Wage Inequality and the Spatial Expansion of Firms[*][†]

Benny Kleinman[‡]
Stanford University

July 2024

Abstract

Multi-location service firms have experienced substantial geographical expansion in recent decades. They also account for most of the increase in U.S. wage inequality. I develop a theory that links these trends, featuring firms that open branches to serve local markets and hire headquarters workers whose output is non-rival across branches. The model rationalizes multiple labor-market trends through changes in firms' geographical scope, including endogenous skill-biased technical change, increasing spatial disparities, and growing wage dispersion across and within firms. I provide supporting evidence for these predictions and the assumption of within-firm non-rivalry, and quantitatively demonstrate the aggregate implications of these mechanisms.

[*]I am grateful to my advisors Gene Grossman, Ezra Oberfield, Stephen Redding, and Richard Rogerson for their continued guidance and support, as well as Esteban Rossi-Hansberg and Ernest Liu for detailed and insightful comments. This paper has also benefited from discussions and comments from Mark Aguiar, Juanma Castro-Vincenzi, Oren Danieli, Xiang Ding, Pablo D. Fajgelbaum, Teresa Fort, Manuel García-Santana, John R. Grigsby, Erik Hurst, Gregor Jarosch, Marta Lachowska, Hugo Lhuillier, Isabela Manelici, Simon Mongey, Joan Monras, Eduardo Morales, Pascual Restrepo, Hannah Rubinton, Jose P. Vasquez, Gianluca Violante, Jonathan Vogel, and Conor Walsh, as well as numerous seminar participants. I thank Shahin Davoudpour, Shirley Liu, and Hannah Rubinton for help with the data. This research benefited from financial support from the International Economics Section (IES) and the Simpson Center for the Study of Macroeconomics in Princeton University, the Washington Center for Equitable Growth, and the Becker Friedman Institute for Economics at University of Chicago.

[†]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. (CBDRB-FY22-P2542-R9788, CBDRB-FY22-P2542-R10063, CBDRB-FY24-P2542-R11261)

[‡]Stanford University, bennyk@stanford.edu.

1 Introduction

The U.S. economy is increasingly dominated by multi-location 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 provide a new theory that links changes in firms' geographical scope to the distribution of wages in the economy. The resulting framework is able to rationalize multiple labor market trends from recent decades, providing new micro-foundations for skill-biased technical change, increasing spatial disparities, and increasing inequality across firms and establishments.

I conceptualize multi-location service firms as networks of local branches linked by national headquarters, in which the output of branch workers is *non-tradable*, and the output of headquarters workers is *non-rival* across the firm's locations.[1] Examples include retail and restaurant chains, banks, healthcare networks, real-estate management companies, and telecommunications firms. I argue that with such structure, changes in firms' spatial scope can have non-trivial and rich distributional implications, and I investigate this idea in several steps. First, I provide motivational 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 model of these firms in a spatial general equilibrium setting, and use it to show how shocks that raise firms' ability to expand in space can lead to endogenous skill-biased technical change, to increases in residual wage inequality, and to increases in spatial disparities. Third, I provide reduced-form evidence for the key mechanisms and implications of the model, including evidence for the assumption of within-firm non-rivalry. Finally, I employ a quantitative version of the model to answer: (a) How far can we go in explaining labor market trends from recent decades by simply lowering spatial frictions to firm expansion? (b) What are the implications of policies that alter firms' expansion opportunities, such as the deregulation of cross-state firm activity?

I begin by documenting key motivating facts about multi-establishment service firms in the U.S. economy since the 1980s. First, I show that these firms have experienced substantial spatial expansion. Second, these firms account for most of the increase in the variance of wages across the universe of U.S. establishments. Third, their sizable role in the rise of inequality reflects both higher variance across multi-establishment service firms and higher variance across different establishments within these firms. Moreover, the increase in variance across establishments within multi-establishment firms is significantly larger than across the universe of stand-alone establishments. Therefore, frameworks that focus solely on inequality across firms or solely on inequality across establishments are not well suited to study these patterns, and a model of the internal structure of multi-

[1] Throughout the paper, I employ a wide notion of headquarters, which encompasses all firm's workers whose output is non-rival across space. I later argue that such workers are indeed highly spatially-concentrated within firms, and that actual headquarters locations provide a good proxy for this notion.

establishment firms is required.

I then develop a spatial general equilibrium model with the above form of organization at its core. The model has four main components. First, firms open branches and hire local workers to provide a non-tradable product across local labor markets (e.g. a bank branch, a restaurant, a maintenance facility, a warehouse, etc.). Second, firms can improve the productivity of their branches by hiring high-skill workers at a national headquarters, such that the output of their headquarters workers is *non-rival* across the firm's locations.[2] Third, firms face a cost of spatial expansion, which is higher when opening more and further-away branches. This cost also serves as a key exogenous determinant of the degree of spatial expansion in the economy. Lastly, firms have wage-setting power, due to upward-sloping, firm-specific labor supply curves in each location. The allocation of workers, headquarters, and firm branches across space are all determined in equilibrium.

I emphasize three main takeaways from the model. The first takeaway is a new microfoundation for skill-biased technical change (SBTC). Similarly to standard models in this literature following Katz and Murphy (1992) and Krusell et al. (2000), firms' relative expenditure on skilled (headquarters) workers depends on relative factor prices and on exogenous biased technical change. However, in the current setting, firms' expenditure on skilled workers also depends on their chosen geographical scope. A higher geographical scope is similar to a technical change that augments factors of production which are used intensively in the firm's non-rival component. Intuitively, due to the non-rivalry assumption, the marginal product of headquarters workers scales with the number of locations in which their output can be applied, whereas the marginal product of branch workers does not. This effect holds even in simple cases of the model with symmetric space and perfectly competitive labor markets, and relies on the assumption that the output of a subset of workers is non-rival across the firm's locations.

The second result that I highlight is a characterization of firm wages when labor supply is firm-specific and upward sloping, e.g. due to monopsonistic labor markets, as in Card et al. (2018) and Berger et al. (2022). In this case, there is residual wage inequality conditional on workers' skill group and location, and different firms and establishments pay different wages. As is standard in this literature, larger firms pay higher wages on average. However, in the current setting, larger firms are also characterized by higher wage dispersion, in particular between their headquarters and branches. Moreover, despite the existence of such upward-sloping labor supply curves, I show that firm expansion does not necessarily lead to higher wages for branch workers. While expansion generates a positive productivity effect that raises demand for workers across the firm's locations, it can also make local production more dependent on the firm's non-rival component. Consequently,

[2] Consider for example the output of a firm's designers, software engineers, marketing specialists, and researchers. These are indeed some of the top occupations that are typically hired in firms' headquarters, as I demonstrate later.

the size gradient of wages is steep for headquarters workers, but can be very flat for branch workers; and firm-level shocks pass-through more to headquarters wages than to branch wages.

The final theoretical takeaway that I highlight is a link between firms' spatial scope and rising spatial disparities, or "the great urban divergence" in Moretti (2012). In the model, regional skill intensity depends on two endogenous determinants. The first is regional specialization in headquarters services (as opposed to specialization in production of tradable goods). This specialization raises regional demand for skill when the non-rival part of the service sector is more skill-intensive than the tradable-goods sector as a whole. The scope for such specialization is tightly linked to firms' ability to expand in space. For example, in the limit economy when cross-region firm activity is prohibited, the non-rival firm output is used only locally, and there is no room for specialization. The second determinant of regional skill intensity is the importance of the non-rival component in the production of local firms, which itself rises with firms' geographical scope. Through these two channels, an increase in firms' spatial scope can lead to greater spatial disparities in the economy.

I provide reduced-form evidence for the model's assumptions and implications. I start by showing evidence for the key assumption of within-firm non-rivalries. An important implication of this assumption is that positive demand shocks in a subset of the firm's locations spillover into greater activity in all its branches, including those in unaffected locations. This spillover arises since firms respond to local demand shocks by increasing spending on non-rival inputs, which raise the productivity across all the firm's locations. I utilize U.S. Census data on firms' activity across space, and test how a firm's activity in each market responds to exogenous demand shocks in its other markets. I find strong positive spillovers across markets within firms, in line with the existence of strong non-rivalries.[3]

I also provide empirical evidence in support of the model's key predictions, leveraging multiple datasets and sources of variation. To investigate disparities between firms' headquarters and branches, I employ two approaches. The first approach is to measure headquarters activity from occupational classifications. To this end, I construct a measure of occupational headquarters intensity by combining data on firms' geography and their job postings for different occupations across space. In line with the model's predictions, I show evidence for within-sector reallocation of economic activity towards more headquarters-intensive tasks, and particularly so in high-expansion sectors. Rel-

[3]The model also includes two forces that can work in the opposing direction. First, I allow for firm-level convexities in the cost of expansion, which implies that more expansion in one market raises the cost of expansion in other markets. Second, if the elasticity of substitution between headquarters and branch workers is particularly high, more expansion opportunities in one market could reallocate production away from all markets to the headquarters. These empirical results suggest that the non-rivalry is strong enough relative to these other forces.

ative wages rise more for headquarters-intensive occupations, in line with the model's predictions. The second approach is to measure the distinction between headquarters and branch activity using the industry classification of individual establishments in the U.S. Census data. In line with the model's predictions, I show that as firms expand in space, firms reallocate expenditure from their branches to their headquarters, and the relative wage gap between them rises. Finally, I test the model's predictions across local labor markets, and show that commuting zones that have experienced greater national expansion of locally-headquartered firms (i.e., more expansion of local firms in other markets), have seen greater increase in their wages, skill-intensity, and share of headquarters activity in total output.

In the last part of the paper, I use a quantitative version of the model to demonstrate the aggregate relevance of the new mechanisms. I estimate the model for 200 local labor markets of the contiguous U.S. in 1980. I estimate frictions to firm spatial expansion from the universe of headquarters-branch linkages in the data. To estimate the main parameters of the production function and the cost of expansion, I employ a Simulated Method of Moments (henceforth SMM) approach, targeting key moments relating to firm structure and inequality from my empirical analysis; and inverting region-level fundamentals from the model's equilibrium conditions.

The key thought experiment that I perform is a realistic reduction in the cost of opening far-away branches. I infer this shock from changes in the network of headquarters-branch linkages over time. I show that these linkages have experienced a big decline in their distance-elasticity between 1980-2017, in line with the narrative that improvements in communication and transportation infrastructure have lowered barriers to operate distant branches. I thus only use information on changes in the *relative* number of far-away branches to nearby branches, and not on their absolute number. In practice, I reduce the same distance elasticity in the model, holding all other primitives constant. This experiment has a few key advantages. First, all the implications to labor markets and welfare materialize via the novel mechanisms in this paper, since this shock affects only the cost of spatial expansion. Second, while many primitives in the model can induce an increase in firms' geographical scope, we can easily discipline this shock from data on the relative number of far-away branches within firms. Lastly, it has a clear interpretation of lower spatial frictions to operate establishments across space.

Quantitatively, I find that such a decline in the cost of spatial expansion leads to rich labor market implications that align with much of the experience of the U.S. economy since the 1980s. First, there is an increase in wage inequality (as measured by the variance of log wages), by around a third of the equivalent within-industry change in the data. Second, as in the data, inequality rises both within firms – across their different establishments – and between different firms. Third, an important component of the increase in inequality within firms is a higher wage dispersion between headquarters and branch workers. Fourth, it leads to an increase in spatial inequality and skill-segregation, rationalizing around a third of the empirical rise in spatial disparities. Fifth, wage gains are greater in markets that ex-ante specialize in providing headquarters services, which tend to be the dense markets, raising the urban wage premium. At the same time, this shock is associated with positive welfare gains of about 4%, reflecting variety gains from firms' greater scope and productivity gains that result from the non-rivalry in production.

As a final exercise, I demonstrate how the model can be used to evaluate the aggregate and distributional consequences of policies that shape firms' ability to expand in space. One such example from recent decades is the deregulation of cross-state firm activity over the 1980s and the 1990s in some key service sectors, notably financial services, transportation, and utilities. This deregulation was one potential reason for the rise in firms' geographical scope over this period. I replicate this experiment in the model by lowering the importance of state border effects, which I measure using a gravity equation for headquarters-branch linkages in the data. I find that a similar reduction in the importance of state borders in the model results in average welfare gains of 0.8%, with a slight increase in firms' scope and wage inequality.

This paper connects to several strands of related literature. First, I relate to the recent literature that documents the spatial expansion of firms, such as Cao et al. (2017), Jiang (2021), Rossi-Hansberg et al. (2021), Aghion et al. (2023), and Hsieh and Rossi-Hansberg (2023). Building on this literature, I turn to study the labor market implications of firm expansion. I argue that changes in firms' geographical scope can have non-trivial distributional implications when combined with the idea of within-firm non-rivalries, especially in an environment with firm wage-setting.

Second, I relate to multiple strands of the vast literature on rising wage inequality. Relative to the literature on skill-biased technical change (henceforth SBTC) – following Katz and Murphy (1992) and Krusell et al. (2000) – I contribute a new micro-foundation for endogenous SBTC that is based on the combination of within-firm non-rivalries and changes in firms' scope. Relative to the literature on "the great divergence" in spatial economics – 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 that relies on spatial firm expansion and connects to other trends in the broad inequality literature. I also provide theory and evidence on the importance of cross-region firm linkages in determining the spatial allocation of economic activity. Relative to the empirical literature that documents rising inequality across firms in the U.S. – Barth et al. (2016), Song et al. (2019) and Haltiwanger et al. (2022) – I provide new evidence on the central role of multi-establishment service firms, emphasizing a new important margin of inequality across different establishments within these firms, and provide a theory that rationalizes these patterns. Finally, relative to papers that study wage inequality in an environment with monopsonistic competition – as in Card et al. (2018) and Berger et al. (2022) – I provide an environment in which firm size matters differently for skilled and unskilled workers and which allows for heterogeneous pass-through of firm-level shocks to different workers.

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. (2024), 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. The key deviation from this literature is to model the firm's non-rival component as a variable labor cost (and not just as fixed costs). This has important distributional implications when firms change their geographical scope, even when space is symmetric. I also deviate from most of this literature by allowing for imperfectly-competitive labor markets, and highlight important interactions of this margin with firms' geographical scope. Finally, I integrate the expansion of these firms into a spatial general equilibrium in which the allocation of headquarters, branches, and workers are all determined in equilibrium. I also relate to the theoretical literature that connects within-firm wage inequality to firms' hierarchical organization, e.g. Garicano and Rossi-Hansberg (2004) and Garicano and Rossi-Hansberg (2006). My notion of headquarters is not a hierarchical layer, but rather it captures tasks for which the output is non-rival across space, which allows me to encompass broader patterns of inequality; and it provides a natural link between firms' organization to the literature on rising spatial disparities.

I also 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. (2023), Ding et al. (2022), and Ding (2023). I explicitly model this knowledge as the output of firms' (headquarters) workers, and highlight the distributional implications that arise when there is a change in firms' ability to spread knowledge across space due to spatial expansion. I also use the structure of the model to overcome the challenge of measuring within-firm knowledge flows and recover frictions to within-firm communication using the empirical network of HQ-branch linkages.

The remainder of the paper is structured as follows. I start with some simple motivational facts in Section 2, highlighting the spatial expansion of firms in recent decades and the centrality of multi-location service firms in the growth of wage inequality. In Section 3, I lay out a model of these firms in spatial general equilibrium, and characterize the link between changes in firms' scope to skill-biased technical change, firm specific wages, and spatial disparities. In Section 4, I provide empirical evidence for the key predictions of the model and for the assumption of within-firm non-rivalries. I estimate the model in Section 5 and use the quantitative model to analyze the aggregate implications of lower frictions to firm expansion in Section 6. Section 7 concludes.

2 Motivating facts

I start by briefly describing key macro-trends for multi-location service firms in the U.S. economy since the 1980s. I demonstrate the substantial expansion of these firms and their key role in the increase of U.S. wage inequality. I use data from the U.S. Census Bureau, mainly the Longitudinal Business Database (LBD) and the Business Dynamics Statistics (BDS) - see additional details in Appendix D.1. I focus on firms in service sectors,[4] which constitute most of the economy. As detailed later, I find very different patterns for goods-producing firms.

2.1 The spatial expansion of service firms

First, I revisit key facts on the expansion of service firms in recent decades.[5] Figure 1 depicts the increase in firms' scope between 1980-2017, as measured by the number of establishments per firm. Subplot (a) presents a simple unweighted average of this metric. Evidently, the average firm operates today around 15% more establishments than in 1980. Note, however, that when we are interested in aggregate labor market outcomes, the employment-weighted expansion might be a more relevant metric, as later demonstrated in the model. Accordingly, subplot (b) divides all firms into four *employment-weighted* bins for each year, and computes the average number of establishments per firm in each bin. This exercise reveals a much more substantial expansion pattern. The majority of U.S. service workers are now employed by firms that are, on average, more than 100% larger than in 1980. Furthermore, about a quarter of workers are employed by firms that are, on average, 500% larger than in 1980. Averaging across these four quartiles, the solid red line of subplot (c) demonstrates that the typical worker is now employed by firms with around three times more establishments than in 1980. Finally, subplot (c) contrasts the change in the average number of establishments per firm with changes in firm size, as measured by total employment per firm (dashed gray line). Evidently, the increase in firms' scope accounts for most of the increase in the size of service firms over this period; firms are now larger because they operate more establishments. More evidence for this point is provided in Appendix D.2.

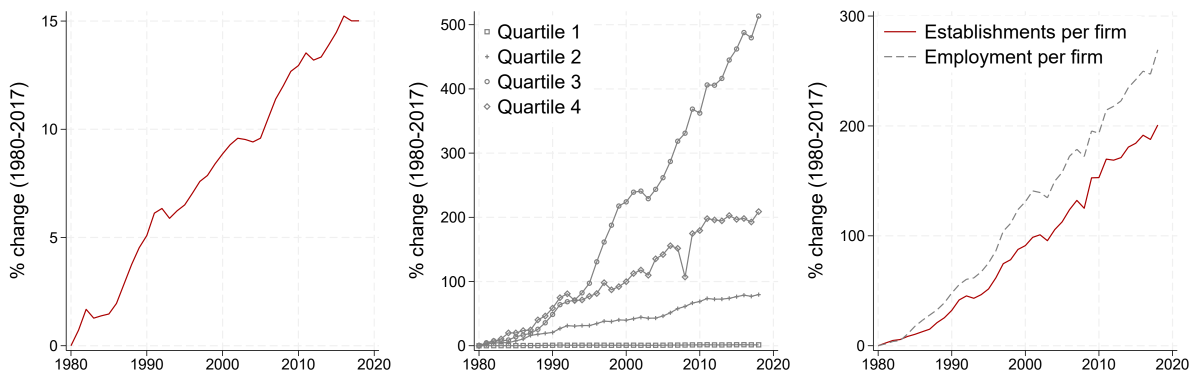
Some additional features of this trend are worth noting. First, it is widespread across the different service sectors and is not driven by any particular sector. The heterogeneous patterns across sectors can be seen in Appendix D.2. Second, over half of the increase in the average number of establishments per firm is due to expansion across local labor

[4] I classify a firm as belonging to the services sector if over half of its payroll across all its establishments is accrued in services-sector establishments. I classify establishments as belonging to the services sector if their 2-digits NAICS code falls under the Bureau of Economic Analysis (BEA) classification of services-producing sectors (i.e. 1 digit NAICS code of 4, 5, 6, 7, or 8.)

[5] Related patterns have been documented also in Cao et al. (2017) and Hsieh and Rossi-Hansberg (2023).

Figure 1: The increase in the number of establishments for service firms (1980-2017)

(a) Change in firm scope **(b)** Change in scope by quartile **(c)** Firm size and scope



Note: Subplot (a) shows the percentage change relative to 1980 in the (unweighted) average number of establishments per firm in the U.S. Census BDS data for service sectors. Subplot (b) shows the same metric when dividing the universe of firms in each year to four employment-weighted quartiles. Subplot (c) shows changes in the average across these four quartiles for the number of establishments per firm (solid red line) and employment per firm (dashed gray line).

markets (i.e., across U.S. commuting zones), though there is non-trivial expansion also within local markets. The model below will allow such firm expansion both across and within local markets. Third, note that in contrast to services, average firm size and average number of establishments per firm in the goods sector have both declined.[6] For the typical goods-producing firm, I find that the average number of establishments per firm has decreased by around 7% between 1980-2017. Further details on the comparison between goods- and service-producing firms are provided in Appendix D.

Finally, the above trend is associated with a substantial reallocation of workers towards multi-establishment service firms. This class of firms has accounted for 29% of the total U.S. workforce in 1980. By 2017, this share increased to 48%. This reallocation reflects both the general transition from goods to services ("structural transformation") and the reallocation from single-establishment to multi-establishment firms within services.

2.2 Multi-location service firms and wage inequality

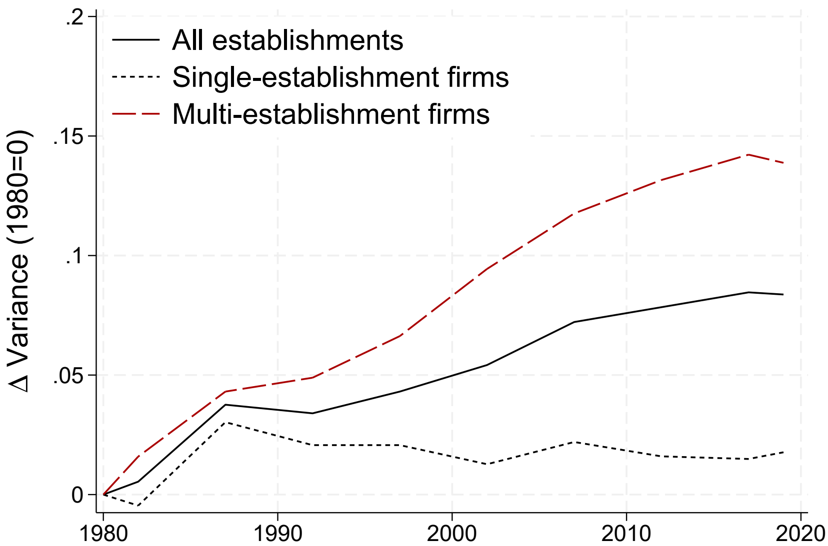
I now show that the same class of firms also accounts for most of the increase in U.S. wage inequality. To this end, I present a simple decomposition of the rise in the variance of log wages across the universe of U.S. establishments. This increase in wage dispersion across establishments has been shown to account for the vast majority of the increase in

[6] 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 local 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.

wage inequality across U.S. workers (see e.g. Barth et al. (2016)), with minimal role for changes in within-establishment variance. I also confirm these findings in Appendix D.3.[7]

I investigate which firms drive the increase in the variance of log wages across establishments in the LBD, starting with a visual representation of the importance of multi-establishment firms. Figure 2 shows the change in within-industry variance of log wages across all establishments (solid black line), and then shows the same change for two separate subgroups: establishments that belong to multi-establishment firms (dashed red line) and establishments that belong to single-establishment firms (dotted gray line). The figure shows a sharp increase in this 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. Note that I focus here on within-industry inequality since this will be the main moment of interest when I turn to the quantitative model, but the same pattern holds when not demeaning industry fixed effects, as can be seen in Appendix D.3.

Figure 2: The role of multi-establishment firms in the rise of wage inequality



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 1980, after demeaning industry fixed effects (4 digits NAICS code) from establishment log wages. The change in variance for the universe of all establishments is given by the solid-black line. The dashed-red line shows this change for establishments in multi-establishment firms, and the dotted-gray line for single-establishment firms.

I formally quantify the contribution of multi-establishment firms to the rise in inequality using a variance decomposition of log wages. To this end, consider a partition of the

[7] As detailed below, it is important to distinguish between wage dispersion across firms and wage dispersion across establishments. In particular, most of the increase in inequality within firms is also across different establishments. Therefore, the fact that almost all of the increase in inequality is across establishments does not contradict the sizable role that has been found for rising wage dispersion within firms, e.g. in Song et al. (2019).

universe of establishments into G distinct groups. The total change in variance equals to the sum of several components:

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

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; $\sigma_{g,t}^2$ 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. establishments belonging to multi-establishment firms, multiplied by its initial share of total employment. The second term captures reallocation of employment across groups, keeping each group's variance constant 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, comprising the rising variance within all other groups $g \in G/g'$ and rising variance between groups.

Table 1 shows the results from this decomposition for the overall rise in within-industry wage inequality in the U.S. economy over 1980-2017. Column (1) considers this decomposition when singling out multi-establishment firms as the group g'. Consistent with the pattern in Figure 2, Over 80% of the rise in overall (within-industry) wage inequality is due to rising variance across establishments that belong to multi-establishment firms.

I now employ the same decomposition to highlight two additional points about the role of multi-establishment firms in the rise of inequality. First, I show that most of the increase in inequality is driven by firms in services-producing sectors. To see this, Column (2) repeats the decomposition in 1 when singling out the group of service-sector firms. In this case, 70% of the increase is accounted by rising variance for services firms, and 16% due to reallocation to services. Second, I highlight the particular importance of multi-establishment service firms. To see this, Column (3) repeats the decomposition in 1 when narrowing down the focus group to only multi-establishment service firms. In this case, 45% of the overall increase in inequality is due to rising inequality across establishments that belong to this group, and 22% is due to the cross effect arising from the fact that this group became both more unequal and a much larger part of the economy. The first term is smaller relative to Column (1) since as mentioned above, multi-unit service firms accounted for only 29% of total employment in 1980. Still, even holding this share

Table 1: Decomposition of the rise in wage inequality

	(1)	(2)	(3)
Group of firms:	**Multi-unit firms**	**Service firms**	**Multi-unit service firms**
Share out of the overall rise in variance of log wages:			
Rising variance within the group of firms	83%	70%	45%
Changes in employment shares b/w groups (reallocation)	-4%	6%	0%
Comovement of variance and employment shares	9%	10%	22%
Residual	12%	13%	33%
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 2.2 for additional details.

constant, the rising variance in this group accounts for 45% of the overall change. Together with the reallocation of employment to this group, it accounts for 67% of the total rise in inequality.

The role of wage dispersion within firms. Having established that multi-establishment service firms play a key role in the rise of wage dispersion in the economy, a natural question is whether their role is driven by rising dispersion across these firms, or alternatively, across different establishments within them? The answer is that both parts are important. To see this, I decompose the above increase in the variance for multi-unit service firms, $\Delta\sigma^2_{\text{MU-serv},t}$, to changes within firms and changes between firms:

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

where w_{fj} is the average wage in establishment j in firm f and $\ln \bar{w}_f$ is the employment-weighted average of log wages across all establishments in firm f. Rising differences within firms – across their different establishments – explain around 45% of the overall change $\Delta\sigma^2_{\text{MU-sserv},t}$.

While the model presented below does not focus exclusively on this part of growing inequality (and addresses rising dispersion across firms as well), establishing this fact is useful for two reasons. First, it highlights that it is not enough to have a theory with single-unit firms, and that one needs to model the firm's different establishments to get at the full picture of rising inequality. Second, it contrasts the theory in this paper relative to some other firm-based theories of rising inequality. For example, an increase in domestic outsourcing has been suggested as one potential explanation for rising dispersion across firms. If anything, rising outsourcing should lead to lower wage dispersion *within* firms, as firms increasingly focus on their core activities. However, in the data, we observe rising inequality both between and within firms. The theory presented in this paper is well suited to address these patterns.

To recap, I have shown that: (a) multi-establishment service firms have experienced substantial spatial expansion between 1980-2017; (b) they account for most of the increase in wage dispersion in the economy over this period; (c) a big part of their role in due to rising dispersion across different establishments within these firms. Taken together, these evidence provide the motivation to model multi-location firms and their expansion in the analysis of growing inequality.[8]

3 Model

In this section I develop a model to formalize the distributional consequences of firm expansion in the presence of within-firm non-rivalries. I start by setting up the firm's problem in partial equilibrium to highlight the novel components of the model. I then add the households side of the economy – which takes a standard form – and complete setting up the spatial general equilibrium.

3.1 General setting

The economy consists of N heterogenous regions (e.g., commuting-zones), indexed by i and j. Each region consists of a unit-mass of identical locations. There are two sectors: services and tradable-goods. 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 workers in local branches; while another part is tradable and generated by hiring workers at a potentially distant headquarters.

[8]Note that in contrast to other decompositions in the literature on wage inequality, the goal of the above decomposition is not to separate between workers' characteristics and residual inequality. In fact, the model below will accommodate both. Rather, it is to demonstrate *which class of firms* accounts for most of the increase in wage dispersion in the economy, and to motivate my focus on the expansion of multi-location firms and on their internal structure.

3.2 The problem a service firm

Firms in the service sector differ by their headquarters location i and baseline productivity z. I use subscripts to denote locations, and refer to the firm's productivity in parentheses. I use $t_{ij}(z)$ to denote the value of any variable t for a i-headquartered firm with productivity z in its region j branches, and $t_i(z)$ to denote variables at the level of the firm.

Demand. Each firm (i, z) faces a downward-sloping isoelastic demand curve in each location with demand elasticity σ, implying that the firm's sales in region j are given by

$$r_{ij}(z) = \Upsilon_j (q_{ij}(z))^{\frac{\sigma-1}{\sigma}}, \tag{2}$$

where Υ_j captures a regional demand shifter, $r_{ij}(z)$ is the firm's sales in each of region j's locations, and $q_{ij}(z)$ is the quantity supplied by the firm.

Production. A firm (i, z) can supply $q_{ij}(z)$ units of its output in region j using the production function

$$A_i z \left(\gamma h_i(z)^{\frac{\eta-1}{\eta}} + \ell_{ij}(z)^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}}, \tag{3}$$

where z is the productivity of the firm across all its markets; A_i is a productivity term common to all firms with headquarters in location i; γ is a parameter capturing the relative productivity of headquarters-level workers; and η is the elasticity of substitution between the headquarters-level and branch-level workers. Importantly, the firm faces this production function in each location j, and the bundle of headquarters workers $h_i(z)$ is common and non-rival across locations.

For example, consider the case of Starbucks branches in San Francisco. In this case, z captures the productivity of Starbucks across all its locations; j stands for San Francisco; i stands for Seattle, the headquarters location of Starbucks; A_i captures the productivity of all Seattle-based firms; $\ell_{ij}(z)$ stands for Starbucks workers in its San Francisco branches; and $h_i(z)$ stands for Starbucks headquarters workers in Seattle, such as Starbucks designers, food-scientists, and programers.

Labor supply. In each location, the firm faces a convex labor supply curve, such that in order to employ $\ell_{ij}(z)$ units of branch-level labor in a region j location, it must pay a wage given by

$$w_{\ell,ij}(z) = W_{\ell j} (\ell_{ij}(z))^{\frac{1}{\epsilon}}, \tag{4}$$

where $W_{\ell j}$ is a labor-supply shifter for branch-level workers in market j; ϵ is the firm-level labor supply elasticity; and $\ell_{ij}(z)$ is the amount of labor employed by the firm. When $\epsilon \to \infty$, the firm's payroll in market j is linear in the number of workers $(W_{\ell j}\ell_{ij}(z))$, in which case $W_{\ell j}$ captures the competitive wage level in market j for branch-level (ℓ)workers. More generally, $W_{\ell j}$ is a local wage index capturing the degree of labor-market competition in market j.

The firm also faces a similar labor supply curve of headquarters-level workers in market i. To hire $h_i(z)$ of headquarters-level workers, it pays a wage $w_{h,i}(z) = W_{hi}(h_i(z))^{\frac{1}{\epsilon}}$, where W_{hi} is the labor-supply shifter for headquarters-level workers in market i.

Below, I micro-found the case of a finite ϵ using a monopsonistic labor market environment in which workers have idiosyncratic preferences for firms, as in Card et al. (2018). However, other micro-foundations for such convex labor supply curves, such as screening of higher ability workers as in Helpman et al. (2010), efficiency wages as in Davis and Harrigan (2011), or fair wages as in Amiti and Davis (2012) would also be compatible with the main results.

Market penetration. I allow firms to decide on the degree of market penetration in each region j, in the spirit of Arkolakis (2010) and Oberfield et al. (2024). Specifically, firms choose to serve a random subset $x_{ij}(z) \in [0,1]$ of the identical locations in each region j. Firms do not serve all locations since expansion is costly, with the cost of expansion given by a generic function $C\left(\{x_{ij}(z)\}_{j=1}^{N}\right)$, which depends on the vector of the firm's market penetration decisions in all regions. The firm's total sales in region j is thus given by $x_{ij}(z)r_{ij}(z)$, and its wage bill in region j is given by $x_{ij}(z)w_{\ell,ij}(z)\ell_{ij}(z)$.

A special case of this formulation that follows much of the literature on trade with heterogenous firms is when firms pay a fixed cost to enter each region. In this case, $C(\cdot)$ is the sum of these fixed costs, and firms either fully serve a region or do not serve it at all ($x_{ij}(z) \in \{0,1\}$). More generally, I allow $x_{ij}(z)$ to vary continuously between 0 and 1. This formulation has a few advantages. First, it features spatial expansion both across and within regions – an important feature of the data – while still preserving cross-region heterogeneity. This allows me to later fit the model to changes in the number of establishments per firm. Second, it allows me to consider marginal changes in firms' geographical scope, which is useful when characterizing the link to wage inequality. Third, it aids with the quantitative computation of the model, since under suitable functional form for $C(\cdot)$, it is possible to employ an iterative algorithm to compute $\{x_{ij}(z)\}$.

For tractability, I assume that $C(\cdot)$ is paid with freely tradable goods, such that the cost function faced by a firm is independent from its headquarters location. This allows me to avoid introducing another source of spatial heterogeneity, although pricing it in terms of the local price index is a straightforward extension. In addition, note that if $C(\cdot)$ includes fixed costs and/or negative externalities across regions (e.g. a span of control in the total number of branches, as I allow in the quantitative section), then firms might choose $x_{ij}(z) = 0$ for some regions. Finally, as I elaborate in Section 5 below, I include spatial frictions to firm expansion in this cost function, to match the gravity pattern of headquarters-branch linkages in the data.

The firm's problem. Combining the above elements, the firm's problem becomes

$$\begin{aligned}\max_{h_i(z),\{\ell_{ij}(z)\},\{x_{ij}(z)\}} &\sum_{j=1}^{N} \underbrace{x_{ij}\left(z\right)\Upsilon_j\left(A_i z\left(\gamma h_i\left(z\right)^{\frac{\eta-1}{\eta}}+\ell_{ij}\left(z\right)^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}}\right)^{\frac{\sigma-1}{\sigma}}}_{\text{Sales in region } j} \\ &-\sum_{j=1}^{N} \underbrace{x_{ij}\left(z\right)W_{\ell j}\ell_{ij}\left(z\right)^{\frac{\epsilon+1}{\epsilon}}}_{\text{Labor costs in region } j} - \underbrace{W_{hi}h_i\left(z\right)^{\frac{\epsilon+1}{\epsilon}}-C\left(\left\{x_{ij}\left(z\right)\right\}_{j=1}^{N}\right)}_{\text{Headquarters labor and expansion costs}}.\end{aligned} \tag{5}$$

Firms maximize the sum of sales across all regions, net of labor costs at all branches, labor costs at their headquarters, and the cost of expansion. They choose the mass of branch-level workers $\ell_{ij}(z)$, headquarters workers $h_i(z)$, and market penetration $x_{ij}(z)$, internalizing that they face a downward sloping demand curve (captured by σ) and an upward sloping labor supply curve (captured by ϵ) in each location. They take as given labor supply shifters $W_{\ell j}, W_{hi}$ and demand shifters Υ_j in all locations – all of which are determined in equilibrium – as well as the exogenous productivity shifter A_i and the cost function $C(\cdot)$.

3.3 Households and other model components

I now layout the households block and the rest of the model's components, which are rather standard in the literature. In doing so, I also provide one particular set of micro-foundations for the labor supply shifters $W_{\ell j}, W_{hi}$ and the demand shifters Υ_j from the firm's problem.

Households. There are S types of workers who differ in skill and are indexed by s, with an aggregate supply of $\bar{L}_s$ workers of type s. For now I set $S = 2$ and assume that headquarters workers correspond to high-skilled workers, and that branch-level workers correspond to low-skilled workers. In the quantification of the model below, I generalize this and allow for more skill-groups and for mixes across skill groups in both branch and headquarters production.

All households consume a bundle of services, Q, and a bundle of tradable-goods, G. I add the tradable-goods sector to allow for regional specialization in headquarters services.[9] Households also choose their region $i \in \{1, ..., N\}$ and employer $\nu \in \mathcal{V}_i$, where $\mathcal{V}_i$ is the set of employers in region i, i.e., the set of headquarters and branches of locally-active firms. The bundle of services Q is a constant elasticity of substitution aggregator across locally provided varieties, i.e., across the set of firms that open branches in region i, taking into account each firm's market penetration in region i. Each household of type s solves the following problem, where to avoid overburdening the notation, I omit indexes for specific

[9] Such specialization is not possible under balanced trade without an additional sector due to the assumption that the output of branches is not-tradable.

households:

$$\max_{i\in\{1,\ldots,N\},\nu\in\mathcal{V}_i,G,\{q(\omega)\}_{\omega\in\Omega_i}} b_{i\nu}Q_{si\nu}^{\beta}G_{si\nu}^{1-\beta}, \quad Q_{si\nu}=\left[\int_{\omega\in\Omega_i}\left(q_{si\nu}\left(\omega\right)\right)^{\frac{\sigma-1}{\sigma}}d\omega\right]^{\frac{\sigma}{\sigma-1}}. \tag{6}$$

A household of type s chooses a region i, employer ν, quantity of tradable goods $G_{si\nu}$, and a quantity of services $Q_{si\nu}$ to maximize its idiosyncratic preference shock for region i and employer ν, $b_{i\nu}$, and a Cobb–Douglas aggregator of goods and services with an expenditure share $\beta\in(0,1)$ on services. In turn, $Q_{si\nu}$ aggregates across locally-provided varieties $\omega\in\Omega_i$ with an elasticity of substitution σ, where Ω_i is the set of varieties provided in region i.[10]

Regional price indices and demand shifters. I denote the price index for local services by P_i, and the price index that aggregates goods and services by $\bar{P}_i$. The services price index is the standard ideal CES price index, given by $P_i=\left(\int_{\omega\in\Omega_i}P_i\left(\omega\right)^{1-\sigma}d\omega\right)^{\frac{1}{1-\sigma}}$. I choose the price of goods as the numeraire, such that $\bar{P}_i=P_i^{\beta}$. 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 $\bar{P}_i$. The demand shifters that firms face in each market from Equation 2 are then given by $\Upsilon_j=E_j^{\frac{1}{\sigma}}P_j^{\frac{\sigma-1}{\sigma}}$, where E_i is regional expenditure on services (a share β of total regional income).

Labor supply. Households draw the set of idiosyncratic shocks $b_{i\nu}$ from a nested Fréchet distribution, which guides their sorting decisions across space and across employers.[11] The upper nest reflects preferences across locations with dispersion given by ξ, capturing a regional labor supply elasticity. The lower nest reflects preferences across employers with dispersion given by ϵ, capturing an employer-level labor supply elasticity.[12]

Consequently, the probability that a household of type s chooses region i is increasing in the effective regional real income for type s, and given by

$$\frac{L_{is}}{\sum_{j=1}^{N}L_{js}}=\frac{\left(W_{si}/\bar{P}_i\right)^{\xi}}{\sum_{j=1}^{N}\left(W_{sj}/\bar{P}_j\right)^{\xi}}, \tag{7}$$

[10]Note that Ω_i is the mass of varieties faced by each household, taking into account the market penetration decisions of firms. Since each region is a unit-mass of identical locations, each household in region i has a probability $x_{oi}(z)$ to access varieties of an o-based firm with productivity z that chooses market penetration $x_{oi}(z)$ in region i. The quantity index in region i can thus also be written as $Q_i=\left[\sum_{n=1}^{N}\int_z x_{ni}\left(z\right)\left(q_{ni}\left(z\right)\right)^{\frac{\sigma-1}{\sigma}}dG_n\left(z\right)\right]^{\frac{\sigma}{\sigma-1}}$, where $G_n\left(z\right)$ is the measure of z-type firms that are headquartered in market n.

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

[12]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.

where W_{si} is the ideal wage index for workers of type s in region i:

$$W_{si} = \left(L_{is}^{-1} \int_{\nu \in \mathcal{V}_i} w_{is} \left(\nu\right)^{\epsilon} d\nu \right)^{\frac{1}{\epsilon}}. \tag{8}$$

This wage index reflects both the level of wages and the variety of jobs in that region. When the employer-specific labor supply elasticity is infinite ($\epsilon \to \infty$), all local employers pay a common wage for type-s workers, equal to the index W_{si}. Otherwise, firms need to pay higher wages to attract more workers, giving rise to upward-sloping firm-specific labor supply curves as given in Equation 4.

Entry. Firms can enter freely in each market by paying f_e units of the tradable good. The entry location determines the headquarters location of a firm. I choose to price entry at the common numeraire (i.e. in units of tradable-goods) to avoid heterogeneity in entry costs across space, though allowing for this margin of heterogeneity (e.g. by using regional price indices) is straightforward.

Production of tradable goods. For simplicity, I assume that tradable-goods are freely traded and produced using constant returns to scale technology. Firms produce using a Cobb-Douglas production function that combines high and low skilled labor, with constant expenditure shares α and $1-\alpha$, respectively. Similar to the services sector, firms can enter freely after paying f_e units of the local final good.[13]

This completes the set-up of the model. I now turn to characterize the link between firms' geographical scope and the wage structure in the economy.

3.4 Firm spatial expansion and skill-biased technical change

The first takeaway from the model is a new micro-foundation for skill-biased technical-change (SBTC). Specifically, I show that changes in firms' scope are isomorphic to technical change that augments the productivity of headquarters workers relative to branch workers. To see this, I study the share of labor costs accrued to headquarters workers in service firms. Recall that this cost-share pins down skill intensity in the services sector, while in the tradable-goods sector skill intensity is exogenous by construction. I denote this cost-share by $\Gamma_i(z)$, and characterize it in the following proposition:

Proposition 1. *Geographical scope as a micro-foundation for skill-biased technical change.*

[13] 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.

(a) The share of costs accrued to headquarters workers in firm (i,z) is given by

$$\Gamma_i(z) = \frac{\gamma^{\bar{\eta}}\left(\frac{W_{hi}}{\sum_{j=1}^{N} x_{ij}(z)}\right)^{1-\bar{\eta}}}{\bar{W}_{\ell i}\left(z\right)^{1-\bar{\eta}} + \gamma^{\bar{\eta}}\left(\frac{W_{hi}}{\sum_{j=1}^{N} x_{ij}(z)}\right)^{1-\bar{\eta}}}, \quad \bar{\eta} \equiv \frac{\eta}{\eta - \frac{\epsilon}{\epsilon+1}\left(\eta - 1\right)}$$

where $\bar{W}_{\ell i}\left(z\right) \equiv \left(\sum_{j=1}^{N} \Omega_{ij}\left(z\right) W_{\ell j}^{\bar{\eta}-1}\right)^{\frac{1}{\bar{\eta}-1}}$ is a weighted power-mean of the labor supply shifters $W_{\ell j}$ across all of the firm's markets. The loadings $\Omega_{ij}\left(z\right)$ capture the share of region j in the firm's total branch-level payroll, as well as its dependency on the headquarters input relative to the firm's average.

(b) Consider the case that $\epsilon \to \infty$, all locations are symmetric, the aggregate endowments of headquarters labor and branch-level labor are given by $\bar{H}$ and $\bar{L}$, respectively, and $\beta = 1$. Then, $W_h = W_{hi}$ and $W_\ell = W_{\ell j}$ are the economy-wide prices of headquarters and branch-level labor, and their ratio is given by

$$\frac{W_h}{W_\ell} = \gamma \bar{x}^{\frac{\eta-1}{\eta}} \left(\frac{\bar{H}}{\bar{L}}\right)^{-\frac{1}{\eta}},$$

where $\bar{x}$ is a weighted power-mean of firms' geographical scope in the economy, with the weight of each firm determined by its total branch-level employment.

Proof: see Appendix B.

The first part of Proposition 1 characterizes the cost-share of headquarters workers at the firm level in terms of four key objects. Similar to the SBTC literature, this cost-share depends on: (1) relative factor prices, as captured by the ideal wage index for headquarters workers in the firm's headquarters region (W_{hi}), and a power-mean of the ideal wage indices for branch-level workers across all markets, ($W_{\ell j}$); (2) an exogenous technological shifter, as captured by the productivity of headquarters labor γ; and (3) the elasticity of substitution between factors, which in this case depends on the raw elasticity between headquarters and branch-level workers in the production function, η, and on the labor supply elasticity ϵ (since the latter affects the firm's ability to substitute across workers in different locations). Departing from the SBTC literature, the firm's geographical scope – captured by the sum of the firm's market penetration terms $\sum_{j=1}^{N} x_{ij}\left(z\right)$ – also enters this expression, in the same way as a reduction in the cost of headquarters labor. Therefore, an increase in firm scope acts as a factor-biased technical change for the factors that are used intensively in the production of the firm's non-rival component.

To see most clearly the implications for aggregate inequality across skill-groups, the second part of Proposition 1 considers a special case of the model with symmetric space, competitive labor markets ($\epsilon \to \infty$), and exogenous aggregate factor endowments of headquarters and branch workers ($\bar{H}$ and $\bar{L}$). In this case, the economy-wide wage differential between headquarters and branch workers is decreasing with the relative factor endowment of headquarters workers ($\bar{H}/\bar{L}$) and increasing with their relative productivity γ, as is standard in the literature. In addition, when $\eta > 1$, it also increases with the employment-weighted average of firms' geographical scope in the economy, captured by $\bar{x}$. Note that in the simple case of symmetric space, market penetration is common across locations $x_{ij}(z) = x(z)$, the firm's geographical scope is given by $x(z)N$, and $\bar{x}$ is the employment-weighted geographical scope in the economy.

Intuitively, the non-rivalry assumption implies that the marginal product of headquarters workers scales with the number of locations in which their output can be applied, whereas the marginal product of branch workers does not. The above result does not take a stance on what drives changes in firms' geographical scope ($\bar{x}$ in the second part of the proposition). Rather, it highlights that conditional on factor endowments and relative productivity γ, shocks that change firms' geographical scope (such as a reduction in the cost of expansion C) would act as a technological change that augments the productivity of headquarters workers.

3.5 Firm spatial expansion and firm-specific wages

The above result highlights the role of firms' spatial scope for inequality across skill-groups, and holds also for the version of the model with $\epsilon \to \infty$. In practice, residual wage inequality is pervasive, and accounts for most of the increase in wage dispersion over time. The model allows for such residual inequality when ϵ is finite. In this case, different firms and establishments pay different wages, conditional on skill group and location. The following Proposition characterizes firm-specific wages in the model:

Proposition 2. *The wages that firm (i, z) pays to its headquarters workers and branch-level workers can be expressed as*

$$\underbrace{\ln w_{\ell,ij}(z)}_{\textit{Wages to local labor}} = const_\ell + \underbrace{\frac{\epsilon}{\epsilon+1}\ln W_{\ell,j}}_{\textit{Local labor market}} + \underbrace{\frac{1}{\epsilon+1}\ln r_{ij}(z)}_{\textit{Local sales}} + \underbrace{\frac{1}{\epsilon+1}\ln\left(1-\Gamma_{ij}(z)\right)}_{\textit{Local labor intensity}}$$

$$\underbrace{\ln w_{h,i}(z)}_{\textit{Wages to headquarters labor}} = const_h + \underbrace{\frac{\epsilon}{\epsilon+1}\ln W_{h,i}}_{\textit{Local labor market}} + \underbrace{\frac{1}{\epsilon+1}\ln\sum_j x_{ij(z)} r_{ij}(z)}_{\textit{Total firm sales}} + \underbrace{\frac{1}{\epsilon+1}\ln\Gamma_i(z)}_{\textit{HQ intensity}}$$

Proof: see Appendix B.

When $\epsilon \to \infty$, all firms pay the same market-level wages, given by the wage indices $W_{\ell,j}$ and $W_{h,i}$. Otherwise, branch level wages increase with local output (as captured by local sales $r_{ij}(z)$) and decrease with the intensity of local output in the non-rival headquarters production, which I label $\Gamma_{ij}(z)$.[14] In contrast to branch-level wages, headquarters wages rise with the firm's *total* sales across all markets, as well as with the firm-level cost-share of headquarters from Proposition 1, $\Gamma_i(z)$.

These expressions help to build-up intuition for the effects of spatial expansion on firm-specific wages and within-firm inequality. For a given set of regional wage-indices, spatial expansion at the firm level pushes up wage dispersion between its headquarters and branches, by raising the ratio of total sales to local sales. This is true in particular when $\eta = 1$, in which case $\Gamma_{ij}(z) = \Gamma_i(z) = \frac{\gamma}{1+\gamma}$ is constant and does not vary with firm scope. If in addition $\eta > 1$, spatial expansion also reallocates production from the branches to the headquarters (a rise in $\Gamma_{ij}(z)$ and $\Gamma_i(z)$, as in Proposition 1), adding another force that pushes up within-firm wage dispersion.[15]

More formally, the following proposition describes how firm-level wages respond when expansion is driven by a small reduction in the cost of expansion C.

Proposition 3. *Consider a firm that experiences a small decline in its cost-of expansion $dC(\cdot)$ that results in non-negative expansion in all markets ($dx_{ij}(z) \geq 0$). Consequently:*

(a) The firm always raises headquarters wages ($w_{h,i}(z)$).

(b) The firm raises branch-level wages ($w_{\ell,ij}(z)$) only if $\sigma > \eta$.

(c) The ratio of headquarters wages to branch-level wages ($w_{h,i}(z)/w_{\ell,ij}(z)$) rises, irrespective of the values for σ and η.

Proof: see Appendix B.

According to Proposition 3, an increase in the geographical scope of a firm due to lower cost of expansion leads to higher headquarters wages and higher dispersion between headquarters and branches. Interestingly, the effect on branch-level wages is ambiguous, as can be seen also in the two opposing terms in the expression for $w_{\ell,ij}(z)$ in Proposition 2. On the one hand, spatial expansion generates a positive *productivity effect* across all branches, following an increase in the firm's non-rival inputs. This is captured by an increase in local sales in Proposition 2. On the other hand, it generates an increase in *headquarters intensity* when $\eta > 1$, which lowers demand and wages for local branch workers. On net, branch-level wages go up if $\sigma > \eta$.

It is worth benchmarking these results relative to the literature on firms in monopsonistic labor markets, as in Card et al. (2018) and Berger et al. (2022). The typical outcome from such models is that larger firms pay higher wages. In the current framework, this

[14] $\Gamma_{ij}(z)$ is the local equivalent of the firm-level cost-share of headquarters from Proposition 1. It is defined as the share of local branch sales that is used to pay headquarters workers, or one minus the ratio of local wage bill to local sales. See Appendix B for more details.

[15] On top of these effects, an increase in firms' geographical scope raises wage-dispersion across branches due to a mechanical effect, as the firm becomes active in more markets with different regional wage indices. This same mechanical effect pushes down inequality between firms.

is still true on average, though it is also associated with larger within-firm dispersion. Moreover, despite the existence of upward-sloping labor supply curves, firm growth does not necessarily lead to higher wages for branch workers, as demonstrated in Proposition 3. Consequently, while there is a clear positive firm size gradient for headquarters wages, it can be very flat (or even negative if $\sigma < \eta$) for branch wages. In addition, it highlights that firm-level shocks pass-through more to headquarters wages than to branch wages, in line with evidence on differential pass-through in Kline et al. (2019).

A similar result to Proposition 3 also applies following a small increase in the firm-level productivity z. Therefore, it is also a characterization of wage dispersion between firms with different productivities in the cross-section. Conditional on local wage indices, higher productivity firms pay higher wages on average, but not necessarily in their branches; and they are characterized by higher within-firm dispersion.

The above propositions also highlight the special role of space for wage inequality across firms and establishments. First, unsurprisingly, part of overall wage dispersion is due to differences in regional wage indices ($W_{\ell j}$ and W_{hi} in Proposition 2). Second, firm wages in each market j are also a function of the firm's headquarters location i.[16] This is reflected in the dependency of regional sales and local labor intensity ($r_{ij}(z)$ and $\Gamma_{ij}(z)$ in Proposition 2) on the firm's headquarters location i, on top of the firm's productivity z and the local market j. Therefore, wages are characterized by a *headquarters market effect*, in line with evidence on wage setting in affiliates of multinational firms, as in Setzler and Tintelnot (2021).[17] An implication of this headquarters market effect is that changes in the spatial distribution of economic activity can affect not only inequality across regions, but also inequality within any region: two branches can pay increasingly different wages from each-other even if they are in the same market j and have the same firm productivity z, if their headquarters market i is not the same.

3.6 Firm spatial expansion and spatial disparities

The model also yields predictions for the connection between firm scope and spatial disparities in skill intensity, yielding a new mechanism for what Moretti (2012) described as "the great divergence".

Proposition 4. *Spatial disparities in skill-intensity.*

[16] The dependency on the headquarters location i is mediated through choices of the firm's non-rival component h, which are affected by headquarters-market characteristics such as the productivity A_i and the cost of headquarters labor W_{hi}.

[17] Through the lens of simple wage decompositions as in Section 2, I find that this effect is quantitatively meaningful. Firms' headquarters locations account for between 30%-50% of the between-firm component of overall wage inequality for multi-region firms.

(a) The share of skilled-labor in aggregate income is given by

$$\left(1-\tilde{\beta}\right)\alpha+\tilde{\beta}\bar{\Gamma},$$

where we define $\tilde{\beta}\equiv\beta\frac{\epsilon}{\epsilon+1}\frac{\sigma-1}{\sigma}$*, and* $\bar{\Gamma}$ *is the sales-weighted average of the headquarters intensity* $\Gamma_i(z)$ *from Proposition 1 across all firms and regions in the economy.*

(b) The share of skilled-labor in region i's income is given by

$$\left(1-\tilde{\beta}\right)\alpha+\tilde{\beta}\left[\left(1-\alpha\right)\frac{R_i}{E_i}\Gamma_{out,i}+\alpha\Gamma_{in,i}\right],$$

where R_i *is total (national) sales by locally-headquartered firms in region i;* E_i *is total expenditure on services in region i; and* $\Gamma_{out,i}$ *and* $\Gamma_{in,i}$ *are the sales-weighted averages of the headquarters intensity for locally-headquartered firms and for locally-active branches, respectively.*

(c) When $\eta=1$*, then* $\bar{\Gamma}=\Gamma_{out,i}=\Gamma_{in,i}=\frac{\gamma}{1+\gamma}$*. In this case, a sufficient statistic for region i's high-skilled income share is its specialization in providing headquarters services, as summarized by the ratio of total sales by locally-headquartered firms,* R_i*, to local expenditure on services,* E_i*.*

(d) In the limit economy with no multi-location firms ($C(\cdot)\to\infty$ *when* $x_{ij}(z)>0$ *for* $i\neq j$*), the share of region i's income accrued to high-skilled labor is summarized the sales-weighted average of the headquarters intensity of its firms,* $\Gamma_i=\Gamma_{out,i}=\Gamma_{in,i}$*. If in addition* $\eta=1$*, there are no spatial differences in high-skilled income shares.*

Proof: see Appendix B.

Proposition 4 starts with characterizing aggregate skill intensity in the economy, as a benchmark for regional skill intensity. The aggregate income-share of skilled labor is a weighted average of the (exogenous) skill intensity in the tradable-goods sector, α, and the average intensity of the non-rival component of firms in services, $\bar{\Gamma}$. The weight $\tilde{\beta}$ reflects the importance of services in the economy. In turn, regional skill intensity depends in addition on two region-specific objects. The first object, captured by R_i/E_i, is regional specialization in headquarters services (as opposed to specialization in production of tradable goods). When the non-rival component of services is more skill-intensive than production of tradables, greater specialization in headquarters services leads to higher relative demand for skill. Note that in the above baseline model this is always the case, since I assumed that only skilled labor is used in headquarters production, though in the quantification below I relax this. When $\eta=1$, this measure of specialization is also a *sufficient statistic* for regional skill intensity (part (c) of Proposition 4). The scope for such specialization is tightly linked to firms' geographical scope. For example, when cross-region firm activity is prohibited (part (d) of Proposition 4), all headquarters services are supplied locally, and there is no room for regions to specialize in this activity: $R_i = E_i$ for all i, and there is no spatial variation in R_i/E_i.

The second endogenous object that shapes regional skill intensity is the share of non-rival factors in the cost structure of local firms ($\Gamma_{out,i}$ and $\Gamma_{in,i}$). This object itself is a function of firms' spatial scope, as summarized in Proposition 1. Therefore, through these two channels, an increase in firms' spatial scope can lead to greater spatial disparities in the economy. An illustrative case is the limit economy with no multi-region firms ($C(\cdot) \to \infty$ when $x_{ij}(z) > 0$ for $i \neq j$) and $\eta \to 1$. In this case, there are no spatial disparities in high-skilled income shares, demonstrating the importance of firms' scope for spatial disparities in an extreme case.

4 Empirical evidence for the main mechanisms

Having established the theoretical link between firms' geographical scope and various dimensions of inequality, I now turn to provide reduced-form evidence for some key assumptions and implications of the above model.

4.1 Evidence for within-firm non-rivalries

A key assumption in the model is that the output of headquarters workers is non-rival across the firm's locations. An important implication of this assumption is that positive demand shocks in a subset of the firm's locations spillover into greater activity in other locations, due to the positive productivity effect arising from spending more on non-rival inputs. This effect declines with the strength of non-rivalries. The model also includes two forces that can work in the opposing direction. First, if the cost of expansion C exhibits convexities (e.g., due a span-of-control cost, as I allow in the quantitative version of the model below), then expansion in one market would raise the marginal cost of expansion in other markets, potentially leading to less firm activity there. Second, if the elasticity of substitution between headquarters and branch workers is very high, more expansion opportunities in one market would reallocate production away from all markets to the headquarters.

I investigate whether such positive spillovers exist in practice using data from the U.S. Census LBD. Denote an outcome of interest for firm f in market j in period t by y_{fjt}, and denote the *aggregate* income in market j in period t by Y_{jt}. We are interested to see whether y_{fjt} responds positively not only to local demand as captured by Y_{jt}, but also to changes in all the other markets in which f operates. I focus on two outcome variables: the log of total firm f payroll in market j, and the log of average firm f wages in market j (the difference between them is the log of f's employment in j). Note that the responsiveness of total payroll does not depend on the firm-level labor supply elasticity ϵ, while how it is split between wages and employment does. I define a region as a U.S. labor market area (LMA, a slight aggregation of commuting zones) and study 5-year changes in these outcomes across years in which the Economic Census is taken. In all specifications, I control for firm f and year t fixed effects, and cluster standard errors at the firm level.

My baseline specification regresses 5-year changes in $\log y_{fjt}$ against 5-year changes in $\log Y_{jt}$, and against a weighted average of changes in $\log Y_{kt}$ across all regions k, weighted by the firm's ex-ante activity in each market k. This specification is valid if changes in Y_{jt} are exogenous demand shifters to the firm, e.g. if the firm is very small in all its markets. There are a few potential concerns with this specification. First, even if the firm is small, its expansion in j might result from some technological shock that affects both market j and all other markets k, e.g. a positive productivity shock to the firm's sector. Therefore, in a richer specification, I control for any trends in market j or in the firm's sector by including $j \times t$ and sector(f) $\times\, t$ fixed effects. These fixed effects absorb regional growth in j, but still allow us to see f's responsiveness in market j to regional growth in its other locations, in line with the idea of within-firm non-rivalries. Second, even controlling for these trends, if f is large in its other markets, changes in $\log y_{fjt}$ and in $\log Y_{kt}$ (for $k \neq j$) could result from a firm-level idiosyncratic productivity shock (e.g. a rise in z in the model). Therefore, in the richest specification, I instrument regional growth in all markets with each region's exposure to national sectoral growth trends, following Bartik (1991). Since I also control for the sectoral trend for firm f, this instrument utilizes variation in regional growth due to national shocks in other sectors.

Table 2 presents results from these regressions. Columns (1) and (4) show the results from the baseline specification for firm f's payroll and average wage in market j, respectively. Unsurprisingly, both respond positively to demand shocks in j. In addition, in line with the idea of within-firm non-rivalries, both respond positively to regional growth in f's other locations. The responsiveness of both payroll and wages also alines with the assumption of firm-specific labor supply curves (finite ϵ). Columns (2) and (5) repeat these estimates when flexibly controlling for time trends in region j and in f's sector. The coefficient on regional growth in j is now absorbed in the time trend, but the coefficient on growth in other markets is still positive and significant for both total payroll and average wages. Finally, Columns (3) and (6) repeat these results when instrumenting growth in other markets with the regional exposure to national industry trends. Again, the results for both total payroll and average wages are positive and significant. The high coefficient for total payroll – reflecting a transmission of around 1-to-1 from regional growth in all of f's markets to its payroll growth in the typical market – is in line with the fact that the typical market is very small for the firm. One should think about this coefficient as the response of a national chain in a very small market to higher demand in all of its other markets.

To conclude, these results align with the idea of within-firm non-rivalries in the model.

Table 2: Within-firm spillovers across markets

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome: 5-year changes firm f, market j:	Δ ln payroll	Δ ln payroll	Δ ln payroll	Δ ln wage	Δ ln wage	Δ ln wage
Regional income growth in market j	0.254***			0.0685***		
	(0.0131)			(0.00733)		
Average income growth in all f's markets	0.395***	0.268***	1.024***	0.178***	0.134***	0.280***
	(0.0234)	(0.0247)	(0.110)	(0.0133)	(0.0140)	(0.0667)
Observations (rounded)	856000	856000	856000	856000	856000	856000
First-stage F stat	.	.	3283	.	.	3283
Year and firm FE	Y	Y	Y	Y	Y	Y
Year $\times$ market FE	N	Y	Y	N	Y	Y
Year $\times$ sector FE	N	Y	Y	N	Y	Y

Note: This table shows how firm outcomes in a particular local market j respond to demand shocks in other markets. Column (1) shows the results from a regression of 5-year changes in the (log) total payroll of firm f in market j against regional income growth in market j over the same time period; and against a weighted-average of regional income growth in all of firm f's markets. Column (2) repeats Column (1) with time-varying fixed effects for the firm's sector and for market j, which also absorbs regional income growth in the focal market j. Column (3) repeats Column (2) by instrumenting regional income growth in all markets using ex-ante exposure to national sectoral trends, as described in the text. Columns (4)-(6) repeat Columns (1)-(3), replacing changes in (log) total payroll with (log) average payroll per employee as the outcome variable. All columns include time and firm fixed effects. Standard errors are clustered at the firm level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

A positive shock that the firm experiences in a subset of its markets results in higher productivity in the firm's other markets, leading to more hiring there as well. In line with the idea of firm-specific upward-sloping labor supply curves, both total payroll and average wages respond positively to these shocks. In the quantification of the model below, I verify that the model indeed replicates these results.

4.2 Dispersion between headquarters and branches

Some of the model's distinctive predictions are for wage and payroll dispersion between headquarters and branch activities, both over time and in the cross-section. Recall that the notion of headquarters in the model is the set of tasks with non-rival output across the firm's locations, and not necessarily the firm's administrative center. This theoretical notion is elusive in the data, so I employ two alternative measurement approaches to capture it, using workers' occupational classifications and establishments' industry classifications.

4.2.1 Evidence from occupational classifications

I first investigate dispersion between headquarters and branches using workers' occupational classifications. Such data is not available at the firm level in the U.S., but we can study the model's predictions at the aggregate and across sectors using data from the Census Bureau's Decennial Census and the American Community Survey. In particular, the model predicts that in service sectors that have experienced an increase firms' geographical scope, we should see: (a) a rise in wage dispersion between headquarters workers and branch workers; (b) if $\eta > 1$, the payroll share of headquarters workers should also increase; (c) these effects should be stronger in sectors with greater spatial expansion.

To test these predictions, I construct a new measure of headquarters-intensity for each occupation. One way to measure this notion is simply to look at which occupations colocate with firms' actual administrative headquarters. Another way to measure this notion is using *within-firm geographical concentration* of occupations: if the output of the firm's branches is non-tradable (as assumed in the model), then occupations in firms' branches should be spatially dispersed relative to those that perform the non-rival component, even if the latter does not exactly coincide with the location of the administrative headquarters. To obtain these measures, I merge data on firms' geography and structure from Dun and Bradstreet with data on firm's hiring of different occupations across locations using online job postings from Lightcast.[18] I merge the Lightcast and Dun & Bradstreet using exact name and location matching. This process results in around 75,000 multi-region firms, out of which around 64,000 are in service sectors. More details on the data and on these procedures are given in Appendix D. I find that both measures for headquarters-intensity – within-firm spatial concentration, and collocation with actual administrative headquarters – are very highly correlated. Table 3 shows examples for the most and least headquarters-intensive occupations resulting from this procedure. Intuitively, occupations in areas such as research, financial analysis, and software development are very headquarters-intensive; while occupations in areas such as retail sales, truck driving, and insurance claims, are not.

Table 3: Examples for the most and least headquarters-intensive occupations

Most HQ-intensive	Least HQ-intensive
Marketing Specialists	Truck and Bus Drivers
Front-End Application Design	Client Support and Sales
Health and Medical Research	Insurance Claims and Sales
Communications and Public Relations	Laborers and Warehouse Workers
Procurement	Non-Technical Sales
Financial Regulation and Compliance	Construction Workers
Financial Analysis	Delivery Drivers and Messengers
Data Analysis and Mathematics	Landscaping and Gardening
General Research	Financial Sales
Software Development	Transmission and Electric Technicians

Note: Examples for the most and least headquarters-intensive occupations. Headquarters-intensity is computed by combining data from firm geography data from Dun and Bradstreet and firm online-job postings from Lightcast, as described in Section 4.2.1.

I then use this measure of headquarters intensity to test the above model's predictions, by merging it to the Census Bureau's Decennial Census and the American Community Survey. Table 4 shows the results from regressions of changes in occupational

[18] See https://lightcast.io/.

outcomes against this measure of headquarters intensity. Specifically, I study changes at the industry-occupation level between 1980-2017.[19] First, Column (1) demonstrates that over this period, there was a relative increase in the total payroll of relatively more headquarters-intensive occupations. Since I control for sectoral fixed effects, this amounts to an increase in the relative sectoral expenditure on headquarters labor relative to branch-level labor. Columns (2) and (3) find similar relationships for log average and for residual wages, after controlling for individual demographics and location, in line with higher wage-dispersion between headquarters and branches. Columns (4)-(6) repeat these estimates, adding an interaction of the headquarters intensity measure with an indicator for sectors that experienced high spatial expansion (defined as sectors in the top quartile of expansion). In line with the model's predictions in Propositions 1-3, sectors with greater spatial expansion have experienced greater increase in relative wages and residual wages for more headquarters-intensive occupations; and in line with the existence of substitutabilities between headquarters and branches ($\eta > 1$), they experienced greater reallocation of total labor costs towards more headquarters-intensive occupations.

Table 4: HQ-branch dispersion based on occupational classifications

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome: 1980-2017 changes in -	Pay share	Log wage	Log r. wage	Pay share	Log wage	Log r. wage
Occupational headquarters intensity	0.0912***	0.519***	0.0809**	0.0401**	0.362***	-0.0325
	(0.0263)	(0.0507)	(0.0390)	(0.0195)	(0.0593)	(0.0491)
Headquarters intensity × high expansion				0.0963**	0.295***	0.214***
				(0.0467)	(0.0936)	(0.0780)
Industry fixed-effects	Y	Y	Y	Y	Y	Y
Number of observations	13325	13325	13325	13325	13325	13325
R-squared	0.206	0.422	0.253	0.213	0.426	0.256

Note: Changes in occupational wages and payroll-shares between 1980-2017 as a function of headquarters intensity. Each cell shows the estimate from a regression of changes in industry-occupation level outcomes against the headquarters intensity of the occupation and against an interaction of this measure with an indicator for high-expansion industry. Each column captures a different industry-occupation-level outcome: changes in within-industry occupational payroll shares (columns 1 and 3); changes in log average wages (columns 2 and 4); and changes in log residual wages (columns 3 and 6). All changes are between 1980-2017. The measurement of occupational headquarters intensity is described in Section 4.2.1. High-expansion industries are those with the greatest increase in average establishments per firm (the top quartile of this distribution). Observations are weighted by total employment. All columns control for industry fixed effects. Occupations and industries are classified in accordance to the *occ1990dd* and *ind1990ddx* classifications from Autor and Dorn (2013) and Autor et al. (2019), which can be found in https://www.ddorn.net/data.htm. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

4.2.2 Evidence from establishments' industry classifications

An alternative way to measure the notion of headquarters activity is using industry-classifications of establishments in the Census LBD data. I map the notion of headquarters

[19] For this exercise, I employ time-invariant classifications of industries and occupations, utilizing the *occ1990dd* and *ind1990ddx* classifications from Autor and Dorn (2013) and Autor et al. (2019). See also ttps://www.ddorn.net/data.htm.

activity in the model to establishments that perform business services within the firm (NAICS sectors 51-55), and the notion of branch activity to establishments in the firm's main subsector (defined based on the NAICS sub-sector with the largest payroll share within the firm). Together, these two classifications capture most of the firm's total payroll for most firms. Naturally, this definition requires me to focus on service firms whose main activity is not business services (i.e., I focus on firms with the majority of their payroll in NAICS sectors 42, 44, 48, 49, 61, 62, 71, 72, 81). While the model also applies to many firms that specialize in business services,[20] I cannot cleanly separate headquarters and branch activity in these firms.[21]

Utilizing establishment-level industry classifications has some benefits and some limitations. A key benefit is that it allows me to study changes within firms as they expand in space using the U.S. Census establishment-level data. One limitation of this approach is that it is not suitable to study long-term changes, due to vast changes in how establishments are classified over time. Accordingly, I use it to study cross-sectional patterns and not long-term trends. Another limitation is that establishments' classifications are often misreported by firms, and many times firms report the same industry classification for all their establishments, especially in some key services sectors. I thus focus on the subset of firms in which there is a clear differentiation across classifications of different establishments.

Armed with these measures of headquarters and branch activities, I study firms' reallocation patterns between these activities as they expand in space. Column 1 in Table 5 regresses the log of the ratio between firm's total payroll in headquarters establishments and payroll in branch establishments, as defined above, against the log number of firm's establishments, controlling for firm and year fixed effects. In line with Proposition 1 for the case of substitutability between headquarters and branch labor ($\eta > 1$), there is a positive and statistically significant relationship between this payroll ratio and changes in firms' scope. Column 3 repeats this exercise when also controlling for sectoral trends and for average wages in the firm's different locations (capturing the dependence of this ratio also on regional wage indices in Proposition 1). The relationship is weaker but still pos-

[20]Consider for example telecommunication firms (NAICS code 517). These firms have a very clear separation of local branches that are responsible for sales, installation, maintenance, and technical support; and a national component that is responsible for developing the firm's technology, in line with the non-rivalry assumption in the model. Another clear example is banking.

[21]An alternative way to identify headquarters activity in the LBD is to use only establishments in the NAICS-55 sector, which is explicitly defined as headquarters services. I found this to be problematic due to multiple reasons. First, many firms do not report any NAICS-55 activity, despite having a clear headquarters location when comparing to other data sources. Indeed, the payroll-share in NAICS-55 establishments is significantly higher when conditioning on reporting any NAICS-55 activity. This is especially common in firms whose main activity is business services (e.g. if the main activity is finance, communications, etc.), further complicating the separation of branch and headquarters activity for these firms. Second, this notion of headquarters activity is more limited than in my model, which refers to tasks that are non-rival at the firm level, and thus could be captured by other NAICS codes. In any case, my results are robust to using this more limited definition of headquarters activity.

itive and statistically-significant. Columns 2 and 4 repeat columns 1 and 3 by changing the outcome of interest to wage dispersion between headquarters and branches (the log ratio of average payroll to employment across the two groups of establishments). In line with Proposition 3, the headquarters-branches wage ratio rises when firms increase their number of establishments. Notwithstanding the above-mentioned limitations of using establishments' industry classifications, these results provide further evidence in support of the model's mechanisms in the cross-section of multi-location service firms. These moments will also be useful in the quantification of the model below.

Table 5: HQ-branch dispersion based on industry classifications

	(1)	(2)	(3)	(4)
Outcome: Log ratio between HQ and branches	Payroll ratio	Wage ratio	Payroll ratio	Wage ratio
Log # of establishments	0.305***	0.0394***	0.153***	0.0403**
	(0.0152)	(0.00526)	(0.0325)	(0.0123)
Observations (rounded)	29500	29500	29500	29500
Year and firm FE	N	N	Y	Y
Year $\times$ sector FE	N	N	Y	Y
Controls	N	N	Y	Y

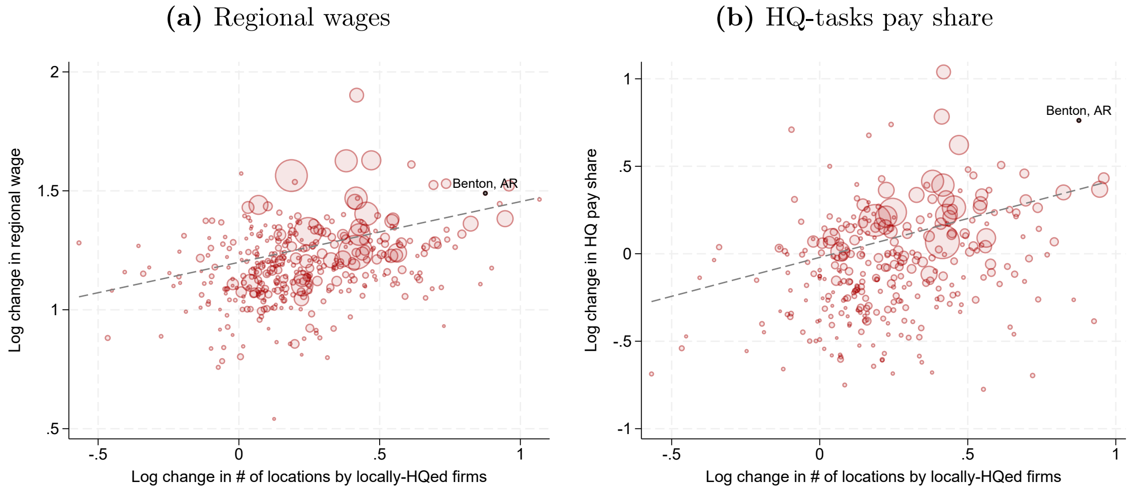
Note: Relationship between firms' scope and within-firm headquarters-branch dispersion. Each observation is a firm at a particular year. Column 1 shows the estimated coefficient from a cross-sectional regression of the log of the firm's payroll ratio between its headquarters and branch establishments, as defined in Section 4.2.2, against the firm's log number of establishments. Column 2 replaces the outcome variable with the log ratio of average payroll per worker between headquarters and branch establishments. Columns 3 and 4 repeat columns 1 and 2 when controlling for firm fixed effects, year fixed effects, year$\times$sector fixed effects, and controls for differences in average regional wages across firms' locations. Standard errors are clustered at the firm level. The sample includes years from Economic-Census years starting from 1997 (i.e. 1997,2002,2007,2012,2017). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

4.3 Spatial disparities

Finally, I investigate the mechanics of the model across local labor markets. In the model, higher relative expansion of locally-headquartered firms across all markets leads to relatively higher regional headquarters payroll share; relatively higher regional skill-intensity; and relatively higher regional reallocation from production of tradable goods to services (due to the increased specialization in providing headquarters services), in line with Proposition 4.

I investigate these patterns across U.S. labor market areas (LMAs), again focusing on changes between 1980-2017. I measure the expansion of locally-headquartered firms by utilizing data on firms' geography from Dun & Bradstreet. This data is more suitable for this exercise than the Census LBD, since it provides an explicit headquarters location for all firms in all years (note that for this exercise, I only use information on firms' locations, and do not data such as employment or payroll). I then compute for each LMA the average

Figure 3: Expansion of locally-headquartered firms and labor market outcomes



Note: local labor market outcomes against expansion of locally-headquartered firms. Subplot (a) shows log-changes in average regional wages between 1980-2017 (on the y-axis) against a measure of external expansion of locally-headquartered firms over the same period (on the x-axis), as defined in Section 4.3. Each circle is an LMA, with the size of the circle indicating overall LMA employment in 1980. Subplot (b) repeats subplot (a), replacing the y-axis with changes in the share of headquarters tasks out of total payroll between 1980-2017, computed using the headquarters intensity of occupations from Section 4.2.1. The highlighted observation in black is Benton county, AR, the headquarters-location of Walmart. The black dashed line in both figures captures a linear fit. Regression results for these relationships can be found in Appendix D.

expansion of its locally-headquartered firms over 1980-2017, taking only firms that were already based there in 1980. For example, the Walmart Corporation was headquartered in Benton county, AR (LMA #303) in 1980, and experienced a substantial increase in the number of locations over 1980-2017. As such, it contributes significantly to the average expansion measure of Benton.

Panel (a) in Figure 3 shows log-changes in average regional wages between 1980-2017 (on the y-axis) against the above measure of expansion by local firms (on the x-axis). Each circle is an LMA, with the size of the circle indicating overall LMA employment in 1980. Evidently, there is a clear positive relationship between regional wage growth over 1980-2017 and external expansion of locally-headquartered firms over the same period. In line with the mechanisms in the model, panel (b) shows a clear increase in the share of headquarters tasks out of total payroll in these markets, computed using the headquarters intensity of occupations from Section 4.2.1. I highlight in black the above example of Benton county, AR: it has experienced particularly high average expansion of local firms, wage growth, and reallocation to headquarters activities, very much due to the expansion of Walmart. I provide the explicit regression results for these relationships in Appendix D, as well as additional results for regional skill-intensity and regional reallocation from manufacturing to services.

Exogenous variation in firm expansion. Figure 3 shows that local labor market outcomes correlate with expansion of locally-headquartered firms in line with the model's predictions, though the direction of causality is not clear. In appendix D.5, I provide casual evidence for these relationships, arguing for causality that goes from firm expansion to local labor market outcomes. To this end, I obtain regional exogenous variation in firm expansion from ex-ante exposure to national expansion trends. The idea is that if a market such as Benton county (AR) already specialized in hosting headquarters of productive retail firms in 1980, and the retail sector as a whole experienced substantial spatial expansion over 1980-2017 at the national level, then Benton county (AR) experienced an "expansion shock" over this period, which is not driven by other local trends. Utilizing variation from these national expansion trends, I show that the results are consistent with those in Figure 3.

5 Model Quantification

I now turn to estimate the model. Using the quantitative model, I demonstrate the importance of the above mechanisms for the overall increase in U.S. wage inequality, and evaluate policies that shape firms' ability to span multiple local markets. I match key features of the U.S. economy in 1980, and study counterfactuals that reflect changes in the economic environment between 1980-2017. I map the model to 200 local labor markets in the U.S., which I construct by combining small commuting zones with neighboring large ones.

5.1 Additional model components

First, I enrich the model with additional components that help to discipline the baseline equilibrium and which are common in spatial equilibrium models.

Multiple skill groups in production. I now assume that households belong to S different skill groups. The headquarters labor bundle $h_i(z)$ and the branch-level bundle $\ell_{ij}(z)$ are now Cobb-Douglas aggregators of all skill groups with skill intensities α_{hs} and $\alpha_{\ell s}$, respectively, such that $\sum_{s=1}^{S}\alpha_{hs} = \sum_{s=1}^{S}\alpha_{\ell s} = 1$. Similarly, the tradables sector also uses a Cobb-Douglas production function with skill-intensities α_{gs}, where $\sum_{s=1}^{S}\alpha_{gs} = 1$.

Regional amenities. Agents of type s choosing location i now enjoy an exogenous amenity component B_{is} that enters multiplicatively in their utility function. Consequently, the probability that an agent of type s chooses location i is given by

$$\frac{L_{is}}{\sum_{j=1}^{N} L_{js}} = \frac{\left(B_{is}\bar{W}_{is}/\bar{P}_i\right)^{\xi}}{\sum_{j=1}^{N}\left(B_{js}\bar{W}_{js}/\bar{P}_j\right)^{\xi}}, \tag{9}$$

where $\bar{P}_i$ is the price index of the local final good, and $\bar{W}_{is} \equiv L_{is}^{\frac{1}{\epsilon}} W_{is}$ is the ideal wage index for type s workers in location i. Households are more likely to choose locations with higher effective real wages $\bar{W}_{js}/\bar{P}_j$ and higher amenities B_{js}.

Housing. I add local housing, which introduces a second dispersion force on top of the idiosyncratic preference shocks. Agents spend a share ζ of their income on housing and a share $1-\zeta$ 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. The aggregate price index, previously given by P_i^{β}, is now equal to $\bar{P}_i = \bar{\beta} P_i^{\beta\left(1-\frac{\zeta}{(1+\varrho)-\zeta\varrho}\right)} Y_i^{\frac{\zeta}{(1+\varrho)-\zeta\varrho}}$, where Y_i is total regional income and $\bar{\beta}$ is a constant that subsumes various parameters. Conditional on the price index of services P_i, local cost of living is increasing in regional income Y_i due to higher cost of housing. The exponent on regional income is higher when agents spend more on housing (higher ζ) or when housing supply is relatively inelastic (lower ϱ).

Regional productivity in tradables. I allow locations to differ in their total factor productivity in the production of tradable goods, given by $A_{i,g}$.

Firm productivity distribution. In line with the literature, I parameterize the firm productivity distribution $G(z)$ as a Pareto distribution, with a shape parameter θ.

5.2 Parameterizing the cost of expansion

I consider a particular structure for market penetration and the cost of expansion $C(\cdot)$ for easy mapping to the data. I assume that market penetration is an increasing function of the number of establishments that a firm opens in a given market, denoted by $n_{ij}(z)$: $x_{ij}(z) = \frac{1}{1+n_{ij}(z)^{-1}}$. Market penetration is zero when $n_{ij}(z)=0$, and approaches 1 when $n_{ij}(z) \to \infty$. The concavity of x_{ij} in n_{ij} reflects potential cannibalization effects across the firm's establishments in a given market: the marginal establishment attracts fewer consumers than the average of existing establishments.[22]

The cost of expansion $C(\cdot)$ captures the cost of setting up establishments $n_{ij}(z)$ in all markets, and is given by

$$C\left(\{n_{ij}(z)\}_{j=1}^{N}\right) = \sum_{j=1}^{N} \tau_{ij} n_{ij}(z) + \delta \left(\sum_{j=1}^{N} n_{ij}(z)\right)^2 .$$

The first term in the cost of expansion captures a constant cost per each establishment: a firm based in location i with presence in location j pays a cost of τ_{ij} for every establishment that it opens there. The second term allows for a potential span-of-control constraint (SoC) for the firm as a whole, reflected in a quadratic cost in the total number of establishments in all markets. The importance of this SoC effect is captured by the

[22]See Arkolakis (2010), Oberfield et al. (2024), and Wenning (2024) for similar specifications.

parameter δ. Note that this SoC effect works in the opposite direction to the non-rivalry assumption in the production function: if δ is high and the productivity of headquarters inputs (γ) is low, then expansion in one market (e.g. following a positive demand shock there) could lead to contraction in other markets.

With an eye toward my main quantitative exercise, I assume a simple parameterization of the bilateral cost per-establishment τ_{ij}, given by

$$\tau_{ij} = \tau_0 \times \text{dist}_{ij}^{\rho} \times \kappa^{\mathbb{I}\{\text{st}_i \neq \text{st}_j\}},$$

where τ_0 is a scaling constant; dist_{ij} is the geographical distance between i and j; ρ is the distance-elasticity in the cost of opening new establishments; and κ is a cost premium when i and j are in different states. The cost of setting up a new establishment is then higher in markets that are far away from the firm's headquarters.

The motivation for this specification is a strong gravity pattern for bilateral firm linkages in the data. To see this, denote by M_{ij} the number of establishments in location j that belong to i-headquartered firms. Table 6 reports the results from a gravity regression of M_{ij} against log bilateral distance, controlling for origin and destination fixed effects, where each observation corresponds to a pair ij of the 200 local markets in my quantification.[23] I also include a dummy variable for whether the two markets share state borders, to account independently for state border effects. Column (1) shows results for 1980 from an OLS regression that sets the log of M_{ij} as the outcome variable, abstracting from ij pairs with zero linkages. Column (2) estimates this relationship using the Poisson Pseudo-Maximum Likelihood estimator from Silva and Tenreyro (2006), allowing for zero bilateral linkages and encompassing all observations. In both specifications, the coefficient on distance is negative and statistically significant, at around unity. Notably, the coefficient on distance has declined over time as can be seen in Columns (3) and (4), a pattern which I return to in Section 6 below.

In addition to distance, I allow for a higher cost per-establishment $\kappa > 1$ when operating establishments in a different state than the firm's headquarters. Indeed, in the data, cross-region firm linkages are lower when operating establishments in different state from the firm's headquarters, as evident in Columns (1) and (2) of Table 6. This state border effect could arise due to the role of state-level regulations and compliance requirements. An extreme example of such regulations is the restriction on cross-state banking activity that characterized the U.S. banking system before a vast deregulation in the 1980s-1990s. In such an extreme case, κ would be set to infinity.

[23] I implement this procedure using the Dun & Bradstreet data, which has the key advantage that every multi-establishment firm has a clear headquarters location. In contrast, in the LBD, headquarters location cannot be inferred for a large share of firms, especially in earlier periods. 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 which it compares well to administrative datasets. See Barnatchez et al. (2017) for additional details.

Table 6: **Gravity for cross-region firm linkages**

	(1)	(3)	(2)	(4)
	1980		2017	
Outcome: number of cross-region establishments	Logs	Levels	Logs	Levels
Log distance	-0.956***	-0.981***	-0.777***	-0.608***
	(0.0135)	(0.0322)	(0.0128)	(0.0165)
Different state	-1.132***	-1.281***	-1.032***	-0.873***
	(0.0263)	(0.0445)	(0.0246)	(0.0334)
Origin and destination fixed effects	Yes	Yes	Yes	Yes
Observations	11451	39800	11451	39800
R-squared	0.688		0.767	
Pseudo R-squared		0.790		0.838

Note: This table shows results from the regression of the number of bilateral headquarters-branch linkages across pairs of local markets against the distance between the markets and a dummy that equals 1 if the markets do not share state borders. Each observation is a pair out of 200 local U.S. labor markets as defined in Section 5.2. Columns (1) and (3) are estimated in logarithms by ordinary least squares (OLS), and Columns (2) and (4) are estimated in levels by Poisson pseudo maximum likelihood (PPML). All specifications include headquarters-location and branch-location (origin and destination) fixed effects. Columne (1) and (3) show results for 1980, Columns (2) and (4) show results for 2017. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

It is worth noting that I allow for spatial frictions in the cost of expansion $C(\cdot)$, but not in the firm's production function, as often assumed in the literature. The reason for this assumption is that while gravity is a very strong feature of firms' location decisions, it is not clearly evident in firm regional wages. Thus, in models with firm-specific wages, one would like to avoid introducing spatial frictions that distort the marginal product of labor across the firm's branches, since it would create a counterfactual systematic relationship between wages and distance.

5.3 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, in particular the services production function and the cost of expansion. This block includes new features relative to the existing literature, and so I estimate it using Simulated Method of Moments (henceforth SMM).

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. (2023). 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. (2022), 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 varieties 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 7 summarizes the calibration of this part of the model.

Table 7: Calibration of the households block of the model

Parameter	Interpretation	Source	Value
ξ	Dispersion of location preference shocks	Galle et al. (2023)	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

Definition of skill groups. I define four skill groups by dividing workers into equally-sized bins in the 1980 Decennial Census based on the average skill requirement of their occupation. The skill requirement is measured using the share of college graduates in each occupation. I use the same allocation of occupations to skill groups when computing statistics for the skill groups in 2017. This division into skill-groups has a few advantages relative to categorizing individuals based on their educational attainment. First, the demographic composition of different education groups has changed dramatically over time; this is not the case for my four groups of occupations. Second, the relative size of these groups remained relatively constant over time, allowing me to abstract from changes in the aggregate supply of skill. Finally, it allows me to consider more skill groups. In support of this approach, the average college intensity of a worker's occupation is a stronger predictor of their wage than their own educational attainment.

Skill intensities. I also recover the skill elasticities in production $\alpha_{hs}, \alpha_{\ell s}, \alpha_{gs}$ directly from the data, by computing the payroll-share of each group in all three activities. To compute α_{hs} and $\alpha_{\ell s}$, I utilize the occupational headquarters-intensity measures from Section 4.2.1, and the fact that each occupation is classified into a single group, delivering the intensity of each skill group s in headquarters and branch production. For α_{gs}, I take the payroll share of each group in goods-producing sectors. The values for all skill intensities can be found in Appendix D.6.

5.4 Model inversion

The set of location fundamentals $\mathcal{A} \equiv \{A_i, B_{is}, A_{g,i}\}_{i,s}$ includes the productivity shifters in the services sector, regional amenities, and regional productivity in tradable-goods. I obtain $\mathcal{A}$ by iterating over the model's equilibrium conditions, conditional on a guess for all the model's parameters. 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 C.

The three sets of regional moments that I match in the data are total regional labor income, Y_i, regional employment in each skill group. L_{is}, and relative regional payroll for skilled workers, which I denote ω_i. I define ω_i as the ratio between total payroll of workers in the upper half of the skill distribution (i.e., the top 2 skill groups defined above) to the income of workers in the bottom half of the skill distribution. Intuitively, B_{is} can be obtained by inverting the location choice equation (9). A_i and $A_{g,i}$ are obtained by jointly matching data on regional income and skill-intensity. Recall from Proposition 4 that regional skill intensity depends on the relative specialization in providing headquarters services, and thus is informative about the relative productivity of hosting services firms relative to producing tradable goods. A benefit of this procedure is that I match regional employment, income, and skill-intensity for all 200 regions in the baseline equilibrium, replicating key features of the U.S. economy.

5.5 Simulated method of moments for the production block

The remaining parameters to estimate are from the production block of the model: the relative productivity of headquarters inputs (γ); the elasticity of substitution between headquarters and branch labor (η); the dispersion of firm productivity draws (θ); the distance-elasticity of the cost of expansion (ρ); the state-border effect in the cost of expansion (κ); the scale of the cost of new establishments (τ_0); the span of control component of the cost function (δ); and the entry cost (f_e). I collect these parameters in the vector Θ, and estimate them by matching a set of empirical moments. Below I provide brief intuition for which empirical moments helps to identify each of these parameters, though recall that in practice all moments are jointly determined in equilibrium by all of these parameters.

The parameter η is directly related to how the expenditure-share on headquarters varies with firm scope, as given in Proposition 1. In particular, when $\eta = 1$, this expenditure share does not commove with firm scope. I therefore target the estimate for this co-movement from Column 3 of Table 5. The productivity shifter of headquarters inputs (γ) affects directly the level of the expenditure share on headquarters workers, and thus it shapes the amount of within-firm inequality. I therefore use it to discipline the importance of within-firm inequality in the baseline equilibrium, targeting the within-firm component of the overall variance of log wages across establishments in the economy (as measured in 1980).[24] These two parameters (η and γ) also affect the strength of within-firm non-rivalries in the economy, as manifested in the within-firm spillover patterns from Section 4.1. I do not target this moment explicitly due to the associated computational burden,[25] but I later verify that the model indeed yields a similar pattern to the estimates in Table 2 as an over-identification check. To match the between-firm component of inequality, I target the shape parameter of the productivity distribution, θ, which determines the magnitude of differences between firms. Note that both a higher γ and a lower θ raise inequality between firms in the economy, but they have opposing effects on inequality within firms.

The other parameters are those in the cost of expansion C and the entry cost. For the distance elasticity in the cost of expansion, ρ, I target the distance elasticity of headquarters-branch linkages from Table 6, by running the gravity regression in the simulated model. For the state-border effect, κ, I target the state-border effect from the same gravity regression. For the scale of the cost of new establishments, τ_0, I target the average number of establishments per firm. For the span of control component of the cost function, δ, I target the employment-weighted number of establishments per firm. Finally, for the entry cost f_e, I target average firm size, i.e. the ratio of total employment (normalized to 1 in the model) to the total mass of firms.

To estimate Θ, I minimize the loss function $\mathcal{L}(\Theta) \equiv (m(\Theta) - \tilde{m})' \boldsymbol{W} (m(\Theta) - \tilde{m})$ where $m(\Theta)$ is the vector of simulated moments from the model; $\tilde{m}$ 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 $\tilde{m}$, expressing the moments in percentage-deviation terms.[26] Results from this estimation procedure are provided in Table 8.

Table 8: Estimated parameters for the production block of the model

Parameter	Interpretation	Estimate	Main targeted moment	Value (model/data)
			Production function and productivity distribution	
η	EoS between headquarters and branches	1.28	Co-movement of relative expenditure on HQ and firm scope (Table 4)	0.15 / 0.15
γ	Productivity shifter for HQ workers	0.19	Within-firm component of (within-sector) variance of log wages across establishments (Census LBD)	0.11 / 0.12
θ	Dispersion of firm productivity	7.89	Between-firm component of (within-sector) variance of log wages across establishments (Census LBD)	0.12 / 0.12
			Cost of expansion (C) and entry cost	
ρ	Distance elasticity in cost of new establishments	0.76	Distance elasticity of HQ-branch linkages (Table 6)	-0.95 / -0.98
κ	State border effect in cost of new establishments	2.50	State border effect in HQ-branch linkages (Table 6)	-1.30 / -1.28
τ_0	Shifter of cost of new establishments	2.15	Average number of establishments per firm (Census BDS)	1.19 / 1.22
δ	Shifter of span of control in the cost of expansion	5.12	Average number of establishments per firm (weighted, Census BDS)	15.4 / 15.2
f_e	Entry cost	10^6	Firm size (employment per firm, Census BDS)	26.5 / 24.5

[24] Within-firm inequality is easier to measure than the expenditure share on headquarters workers (see discussion in Section 4), and especially how it changed over time. Nevertheless, I show that the model also yields reasonable values for the expenditure share on headquarters workers as an over-identification check below.

[25] It requires an additional computation of the equilibrium for every guess of Θ, in addition to the model inversion that I perform for each guess.

[26] 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.

5.6 Over-identification

As one assessment of the model's performance, I now briefly report how the quantified model performs for moments that were not explicitly targeted in the estimation procedure.

Within-firm spillovers. As discussed in Section 4, a key implication of the non-rivalry assumption is the existence of cross-region spillovers within firms. I measured such spillovers in Table 2. Replicating the regression that captures how a firm's expansion in a given market responds to exogenous income growth in the firm's other markets (Column 3) within the model, I find a coefficient of around 0.8. This is slightly lower than the 2SLS point-estimate from Table 2, but within its 95% confidence interval. The model thus replicates well the cross-region spillovers within firms that are implied by the non-rivalry assumption.

Headquarters expenditure share. Due to the challenges of measuring firms' headquarters expenditure share in the data (see discussion in Section 4), I do not target its level explicitly, and prefer to target the (more accurate) variance of log wages within firms. Nevertheless, we can compare the output of the model to different approaches in the data. Consider the approach of measuring headquarters payroll share using establishments' industry classifications in the LBD, as described in Section 4.2.2. According to this approach, the expenditure share on headquarters workers out of total payroll to headquarters and branches is around 16%. A potential upper bound is to treat only establishments in the firm's main NAICS-4 industry as "branches", which yields an headquarters income share of 23%; a potential lower bound is to treat only establishments in the NAICS-55 sector (which are explicitly reported as headquarters activity) as "headquarters", and all other establishments (including other within-firm business services) as "branches", which yields 8%. In the quantified model, the payroll-weighted average of this statistic across firms is 14%, in-between the upper and lower bounds, and very close to my preferred estimate.[27]

Spatial distribution of headquarters. While the allocation of population, wages, and skills was explicitly targeted as part of the model's inversion, I did not use data on the spatial distribution of firms and headquarters. The correlation of the log measure of headquarters between the model and the data is above 0.9. This high correlation results from the fact that this measure scales very strongly with total regional employment. In the model, the tight link to regional employment results from the combination of free entry and high spatial frictions for headquarters-branch linkages.

[27] An alternative approach to measure this moment is the data is to use the notion of SG&A costs (Selling, General, and Administrative Expenses) for publicly-listed firms. This measure is often used to proxy for headquarters and other overhead costs. Depending on the specific data restrictions, the share of these costs in total costs is around 15%-20%, again very close to the share in the model.

6 Quantitative experiments

In this section, I use the quantitative model to evaluate the aggregate and distributional implications of firms' spatial expansion. I focus on a particular set of shocks to the baseline equilibrium: changes in spatial frictions to firm expansion, i.e. on the geographical aspects of the cost of expansion $C(\cdot)$ – the distance elasticity ρ and the state-border effect κ. There are a few key advantages in focusing on these shocks when studying the importance of firm expansion. First, since they are part of the cost of expansion $C(\cdot)$, all their impacts on equilibrium, labor markets, and welfare are materialized through firms' market penetration decisions $x_{ij}(z)$, and thus through the novel mechanisms outlined in this paper. These shocks have no effect when holding constant expansion decisions, in contrast to other primitives such as shocks to demand, firms' production function, or regional fundamentals. Second, while many primitives in the model can induce an increase in firms' geographical scope, we can easily discipline changes in these geographical frictions from data on bilateral headquarters-branch linkages. Moreover, as detailed below, I use only information on changes in the number branches in distant markets *relative* to branches in firms' headquarters market, thus abstracting from any expansion patterns that are unrelated to spatial frictions. Lastly, these shocks have a clear interpretation of lower frictions to operate far-away establishments. Thus, they capture specific technological and policy changes that shape firms' communication frictions across space, such as improvements in information technologies, the expansion of air-travel, and regulatory changes.

6.1 Lower frictions to operate far-away branches

The main shock that I explore is a realistic decline in spatial frictions to firm expansion, driven by an economy-wide decline in the distance elasticity ρ. This shock is motivated by a sharp decline in the distance-elasticity for headquarters-branch linkages in the data, as can be seen in Table 6. Between 1980 and 2017, this distance elasticity has declined (in absolute value) by around 0.2 when using a log-linear specification and by around 0.4 when using a Poisson Pseudo-Maximum Likelihood estimator. I thus change ρ in the model to achieve a decline of around 0.3 in this distance elasticity, holding other primitives constant. Importantly, I use only data on relative distances between headquarters and branches – and no labor market data – to discipline this shock.[28]

Figure 4 displays changes in key aggregate moments of interest in response to this decline in the distance elasticity ρ, where I vary the values of ρ relative to the baseline equilibrium on the x-axis, ranging from zero (the baseline equilibrium) to a difference of 0.2

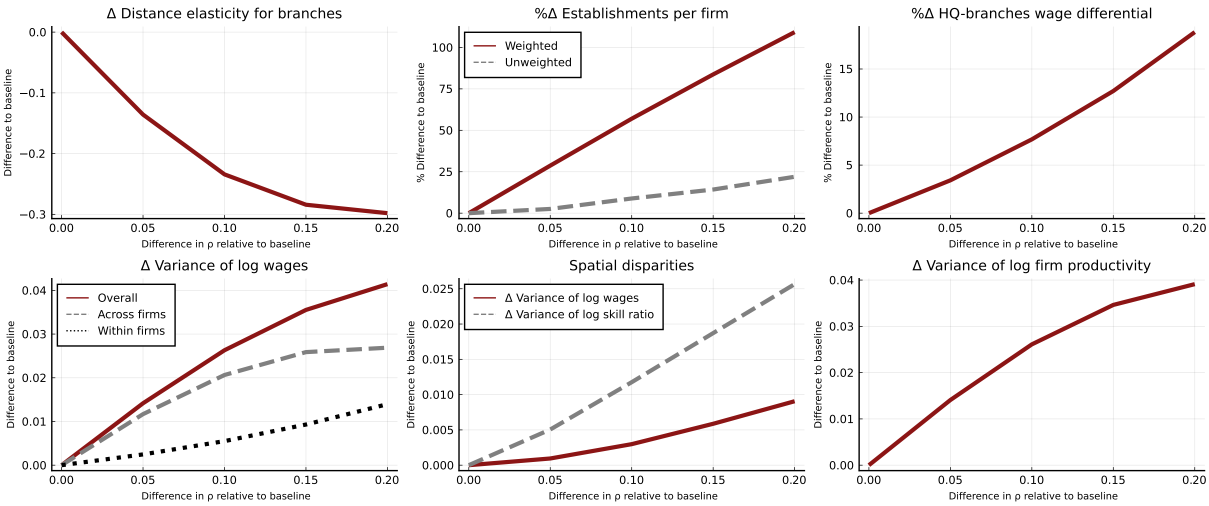
[28]Note that while the model does not admit an exact log-linear gravity equation, I verify that I replicate the change in the empirical distance elasticity by running the exact same gravity regression in the simulated model. The R^2 coefficient from this gravity regression in the model is over 0.9, suggesting that it approximates very closely the log-linear specification that I consider in the data.

relative to the baseline 1980 equilibrium. I then re-compute the equilibrium for every new value of ρ, and plot changes in selected outcomes relative to the baseline equilibrium on the y-axis. By construction, the gravity coefficient of distance (top-left subplot) declines by a similar magnitude to its decline in the data. Since this decline reflects a lower cost of expansion $C(\cdot)$, firms increase their average number of establishments (top-middle subplot). In line with the empirical evidence from Figure 1, this increase is driven by the large firms, such that the employment-weighted expansion is far greater than the unweighted expansion. Overall, the weighted expansion in the model is above 100%, over half of the equivalent increase in the data between 1980-2017.

Turning to labor market outcomes, the above spatial expansion leads to higher economy-wide headquarters-branch wage dispersion (top-right subplot), as in Proposition 1. In the model, the decline in distance elasticity raises this difference by close to 20% . In the data, this moment varies with the exact definition of headquarters and branches, but employing the definitions in Section 2 reveals an increase of about 30%, such that the distance-elasticity shock in the model accounts for around two-thirds of the trend in the data.

When investigating broader measures of inequality, the overall variance in log wages (bottom-left panel) rises by around 0.04 points relative to the baseline equilibrium. Compared to an increase of around 0.09 in the (within-industry) variance of log wages in the data, the decline in spatial frictions to expansion can rationalize above 40% of the observed trend. When considering the overall increase in the variance of log wages (and not just within-industry), the above shock can rationalize close to a quarter of it, though, recall that the model cannot speak to the cross-industry component which is not modeled. Moreover, the model gives rise to higher inequality both across firms and within them, in line with the patterns described in Section 2, and in contrast to theories that highlight only greater differentiation between firms (such as rising outsourcing). To see why inequality across firms rises in the model, recall that the model allows for endogenous changes in the dispersion of firm productivity – defined as sales-per-worker – due to the existence of within-firm non-rivalries. Indeed, in the quantitative model, the dispersion in firm productivity rises following a reduction in the cost of expansion, as depicted in the bottom-right panel. In principle, greater firm scope might also lead to lower dispersion in productivity and wages across firms, if expansion is concentrated in smaller firms. This is not the case in the quantified model or in the data, but it introduces concavity in the relationship between the cost of expansion and dispersion between firms.

Figure 4: Labor market implications of lower spatial frictions to firm expansion

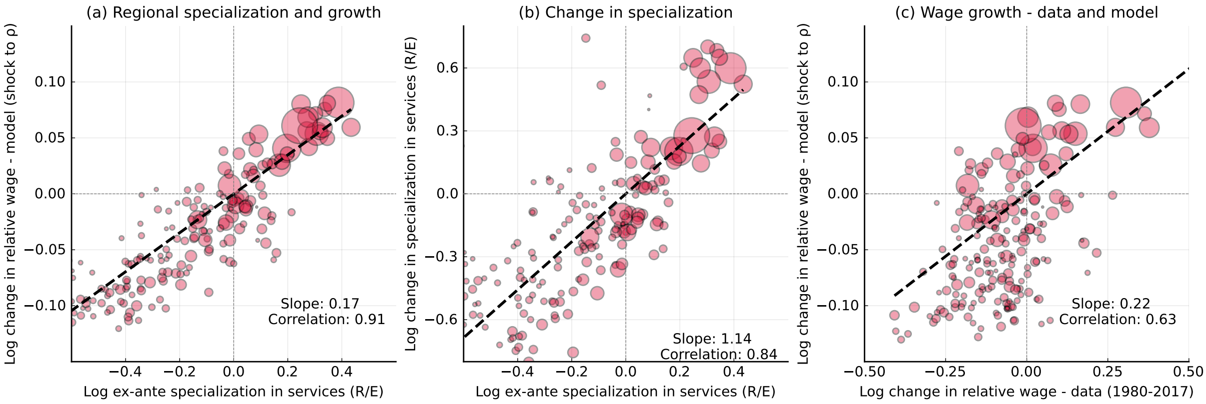


Note: Distributional implications of lower spatial frictions to firm expansion. The x-axis depicts changes in the distance elasticity in the cost of expansion, ρ, expressed in terms of the difference from the baseline equilibrium (with zero representing the baseline equilibrium). Positive values indicate lower distance elasticity in absolute value. The y-axis in each subplot captures changes in selected moments of interest relative to the baseline equilibrium, after re-computing the equilibrium for every new value of ρ: change in the gravity coefficient of headquarters-branch linkages (top-left panel); % change in average establishments per firm (employment-weighted and unweighted, top-middle); % change in mean wage dispersion between headquarters and branches (top-right); change in variance of log wages (overall, across firms, and within firms; bottom-left); change in the variance of log wages and the variance of high-skilled to low-skilled employment (bottom-middle); change in the variance of log firm productivity (bottom-right).

Finally, spatial disparities also increase, in line with Proposition 4. The bottom-middle panel of Figure 4 depicts changes in spatial inequality, as measured by the variance of log wages across regions (solid red line); and changes in spatial segregation, as measured by the variance of log skill-to-unskilled employment ratio across space (dashed gray line). Both rise when the distance elasticity of expansion falls, by a magnitude that equals around a third of the rise in spatial inequality in the data.

To understand which regions gain the most from lower barriers to firm expansion, panel (a) in Figure 5 depicts regional wage growth following the above shock to ρ against a measure of ex-ante regional specialization in providing headquarters services. As in Proposition 4, I measure such specialization using the ratio of total national sales of locally-headquartered firms to regional expenditure on locally-active branches, R_i/E_i. A few patterns emerge. First, specialization in headquarters services (relative to production of tradables) is a strong predictor of regional wage growth following a decline in the cost of expansion, since the markets that benefit the most are those with comparative advantage in providing headquarters services to other regions. Note that this comparative advantage can arise in the model either directly through high regional productivity for local headquarters (higher A_i); or indirectly through greater market access (low outgoing bilateral frictions τ_{ij}) and/or high abundance of relatively skilled labor (e.g. due to high skill-specific amenities B_{is}). Second, it is the relatively larger markets that specialize in such activity, and consequently that gain the most from a national increase in firms' scope. Thus, this shock can partially rationalize the "urban-bias" of regional income growth in recent decades, as in Moretti (2012).

Figure 5: Regional implications of lower spatial frictions to firm expansion



Note: Panel (a) shows regional wage growth in the model following a reduction of the distance elasticity ρ against a measure of regional specialization in headquarters services in the baseline equilibrium. See Section 6 for additional details. Panel (b) shows changes in this measure of specialization following a shock to ρ against its value in the baseline equilibrium. Panel (c) shows regional wage growth in the model following a reduction of the distance elasticity ρ against regional wage growth in the data between 1980 and 2017. In all figures, each circle is one of 200 U.S. local labor markets as defined in Section 5, weighted by total employment in 1980. Dashed black lines capture employment-weighted linear fits. All values are demeaned by the employment-weighted average of each measure.

One key channel through which lower barriers to firm expansion lead to higher spatial disparities is the magnification of the above regional specialization in headquarters services, as summarized in Proposition 4. Indeed, in the quantified model, a reduction in the cost of expansion leads to greater variation in the specialization measure R_i/E_i. This can be seen in Panel (b) of Figure 5, which plots log changes in R_i/E_i across local markets following a decline in ρ against the ex-ante values of this measure in the baseline equilibrium. Large markets with high ex-ante specialization in headquarters services such as New-York City become even more specialized in such activities when firms can open branches in more markets. In return, these regions increasingly import tradable goods from smaller markets, which become increasingly specialized in the production of goods. This aligns with the empirical pattern that regions with greater expansion of locally-headquartered firms are also the ones that see a greater decline in their manufacturing activity, as discussed in Section 4.

Finally, a natural question is how do the above regional manifestations compare to actual patterns of regional growth in the data. Panel (c) of Figure 5 plots log changes in regional average wages in the model, following a shock to ρ, against these changes in the data between 1980-2017 (both demeaned to capture relative changes in wages).[29] Evidently, there is a strong positive relationship between the model and the data. The

[29]To improve visual clarity, the figure omits one notable outlier: the local labor market of San Jose,

scale of changes in the model is smaller than the data, suggesting that a lower cost of firm expansion cannot be the only explanation for the observed patterns. Indeed, as noted above, this shock can explain around a third of the rise in variance of log wages across space that was observed in the data. At the same time, the prediction that larger markets that specialize in headquarters activities should see the greater increase in wages aligns with the experience of the U.S. economy in recent decades, and the correlation between regional growth in the data and in the model is above 0.6. It is also worth recalling that the above experiment considers just a homogeneous increase in the ability to operate far-away branches across all markets. Thus, the model can rationalize much of the regional experiences in the U.S. economy since the 1980s without delving into region-specific shocks and histories.

6.2 Deregulation of cross-state firm activity

Finally, I consider a counterfactual that lowers the cross-state frictions in firms' spatial expansion (κ), holding the importance of distance (ρ) and other primitives constant. This shock is motivated by the decline in the importance of state-border effects for cross-region headquarters-branch linkages, as can be seen in Table 6. Between 1980 and 2017, this state-border effect has declined (in absolute value) by approximately 0.1 when using the log-linear OLS estimator, and by approximately 0.4 when using the Poisson Pseudo-Maximum Likelihood estimator.

One natural interpretation for the decline in state border effects are changes in regulations and compliance requirements across U.S. states. Over the 1980s and 1990s multiple service sectors have undergone substantial deregulation that increased the ability of firms to operate cross-state operations, including in Financial Services (e.g. the Riegle–Neal Interstate Banking and Branching Efficiency Act of 1994, alongside many state-level changes), Telecommunications (The Telecommunications Act of 1996), Air Travel (The Airline Deregulation Act of 1978), Trucking (Motor Carrier Act of 1980), and Electricity (Energy Policy Act of 1992).

I study such changes in the model by shocking κ until the state border effect in the model declines by 0.2. A replication of Figure 4 for this experiment is provided in Appendix D.6. The effect on firm expansion is much smaller relative to the decline in the distance elasticity ρ, raising the (employment-weighted) average number of establishments per firm by less than 10%. This suggests that it is hard to rationalize the vast expansion in the data using the wave of deregulation in the 1980s and the 1990s. Accordingly, the effect on inequality is also minor, with the variance of log wages rising by approximately 0.005. Average welfare increases by 0.8%, capturing gains from more variety in the economy

CA, which includes Silicon Valley. Wage growth in this market has been exceptionally high in the data since the 1980s. While it is also positive and high in the model, it does not constitute a particular outlier. This market is included in the computation of the regression slope and the correlation.

(as each firm can provide its product in more markets); increased productivity, due to the non-rival nature of headquarters inputs (rising by 0.2% for the average firm); and increased regional specialization. Of course, such an experiment cannot capture the full welfare implications of the above deregulations, which are often associated with specific industry-level distortions. Nevertheless, it provides a sense of the equity-efficiency tradeoff that arises from policies that shape firms' expansion opportunities for the typical sector, on top of the role of sector-specific distortions. Note also that the above manifestations are consistent with the empirical findings in Philippon and Reshef (2012), in which the deregulation of the U.S. financial industry is associated with an increase in the skill-premium and top executive compensation in that sector.

7 Conclusion

A key trend in the U.S. economy in recent decades is the increase in firms' spatial scope, particularly in services-producing sectors. This paper offers new theory and evidence that link this trend to changes in the distribution of wages, motivated by the central role of multi-location service firms in the rise of wage inequality. The key idea is that when the output of some workers is non-rival across the firm's locations, changes in firms' scope can generate rich distributional implications. I develop a model that formalizes this idea, yielding new micro-foundations for skill-biased technical-change, rising spatial disparities, and a simultaneous increase in wage dispersion across and within firms – all of which are observed in the data in recent decades. I provide reduced-form evidence for these predictions and for the assumption of within-firm non-rivalries, and quantify their aggregate relevance by estimating the model for the U.S. economy. I show that a reduction in spatial frictions to firm expansion can rationalize much of the observed labor market trends in the U.S. economy since the 1980s. I also demonstrate how this framework can be used to evaluate the implications of policies that shape firms' ability to span multiple markets, such as a deregulation of cross-state firm activity.

I highlight four main benefits of the framework in the paper to the broad question of wage inequality. First, relative to some other theories of rising inequality such as foreign trade or automation, it is well suited to the services sector, which accounts for most of the economy and for most of the increase in inequality. Second, this framework can simultaneously rationalize multiple labor market trends that for the most part have been studied in separate strands of the literature. Third, it aligns with the observed trends in the organization of production and the expansion of multi-establishment firms. Finally, it generates distinctive predictions that are not covered by existing theories of wage inequality, such as the growth of wage dispersion across establishments within firms and the centrality of headquarters-intensity in understanding inequality across occupations.

More broadly, the model in this paper demonstrates the importance of multi-location firms and the network of headquarters-branch linkages for many economic questions, including inequality, the propagation of shocks across space, and the transition of the economy towards services (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," *Review of Economic Studies*, 2023, *90* (6), 2675–2702.

Alviarez, Vanessa, Javier Cravino, and Natalia Ramondo, "Firm-embedded productivity and cross-country income differences," *Journal of Political Economy*, 2023, *131* (9), 2289–2327.

Amiti, Mary and Donald R Davis, "Trade, firms, and wages: Theory and evidence," *The Review of economic studies*, 2012, *79* (1), 1–36.

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

Arkolakis, Costas, "Market penetration costs and the new consumers margin in international trade," *Journal of political economy*, 2010, *118* (6), 1151–1199.

_ , 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.

Autor, David, David Dorn, and Gordon Hanson, "When work disappears: Manufacturing decline and the falling marriage market value of young men," *American Economic Review: Insights*, 2019, *1* (2), 161–178.

Autor, David H and David Dorn, "The growth of low-skill service jobs and the polarization of the US labor market," *American economic review*, 2013, *103* (5), 1553–1597.

Azar, José A, Steven T Berry, and Ioana Marinescu, "Estimating labor market power," Technical Report, National Bureau of Economic Research 2022.

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

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.

Bartik, Timothy J, "Who benefits from state and local economic development policies?," 1991.

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.

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, Donald R and James Harrigan, "Good jobs, bad jobs, and trade liberalization," *Journal of international Economics*, 2011, *84* (1), 26–36.

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

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, *Industry linkages from joint production*, US Census Bureau, Center for Economic Studies, 2023.

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

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," *The Review of Economic Studies*, 2023, *90* (1), 331–375.

Garicano, Luis and Esteban Rossi-Hansberg, "Inequality and the Organization of Knowledge," *American Economic Review*, 2004, *94* (2), 197–202.

— **and** — , "Organization and inequality in a knowledge economy," *The Quarterly journal of economics*, 2006, *121* (4), 1383–1435.

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

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

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

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," *Journal of Political Economy Macroeconomics*, 2023, *1* (1), 3–42.

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.

Kline, Patrick, Neviana Petkova, Heidi Williams, and Owen Zidar, "Who profits from patents? rent-sharing at innovative firms," *The quarterly journal of economics*, 2019, *134* (3), 1343–1404.

Krusell, Per, Lee E Ohanian, José-Víctor Ríos-Rull, and Giovanni L Violante, "Capital-skill complementarity and inequality: A macroeconomic analysis," *Econometrica*, 2000, *68* (5), 1029–1053.

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," *Journal of Political Economy*, 2024, *132* (3), 000–000.

Philippon, Thomas and Ariell Reshef, "Wages and human capital in the US finance industry: 1909–2006," *The Quarterly Journal of Economics*, 2012, *127* (4), 1551–1609.

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.

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.

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.

Wenning, Derek, "Equal prices, unequal access: The effects of national pricing in the us life insurance industry," *Unequal Access: The Effects of National Pricing in the US Life Insurance Industry (February 29, 2024)*, 2024.

Appendices

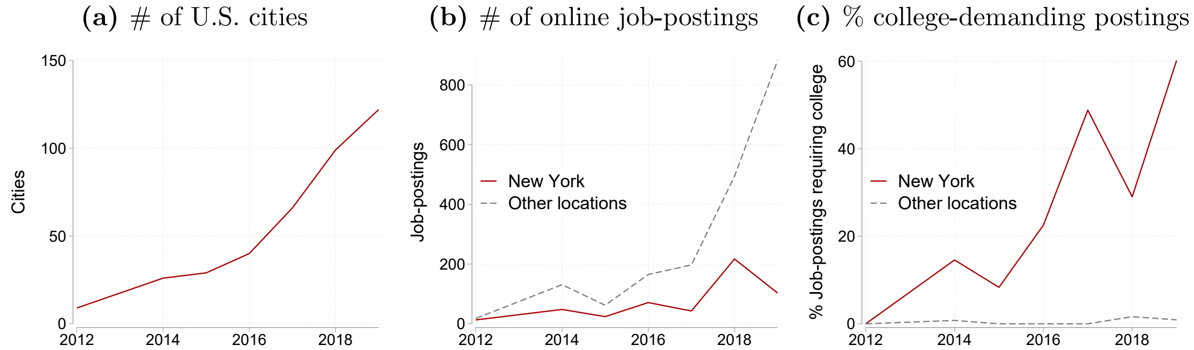
A Case studies for firm spatial expansion

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 Lightcast. Panel (a) in Figure 6 below shows the number of cities with postings by Shake Shack in the Lightcast 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 6: Example: the expansion of Shake Shack (online job-postings data)



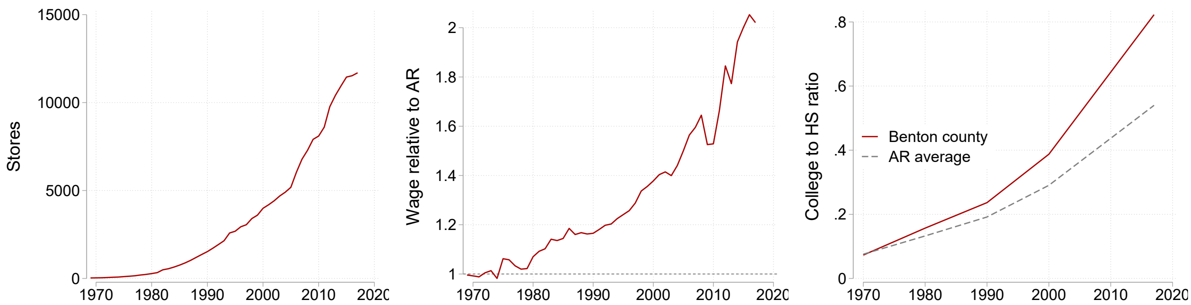
This example demonstrates the heterogenous effects of firm expansion across two dimensions that are considered in the model: first, in the cross-section of establishments, between firm headquarters and its branches; second, across space, between New York City and other locations in the U.S.. This example suggests potentially meaningful distributional implications in the economy as a whole, given that the *average firm* in the economy has exhibited this kind of spatial expansion in recent decades.

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 7 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 resembled 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 a 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 7: 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 Model derivations

B.1 Proof of proposition 1

Recall that $x_{ij}(z) W_{\ell j} \ell_{ij}(z)^{\frac{\epsilon+1}{\epsilon}}$ is the firm's payroll in region j and denote by $x_{ij}(z) r_{ij}(z)$ the firm's sales in region j. The share of $x_{ij}(z) r_{ij}(z)$ that is allocated towards labor compensation (in both the branch and the headquarters) is given by $\psi \equiv \frac{\epsilon}{\epsilon+1}\frac{\sigma-1}{\sigma}$. Define the share of payroll originated from market j ($\psi x_{ij}(z) r_{ij}(z)$) that is allocated towards headquarters labor as $\Gamma_{ij}(z)$, given by

$$\Gamma_{ij}(z) \equiv 1 - \frac{W_{\ell j}\ell_{ij}(z)^{\frac{\epsilon+1}{\epsilon}}}{\psi r_{ij}(z)}.$$

It is also helpful to define $h_{ij}(z) \equiv \left(\frac{\Gamma_{ij}(z)}{1-\Gamma_{ij}(z)}\frac{W_{\ell j}}{W_{hi}}\right)^{\frac{\epsilon}{\epsilon+1}} \ell_{ij}(z)$ as the hypothetical market-specific allocation of headquarters labor. Note that $W_{hi} h_i(z)^{\frac{\epsilon+1}{\epsilon}} = \sum_{j=1}^{N} x_{ij}(z) W_{hi} h_{ij}(z)^{\frac{\epsilon+1}{\epsilon}}$, or $h_i(z) = \left[\sum_{j=1}^{N} x_{ij}(z) h_{ij}(z)^{\frac{\epsilon+1}{\epsilon}}\right]^{\frac{\epsilon}{\epsilon+1}}$.

By the firm's first order condition with respect to $\ell_{ij}(z)$, the ratio of the j-originated headquarters payroll to the branch payroll in j is

$$\begin{aligned}
\frac{x_{ij}(z) W_{hi} h_{ij}(z)^{\frac{\epsilon+1}{\epsilon}}}{x_{ij}(z) W_{\ell j}\ell_{ij}(z)^{\frac{\epsilon+1}{\epsilon}}} &= \frac{x_{ij}\psi r_{ij}(z) - x_{ij}(z) W_{\ell j}\ell_{ij}(z)^{\frac{\epsilon+1}{\epsilon}}}{x_{ij}(z) W_{\ell j}\ell_{ij}(z)^{\frac{\epsilon+1}{\epsilon}}} \\
&= \frac{\Gamma_{ij}(z)}{1-\Gamma_{ij}(z)} \\
&= \gamma\left(\frac{h_i(z)}{\ell_{ij}(z)}\right)^{\frac{\eta-1}{\eta}}
\end{aligned}$$

Rearranging both sides of this expression we obtain

$$\frac{W_{hi}}{W_{\ell j}}\left(\frac{h_{ij}(z)}{\ell_{ij}(z)}\right)^{\frac{\epsilon+1}{\epsilon}} = \left(\frac{W_{hi}}{W_{\ell j}}\right)^{-\frac{\frac{\eta-1}{\eta}}{\frac{\epsilon+1}{\epsilon}-\frac{\eta-1}{\eta}}}\left[\gamma\left(\frac{h_i(z)}{h_{ij}(z)}\right)^{\frac{\eta-1}{\eta}}\right]^{\frac{\frac{\epsilon+1}{\epsilon}}{\frac{\epsilon+1}{\epsilon}-\frac{\eta-1}{\eta}}}.$$

The headquarters payroll share at the firm level can then be obtained as

$$
\begin{aligned}
\frac{\Gamma_i(z)}{1-\Gamma_i(z)} &= \frac{\sum_{j=1}^N x_{ij}(z) W_{hi} h_{ij}(z)^{\frac{\epsilon+1}{\epsilon}}}{\sum_{k=1}^N x_{ik}(z) W_{\ell k} \ell_{ik}(z)^{\frac{\epsilon+1}{\epsilon}}}, \\
&= \sum_{j=1}^N \frac{x_{ij}(z) W_{\ell j} \ell_{ij}(z)^{\frac{\epsilon+1}{\epsilon}}}{\sum_{k=1}^N x_{ik}(z) W_{\ell k} \ell_{ik}(z)^{\frac{\epsilon+1}{\epsilon}}} \frac{W_{hi} h_{ij}(z)^{\frac{\epsilon+1}{\epsilon}}}{W_{\ell j} \ell_{ij}(z)^{\frac{\epsilon+1}{\epsilon}}}, \\
&= \gamma^{\frac{1}{1-\frac{\epsilon}{\epsilon+1}\frac{\eta-1}{\eta}}} \sum_{j=1}^N \frac{x_{ij}(z) W_{\ell j} \ell_{ij}(z)^{\frac{\epsilon+1}{\epsilon}}}{\sum_{k=1}^N x_{ik}(z) W_{\ell k} \ell_{ik}(z)^{\frac{\epsilon+1}{\epsilon}}} \left(\left(\frac{W_{\ell j}}{W_{hi}}\right)\left(\frac{h_i(z)}{h_{ij}(z)}\right)^{\frac{\epsilon+1}{\epsilon}}\right)^{\frac{\frac{\epsilon}{\epsilon+1}\frac{\eta-1}{\eta}}{1-\frac{\epsilon}{\epsilon+1}\frac{\eta-1}{\eta}}}, \\
&= \gamma^{\frac{1}{1-\frac{\epsilon}{\epsilon+1}\frac{\eta-1}{\eta}}} \sum_{j=1}^N \frac{x_{ij}(z) W_{\ell j} \ell_{ij}(z)^{\frac{\epsilon+1}{\epsilon}}}{\sum_{k=1}^N x_{ik}(z) W_{\ell k} \ell_{ik}(z)^{\frac{\epsilon+1}{\epsilon}}} \\
&\times \left(\frac{W_{\ell j}}{W_{hi}} \frac{\Gamma_i(z)}{\Gamma_{ij}(z)} \sum_{m=1}^N \frac{x_{im}(z)}{\sum_{k=1}^N x_{ik}(z)} \frac{r_{im}(z)}{r_{ij}(z)} \sum_{j=1}^N x_{ik}(z)\right)^{\frac{\frac{\epsilon}{\epsilon+1}\frac{\eta-1}{\eta}}{1-\frac{\epsilon}{\epsilon+1}\frac{\eta-1}{\eta}}}, \\
&= \gamma^{\bar{\eta}} \left(\frac{\bar{W}_{\ell i}(z)}{W_{hi}} \left(\sum_{j=1}^N x_{ij}(z)\right)\right)^{\bar{\eta}-1}.
\end{aligned}
$$

where

$$
\begin{aligned}
\bar{W}_{\ell i}(z) &\equiv \left(\sum_{j=1}^N \Omega_{ij}(z) W_{\ell j}^{\bar{\eta}-1}\right)^{\frac{1}{\bar{\eta}-1}}, \\
\bar{\eta} &\equiv \frac{\eta}{\eta - \frac{\epsilon}{\epsilon+1}(\eta-1)},
\end{aligned}
$$

$$
\Omega_{ij}(z) \equiv \frac{x_{ij}(z) W_{\ell j} \ell_{ij}(z)^{\frac{\epsilon+1}{\epsilon}}}{\sum_{k=1}^N x_{ik}(z) W_{\ell k} \ell_{ik}(z)^{\frac{\epsilon+1}{\epsilon}}} \times \left(\frac{\Gamma_i(z)}{\Gamma_{ij}(z)} \times \sum_{m=1}^N \frac{x_{im}(z)}{\sum_{k=1}^N x_{ik}(z)} \frac{r_{im}(z)}{r_{ij}(z)}\right)^{\bar{\eta}-1}.
$$

In the case of symmetric space, the above expressions reduce to:

$$
\frac{\Gamma(z)}{1-\Gamma(z)} = \gamma^{\bar{\eta}} \left(\frac{W_\ell}{W_h} N x(z)\right)^{\bar{\eta}-1}.
$$

where $x(z)$ is the share of locations served by a firm of type z in each of the N regions in the economy. When in addition $\epsilon \to \infty$, we get that W_h and W_ℓ are the economy-wide prices of headquarters labor and branch-level labor, respectively, and that the headquarters payroll share for firm z is

$$
\Gamma(z) = \frac{\gamma^\eta \left(\frac{W_\ell}{W_h} N x(z)\right)^{\eta-1}}{1+\gamma^\eta \left(\frac{W_\ell}{W_h} N x(z)\right)^{\eta-1}}.
$$

Let $\bar{H}$ and $\bar{L}$ be the endowments of headquarters labor and branch-level labor in the economy, and let $F(z)$ be the cumulative distribution function for firms of type z. In this case:

$$\frac{\Gamma(z)}{1-\Gamma(z)} = \frac{W_h}{W_\ell}\frac{h(z)}{Nx(z)\,\ell(z)} = \gamma^\eta \left(\frac{W_\ell}{W_h}Nx(z)\right)^{\eta-1},$$

and

$$h(z) = \gamma^\eta \left(\frac{W_\ell}{W_h}\right)^\eta (Nx(z))^{\eta-1}\, Nx(z)\,\ell(z).$$

Labor market clearing for headquarters workers implies

$$\begin{aligned}\bar{H} &= \gamma^\eta \left(\frac{W_\ell}{W_h}\right)^\eta \int_z (Nx(z))^{\eta-1}\, Nx(z)\,\ell(z)\, dF(z),\\ &= \gamma^\eta \left(\frac{W_\ell}{W_h}\right)^\eta \int_z \frac{Nx(z)\,\ell(z)}{\bar{L}} (Nx(z))^{\eta-1}\, dF(z)\,\bar{L}.\end{aligned}$$

Therefore the ratio of headquarters wages to branch-level wages is given by

$$\frac{W_h}{W_\ell} = \gamma \bar{x}^{\frac{\eta-1}{\eta}} \left(\frac{\bar{H}}{\bar{L}}\right)^{-\frac{1}{\eta}},$$

where $\bar{x}$ is a weighted power-mean of firms' geographical scope in the economy,

$$\bar{x} \equiv \left(\int_z s_\ell(z)\,(Nx(z))^{\eta-1}\, dF(z)\right)^{\frac{1}{\eta-1}}, \quad s_\ell(z) \equiv \frac{Nx(z)\,\ell(z)}{\bar{L}},$$

with branch-level employment of each firm $(s_\ell(z))$ as the weights.

B.2 Proof of propositions 2 and 3

Recall the firm's problem

$$\begin{aligned}\max_{h_i(z),\{\ell_{ij}(z)\},\{x_{ij}(z)\}} &\sum_{j=1}^{N} x_{ij}(z)\,\Upsilon_j \left(A_i z \left(\gamma h_i(z)^{\frac{\eta-1}{\eta}} + \ell_{ij}(z)^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}}\right)^{\frac{\sigma-1}{\sigma}}\\ &- \sum_{j=1}^{N} x_{ij}(z)\, W_{\ell j}\ell_{ij}(z)^{\frac{\epsilon+1}{\epsilon}} - W_{hi}h_i(z)^{\frac{\epsilon+1}{\epsilon}} - C\left(\{x_{ij}(z)\}_{j=1}^{N}\right).\end{aligned}$$

The first order condition with respect to $\ell_{ij}(z)$ yields

$$\ell_{ij}(z) = \left(\psi\,(1-\Gamma_{ij}(z))\,\frac{r_{ij}(z)}{W_{lj}}\right)^{\frac{\epsilon}{\epsilon+1}},$$

where $\psi \equiv \frac{\epsilon}{\epsilon+1}\frac{\sigma-1}{\sigma}$, $r_{ij}(z)$ is the firm's local sales, and we define the intensity of headquarters labor in local production as $\Gamma_{ij}(z) \equiv \frac{\gamma h_i(z)^{\frac{\eta-1}{\eta}}}{\gamma h_i(z)^{\frac{\eta-1}{\eta}} + \ell_{ij}(z)^{\frac{\eta-1}{\eta}}}$. Noting that $w_{\ell,ij}(z) =$

$W_{\ell j}\ell_{ij}(z)^{\frac{1}{\epsilon}}$ and rearranging, we obtain

$$\log w_{\ell,ij}(z) = \log\psi + \frac{\epsilon}{\epsilon+1}\log W_{\ell j} + \frac{1}{\epsilon+1}\log\left(1-\Gamma_{ij}(z)\right) + \frac{1}{\epsilon+1}\log r_{ij}(z).$$

The first order condition with respect to $h_i(z)$ yields

$$W_{hi}h_i(z)^{\frac{\epsilon+1}{\epsilon}} = \psi\sum_{j=1}^{N}\Gamma_{ij}(z)\,x_{ij}(z)\,r_{ij}(z).$$

Noting that $w_{h,i}(z) = W_{hi}h_i(z)^{\frac{1}{\epsilon}}$ and rearranging, we obtain

$$w_{h,i}(z)^{\epsilon+1}W_{hi}^{-\epsilon} = \psi\underbrace{\left(\sum_{j=1}^{N}\frac{x_{ij}(z)\,r_{ij}(z)}{\sum_{k=1}^{N}x_{ik}(z)\,r_{ik}(z)}\Gamma_{ij}(z)\right)}_{=\Gamma_i(z)}\left(\sum_{k=1}^{N}x_{ik}(z)\,r_{ik}(z)\right),$$

which yields

$$\log w_{h,i}(z) = \log\psi + \frac{\epsilon}{\epsilon+1}\log W_{hi} + \frac{1}{\epsilon+1}\log\Gamma_i(z) + \frac{1}{\epsilon+1}\log\left(\sum_{k=1}^{N}x_{ik}(z)\,r_{ik}(z)\right).$$

We now turn to consider a small homogeneous reduction in the cost function $C(\cdot)$ that results in positive expansion in all markets $d\log x_{ij}(z) > 0$. Log-linearizing the first order condition for headquarters labor, we obtain

$$\frac{\epsilon+1}{\epsilon}d\log h_i(z) = \sum_{j=1}^{N}\phi_{ij}(z)\left[d\log\Gamma_{ij}(z) + d\log x_{ij}(z) + d\log r_{ij}(z)\right],$$

where we define

$$\phi_{ij}(z) \equiv \frac{\Gamma_{ij}(z)\,x_{ij}(z)\,r_{ij}(z)}{\sum_{k=1}^{N}\Gamma_{ik}(z)\,x_{ik}(z)\,r_{ik}(z)}.$$

Log-linearizing the first order condition for branch-level labor, we obtain

$$d\log\ell_{ij}(z) = \frac{\epsilon}{\epsilon+1}d\log r_j - \frac{\epsilon}{\epsilon+1}\frac{\Gamma_{ij}(z)}{1-\Gamma_{ij}(z)}d\log\Gamma_{ij}(z),$$

Log-linearizing the local labor intensity, we obtain

$$d\log\Gamma_{ij}(z) = \left(1-\Gamma_{ij}(z)\right)\frac{\eta-1}{\eta}\left(d\log h_i(z) - d\log\ell_{ij}(z)\right).$$

Log-linearizing local sales $r_{ij}(z)$, we obtain

$$d\log r_{ij}(z) = \frac{\sigma-1}{\sigma}\left(\Gamma_{ij}(z)\,d\log h_i(z) + \left(1-\Gamma_{ij}(z)\right)d\log\ell_{ij}(z)\right).$$

Combining the above expressions, we obtain two relationships between $d\log \ell_{ij}(z)$ and $d\log h_i(z)$:

$$
\begin{aligned}
d\log h_i(z) = {} & \frac{\epsilon}{\epsilon+1}\sum_{j=1}^{N}\phi_{ij}(z)\,d\log x_{ij}(z) \\
& +\sum_{j=1}^{N}\phi_{ij}(z)\left((1-\Gamma_{ij}(z))\frac{\eta-1}{\eta}+\Gamma_{ij}(z)\frac{\sigma-1}{\sigma}\right)d\log h_i(z) \\
& +\sum_{j=1}^{N}\phi_{ij}(z)(1-\Gamma_{ij}(z))\left(\frac{\sigma-1}{\sigma}-\frac{\eta-1}{\eta}\right)d\log \ell_{ij}(z),
\end{aligned}
\tag{A.1}
$$

$$
d\log \ell_{ij}(z) = \tilde{\Gamma}_{ij}(z)\,d\log h_i(z), \tag{A.2}
$$

where we denote

$$
\tilde{\Gamma}_{ij}(z) \equiv \frac{\frac{\epsilon}{\epsilon+1}\left(\frac{\sigma-1}{\sigma}-\frac{\eta-1}{\eta}\right)\Gamma_{ij}(z)}{1-\frac{\epsilon}{\epsilon+1}\left(\frac{\sigma-1}{\sigma}(1-\Gamma_{ij}(z))+\frac{\eta-1}{\eta}\Gamma_{ij}(z)\right)}<1.
$$

First, we show that $d\log h_i(z)>0$ following a small reduction in the cost of expansion. Combining expressions (A.1) and (A.2) and rearranging, we obtain

$$
d\log h_i(z) = \frac{\epsilon}{\epsilon+1}\left[1-\left(\left(1-\tilde{\phi}_i(z)\right)\frac{\sigma-1}{\sigma}+\tilde{\phi}_i(z)\frac{\eta-1}{\eta}\right)\right]^{-1}\sum_{j=1}^{N}\phi_{ij}(z)\,d\log x_{ij}(z),
$$

where we define

$$
\tilde{\phi}_i(z) \equiv \sum_{j=1}^{N}\phi_{ij}(z)(1-\Gamma_{ij}(z))\left(1-\tilde{\Gamma}_{ij}(z)\right).
$$

Since $\tilde{\phi}_i(z)<1$, we get that $\left(1-\tilde{\phi}_i(z)\right)\frac{\sigma-1}{\sigma}+\tilde{\phi}_i(z)\frac{\eta-1}{\eta}<1$ and $d\log h_i(z)>0$ when $\sum_{j=1}^{N}\phi_{ij}(z)\,d\log x_{ij}(z)>0$. Consequently, also $d\log w_{h,i}(z)>0$.

To see that headquarters wages increase more than branch-level wages, note that holding constant the labor-supply shifters (or allowing them to change but comparing headquarters and branch labor in the same market such that $i=j$), we get $d\log w_{h,i}(z) - d\log w_{\ell,ij}(z) = \frac{1}{\epsilon}\left(\frac{1}{\epsilon}d\log h_i(z)-d\log \ell_{ij}(z)\right)$. Since $\tilde{\Gamma}_{ij}(z)$ is always smaller than 1, a decline in the cost of expansion that keeps market-specific demand-shifters Υ_j constant results in $d\log \ell_{ij}(z)<d\log h_i(z)$. Consequently, $d\log w_{\ell,ij}(z)<d\log w_{h,i}(z)$.

Finally, whenever $d\log h_i(z)>0$, then $d\log w_{\ell,ij}(z)>0$ if and only if $\tilde{\Gamma}_{ij}(z)>0$. This is the case if $\frac{\sigma-1}{\sigma}-\frac{\eta-1}{\eta}>0$, or $\sigma>\eta$. This completes the proof.

B.3 Proof of proposition 4

Denote by R_i and E_i the total sales of locally-headquartered firms in region i and total expenditure on locally active branches in region i, respectively. R_i and E_i are given by

$$R_i = \sum_{j=1}^{N} M_i \int_z x_{ij}(z) \, r_{ij}(z) \, dG_i(z),$$

$$E_i = \sum_{n=1}^{N} M_n \int_z x_{ni}(z) \, r_{ni}(z) \, dG_n(z),$$

where M_i is the mass of firms headquartered in region i and $G_n(z)$ is the productivity distribution of these firms. In addition, define the headquarters-intensity of locally-headquartered firms, $\bar{\Gamma}_{out,i}$, and of locally active branches, $\bar{\Gamma}_{in,i}$, as

$$\bar{\Gamma}_{out,i} = \sum_{j=1}^{N} \int_z \left(\frac{x_{ij}(z) \, r_{ij}(z)}{\int_{z'} x_{ij}(z') \, r_{ij}(z') \, dG_i(z')} \right) \Gamma_{ij}(z) \, dG_i(z),$$

$$\bar{\Gamma}_{in,i} = \sum_{n=1}^{N} \int_z \left(\frac{x_{ni}(z) \, r_{ni}(z)}{\int_{z'} x_{ni}(z') \, r_{ni}(z') \, dG_n(z')} \right) \Gamma_{ni}(z) \, dG_n(z).$$

We can now write the total headquarters payroll in region i as

$$\psi \bar{\Gamma}_{out,i} R_i,$$

total payroll to branch workers at region i as

$$\psi (1 - \bar{\Gamma}_{in,i}) E_i.$$

and total payroll in the goods producing sector as $\psi R_{g,i}$, where $R_{g,i}$ stands for total sales from tradable goods in region i.

We now assume that the bundles of headquarters labor, branch-level labor, and tradable-goods labor are all Cobb-Douglas aggregators of labor from different skill groups $s \in \{1, ..., S\}$, with skill-intensities $\alpha_{h,s}, \alpha_{\ell,s}, \alpha_{g,s}$, respectively. Define as $\mathcal{S}_{is}$ the share of regional income in i that is paid to skill group s. From the above derivations, it can be written as

$$\mathcal{S}_{is} = \frac{\alpha_{g,s} \psi R_{g,i} + \alpha_{h,s} \bar{\Gamma}_{out,i} \psi R_i + \alpha_{b,s} \left(1 - \bar{\Gamma}_{in,i}\right) \psi E_i}{Y_i}.$$

where Y_i is total regional income.

Since income must equal expenditure, $E_i = \beta Y_i$, the above can be re-written as

$$\mathcal{S}_{is} = \beta\psi\alpha_{g,s}\frac{R_{g,i}}{E_i} + \beta\psi\alpha_{h,s}\bar{\Gamma}_{out,i}\frac{R_i}{E_i} + \beta\psi\alpha_{b,s}\left(1-\bar{\Gamma}_{in,i}\right).$$

Since overall trade in goods and services is balanced in the model, total outflow of headquarters services must equal to the total inflow of tradable goods:

$$R_{g,i} = \left(\frac{1}{\beta\psi} - \left(1-\bar{\Gamma}_{in,i}\right)\right)E_i - \bar{\Gamma}_{out,i}R_i.$$

Combining the above expressions, we obtain

$$\mathcal{S}_{is} = \left(1-\tilde{\beta}\right)\alpha_{g,s} + \tilde{\beta}\left[\alpha_{g,s} + \left(\alpha_{h,s}-\alpha_{g,s}\right)\bar{\Gamma}_{out,i}\frac{R_i}{E_i} + \left(\alpha_{b,s}-\alpha_{g,s}\right)\left(1-\bar{\Gamma}_{in,i}\right)\right], \quad \text{(A.3)}$$

where we denote $\tilde{\beta} = \beta\psi$.

Consider now a few special cases of Equation A.3. Suppose that s is the group of skilled labor. When headquarters production uses only skilled labor ($\alpha_{h,s} = 1$), branch production uses only low-skilled labor ($\alpha_{b,s} = 1$), and the skilled-intensity of tradable goods is given by $\alpha = \alpha_{g,s}$, then the skilled income share in region i is given by

$$\mathcal{S}_{is} = \left(1-\tilde{\beta}\right)\alpha + \tilde{\beta}\left[\left(1-\alpha\right)\bar{\Gamma}_{out,i}\frac{R_i}{E_i} + \alpha\bar{\Gamma}_{in,i}\right],$$

as in Proposition 4.

For the economy as a whole, the income share of skill-groups s is given by

$$\mathcal{S}_s = \sum_{i=1}^{N}\frac{Y_i}{\sum_{j=1}^{N}Y_i}\mathcal{S}_{is} = \left(1-\tilde{\beta}\right)\alpha + \tilde{\beta}\bar{\Gamma},$$

where $\bar{\Gamma}$ is the aggregate intensity of the factors producing the non-rival headquarters input in the production of services.

When $\eta = 1$, $\bar{\Gamma} = \bar{\Gamma}_{out,i} = \bar{\Gamma}_{in,i} = \frac{\gamma}{\gamma+1}$. In this case, the aggregate skill-intensity is invariant to the spatial distribution of economic activity and given by

$$\mathcal{S}_s = \left(1-\tilde{\beta}\right)\alpha + \tilde{\beta}\frac{\gamma}{\gamma+1}.$$

In this case, regional skill-intensity is summarized by regional specialization in providing headquarters services, captured by the ratio $\frac{R_i}{E_i}$:

$$\mathcal{S}_{is} = \left(1-\tilde{\beta}\right)\alpha + \tilde{\beta}\left[\left(1-\alpha\right)\frac{R_i}{E_i} + \alpha\right]\frac{\gamma}{\gamma+1}.$$

In particular, when there is no cross-region firm activity (e.g. since the cost of expansion is infinite when $x_{ij}(z) > 0$ for $i \neq j$), then all spatial differences in skill-intensity are eliminated, and skill-intensity in all regions is given by $\left(1-\tilde{\beta}\right)\alpha + \tilde{\beta}\frac{\gamma}{\gamma+1}$.

C Model inversion

The set of location fundamentals $\mathcal{A} \equiv \{A_i, B_{is}, A_{g,i}\}_{i,s}$ includes the productivity shifters in the services sector, regional amenities, and regional productivity in tradable-goods. I obtain $\mathcal{A}$ by inverting the model's equilibrium conditions, conditional on all other parameters.

To this end, I match three sets of regional outcomes in the data: total regional labor income, Y_i, regional employment in each skill group, L_{is}, and relative payroll for skilled workers, which I label ω_i. I define ω_i as the ratio between total payroll of workers in the upper half of the skill distribution (i.e., the top 2 skill groups that I defined in Section 5) to the income of workers in the bottom half of the skill distribution (the bottom 2 skill groups from Section 5).

C.1 Preliminaries: obtaining regional sales and expenditure

It is first useful to recall the definition of key regional outcomes in the model, which I defined in Sections 3 and B.3. I denote by R_i and E_i total sales of the firms that are headquartered in region i and expenditure on branches that operate in region i, respectively. I denote by $R_{g,i}$ total regional sales of tradable goods in region i. I define by $\bar{\Gamma}_{out,i}$ and $\bar{\Gamma}_{in,i}$ the headquarters-intensity of locally-headquartered firms and of locally active branches, respectively.[30] Finally recall that $\alpha_{hs}, \alpha_{\ell s}, \alpha_{gs}$ denote the intensity of group s in the headquarters labor bundle, branch-level labor bundle, and tradable-goods labor bundle, respectively. With the above definition, total income of group s in region i is given by

$$Y_{is} = \psi \left(\alpha_{hs} \bar{\Gamma}_{out,i} R_i + \alpha_{bs} \left(1 - \bar{\Gamma}_{in,i}\right) E_i + \alpha_{gs} R_{g,i} \right),$$

where recall that $\psi \equiv \frac{\sigma-1}{\sigma}\frac{\epsilon}{\epsilon+1}$ is the payroll share of sales. We then have that ω_i in the model is defined as

$$\omega_i \equiv \frac{Y_{i1} + Y_{i2}}{Y_{i3} + Y_{i4}},$$

where groups 1 and 2 are the top two skill groups and groups 3 and 4 are the bottom two skill groups.

The first step of the inversion is to obtain E_i, R_i, and $R_{g,i}$ that are consistent with the model's structure and with the above observables Y_i and ω_i. Since E_i is total local

[30] I elaborate on these definitions in Appendix B.3:

$$\bar{\Gamma}_{out,i} = \sum_{j=1}^{N} \int_z \left(\frac{x_{ij}(z) r_{ij}(z)}{\int_{z'} x_{ij}(z') r_{ij}(z') dG_i(z')} \right) \Gamma_{ij}(z) dG_i(z),$$

$$\bar{\Gamma}_{in,i} = \sum_{n=1}^{N} \int_z \left(\frac{x_{ni}(z) r_{ni}(z)}{\int_{z'} x_{ni}(z') r_{ni}(z') dG_n(z')} \right) \Gamma_{ni}(z) dG_n(z).$$

expenditure on local branches, it is simply given by $E_i = \beta Y_i$, where recall that β is exogenous expenditure share on services. Now, let s denote the upper half of the skill distribution and k denote the lower half. From the above expression for total income of group s in region i, we have that

$$\left(\omega_i \alpha_{gk} - \alpha_{gs}\right) R_{g,i} = \left(\alpha_{hs} - \omega_i \alpha_{hk}\right) \bar{\Gamma}_{out,i} R_i + \left(\alpha_{\ell s} - \omega_i \alpha_{\ell k}\right) \left(1 - \bar{\Gamma}_{in,i}\right) E_i.$$

In addition, regional balance of payments in the model implies that

$$R_{g,i} = \left(\frac{1}{\beta\psi} - \left(1 - \bar{\Gamma}_{in,i}\right)\right) E_i - \bar{\Gamma}_{out,i} R_i.$$

Combining these two equations, we obtain two equations for R_i and $R_{g,i}$, conditional on parameters, observables, and guesses for $\bar{\Gamma}_{in,i}$ and $\bar{\Gamma}_{out,i}$:

$$R_{g,i} = \frac{\left(\beta\psi\right)^{-1}\left(\alpha_{hs} - \omega_i \alpha_{hk}\right) + \left(\left(\alpha_{\ell s} - \omega_i \alpha_{\ell k}\right) - \left(\alpha_{hs} - \omega_i \alpha_{hk}\right)\right) \bar{\nu}_{in,i}}{\left(\alpha_{hs} - \omega_i \alpha_{hk}\right) - \left(\alpha_{gs} - \omega_i \alpha_{gk}\right)} E_i$$

$$R_i = \frac{1}{\psi \bar{\Gamma}_{out,i}} \left[E_i \left(1/\beta - \psi\left(1 - \bar{\Gamma}_{in,i}\right)\right) - \psi R_{g,i}\right]$$

C.2 Inverting A_i

We now invert A_i using the recovered value for R_i. Recall that sales of a firm with productivity z from region i in region j are given by

$$x_{ij}(z) r_{ij}(z) = x_{ij}(z) E_j Q_j^{-\frac{\sigma-1}{\sigma}} A_i^{\frac{\sigma-1}{\sigma}} z^{\frac{\sigma-1}{\sigma}} \gamma^{\frac{\sigma-1}{\sigma}\frac{\eta}{\eta-1}} \Gamma_{ij}(z)^{-\frac{\sigma-1}{\sigma}\frac{\eta}{\eta-1}} h_i(z)^{\frac{\sigma-1}{\sigma}}$$

Aggregating across all i-based firms, we obtain:

$$R_{ij} = \int_z x_{ij}(z) r_{ij}(z)\, dG_i(z) = E_j Q_j^{-\frac{\sigma-1}{\sigma}} A_i^{\frac{\sigma-1}{\sigma}} \gamma^{\frac{\sigma-1}{\sigma}\frac{\eta}{\eta-1}} \Xi_{ij},$$

where we define

$$\Xi_{ij} \equiv \int_z x_{ij}(z) z^{\frac{\sigma-1}{\sigma}} \Gamma_{ij}(z)^{-\frac{\sigma-1}{\sigma}\frac{\eta}{\eta-1}} h_i(z)^{\frac{\sigma-1}{\sigma}}\, dG_i(z).$$

Total sales by i-headquartered firms are given by

$$R_i = A_i^{\frac{\sigma-1}{\sigma}} \gamma^{\frac{\sigma-1}{\sigma}\frac{\eta}{\eta-1}} \sum_{j=1}^{N} E_j Q_j^{-\frac{\sigma-1}{\sigma}} \Xi_{ij}.$$

Thus, we can update our guess for A_i by iterating over the following equation as part of the solution of equilibrium:

$$A_i = \gamma^{-\frac{\eta}{\eta-1}} \left(\frac{R_i}{\sum_{j=1}^{N} E_j Q_j^{-\frac{\sigma-1}{\sigma}} \Xi_{ij}} \right)^{\frac{\sigma}{\sigma-1}} .$$

Note that we have replaced the equilibrium condition that derives R_i based on knowledge of A_i with a condition that recovers A_i based on knowledge of R_i.

C.3 Inverting $A_{g,i}$

Using the information on $R_{g,i}$, we can similarly obtain a new guess for $A_{g,i}$ from the first order condition of firms in i's tradable sector, yielding:

$$A_{g,i} = \left(\frac{R_{g,i}}{M_{g,i} E_g Q_g^{-\frac{\sigma-1}{\sigma}} \left(\prod_{s=1}^{S} \ell_{g,is}^{\alpha_{gs}} \right)^{\frac{\sigma-1}{\sigma}}} \right)^{\frac{\sigma}{\sigma-1}} ,$$

where $M_{g,i}$ is the guess for the mass of firms in the tradable sector; E_g and Q_g are the economy-wide expenditure and quantity index in the tradable sector; and $\prod_{s=1}^{S} \ell_{g,is}^{\alpha_{gs}}$ is the Cobb-Douglas labor aggregator in the tradable production function.

C.4 Inverting B_{is}

Inverting the location choice equation 9, we obtain a new guess for B_{is}:

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

where $\bar{P}_i$ is the guess for the regional price index (including housing) and $\bar{W}_{is} \equiv L_{is}^{\frac{1}{\epsilon}} W_{is}$ is the regional ideal price index for workers of type s.

D Data and additional empirical and quantitative results

D.1 Additional details on the data

D.1.1 U.S. Census Longitudinal Business Database (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.

Firms and firm-level industry code. 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. A multi-establishment firm is defined as having at least two establishments.

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.[31] 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.[32] 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".[33] I define a firm as a "service firm" if establishments in services-producing sectors account for at least half of its total payroll.

Geography. In all data sources, I define a region or a local labor market according to 1990 commuting zones (CZs) as in Tolbert and Sizer (1996). In the model, I consider an aggregation of these commuting zones that groups together small neighboring commuting zones. I focus on the contiguous U.S., excluding Alaska, Hawaii and the American territories, yielding a total of 722 commuting zones that aggregate into 200 local markets.

[31] 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.

[32] 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.

[33] "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.

D.1.2 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 omit observations with missing firm linkages.

This data source has pros and cons relative to the Census LBD. For data on employment, it is less accurate than the LBD – especially for data on establishment-level employment – and it lacks data on wages. However, a key advantage relative to the LBD is that it clearly distinguishes between firm headquarters and branches, and it has a headquarters identifier for each firm. Therefore, I use this data to compute cross-region headquarters-branch linkages. See Barnatchez et al. (2017) for additional discussion of this dataset, including evidence on its good coverage of the spatial allocation of firms.

Sample selection. I drop observations with missing geographical data; missing industry; and missing data on establishment type and firm linkages. I omit establishments in Public Administration and other selected industries that are beyond the scope of the LBD dataset such as Membership Organizations, U.S. Postal Service, Federal and Federally-Sponsored Credit Agencies, Private Households. I also drop Educational Services, since it is hard to distinguish between privately-owned and government-owned establishments.

D.1.3 Lightcast

Data on online job postings is obtained from Lightcast, 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. Lightcast 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.

In these data, a firm is defined based on the set of job postings that share the same codified employer name.[34] The Lightcast 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-location service firms are structured. To distinguish between firm headquarters and branches, I merge the Lightcast data with firm geography from Dun & Bradstreet using name and location matching. I do so using name and location matching for firms that operate in at least two commuting zones in

[34]Lightcast claim to invest much effort into name codification to ensure that they capture the same entities.

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. This process results in around 75,000 multi-region firms, out of which around 64,000 are in service sectors.

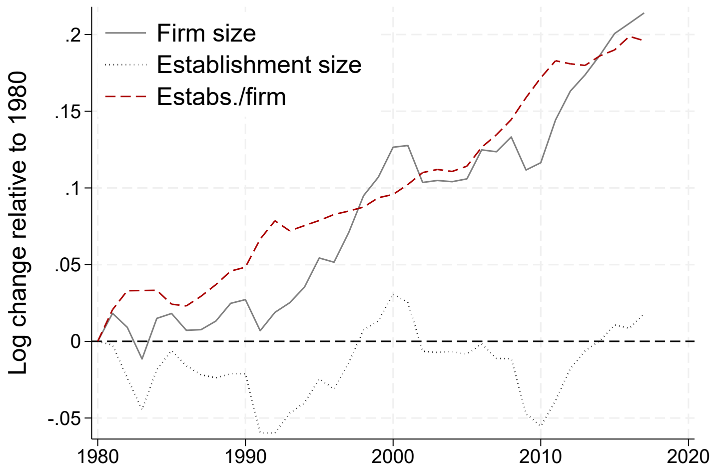
D.2 Additional information on firm expansion (Section 2.1)

Table 10: 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

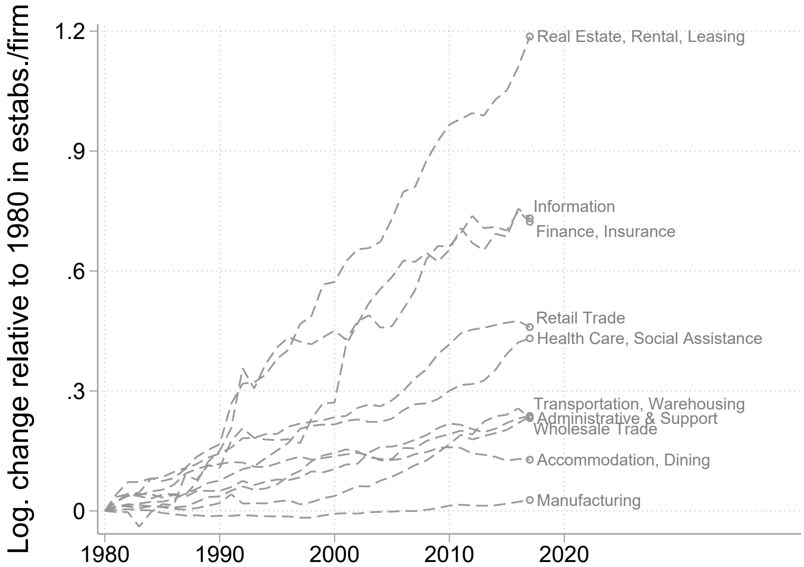
Note: Data from the Census Bureau Longitudinal Business Database. Goods- and services-producing firms stand for firms with at least half of their wage bill across all establishments in goods- and services-producing sectors, respectively. Multi-estab. stands for a firm with at least two establishments,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

Figure 8: 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 9: Firm expansion by sector



Note: 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.

Table 11: 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.

D.3 Additional results on decompositions of wage inequality (Section 2.2)

D.3.1 Comparison of rising inequality across workers and across establishments

I now establish that inequality across establishments accounts for most of the increase in overall inequality across workers in the economy. To this end, Table 12 compares the change in the variance of log wages between 1980 and 2017 across multiple data sources. The first row in Table 12 shows the change in this variance using data on individual-level reported wage earnings in the Decennial Census and the American Community Survey. Between 1980-2017, this variance has increased by 0.17 points. Similar magnitudes have been found by Song et al. (2019) using data from the U.S. Social Security Administration, and by Barth et al. (2016) using data from the BLS Current Population Survey (CPS). The second row in Table 12 repeats this exercise for establishment-level wages in the LBD, weighting different establishments by their employment. Evidently, both sources yield the same increase in the variance of log wages. Formally, let w_i be the log of earnings for individual i, and let $e(i)$ be the establishment that employs individual i. By the law of total variance, the total change in variance equals to the sum of change in variance across establishments and the average change within establishments,

$$\underbrace{\Delta \mathrm{Var}\left[w_i\right]}_{\text{Total change in variance}} = \underbrace{\Delta \mathrm{Var}\left[\mathbb{E}\left[w_i | e\left(i\right)\right]\right]}_{\text{Change in variance across establishments}} + \underbrace{\Delta \mathbb{E}\left[\mathrm{Var}\left[w_i | e\left(i\right)\right]\right]}_{\text{Change in variance within establishments}} .$$

Both the left-hand-side of this equation and the first term on its right-hand-side equal to approximately 0.17. Therefore, the last term that captures rising within-establishment variance is found to be close to zero.

The second Column in Table 12 repeats these findings for within-industry inequality, by demeaning detailed industry fixed-effects from log individual and establishment wages before computing the increase in the variance. Again, I find similar magnitudes across the Census/ACS and the LBD datasets, reinforcing the small role of within-establishment variance in the overall increase in inequality. Overall, rising within-industry wage dispersion accounts for slightly over half of the total increase in inequality.[35] Similar magnitudes for the role of within-industry trends are obtained when utilizing data from the CPS.[36]

Table 12: Rising variance of log wages in the U.S. economy

Change in the variance of log wages (1980-2017)

Data	Total change	Within-industry change
Decennial Census / ACS - across U.S. workers	0.17	0.1
LBD - across all establishments	0.17	0.09

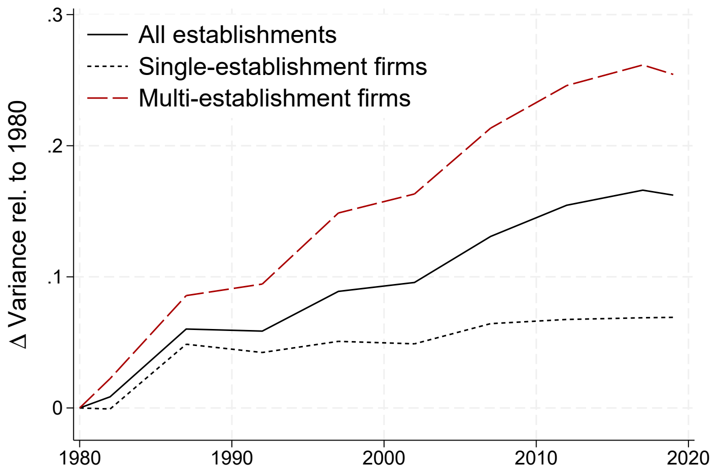
Note: This table summarizes changes in the variance of log wages in the U.S. economy between 1980-2017. The first row computes this measure from individual-level reported wage earnings in the Decennial Census and the American Community Survey. The second row reports this measure for average establishment-level payroll in the LBD. The first column shows the overall change, and the second column reports this change after controlling for industry-level fixed effects.

[35] 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, I find that the importance of differences between industries becomes larger in the latter part of the sample, reconciling the figures reported here with their findings.

[36] Comparing these moments to the Social Security Administration data as in Song et al. (2019) is infeasible due to the lack of high-quality industry identifiers in that data.

D.3.2 The role of multi-establishment firms in the rise of inequality – without industry fixed effects

Figure 10: The role of multi-establishment firms in the rise of inequality - without industry fixed effects



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 2 in the paper, this figure shows this decomposition without first demeaning industry fixed effects.

D.4 Additional information on occupational headquarters intensity (Section 4)

Characterization of occupational headquarters intensity

I utilize firms' online job-postings data from Lightcast, merged with firm headquarters locations from Dun & Bradstreet, as described above in the description of the data. For each firm f and occupation o, I compute the share of job-postings in the firm's headquarters commuting zone, H_{fo}. I then take the occupational fixed effect from the projection of H_{fo} on occupation and firm fixed effects as the occupation-level measure of headquarters intensity. This procedure ensures that my measure of headquarters intensity truly captures differences in the position of an occupation within firms, and not the tendency of some occupations to be hired in particular firms or industries. I compute these measures for 158 distinct occupational codes in the Lightcast data, and then merge them into the U.S. decennial census data by creating a concordance between these 158 occupational codes the *occ1990dd* codes from David Dorn's website.[37]

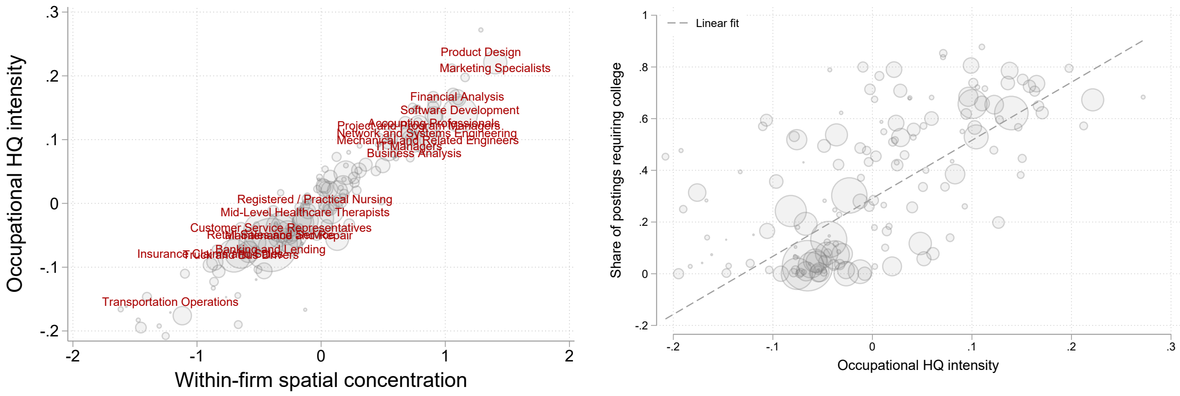
Figure 11 shows basic characteristics of these occupational HQ-intensity measures. The right panel plots the HQ-intensity measure against a simple measure of skill-intensity: the share of online-job postings in an occupation that explicitly require a college degree. A strong positive relationship emerges between these two measures, confirming that headquarters activity is highly skill-intensive. The left panel plots it against a simple measure of within-firm spatial concentration, given by the geographic Herfindahl-Hirshmann index for each occupation, averaged across firms.[38] Evidently, headquarters activity exhibits substantial spatial-concentration within firms.

[37]https://www.ddorn.net/data.htm.

[38]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.

Figure 11: The spatial distribution of occupations within firms

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



Note: This figure shows key characteristics of the occupational headquarters-intensity measures from Section 4. Panel (a) relates the geographical concentration of each occupation within firms to the tendency of the occupation to be hired in firm headquarters. 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. 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.

Occupational headquarters intensity and the rise of wage inequality

I now investigate how much of the overall increase in the variance of log wages can this headquarters-intensity measure account for. To this end, I decompose the overall increase in variance in the Census-ACS data, as in Appendix D.3. I merge the headquarters-intensity measures to the Census-ACS data using workers' occupation codes and the above-mentioned *occ1990dd* classification, and then compute how much of the rise in variance is explained by the increasing importance of this measure between 1980-2017. Specifically, I project log individual wage earnings on my HQ-intensity measure and other selected individual characteristics, and compute the share of the overall rise in variance this measure accounts for. I include in this projection a series of fixed effects capturing individual educational attainment, geography (commuting zones), and their interactions, in order to highlight the strength of my HQ-intensity measures on top of these other factors. Results from this variance decomposition can be seen in Table 13. The overall increase in the variance of log wages in the economy over the covered period stands at 0.17 points, similarly to the increase across the universe of establishments in the LBD. Out of this increase, 46% can be accounted for by my occupational HQ-intensity measure; 18% by the combination of education and geography; and additional 20% by their covariance. Of course, this is only an accounting decomposition, and the substantial role for my HQ-intensity measures could be reflecting other mechanisms that are not related to firm expansion. However, it provides further evidence that the HQ-branch distinction seem to

Table 13: Decomposition of the rise in variance of individual earnings

Component of log individual earnings	**Δ Variance 1980-2017**	**% of total Δ**
HQ-intensity of the occupation	0.08	46%
Commuting zone and college attainment	0.03	18%
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 Appendix D.4 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.

be quantitatively important for the question of rising wage inequality.

D.5 Additional results for Section 4.3

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 9 of the main text. Accordingly, this local market experienced greater income growth and skill deepening than the rest of Arkansas.

Table 14 shows results for local labor market outcomes against expansion of locally-headquartered firms, controlling for year and regional fixed effects. Each column shows the estimated coefficient from a regression of a different labor market outcome for a given local market against the log number of locations that are operated by the firms that were headquartered there in 1980. Columns 1-4 include as outcome variables the logs of average wages, college-to-no-college ratio, headquarters share of total payroll in services, and manufacturing share of total employment, respectively. Each observation combines one of five periods (1980,1990,2000,2010,2017) and one of 301 labor market areas that have the required data for all years.

Columns 5-8 repeat columns 1-4 when instrumenting the expansion of locally-headquartered firms with the exposure to national industry trends. Specifically, for each labor market area i and period t, I define the following regional exposure to firm spatial expansion EE_{it}:

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

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 $\overline{\text{locations}}_{k,t}$ is the *national* average number of locations 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. In line with the rationale of the mechanism in the model, labor market areas 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 regional characteristics such as its size or skill intensity in 1980 are absorbed in the labor market area fixed effect.

Table 14: Expansion of locally-headquartered firms and regional labor market outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS				2SLS			
Outcomes in logs	Wage	H/L	HQ	Manuf	Wage	H/L	HQ	Manuf
Log # markets for locally HQed firms in 1980	0.214***	0.194***	0.360***	-0.180*	0.638***	1.578***	1.722***	-1.849***
	(0.0448)	(0.0665)	(0.0481)	(0.0969)	(0.246)	(0.460)	(0.422)	(0.589)
Year and CZ FE	Y	Y	Y	Y	Y	Y	Y	Y
F-stat					22.18	19.90	22.18	22.18
Observations (5 periods $\times$ 301 LMAs)	1505	1203	1505	1505	1505	1203	1505	1505

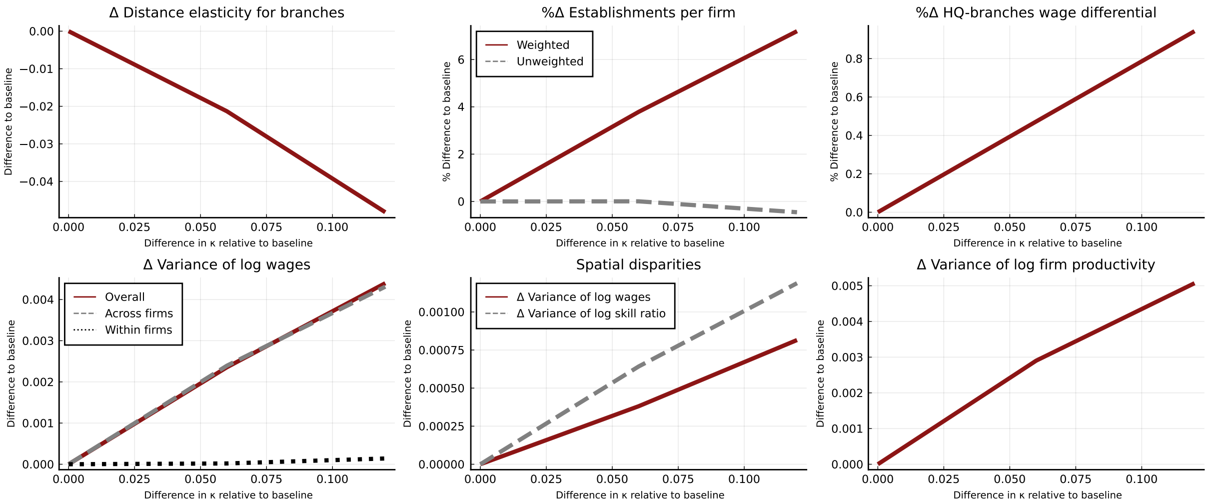
Note: local labor market outcomes against expansion of locally-headquartered firms. Each column shows the estimated coefficient from a regression of a different labor market outcome for a given local market against the log number of locations that are operated by the firms that were headquartered there in 1980. Columns 1-4 include as outcome variables the logs of average wages, college-to-no-college ratio, headquarters share of total payroll in services, and manufacturing share of total employment, respectively. Observations include five years (1980,1990,2000,2010,2017) and 301 labor market areas that have the required data for all years. Columns 5-8 repeat columns 1-4, instrumenting the expansion of locally-headquartered firms with the instrument described in Section 4. Standard errors are clustered at the region level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

D.6 Additional information for the quantified model

Skill group	$\alpha_{h,s}$	$\alpha_{\ell,s}$	$\alpha_{g,s}$
1	0.07	0.16	0.37
2	0.11	0.19	0.19
3	0.22	0.26	0.19
4	0.59	0.39	0.25

Table 15: Calibrated values for skill-intensities in the headquarters labor bundle ($\alpha_{h,s}$), branch-level labor bundle ($\alpha_{\ell,s}$), and tradable-goods labor bundle ($\alpha_{g,s}$).

Figure 12: Labor market implications of lower state-border effects



Note: Distributional implications of lower state border effect. The x-axis depicts changes in the state border effect in the cost of expansion, κ, expressed in terms of the difference from the baseline equilibrium (with zero representing the baseline equilibrium). Positive values indicate lower state border effect in absolute value. The y-axis in each subplot captures changes in selected moments of interest relative to the baseline equilibrium, after re-computing the equilibrium for every new value of ρ: change in the gravity coefficient of headquarters-branch linkages (top-left panel); % change in average establishments per firm (employment-weighted and unweighted, top-middle); % change in mean wage dispersion between headquarters and branches (top-right); change in variance of log wages (overall, across firms, and within firms; bottom-left); change in the variance of log wages and the variance of high-skilled to low-skilled employment (bottom-middle); change in the variance of log firm productivity (bottom-right).