.43) to eliminate the tradable-goods revenues yields

$$L_{is}W_{a,is} = \frac{\epsilon}{\epsilon+1}\left(\frac{\sigma-1}{\sigma}\alpha_s + \alpha_{g,s}\frac{1-\beta}{\beta} + \alpha_{g,s}\frac{\psi}{\theta}\right)E_i + \frac{\epsilon}{\epsilon+1}\left(\frac{\sigma-1}{\sigma}\rho_s\gamma - \alpha_{g,s}\frac{\psi}{\theta}\right)R_i.$$

Summing across all skill groups yields

$$L_i W_{a,i} = \frac{\epsilon}{\epsilon+1}\left(\frac{\sigma-1}{\sigma} + \frac{1-\beta}{\beta} + \frac{\psi}{\theta}\right)E_i + \frac{\epsilon}{\epsilon+1}\left(\frac{\sigma-1}{\sigma}\gamma - \frac{\psi}{\theta}\right)R_i,$$

and taking the ratio of the above expressions yields the share of group s in total regional labor compensation:

$$\frac{L_{is}W_{a,is}}{L_iW_{a,i}} = \frac{\alpha_s + \alpha_{g,s}\frac{\sigma}{\sigma-1}\left(\frac{1-\beta}{\beta} + \frac{\psi}{\theta}\right) + \left(\rho_s\gamma - \alpha_{g,s}\frac{\sigma}{\sigma-1}\frac{\psi}{\theta}\right) R_i/E_i}{\left(1 + \frac{\sigma}{\sigma-1}\frac{1-\beta}{\beta} + \frac{\sigma}{\sigma-1}\frac{\psi}{\theta}\right) + \left(\gamma - \frac{\sigma}{\sigma-1}\frac{\psi}{\theta}\right) R_i/E_i}. \tag{A.45}$$

The above expression increases with the regional specialization in services R_i/E_i when

$$\left(\frac{\sigma-1}{\sigma}\rho_s\gamma - \alpha_{g,s}\frac{\psi}{\theta}\right)\left(\frac{\sigma-1}{\sigma} + \left(\frac{1-\beta}{\beta} + \frac{\psi}{\theta}\right)\right)$$
$$-\left(\frac{\sigma-1}{\sigma}\alpha_s + \alpha_{g,s}\left(\frac{1-\beta}{\beta} + \frac{\psi}{\theta}\right)\right)\left(\frac{\sigma-1}{\sigma}\gamma - \frac{\psi}{\theta}\right) > 0.$$

Rearranging this expression, we get

$$\left(\alpha_{g,s} - \alpha_s\right)\gamma\frac{\sigma-1}{\sigma} + \left[\left(\rho_s - \alpha_{g,s}\right)\gamma - \left(\alpha_{g,s} - \alpha_s\right)\right]\frac{\psi}{\theta} + \left(\rho_s - \alpha_{g,s}\right)\gamma\left(\frac{1-\beta}{\beta} + \frac{\sigma-1}{\sigma}\right) > 0.$$

When type-s intensity in services is greater than in goods, i.e. both $\alpha_s > \alpha_{g,s}$ and $\rho_s > \alpha_{g,s}$, the above always holds. When branch-level type-s intensity is smaller than in the goods sector, $\alpha_s < \alpha_{g,s}$, Expression (A.45) increases in R_i/E_i if the following condition holds:

$$\rho_s > \alpha_{g,s} + \frac{1}{\gamma}\left(\alpha_{g,s} - \alpha_s\right), \tag{A.46}$$

i.e. when ρ_s is large enough relative to $\alpha_{g,s}$. In particular, consider the case that when the services sector and the tradable-goods sector are equally intensive in type s labor at the aggregate, which is obtained when

$$\alpha_{g,s} = \frac{1}{\gamma+1}\alpha_s + \frac{\gamma}{\gamma+1}\rho_s.$$

Then the regional compensation share for type s labor is increasing in R_i/E_i if $\rho_s > \alpha_{g,s}$.

Finally, in the limit equilibrium without multi-region service firms, i.e. $A_{ij} \to 0$ for $i \neq j$, the share of type s in regional labor compensation is equalized across all locations, and given by the aggregate share in Expression (A.44).

G Additional details on model inversion

In this appendix, I provide additional details on the inversion procedure that allows me to recover the set of unilateral regional fundamentals $\{A_i, B_i, A_{g,i}\}_{i=1}^{N}$.

First, market clearing in the services sector implies that the total sales of all i-headquartered firms, R_i, must equal to the total spending on them across all locations

$$R_i = \sum_{j=1}^{N} \frac{(A_i/\tau_{ij})^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \left(\frac{\Lambda_i}{\Gamma_i}\right)^{1-\iota} Z_i M_i}{\sum_{n=1}^{N} (A_n/\tau_{nj})^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \left(\frac{\Lambda_n}{\Gamma_n}\right)^{1-\iota} Z_n M_n} E_j. \tag{A.47}$$

Armed with data on R_i across all locations, I recover a guess for A_i (up to scale) from the above set of equations, and iterate over it as part of the solution for equilibrium.

To get the set of regional amenities $\{B_i\}_{i=1}^{N}$, I invert the equilibrium conditions for worker sorting across space. Recall that the probability that a worker of type s chooses region i is given by

$$\frac{L_{is}}{\bar{L}_s} = \frac{\left(B_i W_{is}/\bar{P}_i\right)^{\xi}}{\sum_{j=1}^{N} \left(B_j W_{js}/\bar{P}_j\right)^{\xi}}, \tag{A.48}$$

where $\bar{L}_s$ is the aggregate exogenous measure of type-s workers in the economy. Multiplying by $\bar{L}_s$ and summing across all skill groups, we obtain an expression for the total equilibrium level of employment in each region,

$$L_i = \sum_{s=1}^{S} \frac{\left(B_i W_{is}/\bar{P}_i\right)^{\xi}}{\sum_{j=1}^{N} \left(B_j W_{js}/\bar{P}_j\right)^{\xi}} \bar{L}_s. \tag{A.49}$$

Armed with the observed allocation of employment across space, $\{L_i\}_{i=1}^{N}$, we can use the above expression to exactly recover $\{B_i\}_{i=1}^{N}$. Then, recovering the group-specific measures of workers from (A.48) is immediate.

The final set of fundamentals to recover is regional productivities in production of tradable-goods, $\{A_{g,i}\}_{i=1}^{N}$. I use this set of model residuals to exactly match the distribution of average wages across space. To this end, I first note that the total regional wage bill in the model can be written as

$$L_i W_{a,i} = \underbrace{\frac{\epsilon}{\epsilon+1}\frac{\sigma-1}{\sigma}(E_i + \gamma R_i)}_{\text{Regional wage bill from services}} + \underbrace{\frac{\epsilon}{\epsilon+1} R_{g,i}}_{\text{Regional wage bill from tradable-goods}}, \tag{A.50}$$

where $W_{a,i}$ is the employment-weighted arithmetic mean of wages in location i, equivalent to the average regional wage measured in the data. Therefore, conditional on parameters and values for R_i and E_i, I recover the implied revenues in the tradable-goods sector $R_{g,i}$. Then, we use the equilibrium condition for tradable-goods revenues from Equation A.37

$$R_{g,i} = M_{g,i} \left(\left(\frac{\epsilon}{1+\epsilon} \frac{1}{C_{g,i}} \right)^{\frac{\epsilon}{1+\epsilon}} A_{g,i} \right)^{1+\epsilon}, \quad C_{g,i} \equiv \prod_{s} \left(\alpha_{g,s} W_{is}^{-1} L_{is}^{\frac{1}{\epsilon}} \right)^{-\alpha_{g,s}},$$

to recover the set of productivities $\{A_{g,i}\}_{i=1}^{N}$.

To conclude, we recover the set of unilateral model fundamentals $\{A_i, B_i, A_{g,i}\}_{i=1}^{N}$, by inverting the equilibrium conditions for services market clearing; worker sorting across space; and aggregate revenues in the tradable-goods sector. To this end, we utilize regional data on employment L_i, average wage $W_{a,i}$, and revenues by locally-headquartered firms R_i.

H Model extensions

H.1 Unobserved worker heterogeneity and worker screening

In this section, I develop another version of the model with alternative micro-foundations for wage inequality across workers in the same region. Instead of idiosyncratic preference shocks, households now differ by unobserved ability. I further assume that this ability is imperfectly observed by potential employers, and that firms need to pay a screening cost to recruit higher ability households. This formulation follows Helpman et al. (2010), and it allows me to demonstrate how the main results about the role of multi-region firms arise with alternative assumptions on the source of wage heterogeneity. In this setting, higher wages no longer reflect compensating differentials, and instead capture the screening of higher ability workers by firms. Nevertheless, the main results from Section 4 go through, as stated in Proposition A.1 below. An additional benefit of this formulation is that differences in skill intensity between headquarters and branches are now endogenous. In Proposition A.2 below, I show that similarly to wages, the difference in average worker ability between headquarters and branches is also higher for larger firms, and rises as firms expand in space.

H.1.1 Households

Households maximize the same bundle of consumption as in the baseline model, but no longer have idiosyncratic preferences for employers. In addition, households are no longer divided to groups. Instead, households are differentiated by ability a, unobserved to firms before households and firms are paired. I assume that a is drawn from a Pareto distribution with shape parameter k and scale parameter a_{min}.

H.1.2 Production function

The production function is similar to the baseline specification, with the following changes. A firm with HQ in location i, HQ productivity z and branch-level productivity ϕ in some market j, can supply

$$q_{ij}\left(z,\phi\right)=\varphi_{ij}\left(z,\phi\right)\bar{a}l^{\varrho} \tag{A.51}$$

units of output in that market. The composite of branch-level inputs from the baseline specification ($\prod_s l_{js}^{\alpha_s}$) is now replaced with $\bar{a}l^{\varrho}$, where $\bar{a}$ is the average ability of workers in that branch, l is the measure of workers in the branch, and ϱ captures branch-level decreasing returns. The productivity term $\varphi_{ij}\left(z,\phi\right)$ is now given by

$$\varphi_{ij}\left(z,\phi\right)=A_{ij}z\phi\left(\bar{b}h^{\vartheta}\right)^{\gamma}. \tag{A.52}$$

The components A_{ij}, z and ϕ remain as in the baseline specification. The fourth component replaces the headquarters bundle of labor ($\prod_s h_{is}^{\rho_s}$). It is now given by $\bar{b}h^{\vartheta}$, where $\bar{b}$ is the average ability of the headquarters workers, h is the measure of headquarters workers, and ϑ captures HQ-level decreasing returns.

H.1.3 Labor market frictions

The labor market is characterized by search and matching frictions. Firms randomly match with workers, and can choose the measure of workers to consider for each branch, n, and for the headquarters, m. The cost of meeting a random worker is c_n. Firms then screen these workers by their ability. The firm can choose a minimal ability level of a_c at a given branch and b_c at its headquarters, such that that only workers with ability $a > a_c$ (or $a > b_c$ in the HQ) are hired. This process is costly, and requires the firm to pay a convex cost of $\delta_0 a_c^{\delta}$ (or $\delta_0 b_c^{\delta}$ in the HQ). For simplicity, I assume that the job-posting costs $c_n n$ and $c_n m$, and the screening costs $\delta_0 a_c^{\delta}$ and $\delta_0 b_c^{\delta}$, are all paid in units of the tradable goods, with price P_g (normalized to 1 in the baseline model). Once the firm hires new workers, wages at any given branch or at the headquarters are determined by the marginal surplus of the firm from employing another worker there. This reflects a solution to a "Rolodex" bargaining game between the firm and its workers as in Brügemann et al. (2019).[74]

H.1.4 Branch-level problem

The firm's problem at the level of the branch is to maximize local sales minus payroll, screening and job-posting costs, and the fixed cost of opening a branch, and is given by

$$\max_{\bar{a},l} E_j^{\frac{1}{\sigma}} P_j^{\frac{\sigma-1}{\sigma}} \left(\varphi_{ij}(z,\phi)\,\bar{a} l^{\varrho}\right)^{\frac{\sigma-1}{\sigma}} - w(\bar{a},l)\,l - P_g\left(c_n n + \delta_0 a_c^{\delta}\right) - \bar{P}_j f$$

Subject to:

$$w(\bar{a},l) = \frac{\partial\left(E_j^{\frac{1}{\sigma}} P_j^{\frac{\sigma-1}{\sigma}} \left(\varphi_{ij}(z,\phi)\,\bar{a} l^{\varrho}\right)^{\frac{\sigma-1}{\sigma}} - w(\bar{a},l)\,l\right)}{\partial l} \quad \text{(Wage setting)} \tag{A.53}$$

$$\bar{a} = \frac{k}{k-1}a_c, \quad l = n\left(\frac{a_{min}}{a_c}\right)^{k} \quad \text{(Worker screening)}.$$

The solution to the bargaining problem yields:

$$w(\bar{a},l) = \frac{\varrho^{\frac{\sigma-1}{\sigma}}}{\varrho^{\frac{\sigma-1}{\sigma}}+1} E_j^{\frac{1}{\sigma}} P_j^{\frac{\sigma-1}{\sigma}} \varphi_{ij}(z,\phi)^{\frac{\sigma-1}{\sigma}} \bar{a}^{\frac{\sigma-1}{\sigma}} l^{\varrho\frac{\sigma-1}{\sigma}-1}.$$

Solving the above problem yields the following expressions for average branch-level ability $\bar{a}_{ij}(z,\phi)$, branch-level employment $l_{ij}(z,\phi)$, and branch-level wages $w_{ij}(z,\phi)$, in terms of

[74] This bargaining protocol is a natural generalization of Nash bargaining to the multiple workers and it modifies the Stole–Zwiebel bargaining protocol so that the resulting wages are Shapley values. See Stole and Zwiebel (1996), Helpman et al. (2010), and Brügemann et al. (2019) for additional details.

branch-level revenues, $r_{ij}(z,\phi)$:

$$\bar{a}_{ij}(z,\phi) = \bar{a}_0 r_{ij}(z,\phi)^{\frac{1}{\delta}}, \quad \bar{a}_0 \equiv \left(\frac{1}{\delta_0}\frac{\varrho^{\frac{\sigma-1}{\sigma}}}{\varrho^{\frac{\sigma-1}{\sigma}}+1}\frac{1-\varrho k}{\varrho\delta}\frac{1}{P_g}\right)^{\frac{1}{\delta}}$$

$$l_{ij}(z,\phi) = l_0 r_{ij}(z,\phi)^{1-\frac{k}{\delta}}, \quad l_0 \equiv \frac{1}{c_n\left(\frac{k-1}{k}\frac{1}{a_{min}}\right)^k \bar{a}_0^k}\frac{\varrho^{\frac{\sigma-1}{\sigma}}}{\varrho^{\frac{\sigma-1}{\sigma}}+1}\left(\frac{1}{P_g}\right)^{1-\frac{k}{\delta}} \tag{A.54}$$

$$w_{ij}(z,\phi) = w_0\left(r_{ij}(z,\phi)\right)^{\frac{k}{\delta}}, \quad w_0 \equiv \frac{1}{l_0}\frac{\varrho^{\frac{\sigma-1}{\sigma}}}{\varrho^{\frac{\sigma-1}{\sigma}}+1}P_g^{1-\frac{k}{\delta}}.$$

And branch-level revenues are given by

$$r_{ij}(z,\phi) = \psi_0 E_j^{\frac{1}{\psi}\frac{1}{\sigma}} P_j^{\frac{1}{\psi}\frac{\sigma-1}{\sigma}} \varphi_{ij}(z,\phi)^{\frac{1}{\psi}\frac{\sigma-1}{\sigma}}, \quad \psi_0 \equiv \left(\bar{a}_0 l_0^{\varrho}\right)^{\frac{1}{\psi}\frac{\sigma-1}{\sigma}}, \tag{A.55}$$

where ψ is a parameter subsuming various constants:

$$\psi \equiv 1 - \frac{\sigma-1}{\sigma}\left(\varrho + \frac{1-\varrho k}{\delta}\right).$$

This yields a similar structure for branch-level revenues in the baseline model. Accordingly, expected firm revenues in market j are now given by

$$\bar{r}_{ij}(z,\phi) = \tilde{\psi} f^{1-\theta}\Upsilon_j\left(A_{ij} z \bar{b}^{\gamma} h^{\vartheta\gamma}\right)^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}}, \tag{A.56}$$

where Υ_j stands a similar market potential term as in the baseline version of the model, and total expected firm sales (conditional on headquarters choices $\bar{b}$ and h) are given by

$$\bar{r}_i(z,\phi) = \tilde{\psi} f^{1-\theta}\left(\sum_j A_{ij}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}}\Upsilon_j\right)\left(z\bar{b}^{\gamma}h^{\vartheta\gamma}\right)^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}}.$$

H.1.5 Firm-level decisions

The firm's problem at the headquarters level becomes to maximize expected operating profits across all markets, minus headquarters payroll, hiring and screening costs:

$$\max_{\bar{b},h} \tilde{\psi} f^{1-\theta} \left(\sum_j A_{ij}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \Upsilon_j \right) \left(z \bar{b}^{\gamma} h^{\vartheta\gamma} \right)^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} - w\left(\bar{b}, h\right) h - P_g \left(c_m m + \delta_0 b_c^{\delta} \right)$$

Subject to:

$$\begin{aligned} w\left(\bar{b}, h\right) &= \frac{\partial \left(\tilde{\psi} f^{1-\theta} \left(\sum_j A_{ij}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \Upsilon_j \right) \left(z \bar{b}^{\gamma} h^{\vartheta\gamma} \right)^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} - w\left(\bar{b}, h\right) h \right)}{\partial h} \qquad \text{(Wage setting)} \\ \bar{b} &= \frac{k}{k-1} b_c, \quad h = m \left(\frac{b_{min}}{b_c} \right)^k \quad \text{(Worker screening)}. \end{aligned} \tag{A.57}$$

Solving the above problem yields the following expressions for average HQ-level ability $\bar{b}_i(z)$, HQ-level employment $h_i(z)$, and HQ-level wages $w_i^{HQ}(z)$, in terms of total firm expected revenues, $\bar{r}_i(z)$:

$$\begin{aligned} \bar{b}_i(z) &= \bar{b}_0 \bar{r}_i(z)^{\frac{1}{\delta}} \\ h_i(z) &= h_0 \bar{r}_i(z)^{1-\frac{k}{\delta}} \\ w_i^{HQ}(z) &= w_0^{HQ} \left(\bar{r}_i(z) \right)^{\frac{k}{\delta}}, \end{aligned} \tag{A.58}$$

where $\bar{b}_0$, h_0, and w_0^{HQ} are constants that subsume model parameters. The firm-wide sales assume a similar expression to the one in the baseline model, $\bar{r}_i(z) = \Lambda_i z^{\frac{\theta}{\iota\psi}\frac{\sigma-1}{\sigma}}$, where now

$$\Lambda_i \equiv \left[\tilde{\psi} f^{1-\theta} \left(\sum_j A_{ij}^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \Upsilon_j \right) \left(\bar{b}_0^{\gamma} h_0^{\vartheta\gamma} \right)^{\frac{\theta}{\psi}\frac{\sigma-1}{\sigma}} \right]^{\frac{1}{\iota}},$$

and

$$\iota \equiv 1 - \gamma \frac{\theta}{\psi} \frac{\sigma-1}{\sigma} \left(\frac{1}{\delta} + \vartheta \left(1 - \frac{k}{\delta} \right) \right).$$

H.1.6 Results

The role of multi-region firms for the structure of wages remains similar in this model to the baseline formulation from Section 3. In the following, I re-state some of the main results from Section 4 in the current setting:

Proposition A.1. *In the model with unobserved worker productivity and costly screening:*

1. *Branch-level wages admit a similar log-linear structure as in Proposition 2 of the baseline model, and depend on a local market effect j, branch-level productivity ϕ,*

firm-level productivity z, headquarters-market effect i, and bilateral HQ-branch frictions ij.

2. *Proposition 3 from the baseline model continues to hold as stated: wage inequality between headquarters and branches is higher for larger firms.*

In addition, this formulation endogenizes branch-level and HQ-level skill intensity, which rise as firms expand in space. I obtain the following result:

Proposition A.2. *The ratio of average worker ability in the headquarters $\bar{b}$ and the average worker ability in branches $\bar{a}$ is higher for larger firms:*

1. *Consider two firms with headquarters in location i and HQ-level productivities z_i' and z_i such that $z_i' > z_i$. Assume that both are active in market j and have the same branch-level productivity shock ϕ_j. Then the ratio of headquarters average skill $\bar{b}_i(z)$ to the average skill in branch j $\bar{a}_{ij}(z, \phi)$ is higher in firm z_i' than in firm z_i.*

2. *Let $r_{ij}(z_i, \phi_j)$ be firm revenues in some market j. The ratio of average worker skill $\bar{b}_i(z)$ to branch wages in the firm's headquarters market $\bar{a}_{ii}(z, \phi)$ is increasing in the ratio of total expected firm revenues across all locations, $r_i(z_i) \equiv \mathbb{E}_{\boldsymbol{\phi}}\left[\sum_j r_{ij}(z, \phi_j)\right]$, to local revenues in the headquarters market $r_{ii}(z_i, \phi_i)$.*