

Spatial Firm Sorting and Local Monopsony Power*

Ilse Lindenlaub Ryungha Oh Michael Peters
Yale University and NBER *University of Chicago* *Yale University and NBER*

August 2025

Abstract

Using administrative data from Germany, we document that high-wage locations have substantially lower labor shares and higher wage dispersion. We show that a parsimonious model, in which firm monopsony power stems from search frictions in local labor markets, can explain these facts as long as “superstar” firms sort into productive locations. This positive sorting, which emerges as the unique equilibrium if firm and location productivity are sufficient complements or labor market frictions are sufficiently large, steepens the local wage ladder in productive locations and leads to not only higher wages, but also greater wage inequality. At the same time, positive firm sorting reduces local labor shares in prosperous places because more productive firms have more monopsony power. Our estimated model indicates that firm sorting can rationalize the lower local labor shares in regions with endogenously higher wages and can account for 40% of their increased wage dispersion. In spatial firm sorting, we thus highlight a new source of disparities in local labor market outcomes.

Keywords: Firm Sorting, Frictional Labor Markets, On-the-Job Search, Monopsony Power, Local Labor Shares, Local Wage/Job Ladders, Inequality.

*Lindenlaub: ilse.lindenlaub@yale.edu. Oh: ryungha.oh@chicagobooth.edu. Peters: m.peters@yale.edu. We thank audiences at UCLA, Berkeley University, Sciences Po University, University of Minnesota, Dartmouth University, Columbia University, Johns Hopkins University, Boston University, Harvard University, Princeton University, Washington University in St. Louis, University of Texas at Austin, the NBER Summer Institute 2022, SED 2022, EEA 2022, the NYU-Columbia Trade Day 2023, the Inequality Conference at the St. Louis Federal Reserve Bank 2024 and SITE Conference 2025 at Stanford University for comments. We are grateful to Ariel Burstein, Hector Chade, Kyle Herkenhoff, Oleg Itskhoki, Rasmus Lentz, Guido Menzio, Simon Mongey and Aysegul Sahin for useful discussions and comments, and to Gabriel Ahlfeldt for sharing his data on German property prices. We thank Juan Gambetta and Meili Wang for outstanding research assistance. The paper uses Linked-Employer-Employee-Data from the Institute for Employment Research (IAB) in Germany. Data access was provided via on-site use at the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the IAB and subsequently also remotely. This is a substantially revised version of our previous paper entitled “Firm Sorting and Spatial Inequality.”

1 Introduction

It is well known that local labor markets vary significantly in terms of income per capita and average wages. What is much less recognized, however, is that this is also true for monopsony power and wage dispersion. Even more surprisingly, these variations follow a systematic pattern. Take Germany as an example: Average wages are negatively correlated with local labor shares, but positively with local wage dispersion. In particular, local labor shares are 6 percent lower in the most prosperous regions, whereas local wage dispersion is 20 percent higher.

The main contribution of this paper is to highlight these novel facts and propose a unifying theory that explains why high-wage locations feature both stronger monopsony power and greater wage inequality based on spatial firm sorting. We develop a parsimonious model of differentiated local labor markets, where in each location heterogeneous firms hire unemployed and employed workers in the presence of search frictions. Because of search frictions, firms have monopsony power: They internalize that they hire more workers if they post higher wages. This leads to a local wage ladder with more productive firms paying more. Different wage policies across firms with local labor market power also imply that firms differentially markdown wages relative to their marginal product and, as a result, have different labor shares. Importantly, the extent of local labor market power and the shape of the local wage ladder directly depend on the local distribution of firm productivity, and therefore on how firms sort into local labor markets.

To understand why firm sorting leads to lower labor shares and more wage inequality in high-wage locations, suppose that more productive, “superstar” firms settle in more productive labor markets. This concentration of productive firms enhances competition for workers, which steepens the wage ladder in those regions. This not only increases the local average wage but, notably, local wage dispersion by driving a wedge between the earnings of those who successfully climb the ladder and those who do not. Perhaps surprisingly in light of this enhanced poaching competition, the same spatial differences in firm composition lead to *lower* local labor shares in prosperous places: Because productive firms have more monopsony power and lower labor shares, their abundance implies a lower average local labor share. Through the lens of our estimated model, firm sorting explains the lower local labor shares in high-wage locations and accounts for 40% of their increased wage dispersion.

Our model rests on two central ingredients. First, to assess whether highly productive firms indeed prefer to locate in productive regions, we need a model in which firm location choices are endogenous. Second, for the local firm composition to shape labor market competition and—through this channel—local wage distributions, a model with on-the-job search is called for: It gives

rise to local wage ladders that are endogenously determined by the firm types that settle there.

In this setting, we show that firms face a basic trade-off when making their location choice. On the one hand, firms “like” productive locations with high TFP. These are locations with good fundamentals, which we interpret broadly; for instance, they can stem from modern infrastructure, productive spillovers, existing input-output networks, and, importantly, workers’ human capital. On the other hand, firms are hesitant to sort into such locations if many highly productive firms also choose to locate there: The presence of other productive firms pushes the firm into a low position on the local wage ladder, which makes it difficult to poach and retain workers and thereby limits firm size. Hence, firms’ location decisions balance two considerations: local productivity, which raises profit per worker, and local competitiveness, which reduces firm size. To our knowledge, this is the first model that integrates on-the-job search with firms’ location choices and highlights this novel trade-off.¹

We derive sufficient conditions for monotone firm sorting across space. Sorting is positive—i.e., better firms locate in more productive locations—if firm and location productivity are sufficiently complementary in production *or* if local labor market frictions are sufficiently large. Productive complementarities ensure that highly productive firms have greater willingness to pay for land in more productive places. In turn, sufficiently large labor market frictions ensure that the competition motive is of limited importance and does not outweigh this productivity consideration. We also show that under the conditions for monotone sorting, an equilibrium exists and is unique.

Our theory makes precise predictions on why firm sorting affects local labor market power and local wage distributions. Positive firm sorting across space results in a stochastically better *firm composition* in more productive locations. As a result, competition for poaching workers increases in those areas, which leads to steeper wage ladders and, in turn, higher average wages and greater wage inequality. At the same time, despite intensified competition among top firms, more productive locations exhibit a *lower* average local labor share. This is because positive firm sorting leads to an abundance of “superstar” firms, which, under mild conditions, have a high degree of monopsony power and low firm-level labor shares. We show that systematic firm sorting is not only sufficient but also necessary to generate spatial variation in local labor shares and within-location wage inequality, and we exploit this result to infer firm sorting from data.

Using administrative data from Germany, we show that our model’s predictions are in line with the evidence: The top quartile of local labor markets in terms of their GDP per capita not only has higher average wages, but also 20% higher wage dispersion and 6% lower labor shares compared

¹Firms often hire from other firms within their local labor market; e.g., in Germany, this accounts for 50% of firms’ hires.

with the bottom quartile. Through the lens of our model, these patterns allow us to infer that firms sort systematically across space. To complement these findings, we provide direct evidence for our model’s mechanism that links firm sorting to spatial wage variation: Wage growth from an employment-to-employment (EE) transition is roughly twice as large in the richest compared with the poorest locations, which indicates that wage ladders are indeed steeper in prosperous places.

Even though we do not explicitly model spatial *worker* sorting, we recognize the importance of worker heterogeneity for spatial wage gaps. In our theory, we capture the spatial variation in worker attributes through local TFP differences, which firms take into account when choosing their location. Under common functional form assumptions, however, local TFP does *not* impact several of our key outcomes that are shaped by firm sorting: the steepness of local wage ladders, local EE returns, within-location wage dispersion, and local labor shares are all independent of local TFP and determined by the productivity distribution of local producers. The documented spatial variation in these outcomes is therefore indicative of *firm* sorting. In practice, we nevertheless control for observed and unobserved worker heterogeneity and show that there is still spatial variation in our outcomes of interest, which we argue firm sorting can help explain.

To gauge the quantitative importance of firm sorting for spatial disparities, we estimate our model using administrative data from Germany. A central aspect of our empirical strategy is to separately identify firm sorting from the fundamental productivity of a location. We prove that we can achieve identification from data on local labor shares and value added. First, because, in our model, local labor shares are entirely determined by the productivity distribution of local firms, the spatial variation in labor shares identifies the extent of firm sorting. Our finding that local labor shares are *lower* in locations with high GDP per capita thus calls for *positive sorting* between firms and locations. Second, the spatial variation in local value added per worker that is *not* due to differences in firm composition pins down the fundamental productivity of each location.

We pursue two objectives with our estimated model. First, to quantify the equilibrium impact of firm sorting, we run a counterfactual that matches firms randomly to locations. Since firm sorting is the sole driver of spatial differences in local labor shares in our model, local labor market power would be equalized in the absence of it. In addition, comparing the poorest and richest regions, positive firm sorting accounts for at least 40% of the spatial gap in within-location wage dispersion and for at least 15% of the gap in average wages. We conclude that firm sorting is a major driver of spatial inequalities, especially in relation to wage dispersion and labor shares.

Second, we conduct several experiments to illustrate the importance of our two central model ingredients: on-the-job search and endogenous spatial firm sorting. We first show that the

presence of *on-the-job search* is crucial for how firm sorting shapes spatial wage disparities: Reestimating our model for lower levels of employed workers’ search efficiency shows that when the poaching rate decreases by 10%, spatial firm sorting contributes 18% less to across-location wage inequality and 8% less to the spatial gap in within-location inequality. This highlights that the *interaction* between firm sorting and on-the-job search is crucial for understanding spatial wage differences. To illustrate the importance of *endogenous* firm sorting (as opposed to exogenous spatial differences in firm productivity), we conduct a stylized policy exercise that reduces place-based subsidies for a set of low-income regions in Germany. This policy leads to the *resorting* of firms across locations with significant effects on the spatial distribution of wages and labor shares.

Related Literature Our framework of frictional local labor markets with on-the-job search is a spatial version of the workhorse job-ladder model of [Burdett and Mortensen \(1998\)](#). The recognition that search frictions, and particularly on-the-job search, are crucial drivers of wage inequality and enable firm heterogeneity to influence wages is a key insight from the literature on labor search and frictional wage dispersion (e.g., [Diamond, 1971](#); [Bontemps et al., 2000](#); [Postel-Vinay and Robin, 2002](#); [Karahan et al., 2017](#), [Bagger and Lentz, 2019](#)). Two important findings are that firm heterogeneity accounts for a sizable share of cross-sectional wage dispersion ($\sim 15\%$ - 30%) and that search frictions and on-the-job search do as well ($\sim 10\%$ - 40%).² Despite this evidence on the importance of firms and search for *cross-sectional* inequality, there has been no attempt to link *spatial* inequality to the sorting of firms into local labor markets that feature on-the-job search. Our paper is the first to conduct such an analysis, both theoretically and quantitatively.³

Because search frictions grant firms labor market power, these theories naturally lend themselves to the analysis of local monopsony power.⁴ Much recent work seeks to measure firms’ labor market power ([Manning, 2011](#); [Langella and Manning, 2021](#); [Yeh et al., 2022](#)) and studies the implications for wages, allocative efficiency, and the aggregate labor share ([Berger et al., 2022a](#); [Jarosch et al., 2024](#)) as well as for the efficacy of policies such as the minimum wage ([Berger et al., 2022b](#)). We add to this literature by showing that monopsony power varies *systematically* across local labor markets and that spatial sorting by firms can account for it. We thus link firm composition to labor shares at the *local* labor-market level and, by adding this spatial dimension, complement studies that made this connection at the economy-wide level over time; see [Autor et al. \(2020\)](#) and [Gouin-Bonenfant \(2022\)](#).

²These estimates come from structural models (e.g., [Postel-Vinay and Robin \(2002\)](#) and [Bagger and Lentz \(2019\)](#)) as well as from non-structural wage regressions that include firm fixed effects (e.g., [Card et al. \(2013\)](#)).

³Subsequent to our work, [Lhuillier \(2025\)](#) builds a model similar to ours to also quantify the role of spatial firm sorting in explaining between- and within-city inequality, but for the case of France.

⁴Search frictions have been found to be the main source of imperfect competition in labor markets ([Berger et al., 2023](#)).

By focusing on spatial disparities, we relate to the urban literature on the sources of spatial wage gaps. This literature has almost exclusively focused on wage-level differences across regions, and especially the city wage premium.⁵ [Schmutz and Sidibé \(2018\)](#), [Heise and Porzio \(2022\)](#), and [Martellini \(2022\)](#) analyze spatial wage gaps using a job ladder model and focus on *worker* sorting as a main source of inequality. By contrast, our work highlights how the spatial sorting of *firms* drives local wage dispersion—by *endogenously* shaping local wage ladders—and local labor shares.

Only a few urban papers analyze firms’ location choices but differ from ours in focus and modeling choices. [Gaubert \(2018\)](#) builds a frictionless model of spatial firm sorting to analyze the efficiency of place-based policies. [Bilal \(2022\)](#) is the first to analyze the effect of firms’ location choices in a model that features labor market frictions. Given his focus on spatial unemployment differences, however, he abstracts from on-the-job search and thus from how firm sorting affects local wage inequality and labor market power by shaping local wage ladders.

Our project thus merges three strands of the literature that have largely existed in isolation: the literature on frictional labor markets and cross-sectional wage dispersion, the literature on monopsonistic firms and labor market power, and the urban literature on spatial inequality.

2 The Model

2.1 Environment

Time is continuous, the horizon infinite and the economy is in steady state. There is a continuum of locations (i.e., local labor markets) and a continuum of firms and workers.

Locations are indexed by ℓ and differ in exogenous productivity $A(\ell)$. We assume that $A(\ell)$ is strictly positive for all ℓ and continuously differentiable, and that locations are ordered by productivity, i.e., $\partial A(\ell)/\partial \ell > 0$. Each location has an exogenous amount of land, distributed with the continuously differentiable cdf R on $[\underline{\ell}, \bar{\ell}]$; $r > 0$ is the corresponding density.

In each location ℓ , there is a unit mass of risk-neutral homogeneous workers who are spatially immobile, something we relax in our estimation below. Unemployed workers in ℓ receive flow utility $b(\ell)$ and search for jobs, while employed workers receive a wage and do on-the-job search (OJS). Importantly, while we do not explicitly model worker heterogeneity across space, it is implicitly contained in local productivity $A(\ell)$ —something we confirm in our estimated model—and thus contributes to a location’s attractiveness. In this sense, more productive workers increase firms’ average output and thus impact spatial firm sorting.

⁵See, e.g., [Glaeser and Maré \(2001\)](#), [Duranton and Puga \(2004\)](#), [Gould \(2007\)](#), [Baum-Snow and Pavan \(2011\)](#), [Moretti \(2011\)](#), or [Behrens et al. \(2014\)](#) for the case of the U.S.; [De La Roca and Puga \(2017\)](#) for Spain; [Combes et al. \(2008\)](#) for France; and [Bamford \(2021\)](#) or [Dauth et al. \(2022\)](#) for Germany.

Firms are risk-neutral and differ in productivity p . We assume $p \in [\underline{p}, \bar{p}]$, distributed with a continuously differentiable cdf Q , with density $q > 0$. We call p the *ex ante productivity* of firms, based on which location choices are made. After settling in location ℓ , each firm with attribute p draws *ex post productivity* $y \in [\underline{y}, \bar{y}]$ from cdf $\Gamma(\cdot|p)$, where Γ is continuously differentiable in both y and p . We assume that the corresponding density, $\gamma(\cdot|p)$, satisfies the strict monotone likelihood ratio property in (y, p) . This implies that $\partial\Gamma(y|p)/\partial p < 0$ for all $y \in (\underline{y}, \bar{y})$, i.e., more productive firms *ex ante* draw their *ex post* productivity from better distributions in the first-order stochastic dominance (FOSD) sense. We distinguish between *ex ante* and *ex post* productivity so that, even with pure sorting between *ex ante* firm types and locations, we obtain a non-degenerate distribution of firm productivity in each location.⁶

In order to produce in location ℓ , firms need to buy one unit of land at price $k(\ell)$ and post a wage to hire local workers. The returns to land accrue to a set of local landowners who operate in the background. Firms have no capacity constraint when employing workers, so they hire any worker who yields a positive profit. Firm y in location ℓ produces output $z(y, A(\ell))$ per worker hired. We assume that z is twice continuously differentiable and strictly increasing in each argument. Note that while the *ex ante* productivity of firms p determines the distribution of *ex post* productivity y , p is irrelevant for production conditional on y . Hence, after entry, firms are fully characterized by their y . We assume that z is the output of the same homogeneous good in all locations, whose price is normalized to one. All agents discount the future at rate ρ .

In each location there is a frictional labor market, in which workers and firms face search frictions and search is random. In the baseline model, we assume that meeting rates are exogenous and constant across locations. Firms meet workers at Poisson rate λ^F . Employed workers' meeting rate is given by λ^E and unemployed workers' meeting rate by λ^U . Matches are destroyed at rate δ . We also denote the meeting rates of employed workers, unemployed workers, and firms *relative* to the job destruction rate by $\varphi^E \equiv \lambda^E/\delta$, $\varphi^U \equiv \lambda^U/\delta$ and $\varphi^F \equiv \lambda^F/\delta$. In our quantitative analysis we endogenize local population size and hence these meeting rates through endogenous labor mobility, which allows them to vary across space.

In terms of wage setting, we assume that firms post wages with commitment as in [Burdett and Mortensen \(1998\)](#). We denote the flow wage paid by firm y in location ℓ by $w(y, \ell)$. Hence, firm y in location ℓ receives flow profit $\pi(y, \ell) = z(y, A(\ell)) - w(y, \ell)$ when employing a worker.

The model timing is such that firms first make a one-time location choice and then the

⁶Alternatively, we could have obtained non-degenerate firm distributions in all ℓ by having firms with different p draw preference shocks over locations or by letting the matching process between p and ℓ be subject to search frictions.

continuous-time economy described above plays out in each local labor market.

To simplify our analytical arguments, we impose the following assumption.

Assumption 1.

1. The distributions of ex post productivity $\Gamma(y|p)$ have a common support: $\forall p, y \in [\underline{y}, \bar{y}]$.
2. In each location ℓ , firms with the lowest ex post productivity, \underline{y} , make zero profits.

An implication of the common support assumption from part 1, which is supported by the evidence,⁷ is that locations inhabited by firms with higher ex ante productivity p have an ex post productivity distribution $\Gamma(y|p)$ that puts more mass on highly productive firms. In particular, it captures the intuition that locations with high- p firms have a relative abundance of “superstar” firms, whose productivity y is high relative to firms at the bottom of the productivity distribution. In turn, part 2 will be guaranteed by making sure that the output of firm \underline{y} equals the reservation wage, i.e., $w^R(\ell) = z(\underline{y}, A(\ell))$ for all ℓ . One way to ensure this property is by appropriately choosing non-employment utility $b(\ell)$ (a primitive) across locations. While Assumption 1 gives us analytical tractability, we show that our quantitative results are robust to dropping it (Section 8).

2.2 Equilibrium

We now discuss agents’ decisions—namely, the job acceptance decisions of workers as well as firms’ location choices and wage-posting decisions. Finally, we specify the steady-state flow balance and market-clearing conditions.

Workers. Workers face a single decision: whether to accept a job offer, both when employed and unemployed. We discuss this job acceptance decision briefly since it is standard.

Consider first a worker who is employed at wage w . The value of being employed at wage w in location ℓ , $V^E(w, \ell)$, solves the recursive equation

$$\rho V^E(w, \ell) = w + \delta(V^U(\ell) - V^E(w, \ell)) + \lambda^E \left[\int_{\underline{w}(\ell)}^{\bar{w}(\ell)} \max\{V^E(t, \ell), V^E(w, \ell)\} dF_\ell(t) - V^E(w, \ell) \right],$$

where F_ℓ is the endogenous wage-offer distribution in location ℓ with support $[\underline{w}(\ell), \bar{w}(\ell)]$ and $V^U(\ell)$ denotes the value of unemployment, given by

$$\rho V^U(\ell) = b(\ell) + \lambda^U \left[\int_{\underline{w}(\ell)}^{\bar{w}(\ell)} \max\{V^E(t, \ell), V^U(\ell)\} dF_\ell(t) - V^U(\ell) \right]. \quad (1)$$

Note that, as is well known (and straightforward to show), $V^E(\cdot, \ell)$ is increasing in w , so the optimal strategy of employed workers is to accept any wage higher than the current one.

⁷Using French data, [Combes et al. \(2012\)](#) find that firm productivity distributions across space do *not* vary in their left truncation, which indicates that the productivity of the least productive firms is similar across locations.

In turn, the optimal strategy of unemployed workers is a reservation wage strategy, pinning down $w^R(\ell)$ for each ℓ from a worker who is indifferent between accepting and rejecting a job,

$$V^E(w^R(\ell), \ell) = V^U(\ell). \quad (2)$$

Firms. Firms face two decisions. First, they choose location ℓ to maximize expected discounted profits, taking competition from other firms and land prices as given. Second, conditional on the location choice, firms post a wage to maximize profits. We solve backward.

Wage Posting. When posting a wage w , a firm in location ℓ trades off profit per worker against firm size, which is given by (see Appendix SA.1.2)

$$l(w, \ell) := \frac{\varphi^F}{(1 + \varphi^E (1 - F_\ell(w)))^2}. \quad (3)$$

Firms with higher rank in the local wage distribution F_ℓ are larger since they poach more and are being poached less. Conversely, holding the firm's wage w fixed, its size is smaller if the local wage distribution, F_ℓ , is stochastically better in a FOSD sense, since the firm faces fiercer competition. Importantly, the (relative) EE rate, φ^E , governs how strongly firm size depends on local competition. If labor market frictions are severe and EE flows are rare, $\varphi^E \rightarrow 0$, the competition channel is mitigated and, in the limit, firm size is independent of the local wage distribution.

Each firm posts the wage that maximizes its net present discounted value of profits, which can be expressed as per-worker profit times size (see Appendix SA.1.1),⁸

$$\tilde{J}(y, \ell) = \max_{w \geq w^R(\ell)} l(w, \ell)(z(y, A(\ell)) - w), \quad (4)$$

whereby a higher wage increases firm size, $l(w, \ell)$, but reduces flow profits, $z(y, A(\ell)) - w$.

Equation (4) highlights the fact that location matters to firms in two distinct ways, which already hints at their trade-off between local productivity and competition. On the one hand, choosing a high ℓ increases location TFP $A(\ell)$ and thus output and flow profits. On the other hand, if many productive firms sort into high- ℓ locations, competition is fierce (the wage offer distribution F_ℓ is stochastically better), and the size of any given firm y becomes compressed.

The firm's objective function (4) is strictly supermodular in (w, y) , which—in combination with a continuum of productivity levels—implies that w is strictly increasing in y . Therefore, the local distribution of wage offers coincides with the local distribution of firm productivity, $F_\ell(w(y, \ell)) = \Gamma_\ell(y)$, where Γ_ℓ is the *endogenous* productivity cdf of firms in location ℓ . Cdf Γ_ℓ encapsulates the spatial sorting of firms and is thus the crucial object in our model. In what follows, we will use Γ_ℓ instead of F_ℓ and denote firm size by $l(y, \ell) := l(w(y), \ell)$.

⁸To simplify exposition, we set $\rho \rightarrow 0$ for the remainder of the analysis.

Making this substitution, we solve the firm's problem to obtain the well-known wage function under wage posting (Burdett and Mortensen, 1998), see Appendix SA.1.3,

$$w(y, \ell) = z(y, A(\ell)) - \int_{\underline{y}}^y \frac{\partial z(t, A(\ell))}{\partial y} \frac{l(t, \ell)}{l(y, \ell)} dt, \quad (5)$$

except that in our setting there is one such wage function in each location ℓ ; and that ℓ matters through both its effect on TFP $A(\ell)$ and firm size $l(y, \ell)$. Note that firms have monopsony power due to search frictions. This creates a wedge between the firm's wage $w(y, \ell)$ and its marginal product $z(y, \ell)$, which depends on the local labor market competition “from below”: If the competitive pressure surrounding firm $t < y$ relative to firm y is strong—captured by a small size of t , $l(t, \ell)$, relative to that of firm y , $l(y, \ell)$ —then it is difficult for firm y to poach workers which bids up its wage. Local labor market competition—and hence monopsony power—is shaped by the local productivity distribution, Γ_ℓ , which affects firm y 's size relative to its lower ranked competitors (see (3)) and thereby wages. Pinning down Γ_ℓ is what we will turn to next.

Location Choice. Given the wage function for each location ℓ , we can now specify the firm's location choice problem. Each firm p considers the expected value from settling in location ℓ , which is the expected present discounted value of profits net of the price of land, $k(\ell)$:

$$\bar{J}(p, \ell) = \int_{\underline{y}}^{\bar{y}} \tilde{J}(y, \ell) d\Gamma(y|p) - k(\ell) = \int_{\underline{y}}^{\bar{y}} \frac{\partial z(y, A(\ell))}{\partial y} l(y, \ell) (1 - \Gamma(y|p)) dy - k(\ell),$$

which we derived using $\tilde{J}(y, \ell)$ from (4) along with the Envelope Theorem (and integration by parts). When choosing their location, firms balance local productivity $A(\ell)$ (which determines output $z(y, A(\ell))$); local competition Γ_ℓ (which determines their size $l(y, \ell)$); and land prices $k(\ell)$. Formally, for each firm of any type p the location choice problem is

$$\max_{\ell} \bar{J}(p, \ell). \quad (6)$$

The solution to (6) describes firms' location decisions and is at the center of our analysis. The FOC of this problem highlights firms' fundamental location choice trade-off:⁹

$$\int_{\underline{y}}^{\bar{y}} \left(\frac{\partial \ln \left(\frac{\partial z(y, A(\ell))}{\partial y} \right)}{\partial \ell} + \frac{\partial \ln l(y, \ell)}{\partial \ell} \right) \frac{\partial z(y, A(\ell))}{\partial y} l(y, \ell) (1 - \Gamma(y|p)) dy = \frac{\partial k(\ell)}{\partial \ell}, \quad (7)$$

where $\frac{\partial \ln l(y, \ell)}{\partial \ell}$ is the (semi-)elasticity of firm size wrt location ℓ and $\frac{\partial \ln (\partial z(y, A(\ell))/\partial y)}{\partial \ell}$ is the (semi-)elasticity of the firm's marginal product wrt ℓ .

FOC (7) reflects firms' trade-off between profitability and firm size when choosing the opti-

⁹We proceed as if Γ_ℓ (and thus $l(y, \ell)$) is continuously differentiable in ℓ , which will be the case under pure sorting below.

mal ℓ . Locations with higher ℓ , by virtue of having higher productivity $A(\ell)$, push up output z and thus firm profits per employee (first term in brackets on the LHS). But if these locations attract many productive firms, competition in high- ℓ locations is fierce; poaching and retaining workers is then difficult, which reduces firm size l (second term in brackets). At the optimal location choice, this marginal (net) benefit of choosing a higher ℓ equals its marginal cost, which is the increase in land price. If high- ℓ locations are more desirable, they command higher land prices, $\partial k(\ell)/\partial \ell > 0$.

This FOC—along with land market clearing—pins down the equilibrium allocation of firms to locations, captured by Γ_ℓ . That is, for all ℓ ,

$$\Gamma_\ell(y) = \int_{\underline{p}}^{\bar{p}} \Gamma(y|\tilde{p}) m_p(\tilde{p}|\ell) d\tilde{p} \quad \forall y \in [\underline{y}, \bar{y}], \quad (8)$$

where we define by $m(\ell, p)$ the endogenous joint matching density between (ℓ, p) with conditional density $m_p(p|\ell)$ (and also $m_\ell(\ell|p)$).¹⁰ In addition, the FOC pins down the land price schedule, $k(\cdot)$, that sustains this allocation. This is obtained by solving (7) for k , when evaluated at the equilibrium assignment (see Appendix SA.1.4).¹¹

Land Market Clearing. The land market clearing condition is given by

$$R(\ell) = \int_{\underline{\ell}}^{\ell} \int_{\underline{p}}^{\bar{p}} m_\ell(\tilde{\ell}|\tilde{p}) q(\tilde{p}) d\tilde{p} d\tilde{\ell}. \quad (9)$$

It ensures that the mass of land with quality below ℓ equals the mass of firms settling in those places. Thus, the mapping between the distributions of firms Q and land R is measure-preserving.

Good Market Clearing. In each location ℓ , workers, firms and land owners consume their entire income. Total income thus equals total consumption, which in turn equals total output, so that the good market clears in each ℓ (where we cancel a multiplicative $r(\ell)$ on both sides),

$$\int_{\underline{y}}^{\bar{y}} z(y, A(\ell)) l(y, \ell) d\Gamma_\ell(y) = \int_{\underline{y}}^{\bar{y}} w(y, \ell) l(y, \ell) d\Gamma_\ell(y) + \bar{J}(\mu(\ell), \ell) + k(\ell). \quad (10)$$

Flow-Balance Conditions. We have two flow-balance conditions in steady state, which pin down the equilibrium unemployment rate and the distribution of employment in each location.

First, the inflow into and outflow out of unemployment must balance, which pins down the unemployment rate, $u(\ell)$ (which in our baseline model does not vary across ℓ):

$$\delta(1 - u(\ell)) = u(\ell)\lambda^U \quad \Rightarrow \quad u(\ell) = \frac{1}{1 + \varphi^U}. \quad (11)$$

¹⁰Under a measure-preserving matching between firms p and locations ℓ , the marginal densities of m are given by r and q .

¹¹In this competitive land market, firms that maximize expected profits and landowners who maximize land prices will result in the same allocation of firms to locations, which is why we detail only one side's decision: the one by firms.

Second, the inflow into and outflow out of employment in firms with productivity below y must balance (for all y), taking into account the optimal job acceptance decisions of employed workers. This determines the cdf of employment in location ℓ , denoted by G_ℓ :

$$u(\ell)\lambda^U\Gamma_\ell(y) = (\delta + \lambda^E(1 - \Gamma_\ell(y)))G_\ell(y)(1 - u(\ell)) \quad \Rightarrow \quad G_\ell(y) = \frac{\Gamma_\ell(y)}{1 + \varphi^E(1 - \Gamma_\ell(y))}. \quad (12)$$

Note that the outflow of workers from firms with productivity below y , $G_\ell(y)(1 - u(\ell))$, has two sources: exogenous job destruction (driven by δ) and endogenous on-the-job search, which induces workers to quit for better jobs when they find them (which happens at rate $\lambda^E(1 - \Gamma_\ell(y))$). Local employment distribution G_ℓ reflects the local firm productivity distribution, Γ_ℓ , but is stochastically better as long as there is an active wage ladder, $\varphi^E > 0$.

Steady-State Equilibrium. We can now define a steady-state equilibrium.

Definition 1. *A steady-state equilibrium is a tuple $(w(\cdot, \ell), k(\ell), m(\ell, p), \Gamma_\ell(\cdot), l(\cdot, \ell), G_\ell(\cdot), u(\ell), w^R(\ell))$, such that for all $\ell \in [\underline{\ell}, \bar{\ell}]$ and $p \in [\underline{p}, \bar{p}]$:*

1. *Walrasian equilibrium in the land market: The pair $(k(\ell), m(\ell, p))$ is a competitive equilibrium of the land market, pinning down $\Gamma_\ell(\cdot)$ and also $l(\cdot, \ell)$;*
2. *Optimal wage posting: $w(y, \ell)$ is consistent with (4) for all firm types $y \in [\underline{y}, \bar{y}]$;*
3. *Optimal worker behavior: Employed workers accept job offers from more productive firms; unemployed workers accept any job y with $w(y, \ell) \geq w^R(\ell)$, where $w^R(\ell)$ is pinned down by (2);*
4. *Flow-balance conditions (11) and (12) hold, pinning down $u(\ell)$ and $G_\ell(\cdot)$;*
5. *Good market clearing (10) holds.*

3 Equilibrium Analysis

3.1 Spatial Firm Sorting

We now analyze the patterns of firm sorting that occur in equilibrium. We provide conditions under which more productive firm types p sort into more productive locations ℓ . This is an allocation with positive assortative matching (PAM), which, as we show below, is the empirically relevant case (see Appendix SA.2.1 for the analysis of negative sorting).

Sufficient Conditions for Positive Sorting. We focus on pure assignments between (p, ℓ) , in which any two firms of the same type are matched to the same location type (and vice versa). Assignment $m_p(p|\ell)$ can then be captured by a matching *function* $\mu : [\underline{\ell}, \bar{\ell}] \rightarrow [\underline{p}, \bar{p}]$. We define positive sorting in a standard way.

Definition 2 (Positive Sorting of Firms to Locations). *There is positive sorting in (p, ℓ) if matching function μ is strictly increasing, $\mu'(\ell) > 0$.*

Under positive sorting, more productive firms sort into more productive locations. Moreover, $m_p(p|\ell)$ has positive mass only at a single point $p = \mu(\ell)$, and we can simplify the endogenous distribution of firms in location ℓ in (8) to

$$\Gamma_\ell(y) = \Gamma(y|\mu(\ell)),$$

so that high- ℓ locations have ex post productivity distributions that are stochastically better.¹²

To obtain *sufficient* conditions for positive sorting, recall that firm p chooses location ℓ to maximize $\bar{J}(p, \ell)$ given in (6). Based on results from the literature on monotone comparative statics (Milgrom and Shannon, 1994), the optimal location choice is increasing in p if $\bar{J}(p, \ell)$ satisfies a strict single-crossing property in (p, ℓ) , for which strict supermodularity in (p, ℓ) is sufficient.¹³ Thus, complementarities of $\bar{J}(p, \ell)$ in (p, ℓ) lead to positive sorting, which echoes familiar insights from the literature on sorting. We now derive conditions on primitives that guarantee this property of $\bar{J}(p, \ell)$. We postulate that firms anticipate positive sorting when making their location choices, and check that their optimal behavior indeed induces PAM.¹⁴

Recalling how $\bar{J}(p, \ell)$ varies with ℓ (see FOC (7)) and using the assumption that p shifts $\Gamma(y|p)$ in the FOSD sense, we note that the supermodularity of $\bar{J}(p, \ell)$ —and thus firm sorting—is controlled by the location choice trade-off between productivity gains and competition:

$$\frac{\partial^2 \bar{J}(p, \ell)}{\partial p \partial \ell} > 0 \quad \text{if} \quad \underbrace{\frac{\partial \ln \left(\frac{\partial z(y, A(\ell))}{\partial y} \right)}{\partial \ell}}_{\text{Productivity Gains}} + \underbrace{\frac{\partial \ln l(y, \ell)}{\partial \ell}}_{\text{Cost of Competition}} > 0. \quad (13)$$

Whereas the local productivity gains from settling into high- ℓ locations are positive if production technology z is supermodular (first term in (13)), the local competition effect is negative under positive sorting since productive firms cluster in the best locations (second term in (13)). Positive sorting thus emerges if the productivity benefits that boost profits per worker outweigh the costs from competition that translate into lower expected firm size.

Productivity gains are large if productivity differences across space are large (the A -schedule is steep) and when complementarities of z in (y, A) are strong. Note that for the multiplicative

¹²Cdf $M_p(\cdot|\ell)$ (corresponding to density $m_p(\cdot|\ell)$) is a Dirac measure that concentrates its mass at $p = \mu(\ell)$ and (8) becomes

$$\Gamma_\ell(y) = \int_{\underline{p}}^{\bar{p}} \Gamma(y|\tilde{p}) m_p(\tilde{p}|\ell) d\tilde{p} = \int_{\underline{p}}^{\bar{p}} \Gamma(y|\tilde{p}) dM_p(\tilde{p}|\ell) = \Gamma(y|\mu(\ell)).$$

¹³Strict single crossing of $\bar{J}(p, \ell)$ guarantees that μ is weakly increasing, while the additional assumption of strictly positive densities r and g , make μ *strictly* increasing.

¹⁴Below we check the case in which firms do not postulate PAM and show that it is not consistent with equilibrium.

technology $z(y, A) = yA$ —the functional form used in our quantitative analysis—these gains become $\partial \ln A(\ell)/\partial \ell$ and (13) reduces to a comparison between local TFP and local competition.

In turn, the cost of local competition depresses firm size and is captured by

$$\frac{\partial \ln l(y, \ell)}{\partial \ell} = -\frac{2\varphi^E}{1 + \varphi^E(1 - \Gamma_\ell(y))} \left(-\frac{\partial \Gamma_\ell(y)}{\partial \ell} \right) \leq 0. \quad (14)$$

It depends on two forces: first, on how endogenous firm distribution Γ_ℓ varies across space, and second, on the degree of labor market frictions $\varphi^E = \lambda^E/\delta$ that determines the impact of changes in Γ_ℓ on firm size. The cost of sorting into a high- ℓ region is low when λ^E —the rate at which employed workers meet firms—is small, since in that case poaching and competition do not matter much. The cost is also low if δ is large, so that match duration is mainly determined by workers who separate into unemployment as opposed to quitting. In this case, hiring predominantly results from unemployment and, again, poaching considerations carry less weight. The ratio φ^E captures both of these forces. A small φ^E weakens the competition channel so that it does not interfere with the productivity motive for positive spatial sorting.

To guarantee the supermodularity of $\bar{J}(p, \ell)$ in (p, ℓ) in terms of *primitives*, we use (13) and (14) to obtain the following sufficient condition for PAM:

Proposition 1 (Spatial Sorting of Firms). *If z is strictly supermodular, and either the productivity gains from sorting into higher ℓ are sufficiently large or the competition forces are sufficiently small (i.e., φ^E is sufficiently small), then any equilibrium features positive sorting of firms p to locations ℓ with $p = \mu(\ell) = Q^{-1}(R(\ell))$.*

The proof is in Appendix A.1, where we make the statements regarding “sufficiently large productivity gains” and “sufficiently small φ^E ” precise.

Under the conditions of Proposition 1, the productivity gain from settling into high- ℓ locations outweighs the cost from competition for firms of *all* y -types. But—due to productive complementarities between (y, A) —the net benefit is especially high for those firms with high y , which high ex ante productivity p yields stochastically. Highly productive firms are thus willing to pay higher land prices, outbidding the less productive firms in the competition for land in high- ℓ locations. As a consequence, positive sorting arises, whereby high- ℓ locations have more productive firms in a FOSD sense, $\partial \Gamma_\ell/\partial \ell = (\partial \Gamma(y|\mu(\ell))/\partial p) \mu'(\ell) \leq 0$.

Note that, in the absence of on-the-job search, $\varphi^E = 0$, $\partial \ln l(y, \ell)/\partial \ell = 0$ and complementarities in production are enough to sustain positive sorting.¹⁵ It may be surprising at first sight

¹⁵Even though frictions (small φ^E) facilitate positive sorting in our context, they alone are not sufficient: Some complementarities of z in (y, A) are needed, otherwise $\partial^2 \bar{J}/\partial \ell \partial p$ —see equation (A.2) in Appendix A.1—cannot be positive.

that larger labor market frictions (lower φ^E) facilitate sorting. Importantly, however, in the frictionless case, $\varphi^E \rightarrow \infty$, competitive pressure becomes all-encompassing: A winner-takes-all allocation takes hold whereby the most productive firm attracts all workers in a given location. Ex ante, this discourages sorting since firms of any p expect zero profits no matter where they go.

Generalizations. We generalize our main Proposition 1 in various ways. First, we can allow for *endogenous productivity spillovers*. Instead of assuming exogenous differences in A , we assume that A depends on the endogenous composition of firms in ℓ ; i.e., $A(\ell) = \tilde{A}(\Gamma_\ell)$, where \tilde{A} is increasing in the location’s average firm productivity; see Proposition SA2, Appendix SA.3.2. Second, we *endogenize vacancy creation*; see Proposition SA3 in Appendix SA.3.3. Third, we *endogenize land distribution* R by allowing it to depend on land demand through price $k(\cdot)$; see Proposition SA4 in Appendix SA.3.4). Finally, we can allow for *labor mobility* and *residential housing* (Proposition A2, Appendix B), endogenizing local population size, which will be our quantitative setting and further discussed below. In all these extended settings, we show that essentially the same trade-off between productivity and competition still determines firm sorting.

3.2 Existence and Uniqueness

We also show that when sorting is positive, a unique equilibrium exists.

Proposition 2 (Existence & Uniqueness). *Assume that the conditions from Proposition 1 hold; then a unique equilibrium (up to a constant of integration in land price function k) exists.*

The proof is in Appendix A.2. We show the existence of a fixed point in Γ_ℓ by construction. In turn, uniqueness arises because, under the conditions on primitives stated in Proposition 1, the impact of endogenous firm composition on the firms’ value function leaves the complementarity properties of \bar{J} unchanged.

To summarize how the equilibrium with positive sorting works in our model, note that spatial firm sorting, local land prices, and location-specific wages are all jointly determined. But it is instructive to think of the equilibrium in two stages: First, the frictionless land market equilibrium, which pins down the upward-sloping matching function, μ (allocating more productive firms ex ante to more productive locations) and thus local firm productivity $\Gamma_\ell(y) = \Gamma(y|\mu(\ell))$; and the land price schedule, k , sustaining this allocation through a steep enough slope to ensure that high- ℓ locations are so expensive that only firm types with the highest willingness to pay—i.e., the most productive ones—settle there. And second, the frictional labor market equilibrium within each ℓ , which pins down the wage and the distribution of employment as a function of firm ex post productivity, Γ_ℓ .

4 The Distributional Implications of Spatial Firm Sorting

In this section, we use our model to study the implications of spatial firm sorting for local monopsony power on the one hand, and for spatial wage disparities on the other. We show that positive firm sorting leads to *higher* monopsony power and thus lower labor shares in more productive locations, as well as *higher* wage dispersion and average wages—predictions will assess empirically and quantitatively below.

4.1 Spatial Firm Sorting, Local Monopsony Power and Local Labor Shares

A central implication of positive firm sorting is that it induces spatial differences in firms' labor market power. This, in turn, implies that local labor shares are lower in productive locations. Formally, the local labor share, $LS(\ell)$, can be expressed as the weighted average of firm-level labor shares $Ls(y, \ell) := w(y, \ell)/z(y, A(\ell))$ in each ℓ ,

$$LS(\ell) \equiv \int_{\underline{y}}^{\bar{y}} Ls(y, \ell) \tilde{g}_\ell(y) dy,$$

where \tilde{g}_ℓ is the value-added weighted employment distribution with corresponding cdf \tilde{G}_ℓ (see Appendix A.3 for the derivation). The spatial gradient of the local labor share is then given by

$$\frac{\partial LS(\ell)}{\partial \ell} = \int_{\underline{y}}^{\bar{y}} \left(\underbrace{\frac{\partial Ls(y, \ell)}{\partial y}}_{\text{Within-Location Heterogeneity in Firm Monopsony Power}} \times \underbrace{\left(-\frac{\partial \tilde{G}_\ell(y)}{\partial \ell} \right)}_{\text{Spatial Variation in Employment Composition}} + \underbrace{\frac{\partial^2 Ls(y, \ell)}{\partial y \partial \ell}}_{\text{Spatial Variation in Firm Monopsony Power}} \times (1 - \tilde{G}_\ell(y)) \right) dy. \quad (15)$$

The main economic forces pushing toward declining local labor shares, $LS(\ell)$, are contained in the first product term of decomposition (15). The derivative, $\partial Ls(y, \ell)/\partial y$, captures how firm-level labor shares vary with productivity within a location. As we will show, under natural conditions on primitives, this term is *negative*, reflecting the fact that more productive firms have more monopsony power and hence lower labor shares. In turn, the derivative, $(-\partial \tilde{G}_\ell(y)/\partial \ell)$, captures how the employment composition varies across space. Positive firm sorting triggers a composition effect and helps ensure that this term is *positive* since high- ℓ locations have more employment concentrated in productive firms. The first term in (15) is thus negative, pushing toward local labor shares, $LS(\ell)$, that are decreasing in ℓ .

The intuition is straightforward and stems from differences in firm monopsony power *within locations* paired with differences in employment composition *across locations*: Under positive sorting, high- ℓ places tend to have a better firm *composition* and thus employment distributions that are tilted toward productive, “superstar” firms. If, additionally, more productive firms in any given ℓ face milder competition than less productive firms, and thus have more monopsony

power and lower labor shares $Ls(y, \ell)$, then locations in which employment is concentrated in top firms have a lower aggregate labor share, $LS(\ell)$.

The second term in (15) reflects an additional force behind a declining local labor share, namely a submodular firm-level labor share $Ls(y, \ell)$, which captures how firm-level labor shares vary across ℓ .¹⁶ We now formalize this result in terms of primitives; see Appendix A.3 for the proof.

Proposition 3 (Firm Sorting & Local Labor Shares). *Suppose $\gamma(\cdot|p)$ is sufficiently decreasing, z sufficiently log-supermodular in (y, ℓ) , and φ^E sufficiently small. If there is positive firm sorting across space, then the local labor share, $LS(\cdot)$, is decreasing in ℓ .*

Our key insight from Proposition 3 is that positive firm sorting pushes toward a lower local labor share in high- ℓ places if the local density of firm productivity, $\gamma(\cdot|p)$, is sufficiently decreasing.¹⁷ This is because, within each location, $\gamma(\cdot|p)$ crucially determines whether the firm-level labor shares are declining in firm productivity y : A decreasing local firm density implies that in each ℓ , high-productivity companies are surrounded by fewer firms locally than low-productivity firms. As a consequence, they face less labor market competition, which translates into more monopsony power and lower labor shares of top firms in all locations, $\partial Ls(y, \ell)/\partial y < 0$, despite the fact that they pay higher wages.

Several densities that are commonly used to model firm heterogeneity satisfy the discussed condition, such as the log-normal and Pareto distributions. Figure 1 illustrates the intuition behind Proposition 3 for both cases: the log-normal productivity distribution in the left panel and Pareto productivity distribution in the middle panel. In each of these panels, we depict the firm-level labor share, $Ls(\cdot, \ell)$, and the local value-added weighted employment distribution $\tilde{g}(\cdot|\mu(\ell))$, for the most and least productive locations $\ell \in \{\underline{\ell}, \bar{\ell}\}$. The fact that the firm-level labor share, $Ls(\cdot, \ell)$, is declining in y within each location, together with positive firm sorting—which puts more mass on high- y firms in location $\bar{\ell}$ and induces employment density $\tilde{g}_{\bar{\ell}}$ to first-order stochastically dominate $\tilde{g}_{\underline{\ell}}$ —implies that average local labor shares are declining in location productivity, $\partial LS(\ell)/\partial \ell < 0$; see Figure 1, right panel.¹⁸

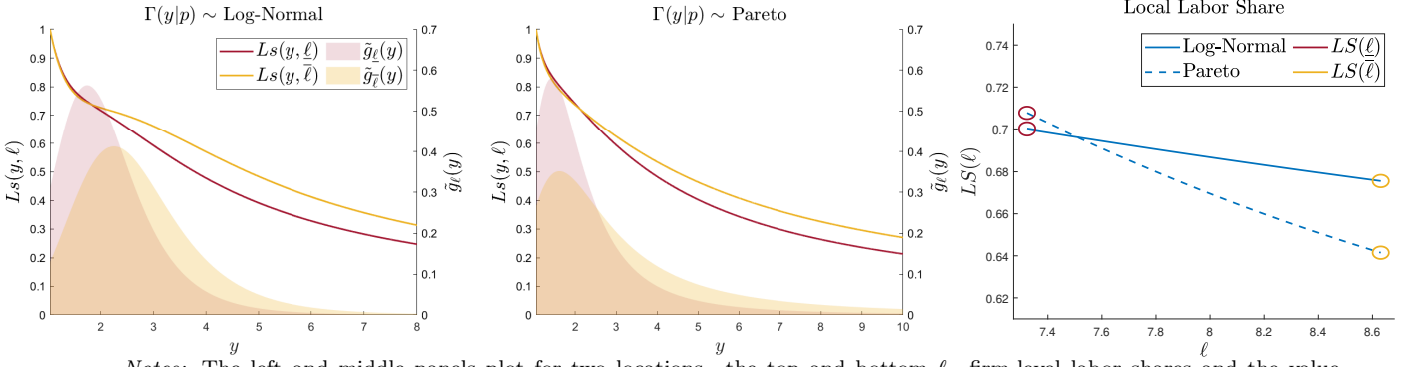
In our quantitative setting, we will assume $\Gamma(y|p) \sim \text{Pareto}(1, 1/p)$. Indeed, for multiplicative z , the local labor share then *only* depends on the Pareto tail coefficient, p , of the local productivity

¹⁶Submodularity of the firm-level labor share means that productive firms in high- ℓ locations have especially strong monopsony power and thus even lower labor shares than their counterparts in low- ℓ areas, which would further lower the average labor share in their location. Clearly, this is sufficient but not necessary for our result to hold.

¹⁷The remaining conditions from the proposition (sufficient complementarities in z and small enough φ^E) guarantee that the cross-partial term, $\partial^2 Ls(y, \ell)/\partial y \partial \ell$, in (15) is negative.

¹⁸Interestingly, Figure 1 shows that the *firm-level* labor share at the upper end of the productivity distribution is *lower* in low- ℓ locations. With positive sorting, productive firms are particularly scarce in low- ℓ labor markets and hence have plenty of monopsony power. But their average *local* labor share is *higher* because they have few of these highly productive firms.

Figure 1: Firm Sorting and Decreasing Firm-Level Labor Shares Induce Decreasing Local Labor Shares



Notes: The left and middle panels plot for two locations—the top and bottom ℓ —firm-level labor shares and the value-added weighted employment distribution. In the left panel, firm ex-post productivity is (truncated) log-normally distributed with p shifting the mean productivity across space. We left-truncate this distribution at the median productivity of the worst location such that its density is sufficiently decreasing, in line with Proposition 3. In the middle panel, firm ex post productivity is Pareto distributed with p shifting up the tail coefficient. In the right panel, we plot for both cases—log normal and Pareto firm productivity—the resulting local labor shares across locations. Source: Authors’ model simulations.

distribution (see Gouin-Bonenfant (2022) for a similar result under Pareto in a single-market economy), which circumvents the need for additional restrictions on φ^E and z from Proposition 3. Under positive sorting, high- ℓ locations have a thicker tail and thus more mass of firms with high monopsony power, which reduces the local labor share; see Appendix A.4 for the proof.

Corollary 1 (Firm Sorting & Local Labor Shares: Pareto). *Suppose $\Gamma(y|p) \sim \text{Pareto}(1, 1/p)$ and $z(y, A(\ell)) = yA(\ell)$. The local labor share is then given by $LS(\ell) = 1 - \mu(\ell)$ (with $\mu(\ell) = p$) and hence decreasing in ℓ if and only if there is positive firm sorting across space.*

Corollary 1 shows the link between firm sorting and local monopsony power in the clearest form: If more productive labor markets are able to attract better firms, their local labor shares are lower. The local labor share here reflects only the thickness of the local productivity distribution, i.e. the relative abundance of “superstar” firms, but is independent of the efficiency of OJS, φ^E . This is because, under the Pareto assumption, a change in φ^E induces changes in the employment composition and firm-level labor shares that cancel out.¹⁹ This corollary plays an important role in our quantitative analysis below because it allows us to infer firm sorting, captured by $\mu(\ell)$, directly from local labor shares.²⁰

We end by emphasizing the central assumptions underlying the negative relationship between local firm productivity and local labor share. Importantly, Proposition 3 does *not* hinge on the

¹⁹Suppose, for example, φ^E was lower. In all locations, the employment composition would shift toward low-productivity firms because high-productivity firms would poach workers at a lower rate. Everything else equal, this would *increase* local labor shares because firm-level labor shares of less productive firms are higher. However, a lower φ^E also reduces labor market competition and therefore firm-level labor shares. This tends to *reduce* local labor shares everywhere. These effects cancel out.

²⁰In the Appendix, we supplement this result on how to detect firm sorting, using the spatial variation in local productivity dispersion (Proposition SA5 and Corollary SA1) or the spatial variation in the relationship between the local and the global productivity rank of firms (Proposition SA6).

assumption of Pareto-distributed firm productivity or fatter tails of firm productivity in high- ℓ locations. Instead, what is central is a sufficiently decreasing local density of firm productivity, something many distributions satisfy. Apart from the decreasing shape of $\gamma(\cdot|p)$ what is important is that spatial firm sorting gives rise to more mass on highly productive firms in some locations relative to others—something we term a “non-neutral firm productivity shift”. This is achieved by assuming that $\gamma(y|p)$ has common support across all p (Assumption 1.1) together with the strict monotone likelihood ratio property of $\gamma(y|p)$ in (y, p) . By contrast, if spatial sorting were to give rise to “neutral firm productivity shifts” across locations, in the sense that firm productivity in some location ℓ' is simply a scaled version of that in some ℓ , $y' = ay, a > 0$ (violating common support), there is the same mass on high-productivity firms in all locations and, consequently, no spatial variation in local labor shares (see Appendix A.8 for details).

As an example of a “neutral shift”, suppose the distribution of firm productivity was Pareto with $\Gamma(y|p) \sim \text{Pareto}(p, 1/\theta)$. Thus, firm sorting, captured by $p = \mu(\ell)$, would not affect the relative mass of superstar firms across locations (tail coefficient $1/\theta$ is a parameter here and does not vary in ℓ) but only the lower bound of firm productivity. By Corollary 1, the local labor share would be given by $LS = 1 - \theta$ and hence constant in ℓ despite firm sorting. However, we will show that empirically, prosperous locations have lower labor shares, which suggests that firm sorting induces, at least partially, non-neutral productivity shifts across space.

4.2 Spatial Firm Sorting, Local Wage Ladders and Wage Inequality

We now turn to the implications of firm sorting for spatial wage disparities. In the previous section, we showed that firm sorting affects local monopsony power by altering the local composition of firms. In this section, we will show that firm sorting also has important implications for spatial wage inequality: Positive firm sorting *steepens* the local wage ladder in productive locations and thereby leads to *higher* wage dispersion and average wages.

To simplify the exposition, we will derive the intuition behind the propositions in this section for the case of multiplicative technology $z(y, A(\ell)) = yA(\ell)$, but none of our results depend on this assumption.²¹ In this case, wage schedule (5) reads

$$w(y, \ell) = A(\ell) \left(y - \int_{\underline{y}}^y \frac{l(t, \ell)}{l(y, \ell)} dt \right). \quad (16)$$

This wage function is instructive because it is log-additive in location TFP $A(\ell)$ and a term that captures firm y ’s competition through its relative firm size, $l(t, \ell)/l(y, \ell)$. Firm size is determined

²¹The proofs in this section are contained in Appendices A.5-A.6.

by firm sorting through local productivity distribution Γ_ℓ and the extent of labor market frictions φ^E , but not by $A(\ell)$ (see (3)). We refer to (16) as the local wage ladder or job ladder.

We first analyze the relationship between firm sorting and wage inequality *within* locations, which we measure by the local *relative wage* between a firm with rank $\mathcal{R} = \Gamma_\ell(y)$ in the local productivity distribution relative to the least productive firm with $\mathcal{R} = 0$, $w(\Gamma_\ell^{-1}(\mathcal{R}), \ell)/w(\Gamma_\ell^{-1}(0), \ell)$. This statistic—which does not depend on $A(\ell)$ and captures the *steepness of the local wage ladder*—proxies in a tractable way standard measures of wage inequality, such as the wage gap between the 95th and 5th percentile of wages when choosing \mathcal{R} sufficiently high.

To show that wage inequality is higher in high- ℓ locations if firm sorting is positive, we analyze the derivative of our wage dispersion measure wrt ℓ , which equals

$$\frac{\partial \frac{w(\Gamma_\ell^{-1}(\mathcal{R}), \ell)}{w(\Gamma_\ell^{-1}(0), \ell)}}{\partial \ell} = \frac{1}{\underline{y}} \int_{\underline{y}}^{\Gamma_\ell^{-1}(\mathcal{R})} 2\varphi^E \frac{(1 + \varphi^E(1 - \mathcal{R}))^2}{(1 + \varphi^E(1 - \Gamma_\ell(t)))^3} \underbrace{\left(-\frac{\partial \Gamma_\ell(t)}{\partial \ell} \right)}_{\text{Spatial Variation in Firm Composition}} dt, \quad (17)$$

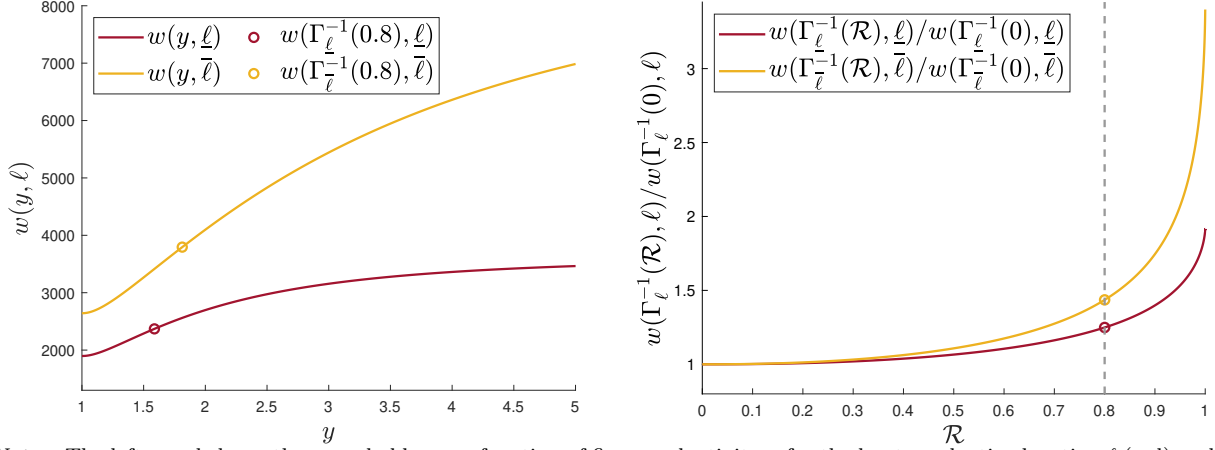
and which we derived from (16), using (3) and (14).

Under positive sorting, $\partial \Gamma_\ell(y)/\partial \ell < 0$, which then directly implies that high- ℓ locations have steeper wage ladders and, as a consequence, more wage dispersion. The economic intuition behind this result is simple: Within each location, the wage ladder is increasing in firm productivity y , i.e., more productive firms pay higher wages. And due to positive firm sorting, firm composition differs across locations, whereby a larger mass of productive firms in high- ℓ locations amplifies competition for workers. This is reflected in the fact that any given firm productivity rank $\mathcal{R} > 0$ is associated with higher productivity level y in high- ℓ compared to low- ℓ places and therefore with higher wages. Note that we can make an additional point: Spatial differences in within-location wage inequality require $\partial \Gamma_\ell(y)/\partial \ell \neq 0$ and therefore that firms sort systematically across space.

Proposition 4 (Firm Sorting & Within-Location Wage Inequality). *Suppose z is weakly log-supermodular in (y, ℓ) . If there is positive firm sorting across space, then wage dispersion, $w(\Gamma_\ell^{-1}(\mathcal{R}), \ell)/w(\Gamma_\ell^{-1}(0), \ell)$, is increasing in ℓ .*

We illustrate this result in Figure 2, which depicts two representations of the local wage ladder in the most and least productive locations. In the left panel, we plot wages as a function of firms' ex post productivity y , based on (16). Two properties stand out. First, the wage intercept, $w(\underline{y}, \ell) = \underline{y}A(\ell)$, is increasing in ℓ , which reflects spatial differences in local productivity $A(\ell)$. Second, while more productive firms pay higher wages everywhere, the local wage ladder is steeper in high- ℓ locations. This differential steepness of the wage ladder is more directly seen in the

Figure 2: Spatial Firm Sorting and Local Wage Ladders



Notes: The left panel shows the wage ladder as a function of firm productivity y for the least productive location $\underline{\ell}$ (red) and most productive location $\bar{\ell}$ (yellow). The right panel shows the wage at firm productivity quantile \mathcal{R} relative to the wage at the lowest ranked firm for the same two locations. Source: Authors' model simulations.

right panel, which plots our statistic of the relative wage as a function of local productivity *rank*, $w(\Gamma_{\ell}^{-1}(\mathcal{R}), \ell) / w(\Gamma_{\ell}^{-1}(0), \ell)$. Positive firm sorting implies that a given firm rank \mathcal{R} translates into a higher productivity level y , and hence higher wages, in productive locations. To illustrate this, we indicate in the left panel for each location the firm productivity level, $y = \Gamma_{\ell}^{-1}(\mathcal{R})$, corresponding to the 80th quantile ($\mathcal{R} = 0.8$) of the local firm distribution and the associated wages, where we obtain $\Gamma_{\bar{\ell}}^{-1}(0.8) > \Gamma_{\underline{\ell}}^{-1}(0.8)$ and, thus, $w(\Gamma_{\bar{\ell}}^{-1}(0.8), \bar{\ell}) > w(\Gamma_{\underline{\ell}}^{-1}(0.8), \underline{\ell})$; this translates into $w(\Gamma_{\bar{\ell}}^{-1}(0.8), \bar{\ell}) / w(\Gamma_{\bar{\ell}}^{-1}(0), \bar{\ell}) > w(\Gamma_{\underline{\ell}}^{-1}(0.8), \underline{\ell}) / w(\Gamma_{\underline{\ell}}^{-1}(0), \underline{\ell})$ in the right panel.

We emphasize that there are two important assumptions that underlie the relationship between firm sorting and local wage dispersion: First, firm sorting affects the local wage ladder and thus local wage dispersion only if there is OJS. Without OJS, $\varphi^E = 0$, local wage ladders collapse everywhere and all workers receive their local reservation wage—a spatial Diamond paradox. Second, as for our results on local labor shares in Section 4.1, firm sorting affects spatial wage inequality only if it induces non-neutral firm productivity shifts across locations—guaranteed by our assumptions that $\gamma(y|p)$ has common support for all p and satisfies the strict monotone likelihood ratio property in (y, p) . By contrast, if firm heterogeneity across space was characterized by neutral productivity shifts, the composition of firms—i.e., the relative amount of productive firms—would not differ across ℓ . As a consequence, wage ladders would be equally steep everywhere, inducing the same within-location inequality in all locations (see Appendix A.8 for details). Thus, the wage structure would be very much the same everywhere, something that we show below is counterfactual vis-a-vis the actual data. Importantly, none of our results in this section hinges on the assumption of Pareto-distributed firm productivity or fatter tails of firm productivity in high- ℓ places.

Finally, we show that the effect of firm sorting on the steepness of local wage ladders is also a central determinant of wage level differences *between* locations. Consider, for example, the *spatial wage premium*, that is the average wage of more productive locations relative to the least productive one, $\mathbb{E}[w(\Gamma_\ell^{-1}(\mathcal{R}), \ell)|\ell]/\mathbb{E}[w(\Gamma_{\underline{\ell}}^{-1}(\mathcal{R}), \underline{\ell})|\underline{\ell}]$, where—consistently with our measure on within-location inequality above—we define this statistic using firm ranks instead of their level of productivity.²² To see that wages are higher in high- ℓ locations, note that the derivative of the spatial wage premium wrt ℓ is equal in sign ($\stackrel{\text{S}}{=}$) to (see Appendix A.6)

$$\frac{\partial \frac{\mathbb{E}[w(\Gamma_\ell^{-1}(\mathcal{R}), \ell)|\ell]}{\mathbb{E}[w(\Gamma_{\underline{\ell}}^{-1}(\mathcal{R}), \underline{\ell})|\underline{\ell}]}}{\partial \ell} \stackrel{\text{S}}{=} \underbrace{A'(\ell)\underline{y}}_{\text{Spatial Variation in Wage Ladder Intercept}} \int_0^1 \frac{w(\Gamma_\ell^{-1}(\mathcal{R}), \ell)}{w(\Gamma_\ell^{-1}(0), \ell)} \tilde{l}(\mathcal{R}) d\mathcal{R} + A(\ell)\underline{y} \int_0^1 \underbrace{\frac{\partial}{\partial \ell} \left(\frac{w(\Gamma_\ell^{-1}(\mathcal{R}), \ell)}{w(\Gamma_\ell^{-1}(0), \ell)} \right)}_{\text{Spatial Variation in Steepness of Wage Ladder}} \tilde{l}(\mathcal{R}) d\mathcal{R}, \quad (18)$$

where $\tilde{l}(\mathcal{R}) = g_\ell(\Gamma_\ell^{-1}(\mathcal{R}))/\gamma_\ell(\Gamma_\ell^{-1}(\mathcal{R}))$ is the size density of a firm with rank \mathcal{R} and independent of ℓ . This expression is positive and highlights two fundamental differences between locations ℓ and $\underline{\ell}$, which boost wages in more productive places: Location ℓ has higher TFP, $A(\ell)$, and—in our equilibrium with positive sorting—a better distribution of firms, Γ_ℓ . Higher location TFP increases the intercept of the wage ladder in high- ℓ locations. In turn, as we have shown in Proposition 4, positive firm sorting leads to steeper wage ladders in high- ℓ locations. We summarize.²³

Proposition 5. *Suppose z is weakly log-supermodular in (y, ℓ) . If there is positive firm sorting across space, then spatial wage premium $\mathbb{E}[w(\Gamma_\ell^{-1}(\mathcal{R}), \ell)|\ell]/\mathbb{E}[w(\Gamma_{\underline{\ell}}^{-1}(\mathcal{R}), \underline{\ell})|\underline{\ell}]$ is increasing in ℓ .*

In our quantitative analysis we provide a structural decomposition of (18) and separate the observed variation in wages across space into the part that stems from the exogenous differences in TFP and the part that is due to endogenous firm sorting.

5 Descriptive Evidence

Turning to our empirical application, we first assess our model’s *qualitative* predictions on the link between firm sorting, monopsony power, and spatial wage disparities. In Sections 6 and 7, we estimate our model to highlight the *quantitative* implications of firm sorting for spatial inequality.

5.1 Data and Measurement

We base our analysis on 257 commuting zones (CZ)—our local labor markets. For graphical illustrations, it is convenient to order locations by index ℓ —or, equivalently, by local TFP $A(\ell)$ —which however are both unobserved. We therefore use local GDP per capita as a proxy for ℓ ,

²²Both formulations (based on level or rank) render the same implications for the spatial wage premium (Appendix A.6).

²³Note that, in contrast to our previous results on local labor share and local wage dispersion, even neutral productivity shifts would generically give rise to a spatial wage premium through spatial differences in $A(\ell)$ and thus the wage intercept.

which, albeit endogenous in our model, is increasing in $A(\ell)$. In our quantitative analysis, we draw on our estimated model to exactly decompose local GDP into A and endogenous firm sorting.

We use three main data sources: (i) regional data from the German Federal Statistical Office on GDP per capita, labor compensation, value added, and unemployment rates for each CZ; (ii) a worker-level panel from linked employer-employee data in Germany (LIAB) provided by the Research Data Centre (FDZ) of the German Federal Employment Agency, which are based on workers’ social security records and contain information on wages and worker flows; and (iii) firm-level data from the Establishment History Panel (BHP), a 50% random sample of all German establishments with at least one employee subject to social security as of June 30 in any given year.²⁴ From the BHP, we construct firm-level average wages—the relevant concept of wages through the lens of our theory—to assess wage inequality within and across local labor markets. We deflate wages using a nationwide CPI.²⁵ We complement our main data sources with information on firm-level sales from the German Establishment Panel (EP). Throughout, we focus on the period 2010-2017 (using an average).

In Appendix C, we describe these data sources in more detail and define important variables. In Appendix SA.5, we report some basic statistics, which show that average wages, firm size, and value added differ substantially across commuting zones (see Table SA.2).

5.2 Spatial Firm Sorting: Evidence

Guided by our model predictions from Section 4, we now provide evidence that more productive firms sort into more productive locations. Obtaining direct evidence on firm sorting is challenging because our dataset—like many others—does not contain a clean measure of physical firm productivity and does not allow us to identify firm movers.²⁶ And deducing firm sorting indirectly from spatial variation in commonly observed outcomes, such as wages or value added, is also difficult because these statistics conflate firm sorting with local TFP differences. We therefore focus on those measures that, according to our theory, are only impacted by firm sorting: local labor shares and within-location wage inequality. We also provide direct evidence on the underlying mechanisms our theory highlights.

²⁴The data are at establishment level, but we use the terms “firm” and “establishment” interchangeably.

²⁵This is consistent with our model, which features nominal wages (the good’s price is normalized to 1). We show in Table A.1 (Appendix D.1) that spatial inequality remains substantial even when adjusting for regional price deflators.

²⁶We deliberately do not use AKM firm fixed effects as a proxy for firm productivity since they conflate firm with location productivity. Moreover, limited mobility bias tends to inflate the variance of these (FDZ-provided) firm fixed effects (e.g., Bonhomme et al., 2022), which makes them unsuitable for assessing local firm productivity *distributions*, especially in small CZs.

Local Labor Shares. Our model highlights that local labor shares identify the extent of spatial firm sorting (Proposition 3 and Corollary 1). As stated in Corollary 1, if firm ex ante productivity is Pareto distributed and technology multiplicative, local labor shares are pinned down by the tail parameter of the local productivity distribution that is shaped by firm sorting, $LS(\ell) = 1 - \mu(\ell)$, independent of A .²⁷ Specifically, there is positive sorting of firms across space, $\mu'(\ell) > 0$, if the local labor share is decreasing in ℓ . The top left panel of Figure 3 shows that richer locations indeed have lower labor shares—which we believe is a new fact—and this spatial variation is economically meaningful: The labor share is about 6% lower in rich compared with poor CZs.²⁸

As highlighted in our discussion of Proposition 3, this negative correlation arises because more productive firms have more monopsony power and thus lower *firm-level* labor shares—something that resonates with empirical findings from a variety of countries (e.g., Lochner and Schulz, 2022 for Germany; Bontemps et al., 2000 and Cahuc et al., 2006 for France; Autor et al., 2020 and Yeh et al., 2022 for the US).²⁹ Rich locations with low average local labor shares must then have a higher concentration of employment in top firms whose labor share is low—a composition effect implied by positive firm sorting.

While our theory implies that local labor shares only depend on firm sorting μ , we now explore whether other mechanisms outside of our model can explain the documented labor share patterns. We summarize this discussion in the first column of Table 1, which reports in row 1 our baseline correlation between local labor shares and GDP per capita from Figure 3. A first concern might be that this correlation does not reflect firm sorting but is driven by regional differences in the industrial composition (Gaubert, 2018). For example, the negative correlation between local GDP and local labor shares could be driven by the greater importance of capital-intensive industries in prosperous locations or industry-specific union power. However, when controlling for regional differences in the industrial composition through sectoral employment shares in row 2, the negative relationship between local labor shares and local GDP becomes even more pronounced.

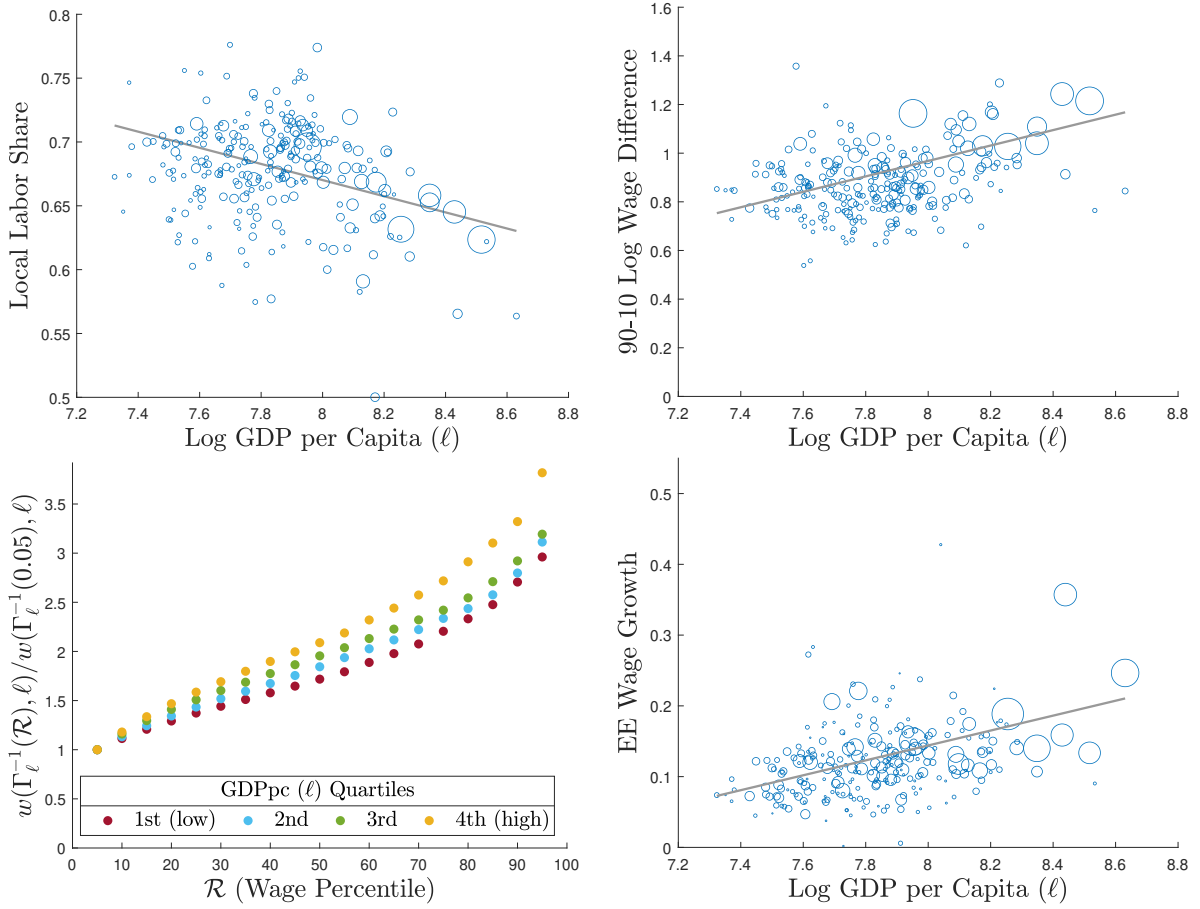
In rows 3 and 4, we control for spatial differences in worker composition. If, for example,

²⁷Appendix SA.4.1 provides empirical support for the Pareto assumption based on tail variation in local productivity.

²⁸Appendix SA.4.1 supplements this evidence of positive firm sorting, using spatial variation in firm sales per worker. We document that the within-location dispersion in sales per worker is larger in productive labor markets, which—based on Corollary SA1, Appendix SA.4.1—indicates positive firm sorting across space. In addition, we show that the empirical relationship between the difference in firms’ global and local productivity ranks and firm productivity is consistent with spatial firm sorting (Proposition SA6, Appendix SA.4.2). However, given that we observe sales for only a small subset of firms, our preferred evidence for spatial firm sorting stems from local labor shares, based on *regional* data from the German Federal Statistical Office.

²⁹Given the link between local labor shares and firm composition in our model, measuring local labor market power based on the labor share is natural in our context. Alternative models/data may suggest a different route, e.g., by estimating firm-level labor supply elasticities (Manning, 2003, Manning, 2011, Hirsch et al., 2020), which are then aggregated to the local level, or by computing local concentration indices (e.g., Berger et al., 2022a, Berger et al., 2023 who build on Atkeson and Burstein, 2008).

Figure 3: Implications of Spatial Firm Sorting: Spatial Variation in Labor Shares, Within-Location Wage Inequality, Wage Ladders and EE Returns



Notes: Data sources: BHP and LIAB. In the top left panel, we plot local labor shares against local log GDP pc. Local labor shares are defined as the ratio between labor compensation and gross value added in each CZ (see Appendix C.1 for details). In the top right panel, we plot the local 90-10 log difference in firm-level wages against local GDP pc. In the bottom left panel, we display wage quantiles relative to the 5% quantile for four groups of CZs, ordered by quartiles of local GDPpc. In the bottom right panel, we plot location-specific EE returns, i.e., β_ℓ^{EE} based on (20), against log GDP pc. The size of the markers either indicates the number of firms or the number of EE moves within each CZ, depending on the panel.

skilled work is complimentary with capital, spatial worker sorting (e.g., [Diamond, 2016](#); [Heise and Porzio, 2022](#)) rather than firm sorting could drive the negative relationship between labor shares and GDP per capita. But when we control for the observed skill composition of the local workforce (row 3) or for differences in unobserved skills via the local average AKM worker fixed effect provided by the FDZ (row 4), the negative relationship between local labor shares and GDP gets reinforced.

Finally, a potential concern may be that land—a factor of production our model omits—is more expensive in rich places, which could mechanically lower their local labor shares. In row 5, we control for local prices of commercial real estate, which does not affect our conclusion.

Wage Inequality Within Labor Markets. As we have shown in Proposition 4, within-location inequality is greater in high- ℓ locations if firm sorting is positive, *independent* of local

Table 1: Labor Shares, EE Returns, and Wage Inequality Within/Across Local Labor Markets

	(1) Labor share	(2) 90-10 Log Wage Diff.	(3) Log Wage	(4) EE Return
Log GDPpc [No Controls]	-0.0632*** (0.0088)	0.3172*** (0.0459)	0.5350*** (0.0271)	0.1094*** (0.0313)
Log GDPpc [Control: Industry]	-0.1024*** (0.0212)	0.2328*** (0.0362)	0.5116*** (0.0253)	0.1057*** (0.0315)
Log GDPpc [Control: Observed Comp.]	-0.1130*** (0.0134)	0.1052*** (0.0203)	0.3560*** (0.0330)	0.1082*** (0.0314)
Log GDPpc [Control: Unobserved Comp.]	-0.0776*** (0.0104)	0.2094*** (0.0267)	0.1880*** (0.0301)	0.1087*** (0.0317)
Log GDPpc [Control: Other]	-0.0997*** (0.0183)	0.2029*** (0.0601)	0.4723*** (0.0418)	0.1152*** (0.0385)
N	257	257	257	257

Notes: Data Sources: German Federal Statistical Office, BHP, and LIAB. We report coefficient β from the regression $\Upsilon_\ell = \beta \ln GDPpc_\ell + \epsilon_\ell$, where $GDPpc_\ell$ denotes GDP per capita in CZ ℓ and Υ_ℓ denotes one of the outcomes in the different columns. The different rows refer to the inclusion of various controls at the regional level (column 1) or to the case in which we residualize firm/worker-level outcomes at the micro level before aggregating them to regional outcome Υ_ℓ (columns 2-4). In row 2, we account for industry impact: by controlling for local employment shares in seven industries (agriculture; mining and utilities; manufacturing; construction; trade and transportation; professional services; public administration and health) in column 1; by residualizing firm-level wages before computing the regional wage outcomes in columns 2 and 3; and by controlling for 1-digit industry fixed effects in (20) when estimating the LHS variable β_ℓ^{EE} in column 4. In row 3, we control for workers' observable skills/traits: through local shares of college graduates (column 1); by residualizing firm-level wages using firms' shares of college graduates (column 2 and 3); and by controlling for age/gender/education when estimating β_ℓ^{EE} from (20) (column 4). In row 4, we control for workers' unobserved skills: by controlling for workers' average local AKM fixed effects (provided by the FDZ) in column 1; or by including individual (not AKM) fixed effects when estimating (20) for column 4; in columns 2 and 3, we use the LIAB and compute workers' *residual* wages as $\tilde{w}_{i\ell,t} \equiv \ln w_{i\ell,t} - \alpha_i$, where $w_{i\ell,t}$ is the wage of worker i in location ℓ and month t , and α_i is the AKM worker fixed effect (from FDZ). We then compute local wage dispersion (column 2) and average wages (column 3) after taking out month fixed effects. In row 5, we control for local corporate real estate prices in column 1 and for log population density in columns 2-4. Due to limited availability of local land prices, the regression in row 5, column 1, has only 118 observation.

productivity $A(\ell)$. The top right panel of Figure 3 displays the relationship between income per capita and local wage inequality, which we measure by the difference between the 90th and 10th quantiles of log firm wages within each CZ. There is a clear positive relationship: The 90-10 log wage gap is about 0.4 log points higher in the most prosperous labor markets.

While Figure 3 simply plots the unconditional correlation, this positive relationship is robust to a variety of alternative explanations that are outside of our model. In column 2 of Table 1, we report the bivariate correlation between local GDP per capita and within-location wage dispersion, which we compute after residualizing firm wages in different ways.

Row 2 shows that prosperous labor markets have more dispersed wages also within industries, which addresses the concern that industries may be systematically sorted across space and, e.g., differ in union coverage (Jäger et al., 2022) and thus the extent of wage compression.

In rows 3 and 4, we address the possibility that local wage dispersion mostly reflects the spatial sorting of workers (rather than firms). This may be the case if observed skills are more dispersed in productive locations (Eeckhout et al., 2014) or if the dispersion of unobserved ability is higher among skilled workers who cluster in those places (Card et al., 2023). In row 3, before computing

local wage dispersion, we thus control for observed worker heterogeneity by residualizing firm-level wages with respect to the share of firms' college graduates. The resulting correlation with local GDP is lower but still positive and significant. To address the concern about unobserved worker heterogeneity, we turn to worker-level (rather than firm-level) wages and purge them from their AKM worker fixed effects (see the *Notes* of Table 1 for details). Row 4 shows that the correlation between the local dispersion of residualized wages and local GDP remains positive and sizeable.

Finally, in row 5, we consider the possibility that rich and dense places feature stronger worker-firm sorting, which would augment local wage inequality. Controlling for local population density, however, does not annul our baseline correlation from row 1.

Spatial Wage Inequality Across Labor Markets. We briefly turn to the implications of our model for wage inequality across locations (Proposition 5). Poor locations are not only hurt by adverse fundamentals but also by the fact that unproductive firms settle there. Indeed, Figure A.1 (Appendix D.2) shows that spatial wage inequality in Germany is pervasive: Average monthly wages paid by firms vary between 2,500EUR in the poorest and 4,500EUR in the richest places.

While our model predicts that positive firm sorting across space is one contributing factor, spatial wage differences clearly also reflect variation in local TFP $A(\ell)$ and its determinants. For completeness, column 3 of Table 1 reports the correlation between local average wages and GDP per capita after controlling for firms' industry (row 2), observed and unobserved heterogeneity of workers (rows 3 and 4) and population density (row 5) and shows that the positive correlation between local wages and GDP remains significant. In Section 6, we will use our structural model to rigorously decompose the spatial wage gap into the contribution of firm sorting and local TFP.

The Mechanism: Firm Sorting Steepens Local Wage Ladders in High- ℓ Locations.

Our theory highlights a common mechanism of how firm sorting affects spatial wage disparities: It steepens the local wage ladder in productive (high- ℓ) locations, thereby raising their wage dispersion and average wages. Using firm- and worker-level micro data, we now provide direct evidence on the spatial variation of wage ladders that is consistent with this channel.

As a first test, in Figure 3, bottom left panel, we plot local wage ladders for four groups of labor markets that are ordered by their local GDP per capita. Specifically, we first compute the local wage ladder in each location ℓ based on the quantiles of the firm-level wage distribution relative to the 5% quantile of that location, $w(\Gamma_\ell^{-1}(\mathcal{R}), \ell)/w(\Gamma_\ell^{-1}(0.05), \ell)$. This allows us to focus on the differential *steepness* of these local wage ladders that is driven by firm sorting and independent of $A(\ell)$. We then average all wage ladders within each GDP quartile. Local wage

ladders differ meaningfully across space: Whereas in rich regions, top wages exceed bottom wages by a factor of almost 4, poor locations only see a rise in wages by a factor of less than 3.

This spatial heterogeneity in wage ladders has direct implications for the spatial variation in wage returns that result from EE transitions. Indeed, it is a force that pushes toward higher EE returns in rich locations. To see this, we show in Appendix A.7 that the spatial variation in expected EE wage growth, $\mathbb{E}[g_w^{EE}|\ell]$, is given by:³⁰

$$\frac{\partial \mathbb{E}[g_w^{EE}|\ell]}{\partial \ell} \stackrel{s}{=} \int_0^1 \int_{\mathcal{S}} \underbrace{\left(\frac{\partial}{\partial \ell} \left(\frac{w(\Gamma_\ell^{-1}(\mathcal{R}), \ell)}{w(\Gamma_\ell^{-1}(\mathcal{S}), \ell)} \right) - 1 \right)}_{\text{Spatial Variation in Steepness of Wage Ladder}} \frac{1}{(1 + \varphi^E(1 - \mathcal{S}))^2} d\mathcal{R} d\mathcal{S}. \quad (19)$$

Average EE wage growth is higher in high- ℓ locations if wages increase more rapidly across firm productivity ranks. This highlights that the steepness of the local wage ladder is the fundamental determinant of local EE returns.³¹ Because the steepness of the local wage ladder is independent of local TFP $A(\ell)$, the same holds for average EE wage growth in ℓ .

While it is difficult to find broad classes of primitives under which positive firm sorting renders (19) positive (we show in Appendix A.7, Proposition A1 and Corollary 2, that for firm productivity distributions with log-supermodular quantile functions $\Gamma_\ell^{-1}(\mathcal{R})$ such as the Pareto, this result indeed holds), we can investigate the spatial variation in EE returns empirically. We run the following regression, which allows us to zero in on *within*-location moves in line with our model:

$$\frac{w_{i\ell,t} - w_{i\ell,t-1}}{w_{i\ell,t-1}} = \sum_{\ell=1}^{257} \beta_\ell + \sum_{\ell=1}^{257} \beta_\ell^{EE} EE_{i\ell,t} + \sum_{\ell=1}^{257} \beta_\ell^{EXT} EXT_{i\ell,t} + \varepsilon_{i\ell,t}. \quad (20)$$

Here, $w_{i\ell,t}$ is the wage of individual i in CZ ℓ and month t , $EE_{i\ell,t}$ indicates whether i made an EE move to a job in CZ ℓ between months t and $t-1$, and $EXT_{i\ell,t}$ indicates an EE transition to a job in CZ ℓ from a job outside of ℓ . Coefficients $(\beta_\ell, \beta_\ell^{EE}, \beta_\ell^{EXT})$ are CZ fixed effects and CZ-specific returns to EE moves from within and outside the CZ, respectively. Our coefficient of interest is β_ℓ^{EE} : the impact of an EE move *within* CZ ℓ on wage growth. Note that coefficients β_ℓ allow for location-specific differences in average wage growth and thus capture, for example, differences in human capital accumulation across space, which might be greater in high- ℓ places.

In the bottom right panel of Figure 3, we plot β_ℓ^{EE} against local GDP per capita. More prosperous locations have substantially higher EE returns: A single job-to-job move in the richest German local labor market increases wages by around 20%, which is more than twice as much

³⁰We define average wage growth achieved via an EE transition in ℓ as $\mathbb{E}[g_w^{EE}|\ell] = \frac{\int_y^{\bar{y}} \int_y^{\bar{y}} \left(\frac{w(t, \ell)}{w(y, \ell)} - 1 \right) d\Gamma_\ell(t) dG_\ell(y)}{\int_y^{\bar{y}} \int_y^{\bar{y}} d\Gamma_\ell(t) dG_\ell(y)}$.

³¹The slight difference to the previous discussion of steeper job ladders is that here we require the wage difference between *any two* ranks $\{\mathcal{R}, \mathcal{S}\}$ of the firm distribution to be increasing in ℓ rather than between some rank \mathcal{R} relative to the lowest.

as in the poorest location.³² This finding is consistent with the spatial heterogeneity in dynamic wage profiles emphasized by [De La Roca and Puga \(2017\)](#) and suggests that such heterogeneity may in part reflect variation in EE returns across space, driven by firm sorting.

Within our model, the spatial variation in wage growth from EE moves, β_ℓ^{EE} , only reflects differences in firm sorting and is not directly impacted by $A(\ell)$ and its empirical determinants. In Table 1 column 4, we nevertheless investigate alternative mechanisms outside of our model. To do so, we estimate β_ℓ^{EE} in (20) with different sets of RHS variables and then project it on local log GDP per capita. Row 1 contains our baseline specification from Figure 3: A rise in GDP per capita by one log point is associated with an increase in EE returns by 11 percentage points. We show in row 2 that this spatial correlation is not due to a high concentration of industries with faster EE wage growth in rich labor markets: When we estimate (20) while controlling for sector fixed effects, our main coefficient hardly changes.

An additional concern is that spatial worker sorting drives regional differences in EE returns, if, e.g., skilled workers predominantly settle in rich labor markets and experience faster learning ([De La Roca and Puga, 2017](#)) or higher EE wage growth. Row 3 estimates (20) while additionally controlling for observable worker characteristics such as age, gender and education fixed effects. And row 4 controls for unobserved heterogeneity in individuals' wage *growth* by including an individual fixed effect in (20). Both leave the correlation between local EE returns and GDP per capita unchanged. Finally, row 5 controls for local population density when regressing β_ℓ^{EE} on local GDP per capita, which only has a mild impact on our coefficient of interest. This addresses the concern that higher EE returns in rich and dense locations may reflect better match quality (see, e.g., [Dauth et al., 2022](#)) if these places benefit from increasing returns to matching.

Finally, to assess the quantitative impact of this spatial variation in EE returns on spatial differences in the level and dispersion of wages, we perform a statistical decomposition of life-time earnings across locations (see also [Baum-Snow and Pavan, 2011](#)), considering the following drivers: spatial differences in wage growth due to EE moves, starting wages, wage growth during continuing job spells, and wage growth of the frequently unemployed. We find that spatial differences in EE returns are a significant factor behind spatial differences in local wage distributions: 22% of the spatial difference in within-location inequality, as measured by the 90-10 log wage difference, and 20% of the spatial earnings gap between the 25% richest and poorest CZs that emerge 15 years into workers' careers are due to differential EE returns. See Appendix D.3 for details.³³

³²We compute an average EE return of 13%, similar to the results in [Heise and Porzio \(2022\)](#) who also use the LIAB.

³³[Baum-Snow and Pavan \(2011\)](#) find that differential EE wage growth across small and large US cities accounts for 13% of the earning gap after 15 years, possibly lower than our estimate because EE returns are higher in Germany ([Engbom, 2021](#)).

6 Estimation

To evaluate the quantitative importance of firm sorting for spatial disparities in labor market power and wages, we structurally estimate our model. To this end, we first enrich our model along four dimensions and then discuss our identification proof, estimation strategy, and results.

6.1 Bringing our Model to the Data

Setting. Our goal is to introduce minimal changes that render our model suitable for estimation while preserving its key mechanism and tight link to our analytical results. First, we relax the assumption of fully immobile labor and allow unemployed workers to settle in any location. This feature is important, since even though we observe a high degree of local hiring, local labor markets are not perfectly segmented.³⁴ Second, we introduce a residential housing market in each location, so that workers now use their flow income to consume not only the final good but also housing. Third, we introduce local amenities that can vary with ℓ and scale individuals' real consumption utility. Last, we allow job separation rates δ to vary (exogenously) across locations to rationalize the observed spatial variation in unemployment.³⁵ Importantly, in Section 8, we show robustness of our quantitative results to additional changes: introducing imperfect labor mobility; dispensing with Assumption 1 so that there can be firm selection at the lower end of the local productivity distributions; and controlling for local capital intensity via industry.

By allowing for spatial mobility among the unemployed, our model endogenizes local population size $L(\ell)$, and thereby also local meeting rates of workers ($\lambda^U(\ell)$, $\lambda^E(\ell)$) and firms $\lambda^F(\ell)$. We assume that in each ℓ there is a labor market matching function with constant returns to scale, so that meeting rates are determined by local market tightness, $\theta(\ell) = \mathcal{V}(\ell)/\mathcal{U}(\ell)$, where the measure of vacancies in each location, $\mathcal{V}(\ell)$, equals the measure of firms that settle there.³⁶ In turn, we let parameter κ be the relative matching efficiency of employed workers (i.e., $\lambda^E(\ell) = \kappa\lambda^U(\ell)$) so that $\mathcal{U}(\ell) = L(\ell)(u(\ell) + \kappa(1 - u(\ell)))$ is the effective measure of searchers in ℓ , impacted by the endogenous $L(\ell)$. An important implication is that firms' and workers' meeting rates, $(\lambda^F(\ell), \lambda^U(\ell), \lambda^E(\ell))$ can vary across locations. These location-specific meeting rates create congestion, which is an additional channel that affects the costs of competition and thus firm sorting.

³⁴Introducing mobility of the unemployed (as opposed to the employed) preserves the structure of our model. Moreover, employed workers are less mobile empirically: $\sim 90\%$ are hired from within a 100 km radius around the firm, where we assess the workers' location based on the last employer's location.

³⁵Importantly, in the estimated model positive firm sorting is also optimal if δ is held constant in ℓ .

³⁶Note that under positive sorting, each ℓ is chosen by a single p , where we assume that for each p , there is a continuum of firms i s.t. $0 \leq i \leq q(p)$ with Lebesgue measure (i.e., a continuum of mass $Q'(p) = q(p)$). In equilibrium $p = \mu(\ell)$, so the mass of firms in ℓ is $Q'(\mu(\ell)) = q(\mu(\ell))\mu'(\ell)$. Combined with the fact that in any ℓ the measure of firms equals the measure of vacancies, we have $\mathcal{V}(\ell) = q(\mu(\ell))\mu'(\ell) = r(\ell)$, where the mass of vacancies *per unit of land* in ℓ is one, i.e., $\mathcal{V}(\ell)/r(\ell) = 1$.

The residential housing market—a second source of congestion—features exogenous supply, $h(\ell)$, in each location. Workers have Cobb-Douglas preferences over the final good and housing. We denote the share of income that is spent on housing (the final good) by ω ($1 - \omega$). The income of employed workers is wage $w(y, \ell)$ and that of unemployed workers is benefit $w^U(\ell)$, financed via taxes on homeowners' income, τ .³⁷ Further, the government budget needs to balance, $\tau d(\ell)h(\ell) = w^U(\ell)u(\ell)L(\ell)$, where $d(\ell)$ is the housing price in ℓ . It adjusts to clear the housing market, balancing housing demand from unemployed and employed workers with housing supply $h(\ell)$.

The population size in each ℓ , $L(\ell)$, is pinned down by the fact that in equilibrium, workers must be indifferent between any two locations—i.e., the value of search is equalized across space,

$$V^U(\ell') = V^U(\ell'') \quad \forall \ell' \neq \ell'',$$

where $V^U(\ell)$, compared with (1) in the baseline model, reflects the fact that high local house prices, $d(\ell)$, low job-finding rates, $\lambda^E(\ell)$, and high separation rates, $\delta(\ell)$, render job search in location ℓ less attractive. In contrast, favorable local amenities, $B(\ell)$, render it more attractive:

$$\rho V^U(\ell) = B(\ell)d(\ell)^{-\omega} \left(z(\underline{y}, A(\ell)) + \lambda^E(\ell) \left[\int_{z(\underline{y}, A(\ell))}^{\bar{w}(\ell)} \frac{1 - F_\ell(t)}{\delta(\ell) + \lambda^E(\ell)(1 - F_\ell(t))} dt \right] \right). \quad (21)$$

If location ℓ' , for instance, has a better wage distribution than location ℓ'' (causing a temporary imbalance $V^U(\ell') > V^U(\ell'')$), workers will move into ℓ' . This puts downward pressure on market tightness (and thus workers' meeting rates) and upward pressure on housing prices in ℓ' until the difference in the locations' attractiveness is arbitrated away.

Importantly, despite these additions to the model, conditions similar to those in our baseline model guarantee the positive sorting of firms to locations; see Proposition A2 (Appendix B). The value that determines firms' location choices, $\bar{J}(p, \ell)$, is analogous to the baseline model with one key difference: Meeting rates are now endogenous. As a consequence, local competition has *two* components. It depends not only on local firm composition, Γ_ℓ (as before), but also on local labor market congestion, captured by $(\lambda^E(\ell), \lambda^F(\ell))$. Both components now affect how the firm size elasticity in (13) varies across space.

Functional Forms. As for local labor markets, we assume that worker-firm meetings are based on a Cobb-Douglas matching function $M(\mathcal{V}(\ell), \mathcal{U}(\ell)) = \mathcal{A} \sqrt{\mathcal{V}(\ell) \mathcal{U}(\ell)}$, where \mathcal{A} is the overall

³⁷The indirect utility of unemployed workers from consuming the final good and housing is given by $w^U(\ell)/d(\ell)^\omega$, where $d(\ell)$ is the housing price in ℓ . Thus, the flow utility of unemployed workers is given by $b(\ell) = B(\ell)w^U(\ell)/d(\ell)^\omega + \tilde{b}(\ell)$, where amenity $B(\ell)$ scales the consumption utility. We interpret \tilde{b} as a non-monetary (possibly negative) utility component that stems from stigma. In practice, function \tilde{b} gives us flexibility to satisfy Assumption 1, so that $w^R(\ell) = z(\underline{y}, A(\ell))$ for all ℓ .

matching efficiency. As far as production is concerned, we assume that production function z is multiplicative and that the ex post productivity distribution is Pareto with tail parameter $1/p$.³⁸

$$z(y, A(\ell)) = yA(\ell) \quad \text{and} \quad \Gamma(y | p) = 1 - y^{-\frac{1}{p}}.$$

Based on this Pareto specification, firms with ex ante higher firm productivity p draw their ex post productivity y from a stochastically better distribution, in line with our theory.

6.2 Identification

Our model is parameterized by a location ranking $[\ell, \bar{\ell}]$, local TFP $A(\ell)$, local amenities $B(\ell)$, local separation rates $\delta(\ell)$, labor market parameters (κ, \mathcal{A}) , and parameters of the housing market $(\omega, \tau, h(\ell))$. We must also identify the extent of spatial firm sorting, captured by $\mu(\ell)$ (which in equilibrium equals p , the (inverse) Pareto tail of ℓ 's productivity distribution).³⁹

The key step in our identification strategy is separating the effect of firm sorting $\mu(\ell)$ from local productivity $A(\ell)$. Intuitively, are locations prosperous because of high fundamental productivity or because of an advantageous firm composition? Our model allows us to separately identify $\mu(\ell)$ and $A(\ell)$ using the average local labor share $LS(\ell)$ (Corollary 1) and firm value added, $\mathbb{E}[z(y, A(\ell))|\ell]$,

$$LS(\ell) = 1 - \mu(\ell) \tag{22}$$

$$\mathbb{E}[z(y, A(\ell))|\ell] = A(\ell)(1 - \mu(\ell))^{-1}, \tag{23}$$

where the expectation is taken over Γ_ℓ . Variation in local labor shares across space $LS(\cdot)$ thus identifies firm sorting $\mu(\cdot)$. Conditional on local firm composition $\mu(\ell)$, location productivity $A(\ell)$ can then be identified from average local value added: The spatial variation in value added that is *not* accounted for by firm sorting must be driven by differences in local TFP.

To identify the parameters of the labor market, we exploit information on job-finding rates and local unemployment. First, the relative matching efficiency of employed workers κ is identified from their job-finding rate, relative to that of unemployed workers. Second, we can identify the local job-separation rate, $\delta(\ell)$, using the steady-state formula for unemployment, as well as data on local unemployment and job-finding rates:

$$\delta(\ell) = \lambda^U(\ell) \frac{u(\ell)}{1 - u(\ell)}. \tag{24}$$

Finally, the overall matching efficiency, \mathcal{A} , is identified from a mix of the job-finding rate of the unemployed, $\lambda^U(\ell)$, the job-destruction rate, $\delta(\ell)$, and the average firm size, $\bar{l}(\ell)$, in *any* ℓ ,

$$\mathcal{A} = \sqrt{\lambda^U(\ell)(\delta(\ell) + \kappa\lambda^U(\ell))\bar{l}(\ell)}. \tag{25}$$

³⁸Note that we normalize the scale parameter, $y = 1$.

³⁹Given matching function μ , we can identify \bar{Q} using $\mu(\ell) = Q^{-1}(R(\ell))$ (where we assume R is given; see below).

To identify the parameters of the housing market, we use the expenditure share of residential housing to pin down ω and the replacement rate of the unemployed to obtain tax rate τ for residential homeowners. And based on observed house prices—along with the government budget constraint and housing market clearing—we can infer housing supply $h(\ell)$.

Last, we identify the amenity schedule $B(\ell)$ using the indifference condition whereby workers' value of search is equalized across space, given by (21) when imposing the normalization $\rho V^U = 1$. We now summarize this discussion:

Proposition 6 (Identification). *Under the assumed functional forms (summarized in Assumption A1, Appendix E), the model is identified.*

We provide more details on the presented derivations in the proof; see Appendix E.

6.3 Estimation: Strategy and Results

For estimation, we rely on regional data from the official records of the German Federal Statistical Office, which we aggregate at commuting-zone level; see Appendix C.1 for details. Specifically, we use employment, value added, and labor compensation, as well as unemployment rates, number of establishments, and GDP. In turn, for model validation, we use worker- and firm-level data from the FDZ as for our empirical analysis in Section 5. The time unit is 1 month.

The identification argument provides us with a concrete estimation protocol that we follow closely. Our implementation proceeds in seven steps. First, as in Section 5, we rank the 257 CZs based on their log GDP per capita. Local log GDP per capita will be our discretized support $\{\ell_1, \ell_2, \dots, \ell_{257}\}$ of the model's land distribution R . We identify R from the number of firms in each CZ, i.e., we assign to each $\ell_j, j \in \{1, 2, \dots, 257\}$ a probability mass $r(\ell_j)$ equal to its share of firms in Germany. Examples of the highest ranked CZs are Munich and Wolfsburg (in West Germany); among the lowest ranked, we have Goerlitz (East) and Mansfeld-Südharz (rural East).

Second, we use (22) to obtain firm matching function $\mu(\ell)$ from the observed labor share in ℓ . Because our model is stylized (e.g., it lacks noise in the firm-location matching process), we smooth any measurement error in the data moments before feeding them into the model. Specifically, we linearly fit each variable we target in estimation as a function of ℓ . Since the labor share is *decreasing* in ℓ (Figure 4, top left), we obtain an *increasing* matching function μ (top right).⁴⁰ This implies positive sorting between firms and locations.

Third, we obtain the overall matching efficiency, \mathcal{A} , from the Germany-wide observed matching rate, separation rate, and average firm size, using (25) (see Table A.3, Appendix F). To obtain

⁴⁰The size of the dots in Figure 4 is proportional to the size of the CZ, as measured by its number of establishments.

the relative matching efficiency of employed workers κ , we take into account only those EE moves in the data that are associated with wage gains (59.7%) and set $\kappa = 0.597 \cdot \frac{\lambda^E}{\lambda^U}$; see Table A.3.

Fourth, to pin down the local separation rates from (24), we use local unemployment and job-finding rates.⁴¹ An observed unemployment rate that is decreasing in ℓ (Figure 4, second row, left), along with a fairly stable job-finding rate, translates into job-separation rates that are lower in richer places (second row, right).

Fifth, we estimate location TFP based on the average value added per worker across locations using (23), except that we weigh each firm type by its employment.⁴² Since value added per worker is strongly increasing in ℓ (Figure 4, bottom left), we obtain an increasing A -schedule, even after controlling for positive firm sorting through μ (Figure 4, bottom right). To better understand the determinants of local TFP, we project the estimated A 's on several location factors. We find that high local TFP is associated with a low corporate tax rate, the quality of infrastructure, and especially the skills of local workers; see Table A.4, Appendix F. This is consistent with our implicit assumption in the baseline model that spatial worker heterogeneity shapes $A(\ell)$ and is therefore taken into account by firms when making their location choices.

Sixth, to pin down the parameters that govern residential housing markets, we target the average rent-to-income ratio of main tenant households (and obtain $\omega = 0.272$) and an average replacement rate of 60%, which implies a proportional tax rate on residential landlords of $\tau = 0.164$ (Table A.3, Appendix F). Finally, we pin down local housing supply $h(\ell)$ using observed location-specific rental rates $d(\ell)$; see Figure A.3 in Appendix F.

Last, given $(\mu(\ell), A(\ell), \lambda^E(\ell), \delta(\ell), d(\ell))$ for each ℓ , we use (21) to back out amenity schedule B , which ensures that unemployed workers are indifferent between all locations. The left panel of Figure A.3 (Appendix F) shows that amenities are decreasing in the location index. Thus, even though residential housing is more expensive in high- ℓ places, this force is not strong enough to dissuade workers from settling in those locations with high TFP *and* better firms, which calls for particularly low amenities in these places.

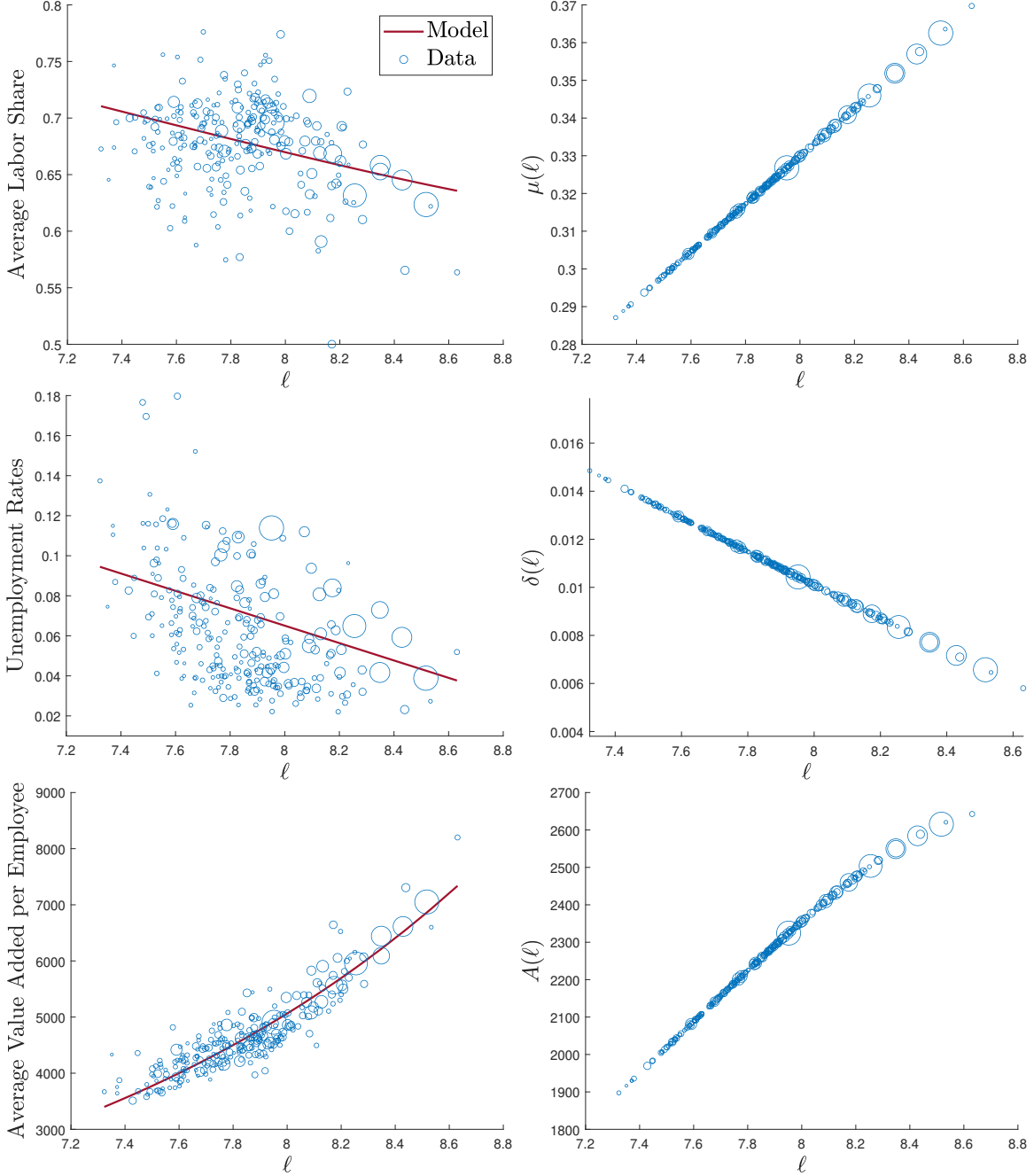
Importantly, while we stipulate that the empirical firm sorting can be captured by a function μ , at no point of the estimation do we impose PAM. Given the estimation output, we verify that the

⁴¹To avoid using noisy CZ-specific job-finding rates from the small FDZ sample, we infer the (endogenous) job finding rate $\lambda^U(\ell)$ in each ℓ from the average firm size $\bar{l}(\ell)$ provided by the German Federal Statistical Office. To do so, note that solving (25) for $\lambda^U(\ell)$ while taking $\delta(\ell) = \lambda^U(\ell)u(\ell)/(1 - u(\ell))$ into account gives $\lambda^U(\ell) = \mathcal{A}(\bar{l}(\ell) \cdot (\kappa + \frac{u(\ell)}{1-u(\ell)}))^{-\frac{1}{2}}$. We then compute $\lambda^E(\ell) = \kappa\lambda^U(\ell)$ and $\lambda^F(\ell) = \lambda^U(\ell)/\theta(\ell)$. Since average firm size is increasing in ℓ , we obtain a slightly increasing $\lambda^U(\cdot)$, which implies higher meeting rates for workers and lower meeting rates for firms in high- ℓ locations (Figure A.2, Appendix F).

⁴²Instead of applying (23), we apply its weighted version $A(\ell) = \mathbb{E}_{g_\ell}[z(y, A(\ell))|\ell]/(\int yg_\ell(y)dy)$, where we observe average value added per employee, $\mathbb{E}_{g_\ell}[z(y, A(\ell))|\ell]$, in the data; and where we compute $\int yg_\ell(y)dy$ in the model, taking density g_ℓ based on (12) into account, which depends on $(\mu(\ell), \lambda^U(\ell), \delta(\ell))$ —all objects that we pinned down above.

value of firm p of settling in ℓ , $\bar{J}(\ell, p)$, is supermodular in (p, ℓ) (in line with Proposition 1), which verifies that the positive sorting of firms into locations is indeed *optimal* in the estimated model.

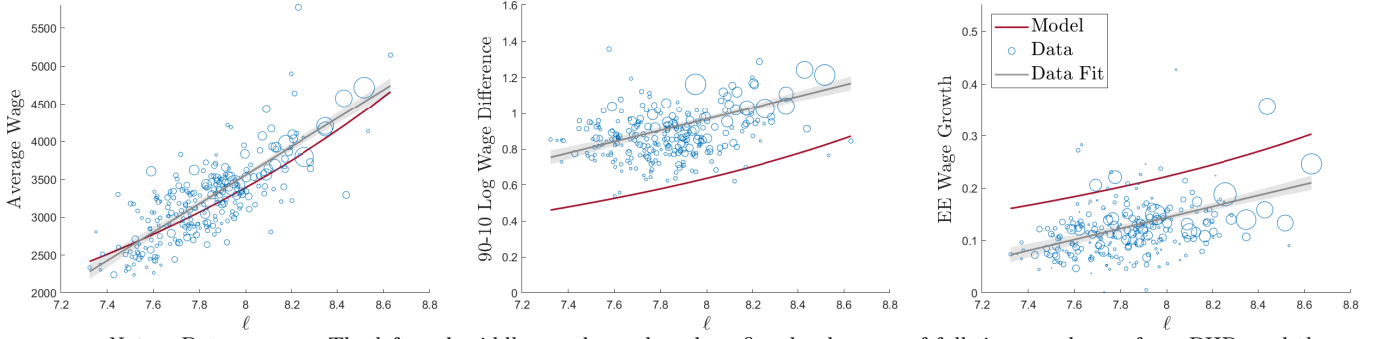
Figure 4: Model Fit of Targeted Moments (left) and Estimated Parameters (right)



6.4 Model Validation

Given our estimation approach, we fit the targeted data series of local labor shares, firm size, unemployment rates, and value added per worker by construction (Figure 4, left column). Importantly, despite its parsimony, our model also performs reasonably well regarding several non-

Figure 5: Model Fit: Non-Targeted Moments



Notes: Data sources: The left and middle panels are based on firm-level wages of full-time employees from BHP; and the right panel is based on worker-level wages from the LIAB. The right panel (data) shows β_{ℓ}^{EE} from regression (20). These local statistics are weighted by the number of firms (left and middle panel) and the number of EE moves within the CZ (right panel), indicated by different marker sizes. 95% confidence intervals are displayed in gray.

targeted features of the data that are related to worker inequality and beyond.

In Figure 5, we confront our model with the findings from our empirical analysis (Section 5). In the left panel, we display the average local wage across space. Because our model perfectly matches value added per capita across ℓ , it is expected that it also fits spatial differences in average wages. Second, and more importantly, our model captures quite well both spatial differences in wage inequality within locations and spatial wage ladder heterogeneity. In the middle panel of Figure 5, we display local 90-10 log wage gaps across space, which is the model counterpart of the top right panel of Figure 3. Although our model underestimates the level of within-location inequality, it almost perfectly rationalizes its spatial variation. The right panel shows that our model also replicates the heterogeneity in EE wage growth across space, generated by steeper wage ladders in high- ℓ locations—the theoretical counterpart to the bottom right panel of Figure 3. We overestimate the level of EE wage growth, which may not be surprising since our model features only positive wage changes due to EE moves while the data also contains a substantial share of negative wage growth. Reassuringly, however, we closely match the variation in EE returns across space.

We further validate the model along the following dimensions (see Appendix F). First, we show that our model matches relatively well the decreasing rate of job loss, $\delta(\cdot)$, as well as the flat EE transition rate, $\mathbb{P}(EE|\cdot)$, and UE transition rate, $\lambda^U(\cdot)$ (Figure A.4). Second, our model replicates the fact that employment is more concentrated among top firms in rich labor markets. Specifically, in both the data and the model, the employment share among the largest 25% of firms is increasing in ℓ (Figure A.5, left panel). Finally, our model is consistent with the fact that commercial land prices are higher in rich labor markets, even though the increase is more pronounced in the data (Figure A.5, right panel).

7 Spatial Firm Sorting: Quantitative Implications

We now use our estimated model to quantify the role of spatial firm sorting for local labor shares, wage ladders, and spatial wage disparities. To do so, we consider a counterfactual in which we allocate firms randomly across locations and let the equilibrium play out. Hence, workers reallocate across space, job-finding rates adjust via local market tightness, and local wage schedules change. Appendix SA.6.1 contains technical details on this exercise.⁴³ We then highlight the importance of OJS and endogenous firm location choices for our conclusions.

7.1 Implications for Local Labor Shares and Spatial Wage Inequality

Our model predicts that positive firm sorting has distinct implications for productive locations: It increases local monopsony power and lowers their labor shares; and it steepens their local job ladders, thereby contributing to higher and more dispersed wages. We now assess the quantitative importance of these implications.

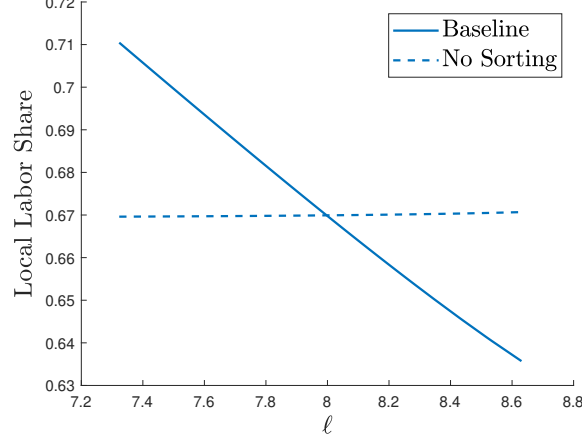
Local Labor Shares. Consider first the effect of firm sorting on local labor shares. In our baseline estimation, rich labor markets have lower labor shares because they have a larger mass of productive, “superstar” firms with low labor shares. In contrast, in the absence of sorting, firm distributions are identical in all locations and local labor shares are essentially equalized. This is illustrated in Figure 6. We display the local labor share as a function of location productivity index ℓ both for the baseline allocation with positive sorting (solid line) and the no-sorting counterfactual (dashed line), which equalizes local labor market power across space.⁴⁴ The implied *changes* in local labor shares are quantitatively sizable. In the absence of firm sorting, labor shares would fall by around four percentage points in the least productive labor markets and increase by the same amount in the most productive locations. At the same time, the economy-wide labor share would not be affected and remains at 0.67.

Our theory sheds light on why local labor shares are equalized when firm sorting is absent; see decomposition (15). While the absence of sorting still results in lower *firm*-level labor shares for productive firms, it causes significant shifts in the local employment composition: Equalizing firm productivity distributions across space shifts employment towards less productive, high labor-share firms in rich labor markets and towards more productive, low labor-share firms in

⁴³Throughout, we maintain Assumption 1. As a result, the *economy-wide* firm distribution stays the same as in the baseline model. In Section 8, we explicitly allow for firm selection as robustness.

⁴⁴Note that, in the absence of firm sorting, labor shares still vary slightly across space since each local productivity distribution becomes a mixture of Pareto distributions, which itself is not Pareto. Hence, Corollary 1 does not directly apply and differences in the efficiency of OJS, $\varphi^E(\ell)$, across ℓ impact local labor shares.

Figure 6: The Effect of Spatial Firm Sorting on Local Labor Shares



Notes: The figure plots the local labor share in each location ℓ under the baseline estimation (solid line) and the no-sorting counterfactual (dashed line).

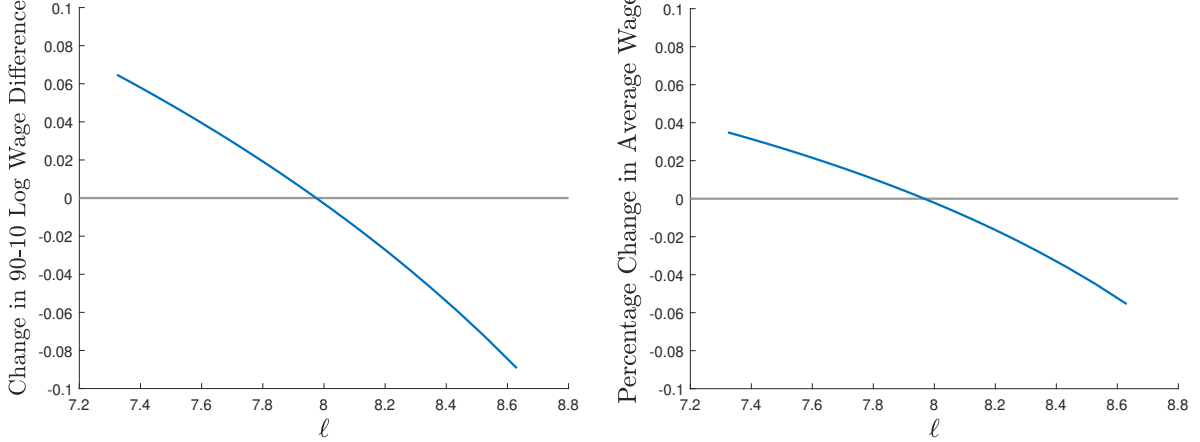
poor ones. As a result $\partial \tilde{G}_\ell(y)/\partial \ell \approx 0$, which has a levelling effect on spatial labor share gaps.⁴⁵

Wage Inequality Within and Across Locations. We now turn to the implications of firm sorting for the distribution of wages both within and across labor markets. We begin with the implications of firm sorting for wage inequality *within* locations, which, as before, we measure by the 90-10 difference in local log wages. The left panel of Figure 7 shows how within-location inequality would change if firm sorting were removed: Wage inequality would decline by almost 0.1 log points in productive labor markets and increase by 0.06 log points in unproductive locations. This shows that firm sorting is a quantitatively important determinant of the spatial variation in within-location inequality. Indeed, Table 2 (left panel) shows that firm sorting accounts for 40% of the higher inequality within rich labor markets: Whereas the difference in within-location inequality between rich and the poor German labor markets (i.e., between locations in the top and bottom quartiles of the distribution of local GDP per capita) is 0.2 log points in our baseline model with sorting, this gap would drop to 0.12 log points in the absence of it. While differences in local TFP A do *not* affect the spatial gap in within-location inequality (see (17)), spatial heterogeneity in job-destruction rates δ and job-finding rates λ^E does impact local wage ladders and account for the remaining spatial variation in our quantitative model.

In the right panel of Figure 7, we show the link between firm sorting and wage inequality *across* locations by plotting the change in average local wages in the absence of sorting relative to the baseline equilibrium. Firm sorting is a quantitatively important determinant of spatial differences: Without sorting, wages would be 6% lower in the richest labor markets and around 4% higher in the poorest locations. Hence, unproductive labor markets are not only disadvantaged because of

⁴⁵The lack of sorting (and thus the same local firm productivity distribution everywhere) also ensures $\partial^2 Ls(y, \ell)/\partial y \partial \ell \approx 0$.

Figure 7: The Effect of Spatial Firm Sorting on Spatial Wage Inequality



Notes: The left panel shows the percentage point change in the 90-10 difference in log wages within locations if firm sorting is abolished. The right panel shows the percentage change in average local wages if firm sorting is abolished.

Table 2: The Effect of Spatial Firm Sorting on Spatial Wage Inequality

Within-Location Inequality (90-10 Difference in log Wages)			Across-Location Inequality (Spatial Wage Premium)		
Baseline	No Sorting	Contribution of Sorting	Baseline	No Sorting	Contribution of Sorting
0.20	0.12	40%	40%	34%	15%

Notes: Columns 1-3 pertain to the spatial gap in the 90-10 difference of log wages within labor markets, where we first assess the gap between the 90th and 10th quantiles of log wages in each location and then compute the difference of this measure across poor and rich locations. Column 1 (2) reports this statistic for the baseline model (for the counterfactual equilibrium without sorting). Column 3 reports the contribution of sorting as the percentage difference between columns 1 and 2. Columns 4-6 report the analogue for the spatial wage premium between rich and poor locations. Throughout, we define poor (rich) locations as the bottom (top) quartile of commuting zones ranked by their local GDP per capita.

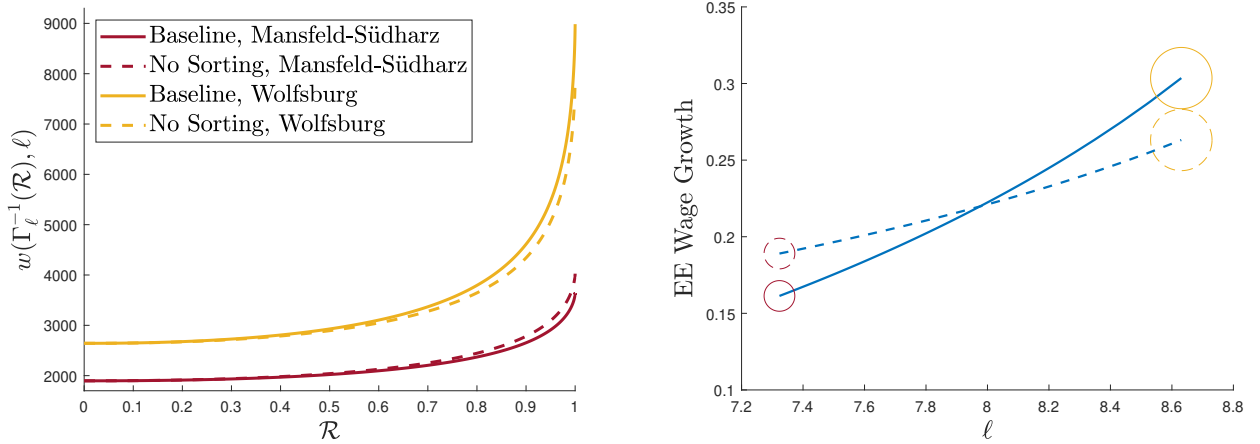
poor economic fundamentals but also because workers lack access to the most productive firms. In the right panel of Table 2, we decompose the observed wage premium between poor and rich locations. In our no-sorting counterfactual, this wage premium drops from 40% to 34%. Firm sorting thus accounts for $6/40 = 15\%$ of the observed spatial wage gap.

To put the quantitative role of sorting into perspective, we can also shut down spatial differences in local TFP and set $A = \mathbb{E}[A(\ell)]$ in all locations while keeping firm sorting at its baseline level. Doing so would cut the spatial wage gap in half—that is, reduce it to 21%. Hence, the effect of firm sorting on spatial inequality is about one-third as large as the effect of differences in location TFP. As we show in Section 8, where we discuss alternative model assumptions, we view the effect of firm sorting on the spatial wage premium reported here as a lower bound.

To better understand *why* changes in firm sorting affect wage distributions across space, we draw on our theoretical analysis on the drivers of within-location inequality, captured by (17), and of across-location inequality, given by (18). There we have shown that firm sorting affects within- and across-location inequality through its effect on the steepness of local wage ladders.

Figure 8 illustrates this point by comparing two locations at opposite ends of the spectrum

Figure 8: No-Sorting Counterfactual: Wage Ladders and EE Returns



Notes: The left panel shows the local wage ladders in Wolfsburg (WB, yellow) and Mansfeld-Südharz (MS, red) with and without sorting. The right panel reports wage growth due to EE moves with sorting (solid line) and without sorting (dashed line) as a function of ℓ . WB and MS are denoted by yellow and red circles, which are proportional to their relative size.

of the local GDP per capita distribution, Wolfsburg at the top and Mansfeld-Südharz at the bottom. The left panel stresses that random matching of firms to locations reduces wage ladder differences across space. The random allocation improves firm productivity in bottom locations and deteriorates it in top ones. More specifically, the firm productivity y —and therefore the wage—associated with any given productivity rank $\mathcal{R} > 0$ *declines* in rich labor markets and *increases* in poor labor markets. As a consequence, while in our baseline estimation the wage ladder was considerably steeper in the top location (yellow solid) compared with the bottom location (red solid), this differential steepness shrinks when firm sorting is absent (dashed lines).

These differential changes in local wage ladders are also reflected in changes of local returns from EE moves. The right panel of Figure 8 shows the effect of firm sorting on EE returns. Relative to our baseline estimation (solid line), spatial differences in EE returns are substantially lower without sorting (dashed line). For our two locations, Wolfsburg and Mansfeld-Südharz, shown as red and yellow circles, the difference in EE returns falls by around 75%. This analysis suggests that firms sorting is a central determinant of the observed spatial variation in wage ladders and EE wage growth and—through this channel—of the variation in local wage distributions.

7.2 The Role of On-the-Job Search

In our theory, firm sorting affects spatial inequality by shaping the returns to on-the-job search—and the associated wage ladders—within spatially segmented labor markets. This suggests that firm sorting would have *smaller* effects on local wage distributions if OJS was less pronounced.

To illustrate this complementarity between OJS and firm sorting, we reduce the importance of OJS by lowering the relative matching efficiency of employed workers, κ . To this end, we *rees-*

estimate our model for a range of different κ by targeting the same moments as in our baseline estimation except for the relative job-finding rate of employed workers that used to pin down κ . We therefore still match local labor shares (which leaves spatial differences in firm productivity, captured by Γ_ℓ , unaltered), local value added per worker, unemployment rates, and average firm size by appropriate choices of location TFP, $A(\ell)$, separation rates, $\delta(\ell)$, and job-finding rates, $\lambda^U(\ell)$. By extension, we also match local wage levels and the spatial wage premium across ℓ .

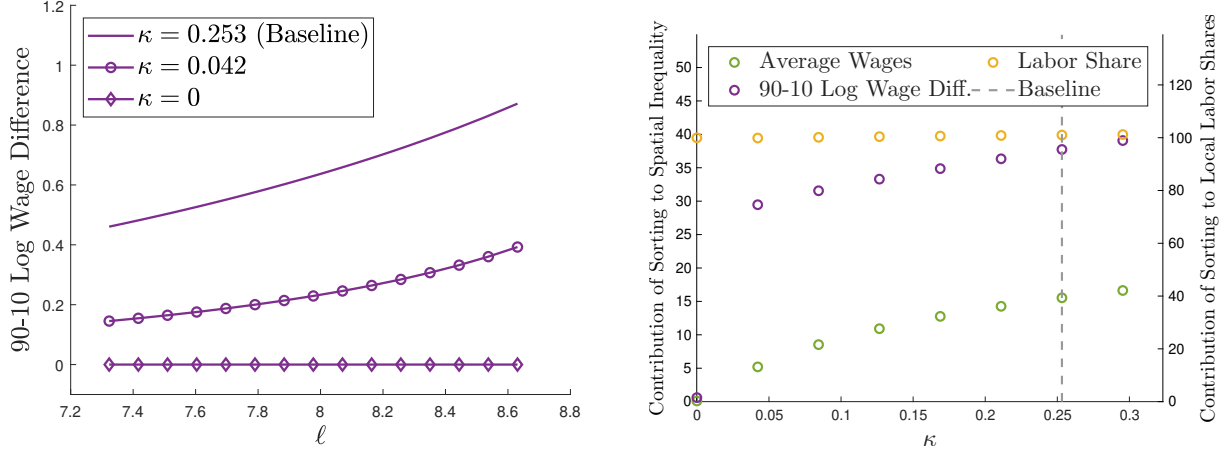
By contrast, a key untargeted moment affected by lower OJS is *within*-location wage inequality and its variation across space. In our baseline estimation, within-location inequality is 0.2 log points higher in the 25% richest relative to the 25% poorest labor markets, resembling the spatial gap in the data. Lowering OJS by lowering κ not only reduces the level of within-location wage dispersion everywhere; it especially reduces spatial differences therein. This is illustrated in the left panel of Figure 9: We see a gradual flattening of spatial differences in 90-10 log wage dispersion when we go from our baseline κ (solid line), to a lower κ (circles), and finally to $\kappa = 0$ (diamonds), which completely eliminates differences in local wage inequality. When there is no OJS, all firms pay the local reservation wage, which makes local wage ladders collapse irrespective of the local firm composition—a spatial version of the Diamond paradox (Diamond, 1971).

The reason for these *differential* changes in within-location inequality across space is the *interaction* between OJS and firm sorting. The right panel of Figure 9 shows the %-contribution of firm sorting to spatial disparities in our central outcomes for varying degrees of κ and thus OJS. To obtain these numbers we use our reestimated models for each κ and run the no-sorting counterfactual from Section 7.1. We find that firm sorting contributes significantly less to the spatial gap in *within*-region inequality when OJS is muted (purple markers): Its contribution to spatial differences in within-location inequality drops from 40% at the baseline κ to zero when κ is zero. In terms of the corresponding poaching share—a proxy for the magnitude of OJS in the economy—this implies that on average, if the poaching share was 10% lower, firm sorting would contribute about 8% less to spatial differences in local wage dispersion.⁴⁶ Intuitively, local firm productivity affects spatial wage inequality by molding local wage ladders. If workers were to climb their local wage ladder at a slower pace, spatial wage ladder heterogeneity would be less significant and the effect of firm sorting on spatial wage disparities would be mitigated.

This mechanism does not only lead to a lower contribution of firm sorting to spatial differences in within-location inequality; it also reduces the contribution of firm sorting to *between*-location differences in wage levels (green markers). In our baseline estimation, firm sorting accounts for

⁴⁶A firm’s poaching share is defined as the ratio of EE inflows relative to all (EE plus UE) inflows.

Figure 9: The Importance of On-The-Job Search and its Interaction with Firm Sorting



Notes: The left panel shows the local 90-10 log wage difference across ℓ under three levels of relative matching efficiency for employed workers, κ , where we reestimate the model for each κ . The right panel shows the %-contribution of sorting to spatial gaps in local labor shares (yellow, right axis), within-location wage inequality (purple, left axis), and across-location wage inequality (green, left axis) between poor and rich labor markets. We define poor (rich) locations using local GDP per capita; see Table 2. We indicate our baseline estimate $\kappa = 0.253$ with a dashed vertical line in the right panel.

about 15% of the spatial wage premium. In the limiting case of no OJS ($\kappa = 0$), firm sorting explains *none* of the spatial wage premium, which is instead entirely driven by local TFP, $A(\ell)$.⁴⁷

In sum, firm sorting and OJS are complementary in nature. Firm sorting affects spatial wage inequality by shaping local wage ladders. In turn, OJS determines the fluidity of local labor markets and hence the incentives of “superstar” firms to bid up wages in order to attract workers. For firm sorting to be quantitatively important for spatial inequality, local labor markets have to be sufficiently dynamic. Our analysis suggests that this is the case in Germany.

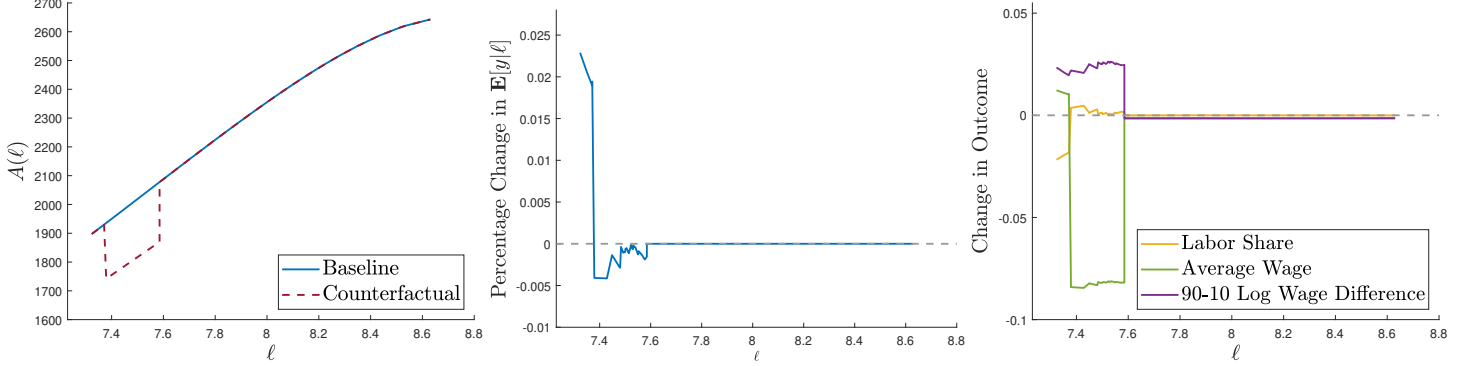
7.3 The Role of Endogenous Firm Sorting

In our theory, firm sorting is determined in equilibrium. Explicitly modelling firms’ location choices is not only important for understanding whether positive sorting is consistent with optimal firm behavior, but is essential for analyzing the effects of place-based policies. Consider, for example, a change in regional subsidies. Such a policy not only affects the targeted locations, but also impacts non-targeted ones through the re-sorting of firms. To make this point, we conduct a stylized exercise inspired by the German initiative for the improvement of local economies (GRW), which since 1969 provides discretionary investment grants to economically disadvantaged regions. Since GRW covers a wide range of CZs during our sample period, its impact is reflected in the estimated local productivity, $A(\ell)$.

Consider a hypothetical reduction of these subsidies in some structurally weak CZs (Appendix SA.6.2 gives details). For illustration, we decrease local TFP $A(\ell)$ by 10% in locations $\{\ell_5, \dots, \ell_{33}\}$,

⁴⁷For completeness, Figure 9 also highlights the fact that firm sorting explains the same share (100%) of the spatial variation in labor shares (yellow markers), irrespective of the level of OJS, simply because local labor shares only depend on $\mu(\ell)$ here.

Figure 10: The Impact of Place-Based Policies on Local TFP, Firm Sorting and Spatial Inequality



Notes: The left panel shows changes in local TFP $A(\ell)$ induced by the counterfactual policy. The middle panel shows the percentage changes of average firm productivity across ℓ induced by the counterfactual. The right panel plots percentage changes in local labor share (yellow), percentage changes in local average log wages (green), and changes in the local 90-10 log wage difference (purple) across ℓ .

which all received GRW subsidies during the time period of our study, and keep TFP in all remaining localities unchanged (see Figure 10, left panel).

By altering the relative attractiveness of local labor markets, this policy affects firm's location choices (Figure 10, middle panel). The composition of firms worsens in the “treated” locations $\{\ell_5, \dots, \ell_{33}\}$, where average firm productivity declines by 0.13%, and improves in the originally least developed, “untreated” CZs, $\{\ell_1, \dots, \ell_4\}$, where firm productivity increases by 2.1%.

The right panel of Figure 10 displays the distributional implications of this policy. Consider first locations $\{\ell_1, \dots, \ell_4\}$, which indirectly benefit from the policy change because they are able to attract better firms, which increases average wages (shown in green). This comes at the cost though of increased wage dispersion (purple) in these places and lower local labor shares (yellow). The equilibrium impacts on the treated locations $\{\ell_5, \dots, \ell_{33}\}$ are subtle. The direct productivity effect of this policy reduces local wages but the overall decline is less than the subsidy reduction of 10%. The reason is that, like productive firms, workers *also* leave these labor markets. This increases the meeting rates of the remaining workers, which counteracts the initial wage decline and increases local wage dispersion. In turn, local labor shares increase due to the presence of less productive firms with low monopsony power.

These pattern highlight the subtle effects of place-based policies on treated and non-treated areas once equilibrium effects due to endogenous firm (re)-sorting are taken into account. In fact, *neglecting* the response of firm sorting to the change in policy would lead to wrong conclusions, especially in the untreated regions that are only indirectly affected by the influx of more productive firms: Keeping firm sorting fixed would leave their local labor shares, wage dispersion and average wages essentially unchanged—at odds with the equilibrium responses our model predicts.

8 Robustness

Our quantitative results are robust to a variety of different measurement and modeling choices. For all of these alternatives, we re-estimate our model and perform the *no-sorting counterfactual*, which allows us to compare the role of firm sorting in spatial inequality to our baseline model. We conclude that our baseline estimation provides a lower bound of the role of sorting in spatial inequality. Table 3 summarizes the results and focuses on counterfactual wage outcomes since, in all robustness checks, the effect of firm sorting on local labor shares is nearly identical to our baseline estimation;⁴⁸ Appendix SA.7 contains details.

Measurement. A first concern may be that different industries operate under different technologies with different labor and capital intensities. If industries sort across space, this could drive spatial labor share differences even in the absence of firm sorting. We address this concern in two ways. First, we re-estimate our model after controlling for the local industry composition in the labor share data. As shown in column 2 of Table 3, focusing on the within-industry variation, if anything, amplifies the role of sorting: Sorting now accounts for, respectively, 48% and 26% of inequality within and across locations.⁴⁹ Second, our model may be a better description of tradable industries, where the local costumer base—a feature absent from our model—is unlikely to contribute to the attractiveness of a location. We therefore repeat our estimation using data on the manufacturing sector only. As shown in column 3 of Table 3, doing so strengthens the impact of firm sorting on spatial wage disparities. The reason is that there is more regional variation in local labor shares *within* manufacturing. As a consequence, our estimation infers stronger spatial firm sorting. See Appendix SA.7.1 for details.

Further, in our baseline estimation we relied on regional statistics from the German Federal Statistical Office. Since these are official records, we view them as more reliable than any regional aggregation we could perform on a smaller sample of firms in the FDZ data. As a consequence, our analysis draws from two different data sources, since the worker-level analysis (e.g., for model validation) requires individual-level data from the FDZ. For robustness, we therefore also perform our estimation using only data from the FDZ (Appendix SA.7.2). As seen in column 4, this yields a more prominent role of firm sorting in spatial inequality.

⁴⁸Across all robustness checks, the absence of firm sorting implies that local labor shares become nearly identical across space (as in Figure 6), which suggests that firm sorting is a robust driver behind differential local labor market power.

⁴⁹This is consistent with our regression analysis in Table 1, which revealed that the within-industry relationship between labor shares and log GDP per capita is steeper than the unconditional one.

Alternative Model Assumptions. We also probed the robustness of our results with respect to two model dimensions.

First, so far we have maintained Assumption 1, i.e., all firm types $y \in [\underline{y}, \bar{y}]$ are active in each market and the least productive firm makes zero profit. This assumption, ensured by a flexible b -schedule, ruled out spatial differences in firm selection at the lower end of local productivity distributions. In Appendix SA.7.3, we implement a version of our model without this assumption. In that model, each labor market has an endogenous productivity cutoff $\underline{y}(\ell)$, and all firms with $y < \underline{y}(\ell)$ choose not to produce. In the estimated model, production threshold $\underline{y}(\ell)$ is increasing in ℓ , which implies that firm sorting is stronger than in our baseline estimation. Thus, allowing for this form of firm selection amplifies the impact of sorting on inequality: Sorting now accounts for 57% of differences in within-location inequality and 22% of spatial wage gaps (column 5 of Table 3).

Second, our estimation was based on the simplifying assumption of full mobility of unemployed workers. As an alternative, we estimate a model with location preference shocks that hamper labor mobility. We discipline workers' propensity to migrate by targeting a spatial labor supply elasticity that is consistent with the empirical literature; see Appendix SA.7.4. Column 6 shows that this model with imperfect worker mobility does not change our quantitative conclusions about the role of sorting in within- and across-location inequality.

Table 3: The Effect of Spatial Firm Sorting and Spatial Wage Inequality: Robustness

	Baseline	Measurement Choices			Model Assumptions	
		Within Industry	Only Manufac.	FDZ Data	Selection	Imperfect Mobility
90-10 Log Wage Difference	40 %	48 %	52 %	66 %	57 %	39 %
Spatial Wag Premium	15 %	26 %	22 %	34 %	22 %	16 %

Notes: This table reports the contribution of firm sorting to the spatial gap in the 90-10 difference of log wages within local labor markets (row 1) and average local wages (row 2) for various measurement choices and specifications of our model. For both measures, we compare 'poor' (bottom quartile of CZs in terms of GDPpc) and 'rich' (top quartile) locations.

9 Conclusion

Why do high-wage locations feature more wage dispersion and lower labor shares? In this paper, we propose a unifying theory based on spatial firm sorting that rationalizes these novel facts. We develop a parsimonious model, in which heterogeneous firms first decide where to produce, and then compete in a frictional local labor market that features on-the-job search. Because of search frictions, firms have monopsony power: They internalize that they poach more workers if they post higher wages. This leads to a local wage ladder with more productive firms paying

more. At the same time, due to monopsony power, firms pay wages below their marginal product and differ in their wage markdowns. Importantly, the extent of local labor market power and the shape of the local wage ladder depend on the local distribution of firm productivity, and therefore on how firms sort across space.

We show analytically that positive firm sorting—i.e., an allocation in which productive firms settle in productive locations—emerges as the unique equilibrium if firm and location productivity are sufficient complements or labor market frictions are sufficiently large. Our theory implies that positive sorting steepens the local wage ladder, raises the returns from an employment-to-employment transition, and concentrates employment among top, high-wage firms in productive locations. Positive sorting therefore amplifies spatial disparities not only in average wages but, more notably, in wage dispersion. At the same time, this concentration of highly productive firms—who have a high degree of monopsony power and low labor shares—leads to a low average labor share in their location.

Using administrative data from Germany, we first provide empirical support for these predictions. We then estimate our model to identify the degree of firm sorting in the data and quantify its role in shaping local labor market power as well as local wage distributions. Our estimated model shows that firm sorting is quantitatively important: Firm sorting can explain the lower local labor shares in high-wage locations, and it accounts for 40% of their increased wage dispersion. In spatial firm sorting, we thus highlight a novel source of disparities in local labor market outcomes.

References

- Ahlfeldt, G. M., S. Heblich, and T. Seidel: 2022, ‘Micro-Geographic Property Price and Rent Indices’. *Regional Science and Urban Economics*.
- Allen, T. and D. Donaldson: 2020, ‘Persistence and Path Dependence in the Spatial Economy’. NBER Working Paper 28059.
- Atkeson, A. and A. Burstein: 2008, ‘Pricing-to-Market, Trade Costs, and International Relative Prices’. *American Economic Review* **98**(5), 1998–2031.
- Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen: 2020, ‘The Fall of the Labor Share and the Rise of Superstar Firms’. *The Quarterly Journal of Economics* **135**(2), 645–709.
- Bagger, J. and R. Lentz: 2019, ‘An Empirical Model of Wage Dispersion with Sorting’. *Review of Economic Studies* **86**(1), 153–190.
- Bamford, I.: 2021, ‘Monopsony Power, Spatial Equilibrium, and Minimum Wages’. Working Paper.
- Baum-Snow, N. and R. Pavan: 2011, ‘Understanding the City Size Wage Gap’. *The Review of Economic Studies* **79**(1), 88–127.
- Behrens, K., G. Duranton, and F. Robert-Nicoud: 2014, ‘Productive Cities: Sorting, Selection, and Agglomeration’. *Journal of Political Economy* **122**(3), 507–553.

- Berger, D., K. Herkenhoff, and S. Mongey: 2022a, ‘Labor Market Power’. *American Economic Review* **112**(4), 1147–93.
- Berger, D. W., K. F. Herkenhoff, A. R. Kostøl, and S. Mongey: 2023, ‘An Anatomy of Monopsony: Search Frictions, Amenities, and Bargaining in Concentrated Markets’. In: *NBER Macroeconomics Annual 2023, volume 38*, NBER Chapters. National Bureau of Economic Research, Inc.
- Berger, D. W., K. F. Herkenhoff, and S. Mongey: 2022b, ‘Minimum wages, efficiency and welfare’. Technical report, National Bureau of Economic Research.
- Bilal, A.: 2022, ‘The Geography of Unemployment’. NBER Working Paper 29269.
- Bonhomme, S., K. Holzheu, T. Lamadon, E. Manresa, M. Mogstad, and B. Setzler: 2022, ‘How Much Should we Trust Estimates of Firm Effects and Worker Sorting?’. *Journal of Labor Economics*.
- Bontemps, C., J.-M. Robin, and G. J. Van Den Berg: 2000, ‘Equilibrium Search with Continuous Productivity Dispersion: Theory and Nonparametric Estimation’. *International Economic Review* **41**(2), 305–358.
- Bruns, B.: 2019, ‘Changes in Workplace Heterogeneity and How They Widen the Gender Wage Gap’. *American Economic Journal: Applied Economics* **11**(2), 74–113.
- Burdett, K. and D. T. Mortensen: 1998, ‘Wage Differentials, Employer Size, and Unemployment’. *International Economic Review* **39**(2), 257–273.
- Cahuc, P., F. Postel-Vinay, and J.-M. Robin: 2006, ‘Wage Bargaining with On-the-Job Search: Theory and Evidence’. *Econometrica* **74**(2), 323–364.
- Card, D., J. Heining, and P. Kline: 2013, ‘Workplace Heterogeneity and the Rise of West German Wage Inequality’. *The Quarterly Journal of Economics* **128**(3), 967–1015.
- Card, D., J. Rothstein, and M. Yi: 2023, ‘Location, Location, Location’. Working Paper.
- Combes, P.-P., G. Duranton, and L. Gobillon: 2008, ‘Spatial Wage Disparities: Sorting Matters!’. *Journal of Urban Economics* **63**(2), 723–742.
- Combes, P.-P., G. Duranton, L. Gobillon, D. Puga, and S. Roux: 2012, ‘The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection’. *Econometrica* **80**(6), 2543–2594.
- Dauth, W., S. Findeisen, E. Moretti, and J. Suedekum: 2022, ‘Matching in Cities’. *Journal of the European Economic Association*.
- De La Roca, J. and D. Puga: 2017, ‘Learning by Working in Big Cities’. *The Review of Economic Studies* **84**(1 (298)), 106–142.
- Diamond, P. A.: 1971, ‘A Model of Price Adjustment’. *Journal of Economic Theory* **3**(2), 156–168.
- Diamond, R.: 2016, ‘The Determinants and Welfare Implications of US Workers’ Diverging Location Choices by Skill: 1980–2000’. *American Economic Review* **106**(3), 479–524.
- Duranton, G. and D. Puga: 2004, ‘Chapter 48 - Micro-Foundations of Urban Agglomeration Economies’. In: J. V. Henderson and J.-F. Thisse (eds.): *Cities and Geography*, Vol. 4 of *Handbook of Regional and Urban Economics*. Elsevier, pp. 2063–2117.
- Eeckhout, J., R. Pinheiro, and K. Schmidheiny: 2014, ‘Spatial Sorting’. *Journal of Political Economy* **122**(3), 554–620.
- Engbom, N.: 2021, ‘Labor Market Fluidity and Human Capital Accumulation’. Working Paper.
- Gaubert, C.: 2018, ‘Firm Sorting and Agglomeration’. *American Economic Review* **108**(11), 3117–53.
- Glaeser, E. L. and D. C. Maré: 2001, ‘Cities and Skills’. *Journal of Labor Economics* **19**(2), 316–342.
- Gouin-Bonenfant, É.: 2022, ‘Productivity Dispersion, Between-Firm Competition, and the Labor Share’. *Econometrica* **90**(6), 2755–2793.

- Gould, E. D.: 2007, ‘Cities, Workers, and Wages: A Structural Analysis of the Urban Wage Premium’. *The Review of Economic Studies* **74**(2), 477–506.
- Heise, S. and T. Porzio: 2022, ‘Labor Misallocation Across Firms and Regions’. NBER Working Paper 30298.
- Hirsch, B., E. J. Jahn, A. Manning, and M. Oberfichtner: 2020, ‘The Urban Wage Premium in Imperfect Labor Markets’. *Journal of Human Resources*.
- Jäger, S., S. Noy, and B. Schoefer: 2022, ‘The German Model of Industrial Relations: Balancing Flexibility and Collective Action’. *Journal of Economic Perspectives* **36**(4), 53–80.
- Jarosch, G., J. S. Nimczik, and I. Sorkin: 2024, ‘Granular search, market structure, and wages’. *Review of Economic Studies* p. rdae004.
- Karahan, F., R. Michaels, B. Pugsley, A. Şahin, and R. Schuh: 2017, ‘Do Job-to-Job Transitions Drive Wage Fluctuations over the Business Cycle?’. *American Economic Review* **107**(5), 353–57.
- Langella, M. and A. Manning: 2021, “‘Marshall Lecture 2020: The Measure of Monopsony’”. *Journal of the European Economic Association* **19**(6), 2929–2957.
- Lhuillier, H.: 2025, ‘The local root of wage inequality’. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2025-78).
- Lochner, B. and B. Schulz: 2022, ‘Firm Productivity, Wages, and Sorting’. *Journal of Labor Economics* **forthcoming**.
- Manning, A.: 2003, *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton University Press.
- Manning, A.: 2011, ‘Chapter 11 - Imperfect Competition in the Labor Market’. Vol. 4 of *Handbook of Labor Economics*. Elsevier, pp. 973–1041.
- Martellini, P.: 2022, ‘Local Labor Markets and Aggregate Productivity’. Working Paper.
- Milgrom, P. and C. Shannon: 1994, ‘Monotone Comparative Statics’. *Econometrica* **62**(1), 157–180.
- Moretti, E.: 2011, ‘Local Labor Markets’. In: *Handbook of labor economics*, Vol. 4. Elsevier, pp. 1237–1313.
- Moscarini, G. and F. Postel-Vinay: 2018, ‘The Cyclical Job Ladder’. *Annual Review of Economics* **10**, 165–188.
- Nelson, K., D. Fredriksson, T. Korpi, W. Korpi, J. Palme, and O. Sjöberg: 2020, ‘The Social Policy Indicators (SPIN) Database’. *International Journal of Social Welfare* **29**(3), 285–289.
- Postel-Vinay, F. and J.-M. Robin: 2002, ‘Equilibrium Wage Dispersion with Worker and Employer Heterogeneity’. *Econometrica* **70**(6), 2295–2350.
- Ruf, K., L. Schmidtlein, S. Seth, H. Stüber, and M. Umkehrer: 2021a, ‘Linked Employer-Employee Data from the IAB: LIAB Longitudinal Model (LIAB LM) 1975 – 2019.’. FDZ-Datenreport, 06/2021 (en), Nuremberg. DOI: 10.5164/IAB.FDZD.2106.en.v1.
- Ruf, K., L. Schmidtlein, S. Seth, H. Stüber, M. Umkehrer, S. Griefemer, and S. Kaimer: 2021b, ‘Linked-Employer-Employee-Data of the IAB (LIAB): LIAB longitudinal model 1975-2019, version 1’. Research Data Centre of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB). DOI: 10.5164/IAB.LIABLM7519.de.en.v1.
- Schmutz, B. and M. Sidibé: 2018, ‘Frictional Labour Mobility’. *The Review of Economic Studies* **86**(4), 1779–1826.
- Villani, C.: 2009, *Optimal Transport: Old and New*, Grundlehren der Mathematischen Wissenschaften. Springer London, Limited.
- Yeh, C., C. Macaluso, and B. Hershbein: 2022, ‘Monopsony in the US Labor Market’. *American Economic Review* **112**(7), 2099–2138.

Appendix

A Baseline Model: Proofs

Mostly for expositional reasons, we maintain the assumption $\rho = 0$ (as stated in footnote 8).

A.1 Proof of Proposition 1

We proceed in three steps. In Step 1, we show that under the premise of the proposition, an assignment with positive sorting is among the optimal ones. In Step 2, we show that any pure assignment that is not increasing cannot satisfy firm optimality. In Step 3, we show that given our assumptions on firm productivity and land distributions, there exists a unique optimal assignment that is strictly increasing and thus satisfies positive sorting. Taken together, any equilibrium features positive sorting.

Step 1. *An assignment with positive sorting satisfies firms' optimality.* To this end, we show that, if firms conjecture that there is positive sorting, then value $\bar{J}(p, \ell)$ is indeed globally supermodular in (p, ℓ) under the premise, inducing positive sorting based on firms' optimal choices.

We use $\tilde{J}(y, \ell)$ from (4) to obtain $\bar{J}(p, \ell)$ as:

$$\begin{aligned}\bar{J}(p, \ell) &= \int \tilde{J}(y, \ell) d\Gamma(y|p) - k(\ell) \\ &= \int_{\underline{y}}^{\bar{y}} \frac{\partial \tilde{J}(y, \ell)}{\partial y} (1 - \Gamma(y|p)) - k(\ell) \\ &= \int_{\underline{y}}^{\bar{y}} \frac{\partial z(y, A(\ell))}{\partial y} l(y, \ell) (1 - \Gamma(y|p)) dy - k(\ell),\end{aligned}\tag{A.1}$$

where the second line uses integration by parts and the third line the Envelope Theorem.

To assess the conditions under which $\bar{J}(p, \ell)$ is supermodular in (p, ℓ) , which is sufficient for the single-crossing property of $\bar{J}(p, \ell)$ in (p, ℓ) , we differentiate wrt (p, ℓ) :

$$\frac{\partial^2 \bar{J}(p, \ell)}{\partial p \partial \ell} = \int_{\underline{y}}^{\bar{y}} \left(\frac{\partial \ln \left(\frac{\partial z(y, A(\ell))}{\partial y} \right)}{\partial \ell} + \frac{\partial \ln l(y, \ell)}{\partial \ell} \right) \frac{\partial z(y, A(\ell))}{\partial y} l(y, \ell) \left(-\frac{\partial \Gamma(y|p)}{\partial p} \right) dy. \tag{A.2}$$

In order for this expression to be (strictly) positive, it suffices that the integrand is positive for all $y \in [\underline{y}, \bar{y}]$ and strictly so for some set of y of positive measure. In turn, for this it is sufficient

that for all (y, ℓ) (recall that we assume $\frac{\partial \Gamma(y|p)}{\partial p} < 0$ for all $y \in (\underline{y}, \bar{y})$):

$$\frac{\frac{\partial^2 z(y, A(\ell))}{\partial y \partial A(\ell)} \frac{\partial A(\ell)}{\partial \ell}}{\frac{\partial z(y, A(\ell))}{\partial y}} > \frac{2\lambda^E}{\delta + \lambda^E(1 - \Gamma_\ell(y))} \left(-\frac{\partial \Gamma_\ell(y)}{\partial \ell} \right),$$

where we used the expression for firm size $l(y, \ell)$ in (3).

Under (the conjecture of) positive sorting, land market clearing is given by $Q(\mu(\ell)) = R(\ell)$; see Appendix SA.1.5. Further, based on our assumption of strictly positive densities (r, q) , this PAM assignment is one-to-one, i.e., μ is strictly increasing, whereby the firm type p assigned to location ℓ is $p = \mu(\ell) = Q^{-1}(R(\ell))$. Positive sorting then implies that the endogenous firm distribution in location ℓ is given by $\Gamma_\ell(y) = \Gamma(y|Q^{-1}(R(\ell)))$ (see also Footnote 12). Hence, $\frac{\partial \Gamma_\ell(y)}{\partial \ell} = \frac{\partial \Gamma(y|Q^{-1}(R(\ell)))}{\partial p} \frac{r(\ell)}{q(\mu(\ell))}$ and so to guarantee supermodularity of $\bar{J}(p, \ell)$ in (p, ℓ) , we need to ensure that

$$\frac{\frac{\partial^2 z(y, A(\ell))}{\partial y \partial A(\ell)} \frac{\partial A(\ell)}{\partial \ell}}{\frac{\partial z(y, A(\ell))}{\partial y}} > \frac{2\lambda^E}{\delta + \lambda^E(1 - \Gamma(y|Q^{-1}(R(\ell))))} \left(-\frac{\partial \Gamma(y|Q^{-1}(R(\ell)))}{\partial p} \right) \frac{r(\ell)}{q(Q^{-1}(R(\ell)))}, \quad (\text{A.3})$$

which is a condition in terms of primitives. To specify bounds that make this condition hold uniformly in (ℓ, y) , let

$$\varepsilon^P \equiv \min_{\ell, y} \frac{\frac{\partial^2 z(y, A(\ell))}{\partial y \partial A(\ell)} \frac{\partial A(\ell)}{\partial \ell}}{\frac{\partial z(y, A(\ell))}{\partial y}} \quad \text{and} \quad t^P \equiv \max_{\ell, y} \left(-\frac{\partial \Gamma(y|Q^{-1}(R(\ell)))}{\partial p} \right) \frac{r(\ell)}{q(Q^{-1}(R(\ell)))}.$$

Note that under our assumptions and the premise of the proposition, ε^P exists: It is strictly positive and bounded. In turn, t^P exists (and it is also strictly positive and bounded) since we assume that $\Gamma(y|p)$ is continuously differentiable in p where both p and y are defined over compact sets, and that cdf's Q and R are continuously differentiable on the intervals $[\underline{p}, \bar{p}]$ and $[\underline{\ell}, \bar{\ell}]$, respectively, with strictly positive densities (q, r) .

A sufficient condition for \bar{J} to be supermodular in (ℓ, p) is therefore $\varepsilon^P > 2\varphi^E t^P$. So, PAM is optimal, either if the productivity gains, ε^P , are sufficiently large or if φ^E is sufficiently small.

Step 2. *Any pure assignment that is not increasing cannot satisfy firm optimality.* To show this, we proceed by contradiction. That is, suppose there exists $\{\ell', \ell''\} \in [\underline{\ell}, \bar{\ell}]$ with $\ell' < \ell''$ such that the assigned firm types satisfy $\mu(\ell') > \mu(\ell'')$, where $\mu(\ell') = p''$, $\mu(\ell'') = p'$ and $p' < p''$, i.e., there exists a negatively assorted pair of agents. For this negatively assorted pair to be consistent with the optimal choices of firms p' and p'' , it must be that if firms conjecture this assignment then their objective function $\bar{J}(p, \ell) - k(\ell)$ satisfies decreasing differences when evaluated at this

pair, i.e., $\bar{J}(p', \ell') + \bar{J}(p'', \ell'') < \bar{J}(p', \ell'') + \bar{J}(p'', \ell')$. Otherwise there exists a blocking pair, which renders the assignment non-optimal. This requirement of decreasing differences translates into the requirement that the following expression, based on (A.1), be positive:

$$\int_{\underline{y}}^{\bar{y}} \underbrace{\left(\frac{\frac{\partial z(y, A(\ell''))}{\partial y}}{[\delta + \lambda^E(1 - \Gamma_{\ell''}(y))]^2} - \frac{\frac{\partial z(y, A(\ell'))}{\partial y}}{[\delta + \lambda^E(1 - \Gamma_{\ell'}(y))]^2} \right)}_{(1)} \underbrace{(1 - \Gamma(y|p') - (1 - \Gamma(y|p'')))}_{(2)} dy.$$

Note that the first term of the integrand (1) is strictly positive since: (i) $\frac{\partial z(y, A(\ell''))}{\partial y} > \frac{\partial z(y, A(\ell'))}{\partial y}$ due to strict monotonicity of A and strict supermodularity of z in (y, A) ; and moreover, (ii) $[\delta + \lambda^E(1 - \Gamma_{\ell''}(y))]^2 < [\delta + \lambda^E(1 - \Gamma_{\ell'}(y))]^2$ based on the conjecture of negative sorting, which implies that the firm productivity distribution in ℓ'' is stochastically worse than in ℓ' . In turn, the second term (2) is negative under our assumption on Γ . Thus, the integral is *negative*, i.e., $\bar{J}(p, \ell)$ satisfies *increasing* differences when evaluated at this pair of agents, which implies that negative sorting violates firm optimality: Any optimal assignment must be strictly increasing.

Step 3. *There is a unique assignment with positive sorting, given by $\mu(\ell) = Q^{-1}(R(\ell))$.* We conclude from Step 2 that, under the premise, the only conjecture about sorting that is consistent with firm optimality is that of positive sorting; and we have shown in Step 1 that in this case the optimal assignment will indeed be positive.

Finally, there is a unique way of matching up two one-dimensional, continuously differentiable distributions (Q, R) in a positively assortative way that preserves the measures, given by $Q(\mu(\ell)) = R(\ell) \Leftrightarrow \mu(\ell) = Q^{-1}(R(\ell))$. We conclude that under the premise, *any* equilibrium satisfies positive sorting with strictly increasing matching function, $\mu'(\ell) = r(\ell)/q(\mu(\ell)) > 0$. \square

Remark 1. In Step 2 of the proof of Proposition 1, the only deviation from positive sorting we consider is that of *strictly* negative sorting for a set of agents. Note that since we restrict attention to pure assignments, we do not need to consider the case of weakly negative sorting.

Remark 2. Clearly, a weaker sufficient condition for (A.2) to be positive would be based on the entire integral, which is different from our approach in the proof of Proposition 1 in which we aim for (more demanding) sufficient conditions that the integrand be positive *for all* y . Working with the integrand proved tractable, while working with the integral did not: First, none of the standard integral inequalities, which could have helped signing (A.2), apply; second, even when making functional form assumptions, the integral does not admit an explicit solution.

We give the detailed interpretation of the condition for positive sorting on the integrand (A.3) in the text, after stating condition (13). We note that the RHS of (A.3) is equivalent to (the negative of) the firm size elasticity in ℓ in (13) (see also (14)).

A.2 Proof of Proposition 2

We first show that a fixed point in Γ_ℓ exists and we will do so by construction. Suppose the conditions of Proposition 1 hold, i.e., any equilibrium features PAM of firms to locations.

We therefore consider the assignment $\mu(\ell) = Q^{-1}(R(\ell))$, which yields a unique firm distribution across locations $\Gamma_\ell = \Gamma(y|\mu(\ell))$ and a unique wage function $w(y, \ell)$ in (5). We will show that the pair (μ, k) is a Walrasian equilibrium of the land market, where

$$k(\ell) = \bar{k} + \delta\lambda^F \int_{\underline{\ell}}^{\ell} \int_{\underline{y}}^{\bar{y}} \frac{\partial \left(\frac{\frac{\partial z(y, A(\hat{\ell}))}{\partial y}}{[\delta + \lambda(1 - \Gamma_{\hat{\ell}}(y))]^2} \right)}{\partial \ell} (1 - \Gamma(y|\mu(\hat{\ell}))) dy d\hat{\ell}$$

is the land price schedule supporting assignment μ ; see Appendix SA.1.4 for the derivation.

By construction, μ clears the land market. To see that it is also *globally* optimal, we analyze firms' optimal behavior. Consider a firm with attribute p . It solves (6), i.e.,

$$\max_{\ell} \bar{J}(p, \ell) = \delta\lambda^F \int_{\underline{y}}^{\bar{y}} \frac{\frac{\partial z(y, A(\ell))}{\partial y}}{[\delta + \lambda(1 - \Gamma_{\ell}(y))]^2} (1 - \Gamma(y|p)) dy - \delta\lambda^F \int_{\underline{\ell}}^{\ell} \int_{\underline{y}}^{\bar{y}} \frac{\partial \left(\frac{\frac{\partial z(y, A(\hat{\ell}))}{\partial y}}{[\delta + \lambda(1 - \Gamma_{\hat{\ell}}(y))]^2} \right)}{\partial \ell} (1 - \Gamma(y|\mu(\hat{\ell}))) dy d\hat{\ell} - \bar{k}.$$

To reduce notation, we define

$$\mathcal{J}(p, \ell) := \delta\lambda^F \int_{\underline{y}}^{\bar{y}} \frac{\frac{\partial z(y, A(\ell))}{\partial y}}{[\delta + \lambda(1 - \Gamma_{\ell}(y))]^2} (1 - \Gamma(y|p)) dy,$$

which is supermodular in (p, ℓ) under the conditions specified in Proposition 1. Firm p thus solves

$$\max_{\ell} \mathcal{J}(p, \ell) - \int_{\underline{\ell}}^{\ell} \frac{\partial \mathcal{J}(\mu(\hat{\ell}), \hat{\ell})}{\partial \ell} d\hat{\ell} - \bar{k},$$

with solution $p = \mu(\ell)$. To show that $\mu(\ell)$ is a *global* optimum, note that for any $\ell' < \ell$

$$\mathcal{J}(p, \ell) - \int_{\underline{\ell}}^{\ell} \frac{\partial \mathcal{J}(\mu(\hat{\ell}), \hat{\ell})}{\partial \ell} d\hat{\ell} \geq \mathcal{J}(p, \ell') - \int_{\underline{\ell}}^{\ell'} \frac{\partial \mathcal{J}(\mu(\hat{\ell}), \hat{\ell})}{\partial \ell} d\hat{\ell}$$

if and only if

$$\mathcal{J}(p, \ell) - \mathcal{J}(p, \ell') \geq \int_{\ell'}^{\ell} \frac{\partial \mathcal{J}(\mu(\hat{\ell}), \hat{\ell})}{\partial \ell} d\hat{\ell}. \quad (\text{A.4})$$

Since $p = \mu(\ell)$ and since $\mathcal{J}(p, \ell) - \mathcal{J}(p, \ell') = \int_{\ell'}^{\ell} \frac{\partial \mathcal{J}(\mu(\hat{\ell}), \hat{\ell})}{\partial \ell} d\hat{\ell}$, it follows that (A.4) is equivalent to

$$\int_{\ell'}^{\ell} \frac{\partial \mathcal{J}(\mu(\ell), \hat{\ell})}{\partial \ell} d\hat{\ell} \geq \int_{\ell'}^{\ell} \frac{\partial \mathcal{J}(\mu(\hat{\ell}), \hat{\ell})}{\partial \ell} d\hat{\ell}$$

and this holds due to the (strict) supermodularity of $\mathcal{J}(p, \ell)$ and $\mu(\ell) \geq \mu(\hat{\ell})$ for all $\hat{\ell} \in [\ell', \ell]$. Moreover, it holds strictly if $\hat{\ell} \neq \ell$. Hence, firm p strictly prefers ℓ over $\hat{\ell} < \ell$. A similar argument holds for $\hat{\ell} > \ell$, and hence choosing ℓ is the unique global optimum for p . Since p was arbitrary, all firm types behave globally optimally. We have shown that the optimal μ (and thus Γ_{ℓ}) coincides with the postulated μ (and thus Γ_{ℓ}) from above, i.e., we have constructed an equilibrium. Note that all land is occupied, and that, for each ℓ , land (owner) ℓ obtains $k(\ell) \geq 0$.

To see that this equilibrium is unique, we first note that by Proposition 1, there exists a unique optimal assignment μ , which is deterministic (and satisfies PAM). Second, the uniqueness of $k(\ell)$ (up to a constant of integration) then follows from Remarks 10.29 and 10.30 in Villani (2009). \square

A.3 Proof of Proposition 3

Preliminaries. Denote the firm-level labor share by $Ls(y, \ell) := w(y, \ell)/z(y, A(\ell))$ and let the value-added-weighted employment density be given by $\tilde{g}_{\ell}(y) := \frac{z(y, A(\ell))g_{\ell}(y)}{\int_{\underline{y}}^{\bar{y}} z(y', A(\ell))g_{\ell}(y')dy'}$, with corresponding cdf $\tilde{G}_{\ell}(y)$. The local labor share in each ℓ is then

$$LS(\ell) = \frac{\int_{\underline{y}}^{\bar{y}} w(y, \ell)g_{\ell}(y)dy}{\int_{\underline{y}}^{\bar{y}} z(y, A(\ell))g_{\ell}(y)dy} = \int_{\underline{y}}^{\bar{y}} \frac{w(y, \ell)}{z(y, A(\ell))} \frac{z(y, A(\ell))g_{\ell}(y)}{\int_{\underline{y}}^{\bar{y}} z(y', A(\ell))g_{\ell}(y')dy'} dy = \int_{\underline{y}}^{\bar{y}} Ls(y, \ell)\tilde{g}_{\ell}(y)dy.$$

Thus, for the local labor share to be decreasing in ℓ , the following must hold:

$$\frac{\partial LS(\ell)}{\partial \ell} = \frac{\partial Ls(\underline{y}, \ell)}{\partial \ell} + \int_{\underline{y}}^{\bar{y}} \frac{\partial Ls(y, \ell)}{\partial y} \left(-\frac{\partial \tilde{G}_{\ell}(y)}{\partial \ell} \right) + \frac{\partial^2 Ls(y, \ell)}{\partial y \partial \ell} (1 - \tilde{G}_{\ell}(y)) dy < 0, \quad (\text{A.5})$$

which uses integration by parts and equals (15) since the first term is zero, $\partial Ls(\underline{y}, \ell)/\partial \ell = 0$.

The proof proceeds in three steps. First, we show that for $\varphi^E \rightarrow 0$ and z sufficiently complementary in (ℓ, y) , the firm-level labor share is submodular in (ℓ, y) , which renders the third term in (A.5) negative. Second, we show that a sufficiently decreasing $\gamma_{\ell}(\cdot) = \gamma(\cdot|\mu(\ell))$ guarantees that the firm-level labor share is downward sloping. Third, we show that PAM and $\varphi^E \rightarrow 0$ imply that \tilde{G}_{ℓ} is shifted by ℓ in the FOSD sense. Steps 2 and 3 ensure that the second term in (A.5) is

negative. Thus, under PAM and the specified conditions on z , φ^E and γ , $LS(\cdot)$ is decreasing in ℓ . We now provide the details.

Step 1. We focus on the third term in (A.5) and show that $LS(y, \ell)$ is submodular in (y, ℓ) under the stated conditions. First, we spell out the condition we have to sign:

$$\begin{aligned} \frac{\partial^2 LS(y, \ell)}{\partial y \partial \ell} &\stackrel{s}{=} z^2 \left(\frac{\partial^2 w}{\partial y \partial \ell} z + \frac{\partial w}{\partial y} \frac{\partial z}{\partial \ell} - \frac{\partial w}{\partial \ell} \frac{\partial z}{\partial y} - w \frac{\partial^2 z}{\partial y \partial \ell} \right) - 2z \frac{\partial z}{\partial \ell} \left(\frac{\partial w}{\partial y} z - w \frac{\partial z}{\partial y} \right) \\ &= -zw \left(\frac{\partial^2 z}{\partial y \partial \ell} z - 2 \frac{\partial z}{\partial y} \frac{\partial z}{\partial \ell} \right) + \frac{\partial^2 w}{\partial y \partial \ell} z^3 - z^2 \frac{\partial w}{\partial y} \frac{\partial z}{\partial \ell} - z^2 \frac{\partial w}{\partial \ell} \frac{\partial z}{\partial y} \end{aligned}$$

Note that, using (5), for large labor market frictions ($\varphi^E \rightarrow 0$):

$$\left. \frac{\partial^2 w}{\partial y \partial \ell} \right|_{\varphi^E=0} = 0; \quad \left. \frac{\partial w}{\partial y} \right|_{\varphi^E=0} = 0; \quad \left. \frac{\partial w}{\partial \ell} \right|_{\varphi^E=0} = \frac{\partial z(y, A(\ell))}{\partial \ell}; \quad \left. w \right|_{\varphi^E=0} = z(y, A(\ell)).$$

Hence, if z is sufficiently log-supermodular, i.e., $\frac{\partial^2 z}{\partial y \partial \ell} z / \frac{\partial z}{\partial y} \frac{\partial z}{\partial \ell} > 2$, then

$$\left. \frac{\partial^2 LS(y, \ell)}{\partial y \partial \ell} \right|_{\varphi^E=0} = -z(y, A(\ell))z \left(\frac{\partial^2 z}{\partial y \partial \ell} z - 2 \frac{\partial z}{\partial y} \frac{\partial z}{\partial \ell} \right) - z^2 \frac{\partial z(y, A(\ell))}{\partial \ell} \frac{\partial z}{\partial y} < 0,$$

where, if omitted, the arguments of z are $(y, A(\ell))$.

Step 2. As for the second term in (A.5), we show that under the premise, firm-level labor share,

$$LS(y, \ell) = \frac{w(y, \ell)}{z(y, A(\ell))} = 1 - \frac{[1 + \varphi^E(1 - \Gamma_\ell(y))]^2 \int_y^y \frac{\frac{\partial z(t, A(\ell))}{\partial y}}{[1 + \varphi^E(1 - \Gamma_\ell(t))]^2} dt}{z(y, A(\ell))},$$

is *decreasing* in firm productivity y in each location ℓ . Differentiation and some algebra yield:

$$\frac{\partial LS(y, \ell)}{\partial y} = (1 - LS(y, \ell)) \frac{2\varphi^E \gamma_\ell(y)}{1 + \varphi^E(1 - \Gamma_\ell(y))} - LS(y, \ell) \frac{\frac{\partial z(y, A(\ell))}{\partial y}}{z(y, A(\ell))}. \quad (\text{A.6})$$

We will show that this expression is negative for sufficiently decreasing γ_ℓ . It suffices that

$$\underbrace{\frac{1 - LS(y, \ell)}{LS(y, \ell)} \frac{2\varphi^E \gamma_\ell(y)}{1 + \varphi^E(1 - \Gamma_\ell(y))}}_{:=LHS(y)} < \underbrace{\frac{\frac{\partial z(y, A(\ell))}{\partial y}}{z(y, A(\ell))}}_{:=RHS(y)}, \quad \forall (y, \ell).$$

Toward a contradiction, suppose that $\frac{\partial LS(y, \ell)}{\partial y} > 0$ for some $y \in [\underline{y}, \bar{y}]$, that is, (A.6) is strictly positive at y .

First, note that, at $y = \underline{y}$, (A.6) is negative, since the first term is zero while the second one

is strictly negative, or

$$\underbrace{\frac{1 - Ls(\underline{y}, \ell)}{Ls(\underline{y}, \ell)} \frac{2\varphi^E \gamma_\ell(\underline{y})}{1 + \varphi^E(1 - \Gamma_\ell(\underline{y}))}}_{=0} < \underbrace{\frac{\frac{\partial z(\underline{y}, A(\ell))}{\partial y}}{z(\underline{y}, A(\ell))}}_{=RHS(\underline{y})}.$$

Second, in order for $\frac{\partial Ls(\underline{y}, \ell)}{\partial y} > 0$ for some $y \in (\underline{y}, \bar{y})$ on a set $[y - \epsilon, y + \epsilon], \epsilon > 0$, it must be that $RHS(y)$ is *smaller* than $LHS(y)$, i.e., $RHS(\cdot)$ must cross $LHS(\cdot)$ at least once and the first crossing at point $\hat{y} \in (\underline{y}, \bar{y})$ is such that $RHS(\hat{y})$ crosses $LHS(\hat{y})$ *from above*. That is, there exists $\hat{y} \in (\underline{y}, \bar{y})$ such that the slope of $RHS(\hat{y})$ is smaller than the slope of $LHS(\hat{y})$. We now investigate these slopes.

The slope of $LHS(y)$ is given by

$$\begin{aligned} \frac{\partial LHS(y)}{\partial y} &= -\frac{\frac{\partial Ls(y, \ell)}{\partial y}}{Ls(y, \ell)^2} \frac{2\varphi^E \gamma_\ell(y)}{1 + \varphi^E(1 - \Gamma_\ell(y))} \\ &\quad + \frac{(1 - Ls(y, \ell))}{Ls(y, \ell)} \frac{2\varphi^E}{(1 + \varphi^E(1 - \Gamma_\ell(y)))^2} \left(\frac{\partial \gamma_\ell(y)}{\partial y} (1 + \varphi^E(1 - \Gamma_\ell(y))) + \varphi^E \gamma_\ell(y)^2 \right). \end{aligned} \quad (\text{A.7})$$

Next, the slope of $RHS(y)$ is given by

$$\frac{\partial RHS(y)}{\partial y} = \frac{\frac{\partial z(\underline{y}, A(\ell))}{\partial y}}{z(\underline{y}, A(\ell))}.$$

Consider the first crossing of $RHS(\cdot)$ and $LHS(\cdot)$, i.e., the point $\hat{y} \in (\underline{y}, \bar{y})$ at which $RHS(\cdot)$ crosses $LHS(\cdot)$. Noticing that the slope of $LHS(\cdot)$ at crossing $y = \hat{y}$ simplifies due to $\partial Ls(\hat{y}, \ell)/\partial y = 0$ (by construction, see (A.7)), the difference between the slopes on the LHS and the RHS at $y = \hat{y}$ is given by

$$\frac{\partial LHS(\hat{y})}{\partial y} - \frac{\partial RHS(\hat{y})}{\partial y} = \frac{\frac{\partial z(\hat{y}, A(\ell))}{\partial y}}{z(\hat{y}, A(\ell))} \left(\frac{\frac{\partial \gamma_\ell(\hat{y})}{\partial y}}{\gamma_\ell(\hat{y})} + \frac{\varphi^E \gamma_\ell(\hat{y})}{1 + \varphi^E(1 - \Gamma_\ell(\hat{y}))} \right) - \frac{\partial}{\partial y} \left(\frac{\frac{\partial z(\hat{y}, A(\ell))}{\partial y}}{z(\hat{y}, A(\ell))} \right), \quad (\text{A.8})$$

where we simplified $(1 - Ls(\hat{y}, \ell))/Ls(\hat{y}, \ell)$ using (A.6) when evaluated at $y = \hat{y}$.

Difference (A.8) is strictly negative at any $\hat{y} \in (\underline{y}, \bar{y})$ and $\ell \in [\underline{\ell}, \bar{\ell}]$ —so that the slope of $RHS(\cdot)$ is *larger* than the slope of $LHS(\cdot)$ at \hat{y} and thus $RHS(\cdot)$ cannot cross $LHS(\cdot)$ from above at that point—if for all (y, ℓ)

$$\frac{\frac{\partial \gamma_\ell(y)}{\partial y}}{\gamma_\ell(y)} < -\varphi^E \gamma_\ell(y) + \frac{\partial}{\partial y} \log \left(\frac{\frac{\partial z(y, A(\ell))}{\partial y}}{z(y, A(\ell))} \right). \quad (\text{A.9})$$

We will show that condition (A.9) holds if $\gamma_\ell(\cdot) = \gamma(\cdot|\mu(\ell))$ is sufficiently decreasing in y .

To do so, consider two cases:

First, assume that $\min_{y,\ell} \left(\frac{\partial}{\partial y} \log \left(\frac{\partial z(y,A(\ell))}{\partial y} / z(y,A(\ell)) \right) \right) \geq 0$, which holds, for example, when z is log-convex. Then, (A.9) holds for all (y,ℓ) if $\gamma_\ell(\cdot)$ is sufficiently decreasing in the sense that

$$\max_{y,\ell} \frac{\frac{\partial \gamma_\ell(y)}{\partial y}}{\gamma_\ell(y)^2} < -\varphi^E. \quad (\text{A.10})$$

Second, assume that $\min_{y,\ell} \left(\frac{\partial}{\partial y} \log \left(\frac{\partial z(y,A(\ell))}{\partial y} / z(y,A(\ell)) \right) \right) < 0$, a condition that is satisfied when z is log-concave. In this case, we consider two sub-cases:

1. Suppose that $\max_{y,\ell} \gamma_\ell(y) < 1$. In this case, (A.9) holds for all (y,ℓ) if $\gamma_\ell(\cdot)$ is sufficiently decreasing in the sense that

$$\max_{y,\ell} \frac{\frac{\partial \gamma_\ell(y)}{\partial y}}{\gamma_\ell(y)} < -\varphi^E + \min_{y,\ell} \left(\frac{\partial}{\partial y} \log \left(\frac{\frac{\partial z(y,A(\ell))}{\partial y}}{z(y,A(\ell))} \right) \right), \quad (\text{A.11})$$

since $-\varphi^E < -\varphi^E \max_{y,\ell} \gamma_\ell(y)$.

2. Suppose that $\max_{y,\ell} \gamma_\ell(y) \geq 1$. In this case, (A.9) holds for all (y,ℓ) if $\gamma_\ell(\cdot)$ is sufficiently decreasing in the sense that

$$\max_{y,\ell} \frac{\frac{\partial \gamma_\ell(y)}{\partial y}}{\gamma_\ell(y)^2} < -\varphi^E + \min_{y,\ell} \left(\frac{\partial}{\partial y} \log \left(\frac{\frac{\partial z(y,A(\ell))}{\partial y}}{z(y,A(\ell))} \right) \right). \quad (\text{A.12})$$

To see that this suffices for (A.9) to hold, observe that

$$\begin{aligned} \frac{\frac{\partial \gamma_\ell(y)}{\partial y}}{\gamma_\ell(y)} &\leq \frac{\max_{y,\ell} \gamma_\ell(y)}{\gamma_\ell(y)} \frac{\frac{\partial \gamma_\ell(y)}{\partial y}}{\gamma_\ell(y)} < \max_{y,\ell} \gamma_\ell(y) \left(-\varphi^E + \min_{y,\ell} \frac{\partial}{\partial y} \log \left(\frac{\frac{\partial z(y,A(\ell))}{\partial y}}{z(y,A(\ell))} \right) \right) \\ &\leq -\varphi^E \gamma_\ell(y) + \min_{y,\ell} \frac{\partial}{\partial y} \log \left(\frac{\frac{\partial z(y,A(\ell))}{\partial y}}{z(y,A(\ell))} \right) \leq -\varphi^E \gamma_\ell(y) + \frac{\partial}{\partial y} \log \left(\frac{\frac{\partial z(y,A(\ell))}{\partial y}}{z(y,A(\ell))} \right), \end{aligned}$$

where the second inequality follows from condition (A.12); the first and the third inequalities follow from the premise $\max_{y,\ell} \gamma_\ell(y) \geq 1$ in this case; and the fourth inequality follows from $\min_{y,\ell} \left(\frac{\partial}{\partial y} \log \left(\frac{\partial z(y,A(\ell))}{\partial y} / z(y,A(\ell)) \right) \right) \leq \left(\frac{\partial}{\partial y} \log \left(\frac{\partial z(y,A(\ell))}{\partial y} / z(y,A(\ell)) \right) \right)$.

Summarizing the different conditions (A.10)-(A.12), we have shown that if $\gamma_\ell(\cdot) = \gamma(\cdot|\mu(\ell))$

is sufficiently decreasing in y , i.e.,

$$\max \left\{ \max_{y,\ell} \frac{\frac{\partial \gamma_\ell(y)}{\partial y}}{\gamma_\ell(y)^2}, \max_{y,\ell} \frac{\frac{\partial \gamma_\ell(y)}{\partial y}}{\gamma_\ell(y)} \right\} < -\varphi^E + \max \left\{ 0, \min_{y,\ell} \left(\frac{\partial}{\partial y} \log \left(\frac{\frac{\partial z(y, A(\ell))}{\partial y}}{z(y, A(\ell))} \right) \right) \right\},$$

then (A.9) holds. This implies that $RHS(\cdot)$ cannot cross $LHS(\cdot)$ from above at \hat{y} —a contradiction. As a result, the specified assumption on $\gamma_\ell(\cdot)$ guarantees that the firm-level labor share, $Ls(\cdot, \ell)$, is decreasing in firm productivity y , for each ℓ .

Step 3. As for the second term in (A.5), we now show that sufficiently large labor market frictions and PAM guarantee that ℓ shifts \tilde{G}_ℓ in a FOSD sense.

First note that for any $\ell' < \ell''$ and $y < \bar{y}$, $\tilde{G}_{\ell''} \leq \tilde{G}_{\ell'}$ iff

$$\begin{aligned} & \frac{\int_{\underline{y}}^y g_{\ell''}(\tilde{y}) z(\tilde{y}, A(\ell'')) d\tilde{y}}{\int_{\underline{y}}^{\bar{y}} g_{\ell''}(\tilde{y}) z(\tilde{y}, A(\ell'')) d\tilde{y}} \leq \frac{\int_{\underline{y}}^y g_{\ell'}(\tilde{y}) z(\tilde{y}, A(\ell')) d\tilde{y}}{\int_{\underline{y}}^{\bar{y}} g_{\ell'}(\tilde{y}) z(\tilde{y}, A(\ell')) d\tilde{y}} \\ \Leftrightarrow & \int_{\underline{y}}^y g_{\ell''}(\tilde{y}) z(\tilde{y}, A(\ell'')) d\tilde{y} \int_{\underline{y}}^{\bar{y}} g_{\ell'}(\tilde{y}) z(\tilde{y}, A(\ell')) d\tilde{y} \leq \int_{\underline{y}}^y g_{\ell'}(\tilde{y}) z(\tilde{y}, A(\ell')) d\tilde{y} \int_{\underline{y}}^{\bar{y}} g_{\ell''}(\tilde{y}) z(\tilde{y}, A(\ell'')) d\tilde{y}, \end{aligned}$$

i.e., if $\int_{\underline{y}}^y g_\ell z(y, A(\ell))$ is log-supermodular in (y, ℓ) , which is guaranteed if both z and g_ℓ are log-supermodular. The former is assumed under the premise. To establish the latter, note that

$$g_\ell(y) = \frac{\gamma_\ell(y)(1 + \varphi^E)}{(1 + \varphi^E(1 - \Gamma_\ell(y)))^2}.$$

Cross-differentiating $\log(g_\ell)$ shows that g_ℓ is log-supermodular if for all (y, ℓ)

$$\begin{aligned} \frac{\partial^2 \log \gamma_\ell(y)}{\partial y \partial \ell} & \geq -\frac{2\varphi^E}{(1 + \varphi^E(1 - \Gamma_\ell(y)))^2} \left(\frac{\partial^2 \Gamma_\ell(y)}{\partial \ell \partial y} (1 + \varphi^E(1 - \Gamma_\ell(y))) + \varphi^E \frac{\partial \Gamma_\ell(y)}{\partial y} \frac{\partial \Gamma_\ell(y)}{\partial \ell} \right) \\ \frac{\partial^2 \log \gamma_\ell(y)}{\partial y \partial \ell} & \geq -\left(\frac{2\varphi^E}{(1 + \varphi^E(1 - \Gamma_\ell(y)))} \frac{\partial^2 \Gamma_\ell(y)}{\partial \ell \partial y} + \frac{2(\varphi^E)^2}{(1 + \varphi^E(1 - \Gamma_\ell(y)))^2} \frac{\partial \Gamma_\ell(y)}{\partial y} \frac{\partial \Gamma_\ell(y)}{\partial \ell} \right) \\ \frac{\partial^2 \log \gamma(y|p)}{\partial y \partial p} \mu'(\ell) & \geq -\left(\frac{2\varphi^E}{(1 + \varphi^E(1 - \Gamma_\ell(y)))} \frac{\partial^2 \Gamma(y|p)}{\partial p \partial y} \mu'(\ell) + \frac{2(\varphi^E)^2}{(1 + \varphi^E(1 - \Gamma_\ell(y)))^2} \frac{\partial \Gamma(y|p)}{\partial y} \frac{\partial \Gamma(y|p)}{\partial p} \mu'(\ell) \right) \\ \frac{\partial^2 \log \gamma(y|p)}{\partial y \partial p} & \geq -\left(\frac{2\varphi^E}{(1 + \varphi^E(1 - \Gamma_\ell(y)))} \frac{\partial^2 \Gamma(y|p)}{\partial p \partial y} + \frac{2(\varphi^E)^2}{(1 + \varphi^E(1 - \Gamma_\ell(y)))^2} \frac{\partial \Gamma(y|p)}{\partial y} \frac{\partial \Gamma(y|p)}{\partial p} \right), \end{aligned}$$

where the last inequality follows under PAM, $\mu'(\ell) > 0$. We want the following to hold for all (y, p) :

$$\frac{\partial^2 \log \gamma(y|p)}{\partial y \partial p} \geq 2\varphi^E \max_{y,p} \left(-\frac{\partial^2 \Gamma(y|p)}{\partial p \partial y} \right) + 2(\varphi^E)^2 \max_{y,p} \left(\frac{\partial \Gamma(y|p)}{\partial y} \left(-\frac{\partial \Gamma(y|p)}{\partial p} \right) \right), \quad (\text{A.13})$$

where the maxima on the RHS are positive and well-defined as we maximize continuous functions

over compact sets. The LHS is strictly positive for all (y, p) under our assumption that $\gamma(y|p)$ satisfies the strict monotone likelihood ratio property (i.e., that $\gamma(y|p)$ is log-supermodular in (y, p)). For (A.13) to hold, it therefore suffices that φ^E is small enough: The LHS is constant in φ^E and strictly positive; and the RHS is increasing in φ^E , starting at 0 and ending at ∞ . Then, by the Intermediate Value Theorem, there is an intersection of the two and so there exists a point $\hat{\varphi}^E$ such that for $\varphi^E < \hat{\varphi}^E$, (A.13) holds strictly. It follows that g_ℓ is log-supermodular, implying that ℓ shifts \tilde{G}_ℓ in the FOSD sense, i.e., $\partial \tilde{G}_\ell / \partial \ell \leq 0$, provided that φ^E is small enough. As a result of Step 2 and Step 3, the second term of (A.5) is negative for sufficiently small φ^E .

Thus, based on Steps 1, 2 and 3, the integrand of (A.5) is strictly negative under PAM for sufficiently small φ^E , sufficiently complementary z and sufficiently decreasing $\gamma(\cdot|\mu(\ell))$ with $p = \mu(\ell)$, rendering (A.5) negative. \square

A.4 Proof of Corollary 1

Under goods' market clearing, aggregate output in ℓ equals aggregate wages plus profits and land prices in equilibrium

$$\begin{aligned} \int_{\underline{y}}^{\bar{y}} z(y, A(\ell)) l(y, \ell) d\Gamma(y|\mu(\ell)) &= \int_{\underline{y}}^{\bar{y}} w(y, \ell) l(y, \ell) d\Gamma(y|\mu(\ell)) + \varphi^F \int_{\underline{y}}^{\bar{y}} \int_{\underline{y}}^y \frac{A(\ell)}{[1 + \varphi^E(1 - \Gamma(t|\mu(\ell)))]^2} dt d\Gamma(y|\mu(\ell)) \\ &= \int_{\underline{y}}^{\bar{y}} w(y, \ell) l(y, \ell) d\Gamma(y|\mu(\ell)) + \varphi^F \int_{\underline{y}}^{\bar{y}} \frac{A(\ell)}{[1 + \varphi^E(1 - \Gamma(y|\mu(\ell)))]^2} (1 - \Gamma(y|\mu(\ell))) dy, \end{aligned}$$

where we used integration by parts in the second line. Thus, the labor share is given by:

$$LS(\ell) := \frac{\int_{\underline{y}}^{\bar{y}} w(y, \ell) l(y, \ell) d\Gamma(y|\mu(\ell))}{\int_{\underline{y}}^{\bar{y}} z(y, A(\ell)) l(y, \ell) d\Gamma(y|\mu(\ell))} = 1 - \frac{\varphi^F \int_{\underline{y}}^{\bar{y}} \frac{A(\ell)}{[1 + \varphi^E(1 - \Gamma(y|\mu(\ell)))]^2} (1 - \Gamma(y|\mu(\ell))) dy}{\int_{\underline{y}}^{\bar{y}} z(y, A(\ell)) l(y, \ell) d\Gamma(y|\mu(\ell))}.$$

At the same time, aggregate output in the denominator can be expressed as follows, using that Γ is Pareto and firm size expression (3):

$$\begin{aligned} \int_{\underline{y}}^{\bar{y}} z(y, A(\ell)) l(y, \ell) d\Gamma(y|\mu(\ell)) &= \int_{\underline{y}}^{\bar{y}} A(\ell) y \cdot l(y, \ell) \frac{1}{\mu(\ell)} y^{-\frac{1}{\mu(\ell)} - 1} dy \\ &= \frac{1}{\mu(\ell)} \varphi^F \int_{\underline{y}}^{\bar{y}} \frac{A(\ell)}{[1 + \varphi^E(1 - \Gamma(y|\mu(\ell)))]^2} (1 - \Gamma(y|\mu(\ell))) dy \end{aligned}$$

Plugging aggregate output back into $LS(\ell)$ above, we obtain $LS(\ell) = 1 - \mu(\ell)$. Thus, $LS'(\ell) < 0$ if and only if $\mu'(\ell) > 0$. \square

A.5 Proof of Proposition 4

Consider a firm in ℓ with rank $\mathcal{R} := \Gamma_\ell(y) \in [0, 1]$. Its wage is given by:

$$w(y, \ell) = z(y, A(\ell)) - [1 + \varphi^E(1 - \Gamma_\ell(y))]^2 \int_{\underline{y}}^y \frac{\frac{\partial z(t, A(\ell))}{\partial y}}{[1 + \varphi^E(1 - \Gamma_\ell(t))]^2} dt, \quad \Gamma_\ell(t) = x, \gamma_\ell(t) dt = dx$$

$$w(\Gamma_\ell^{-1}(\mathcal{R}), \ell) = z(\Gamma_\ell^{-1}(\mathcal{R}), A(\ell)) - [1 + \varphi^E(1 - \mathcal{R})]^2 \int_0^{\mathcal{R}} \frac{\frac{\partial z(\Gamma_\ell^{-1}(x), A(\ell))}{\partial y}}{[1 + \varphi^E(1 - x)]^2} \frac{1}{\gamma_\ell(\Gamma_\ell^{-1}(x))} dx \quad (\text{A.14})$$

where we used a change of variable and we denote by $\gamma_\ell(\Gamma_\ell^{-1}(x))$ the pdf at the x th quantile. Also, for later use, let $\tilde{l}(\mathcal{R}) := g_\ell(\Gamma_\ell^{-1}(\mathcal{R})) / \gamma_\ell(\Gamma_\ell^{-1}(\mathcal{R})) = (1 + \varphi^E) / (1 + \varphi^E(1 - \Gamma_\ell(\Gamma_\ell^{-1}(\mathcal{R}))))^2$ be the firm size density at a firm with productivity rank $\mathcal{R} = \Gamma_\ell(y)$, which only depends on firm's local rank \mathcal{R} , not on ℓ .

First, using this notation, the wage as a function of local firm productivity rank \mathcal{R} is given by

$$w(\Gamma_\ell^{-1}(\mathcal{R}), \ell) = z(\Gamma_\ell^{-1}(\mathcal{R}), A(\ell)) - \int_0^{\mathcal{R}} \frac{\tilde{l}(x)}{\tilde{l}(\mathcal{R})} \frac{\partial z(\Gamma_\ell^{-1}(x), A(\ell))}{\partial \mathcal{R}} dx$$

$$= \frac{\tilde{l}(0)}{\tilde{l}(\mathcal{R})} z(\Gamma_\ell^{-1}(0), A(\ell)) + \int_0^{\mathcal{R}} \frac{\tilde{l}'(x)}{\tilde{l}(\mathcal{R})} z(\Gamma_\ell^{-1}(x), A(\ell)) dx$$

where we use integration by parts in the second line. We obtain the following relative wage between a firm with rank \mathcal{R} and the least productive firm

$$\frac{w(\Gamma_\ell^{-1}(\mathcal{R}), \ell)}{w(\Gamma_\ell^{-1}(0), \ell)} = \frac{\tilde{l}(0)}{\tilde{l}(\mathcal{R})} + \int_0^{\mathcal{R}} \frac{\tilde{l}'(x)}{\tilde{l}(\mathcal{R})} \frac{z(\Gamma_\ell^{-1}(x), A(\ell))}{z(\Gamma_\ell^{-1}(0), A(\ell))} dx.$$

Second, differentiating this wage ratio wrt ℓ gives,

$$\frac{\partial}{\partial \ell} \left(\frac{w(\Gamma_\ell^{-1}(\mathcal{R}), \ell)}{w(\Gamma_\ell^{-1}(0), \ell)} \right) = \int_0^{\mathcal{R}} \frac{\tilde{l}'(x)}{\tilde{l}(\mathcal{R})} \frac{\partial}{\partial \ell} \left(\frac{z(\Gamma_\ell^{-1}(x), A(\ell))}{z(\Gamma_\ell^{-1}(0), A(\ell))} \right) dx. \quad (\text{A.15})$$

To show that this expression is positive, observe that

$$\frac{\partial}{\partial \ell} \left(\frac{z(\Gamma_\ell^{-1}(x), A(\ell))}{z(\Gamma_\ell^{-1}(0), A(\ell))} \right) = \frac{\frac{\partial \Gamma_\ell^{-1}(x)}{\partial \ell} \frac{\partial z(\Gamma_\ell^{-1}(x), A(\ell))}{\partial y}}{z(\Gamma_\ell^{-1}(0), A(\ell))} + A'(\ell) \left(\frac{\frac{\partial z(\Gamma_\ell^{-1}(x), A(\ell))}{\partial A}}{z(\Gamma_\ell^{-1}(x), A(\ell))} - \frac{\frac{\partial z(\Gamma_\ell^{-1}(0), A(\ell))}{\partial A}}{z(\Gamma_\ell^{-1}(0), A(\ell))} \right) \frac{z(\Gamma_\ell^{-1}(x), A(\ell))}{z(\Gamma_\ell^{-1}(0), A(\ell))} > 0. \quad (\text{A.16})$$

The first term is positive under positive firm sorting since $\frac{\partial}{\partial \ell} \Gamma_\ell^{-1}(\mathcal{R}) = -\frac{\partial}{\partial \ell} \Gamma_\ell(y) / \gamma_\ell(y)$ (using the Implicit Function Theorem together with the fact given that under positive firm sorting $\frac{\partial}{\partial \ell} \Gamma_\ell(y) \leq 0$); and the second term is positive under weakly log-supermodular z in (y, A) . \square

Remark. Proposition 4 and its proof nest the multiplicative production function $z(y, A(\ell))$ discussed in Section 4.2. To see this, note that the second part of derivative (A.16) is zero under multiplicative z . Plugging only the first part of (A.16) into (A.15), we obtain—under a change of variable—equation (17) from Section 4.2 and thus, under multiplicative z , positive firm sorting is sufficient for wage inequality to be higher in productive locations.

A.6 Proof of Proposition 5

First, the denominator of the spatial wage premium, $\mathbb{E}[w(\Gamma_\ell^{-1}(\mathcal{R}), \ell) | \ell] / \mathbb{E}[w(\Gamma_{\underline{\ell}}^{-1}(\mathcal{R}), \underline{\ell}) | \underline{\ell}]$, is independent of ℓ , so we focus on how the numerator varies with location index ℓ .

Second, to derive $\mathbb{E}[w(\Gamma_\ell^{-1}(\mathcal{R}), \ell) | \ell]$ consider the average wage across firms in location ℓ :

$$\begin{aligned} \mathbb{E}[w(y, \ell) | \ell] &= \int_{\underline{y}}^{\bar{y}} w(y, \ell) g_\ell(y) dy \\ &= \int_{\underline{y}}^{\bar{y}} w(y, \ell) \frac{g_\ell(y)}{\gamma_\ell(y)} \gamma_\ell(y) dy \\ &= \int_0^1 \frac{w(\Gamma_\ell^{-1}(\mathcal{R}), \ell)}{w(\Gamma_\ell^{-1}(0), \ell)} w(\Gamma_\ell^{-1}(0), \ell) \tilde{l}(\mathcal{R}) d\mathcal{R} \\ &= z(\underline{y}, A(\ell)) \int_0^1 \frac{w(\Gamma_\ell^{-1}(\mathcal{R}), \ell)}{w(\Gamma_\ell^{-1}(0), \ell)} \tilde{l}(\mathcal{R}) d\mathcal{R} \\ &= \mathbb{E}[w(\Gamma_\ell^{-1}(\mathcal{R}), \ell) | \ell], \end{aligned}$$

where we used the change of variable $\mathcal{R} = \Gamma_\ell(y)$ in the third line.

Third, differentiating $\mathbb{E}[w(\Gamma_\ell^{-1}(\mathcal{R}), \ell) | \ell]$ wrt ℓ yields

$$\frac{\partial \mathbb{E}[w(\Gamma_\ell^{-1}(\mathcal{R}), \ell) | \ell]}{\partial \ell} = A'(\ell) \frac{\partial z(\underline{y}, A(\ell))}{\partial A} \int_0^1 \frac{w(\Gamma_\ell^{-1}(\mathcal{R}), \ell)}{w(\Gamma_\ell^{-1}(0), \ell)} \tilde{l}(\mathcal{R}) d\mathcal{R} + z(\underline{y}, A(\ell)) \int_0^1 \frac{\partial}{\partial \ell} \left(\frac{w(\Gamma_\ell^{-1}(\mathcal{R}), \ell)}{w(\Gamma_\ell^{-1}(0), \ell)} \right) \tilde{l}(\mathcal{R}) d\mathcal{R},$$

which is positive under positive firm sorting, where we rely on the result established in Proposition 4. Note that this derivative is equal to (18) when $z(y, A) = yA$. \square

A.7 Spatial Heterogeneity in EE Wage Growth

We will show conditions on primitives under which EE wage growth is higher in high- ℓ locations.

Preliminaries. Average wage growth due to an EE transition in location ℓ is given by

$$\mathbb{E}[g_w^{\text{EE}} | \ell] = \frac{\int_{\underline{y}}^{\bar{y}} \int_{\underline{y}}^{\bar{y}} \left(\frac{w(t, \ell)}{w(y, \ell)} - 1 \right) d\Gamma_\ell(t) dG_\ell(y)}{\int_{\underline{y}}^{\bar{y}} \int_{\underline{y}}^{\bar{y}} d\Gamma_\ell(t) dG_\ell(y)}.$$

We can re-express this as

$$\begin{aligned}
\mathbb{E}[g_w^{\text{EE}}|\ell] &= \frac{\int_{\underline{y}}^{\bar{y}} \int_{\underline{y}}^{\bar{y}} \left(\frac{w(t, \ell)}{w(y, \ell)} - 1 \right) d\Gamma_\ell(t) dG_\ell(y)}{\int_{\underline{y}}^{\bar{y}} \int_{\underline{y}}^{\bar{y}} d\Gamma_\ell(t) dG_\ell(y)} \\
&= \int_{\underline{y}}^{\bar{y}} \int_{\underline{y}}^{\bar{y}} \left(\frac{w(t, \ell)}{w(y, \ell)} - 1 \right) \frac{\gamma_\ell(t) \frac{\gamma_\ell(y)}{(1+\varphi^E(1-\Gamma_\ell(y)))^2}}{\int_{\underline{y}}^{\bar{y}} (1-\Gamma_\ell(y)) \frac{\gamma_\ell(y)}{(1+\varphi^E(1-\Gamma_\ell(y)))^2} dy} dt dy \\
&= \int_0^1 \int_{\mathcal{S}} \left(\frac{w(\Gamma_\ell^{-1}(\mathcal{R}), \ell)}{w(\Gamma_\ell^{-1}(\mathcal{S}), \ell)} - 1 \right) \frac{\frac{1}{(1+\varphi^E(1-\mathcal{S}))^2}}{\int_0^1 \frac{1-\mathcal{T}}{(1+\varphi^E(1-\mathcal{T}))^2} d\mathcal{T}} d\mathcal{R} d\mathcal{S}, \tag{A.17}
\end{aligned}$$

where, to go from the second to the third line, we use the following changes of variables $\mathcal{R} = \Gamma_\ell(t)$, $\mathcal{S} = \Gamma_\ell(y)$ and $\mathcal{T} = \Gamma_\ell(y)$.

Proposition A1 (Spatial Heterogeneity in EE Wage Growth). *Suppose $z(y, A(\ell)) = yA(\ell)$. EE wage growth is higher in high- ℓ locations, or $\partial \mathbb{E}[g_w^{\text{EE}}|\ell]/\partial \ell > 0$, if the quantile function, $\Gamma_\ell^{-1}(\mathcal{R})$, is log-supermodular in (\mathcal{R}, ℓ) .*

Proof. From (A.17), $\mathbb{E}[g_w^{\text{EE}}|\ell]$ is increasing in ℓ if the wage ladder is increasingly steep in ℓ , i.e., if $w(\Gamma_\ell^{-1}(\mathcal{R}), \ell)/w(\Gamma_\ell^{-1}(\mathcal{S}), \ell)$ is increasing in ℓ for all $\mathcal{S} < \mathcal{R}$.

We will show that the relative wage between two firms with local ranks $\mathcal{S} < \mathcal{R}$, $\frac{w(\mathcal{R}, \ell)}{w(\mathcal{S}, \ell)}$, is increasing in ℓ if the quantile function, $\Gamma_\ell^{-1}(\mathcal{R})$, is log-supermodular in (\mathcal{R}, ℓ) . From (A.14), we obtain the following expression:

$$\frac{w(\mathcal{R}, \ell) \tilde{l}(\mathcal{R})}{w(\mathcal{S}, \ell) \tilde{l}(\mathcal{S})} = \frac{\tilde{l}(0)z(\underline{y}, A(\ell)) + \int_0^{\mathcal{R}} \tilde{l}'(x)z(\Gamma_\ell^{-1}(x), A(\ell))dx}{\tilde{l}(0)z(\underline{y}, A(\ell)) + \int_0^{\mathcal{S}} \tilde{l}'(x)z(\Gamma_\ell^{-1}(x), A(\ell))dx} = 1 + \frac{\int_{\mathcal{S}}^{\mathcal{R}} \tilde{l}'(x)\Gamma_\ell^{-1}(x)dx}{\tilde{l}(0)\underline{y} + \int_0^{\mathcal{S}} \tilde{l}'(x)\Gamma_\ell^{-1}(x)dx},$$

where $\tilde{l}(\mathcal{R}) = g_\ell(\Gamma_\ell^{-1}(\mathcal{R}))/\gamma_\ell(\Gamma_\ell^{-1}(\mathcal{R}))$ is the size density of a firm with rank \mathcal{R} and $A(\ell)$ cancels out in the last term. Using the above formula, we consider the derivative of relative wage wrt ℓ ,

$$\begin{aligned}
\frac{\partial}{\partial \ell} \frac{w(\mathcal{R}, \ell)}{w(\mathcal{S}, \ell)} &\stackrel{\text{s}}{=} \frac{\partial}{\partial \ell} \left(\frac{w(\mathcal{R}, \ell) \tilde{l}(\mathcal{R})}{w(\mathcal{S}, \ell) \tilde{l}(\mathcal{S})} \right) \\
&\stackrel{\text{s}}{=} \left(\tilde{l}(0)\underline{y} + \int_0^{\mathcal{S}} \tilde{l}'(x)\Gamma_\ell^{-1}(x)dx \right) \int_{\mathcal{S}}^{\mathcal{R}} \tilde{l}'(x) \frac{\partial \Gamma_\ell^{-1}(x)}{\partial \ell} dx - \left(\int_{\mathcal{S}}^{\mathcal{R}} \tilde{l}'(x)\Gamma_\ell^{-1}(x)dx \right) \int_0^{\mathcal{S}} \tilde{l}'(x) \frac{\partial \Gamma_\ell^{-1}(x)}{\partial \ell} dx \\
&= \int_{\mathcal{S}}^{\mathcal{R}} \tilde{l}'(x) \frac{\partial \Gamma_\ell^{-1}(x)}{\partial \ell} dx \left(\tilde{l}(0)\underline{y} + \int_0^{\mathcal{S}} \tilde{l}'(x)\Gamma_\ell^{-1}(x)dx - \frac{\int_{\mathcal{S}}^{\mathcal{R}} \tilde{l}'(x)\Gamma_\ell^{-1}(x)dx \int_0^{\mathcal{S}} \tilde{l}'(x) \frac{\partial \Gamma_\ell^{-1}(x)}{\partial \ell} dx}{\int_{\mathcal{S}}^{\mathcal{R}} \tilde{l}'(x) \frac{\partial \Gamma_\ell^{-1}(x)}{\partial \ell} dx} \right).
\end{aligned}$$

Under positive firm sorting, this expression is positive if

$$\frac{\int_{\mathcal{S}}^{\mathcal{R}} \tilde{l}'(x) \frac{\partial \Gamma_{\ell}^{-1}(x)}{\partial \ell} dx}{\int_{\mathcal{S}}^{\mathcal{R}} \tilde{l}'(x) \Gamma_{\ell}^{-1}(x) dx} \geq \frac{\int_0^{\mathcal{S}} \tilde{l}'(x) \frac{\partial \Gamma_{\ell}^{-1}(x)}{\partial \ell} dx}{\int_0^{\mathcal{S}} \tilde{l}'(x) \Gamma_{\ell}^{-1}(x) dx} = \frac{\int_{\mathcal{S}}^0 \tilde{l}'(x) \frac{\partial \Gamma_{\ell}^{-1}(x)}{\partial \ell} dx}{\int_{\mathcal{S}}^0 \tilde{l}'(x) \Gamma_{\ell}^{-1}(x) dx} \geq 0.$$

This inequality holds if $\int_{\mathcal{S}}^{\mathcal{R}} \tilde{l}'(x) \Gamma_{\ell}^{-1}(x) dx$ is log-supermodular in (\mathcal{R}, ℓ) , which is satisfied if the quantile function, $\Gamma_{\ell}^{-1}(\mathcal{R})$, is log-supermodular in (\mathcal{R}, ℓ) . \square

We will now show that if $\Gamma(y|p)$ is a Pareto distribution then EE wage growth is higher in high- ℓ locations:

Corollary 2 (Spatial Heterogeneity in EE Wage Growth: Pareto). *Suppose $z(y, A(\ell)) = yA(\ell)$ and $\Gamma(y|p) \sim \text{Pareto}(1, 1/p)$. If there is positive firm sorting across space, then EE wage growth is higher in high- ℓ locations, $\partial \mathbb{E}[g_w^{EE} | \ell] / \partial \ell > 0$.*

Proof. By Proposition A1, it suffices to show that the quantile function of Pareto distribution, $\Gamma_{\ell}^{-1}(\mathcal{R}) = (1 - \mathcal{R})^{-\mu(\ell)}$, is log-supermodular in (\mathcal{R}, ℓ) . Under PAM, this condition is satisfied, as $\frac{\partial^2 \log(1 - \mathcal{R})^{-\mu(\ell)}}{\partial \ell \partial \mathcal{R}} = \frac{\mu'(\ell)}{1 - \mathcal{R}} > 0$. \square

A.8 Neutral Productivity Shifts & Implications for Labor Share and Wages

Neutral Productivity Shifts: Definition. We say that spatial firm sorting induces “neutral firm productivity shifts” across locations if firm productivity in one location is simply a scaled version of firm productivity of another location.

To formalize this notion, suppose $y \sim \Gamma_{\ell}(y)$ in location ℓ and $y' \sim \Gamma_{\ell'}(y')$ in location ℓ' , where $y' = ay$ for some constant $a > 0$ (which violates Assumption 1.1 of common support of $\Gamma(\cdot|p)$ for all p). Then, since $y \sim \Gamma_{\ell}(y)$ we have $\frac{y'}{a} \sim \Gamma_{\ell}(y)$. Moreover, $\Gamma_{\ell'}(y') = \mathbb{P}[Y' \leq y'] = \mathbb{P}[\frac{Y'}{a} \leq \frac{y'}{a}] = \Gamma_{\ell}(\frac{y'}{a})$ and $\gamma_{\ell'}(y') = \frac{1}{a} \gamma_{\ell}(\frac{y'}{a})$. If $a > 1$, then $\Gamma_{\ell'}$ first-order stochastically dominates Γ_{ℓ} .⁵⁰ We also assume $z(y, A(\ell)) = yA(\ell)$.

Scaling Property of Wages. To show that under neutral firm productivity shifts across locations, the wage structure is almost the same everywhere, it will be useful to first establish the following scaling property of wages, where we can express the wage in location ℓ' as:

⁵⁰One example of neutral productivity shifts is a scenario in which regions differ in the scale (and not the tail) parameter of the Pareto distribution.

$$\begin{aligned}
w(y', \ell') &= A(\ell')y' - (1 + \varphi^E(1 - \Gamma_{\ell'}(y')))^2 \int_{\underline{y}'}^{y'} \frac{A(\ell')}{[1 + \varphi^E(1 - \Gamma_{\ell'}(t'))]^2} dt' \\
&= A(\ell')y' - (1 + \varphi^E(1 - \Gamma_{\ell}(y'/a)))^2 \int_{\underline{y}'}^{y'} \frac{A(\ell')}{[1 + \varphi^E(1 - \Gamma_{\ell}(t'/a))]^2} dt' \\
&= aA(\ell')y'/a - a(1 + \varphi^E(1 - \Gamma_{\ell}(y'/a)))^2 \int_{\underline{y}}^{y'/a} \frac{A(\ell')}{[1 + \varphi^E(1 - \Gamma_{\ell}(t))]^2} dt \quad (\text{change of var. } t' = at) \\
&= aw(y'/a, \ell) \frac{A(\ell')}{A(\ell)}. \tag{A.18}
\end{aligned}$$

No Spatial Variation in Local Labor Shares. We first show that under neutral firm productivity shifts across locations, there is no spatial variation in local labor shares (in contrast to Proposition 3).

Recall that $l(y, \ell)$ is a function of $\Gamma_{\ell}(y)$ and from our observations, $l(y'/a, \ell) = l(y', \ell')$. Then:

$$\begin{aligned}
\text{LS}(\ell') &= \frac{\int_{\underline{y}'}^{\bar{y}'} w(y', \ell') l(y', \ell') \gamma_{\ell'}(y') dy'}{\int_{\underline{y}'}^{\bar{y}'} z(y', A(\ell')) l(y', \ell') \gamma_{\ell'}(y') dy'} = \frac{\int_{\underline{y}}^{\bar{y}} w(ay, \ell') l(ay, \ell') \gamma_{\ell'}(ay) ady}{\int_{\underline{y}}^{\bar{y}} z(ay, A(\ell')) l(ay, \ell') \gamma_{\ell'}(ay) ady} \quad (\text{change of var. } y' = ay) \\
&= \frac{\int_{\underline{y}}^{\bar{y}} w(ay, \ell') l(y, \ell) \gamma_{\ell}(y) dy}{\int_{\underline{y}}^{\bar{y}} z(ay, A(\ell')) l(y, \ell) \gamma_{\ell}(y) dy} \\
&= \frac{\int_{\underline{y}}^{\bar{y}} aw(y, \ell) \frac{A(\ell')}{A(\ell)} l(y, \ell) \gamma_{\ell}(y) dy}{\int_{\underline{y}}^{\bar{y}} z(y, A(\ell)) a \frac{A(\ell')}{A(\ell)} l(y, \ell) \gamma_{\ell}(y) dy} = \text{LS}(\ell),
\end{aligned}$$

where we made use of the definition of a neutral productivity shift and the scaling property of wages (A.18).

No Spatial Variation in Local Wage Ladders. We aim to show that local wage ladders are equally steep in all locations, or for any two locations $\ell \neq \ell'$

$$\frac{w(\Gamma_{\ell'}^{-1}(\mathcal{R}), \ell')}{w(\Gamma_{\ell'}^{-1}(\mathcal{S}), \ell')} = \frac{w(\Gamma_{\ell}^{-1}(\mathcal{R}), \ell)}{w(\Gamma_{\ell}^{-1}(\mathcal{S}), \ell)},$$

where \mathcal{R} and \mathcal{S} satisfy $\mathcal{R} > \mathcal{S}$ and are arbitrary quantiles of the local firm distributions.

Using the scaling property of wages (A.18), the LHS can be expressed as

$$\frac{w(\Gamma_{\ell'}^{-1}(\mathcal{R}), \ell')}{w(\Gamma_{\ell'}^{-1}(\mathcal{S}), \ell')} = \frac{\frac{A(\ell')}{A(\ell)} aw\left(\frac{\Gamma_{\ell'}^{-1}(\mathcal{R})}{a}, \ell\right)}{\frac{A(\ell')}{A(\ell)} aw\left(\frac{\Gamma_{\ell'}^{-1}(\mathcal{S})}{a}, \ell\right)} = \frac{w(\Gamma_{\ell}^{-1}(\mathcal{R}), \ell)}{w(\Gamma_{\ell}^{-1}(\mathcal{S}), \ell)},$$

where the last equality uses $y = \Gamma_{\ell}^{-1}(\mathcal{R})$ and our definition of the neutral productivity shift,

$y' = ay$. Thus, the wage ratio of two firms with productivity ranks \mathcal{R} and \mathcal{S} is the same in any two locations ℓ and ℓ' .

No Spatial Variation in Within-Location Wage Inequality. Replacing $\mathcal{S} = 0$ in the last paragraph immediately implies that, under neutral firm productivity shifts, there are no spatial differences in within-location inequality (in contrast to Proposition 4).

No Spatial Variation in Local EE Wage Growth. Since we have established above that under neutral firm productivity shifts, the wage ratio of two firms with ranks \mathcal{R} and \mathcal{S} is the same in any two locations ℓ and ℓ' , an EE transition leads on average to the same wage gains in *all* locations, see (A.17) (in contrast to Proposition A1).

B Quantitative Model: Labor Mobility & Residential Housing

The firms' location choice problem has the same structure as in the baseline model, only that in (6) they now take into account that their meetings rates vary across locations.

From the firms' perspective, congestion—which can be measured by market tightness—is decreasing in the endogenous population size. If the local population is large, then market tightness is small and firms' meeting rate $\lambda^F(\ell)$ is high, benefiting firms. In addition, competition that stems from poaching risk is mitigated in places with a large population: The job arrival rate for employed workers, $\lambda^E(\ell)$, decreases as the population gets larger and so the probability that firms retain workers rises.

Thus, an important question is how population size varies with ℓ . When agents conjecture that there is positive sorting between firms and locations, high- ℓ locations are more attractive (due to a stochastically better wage distribution), and draw in more workers. Labor market congestion therefore benefits firms in high- ℓ locations, $\partial\lambda^F/\partial\ell > 0$ and $\partial\lambda^E/\partial\ell < 0$, alleviating the competition channel and strengthening their desire to settle there (although this mechanism is mitigated by congestion in the residential housing market, which prevents a massive inflow of workers into high- ℓ locations). As a result, positive sorting materializes more easily than in the baseline model with exogenous meeting rates that are constant across space.

We now state our result on firm sorting under labor mobility formally. To do so, again denote the minimum of the first term on the LHS of (13) (over ℓ, y) by ε^P , i.e.,

$$\varepsilon^P \equiv \min_{\ell, y} \frac{\frac{\partial^2 z(y, A(\ell))}{\partial y \partial A(\ell)} \frac{\partial A(\ell)}{\partial \ell}}{\frac{\partial z(y, A(\ell))}{\partial y}}.$$

Assume that the labor market matching function is given by $M(\mathcal{V}(\ell), \mathcal{U}(\ell)) = \mathcal{A}\sqrt{\mathcal{V}(\ell)\mathcal{U}(\ell)}$ and that workers' flow utility function over housing and consumption is Cobb Douglas with share parameters ω and $1 - \omega$, respectively. We assume that the exogenous functions $B(\cdot)$ and $\delta(\cdot)$ do not vary with ℓ (i.e., $B(\ell) = B = 1$ and $\delta(\ell) = \delta$). This is to tie our hands since there certainly exist some assumptions on how these exogenous objects depend on ℓ that would deliver the result of positive sorting but those conditions would have little to do with the specific mechanism of our model. Moreover, to allow for a more general setting than in the text, we assume that local housing supply is endogenous and given by $h(\ell) = d(\ell)^\xi$, where ξ is the housing supply elasticity.

Proposition A2. *Suppose $B(\ell) = B = 1$ and $\delta(\ell) = \delta$ and local housing supply is given by $h(\ell) = d(\ell)^\xi$. If (i) z is strictly supermodular and either the productivity gains from sorting into higher ℓ , ε^P , are sufficiently large or the competition forces, $1/\delta$, are sufficiently small, and (ii) housing supply elasticity ξ is sufficiently large, then there exists an equilibrium with positive sorting in (p, ℓ) .*

Proof. See Appendix SA.3.1. □

C Data and Sample Restrictions

C.1 Administrative Regional-Level Data from the GFSO

Data Description. We obtain regional-level data from the German Federal Statistical Office (GFSO). To be consistent with our sample from the FDZ below, we focus on the years 2010-2017. We obtain district-level data (for 401 districts) for all years and aggregate them to the commuting-zone level (there are 257 CZs), using a crosswalk provided by the Federal Office for Building and Regional Planning of Germany (*Bundesinstitut für Bau-, Stadt- und Raumforschung*—BBSR). Finally, we take (simple) averages across years to obtain one value for each variable per CZ. If applicable, we adjust the variables to the monthly level, for consistency with our FDZ sample.

Defining Important Variables.

Labor Compensation. Total labor compensation in a commuting zone at year t is defined as

$$\text{Labor Comp}_t = \frac{\text{Total Hours Worked by Total Workforce}_t}{\text{Total Hours Worked by Employees}_t} \times \text{Comp of Employees}_t,$$

and *compensation of employees* consists of gross wages and salaries as well as employers' actual and imputed social contributions. We divide by 12 to obtain the monthly statistic.

Value Added per Worker. The monthly gross value added per worker in any given CZ is calculated as the ratio of (annual) gross value added and total employment in a location, divided by 12.

Labor Share. We construct the local *labor share* as the ratio between *labor compensation* and gross value added in each commuting zone.

Average Wage. The *average monthly wage* of a commuting zone is defined by total (annual) *labor compensation* divided by total employment, divided by 12.

Average Firm Size. We define *average firm size* of a commuting zone by the total number of employees over the total number of establishments.

GDP per Capita. We take the ratio of (annual) GDP and population in each commuting zone; then divide by 12 to get the monthly figure. GDP corresponds to the gross value added of all sectors of the economy plus taxes on products, but excluding subsidies on products.

Unemployment Rate. We first use unemployment rates and number of unemployed workers at the district level to obtain the number of people who are in the labor force in each district. We then sum by commuting zone the number of unemployed workers as well as the number of people in the labor force and divide them to obtain the local *unemployment rate*.

Rent-to-Income Ratio. Germany-wide *rent-to-income ratio* of the main tenant household.

Trade Tax Rate. The trade tax (*Gewerbesteuer*) is levied on the adjusted profit of corporations. It combines a base rate (universal to all municipalities, 3.5%) and a municipal tax rate (which is a multiplier to the base rate and at the discretion of each municipality). We focus on the municipal tax rate and refer to it as *trade tax rate*. We first aggregate municipal tax rates to the district level and then to the CZ level using population weights.

Share of Employees with a Degree. In each CZ, we take the ratio of employees with an academic degree and all employees subject to social security contributions at the place of work.

Net Business Registration Intensity. We define *net business registration intensity* at the CZ level as the balance between business registrations and de-registrations per 1,000 inhabitants.

C.2 Administrative Worker- and Firm-Level Data from the FDZ

Data Description. We use worker-/firm-level data provided by the Research Data Centre (FDZ) of the German Federal Employment Agency at the Institute for Employment Research.

We use three datasets from FDZ: LIAB (Linked Employer-Employee Data), BHP (Establishment History Panel 7518), and EP (Establishment Panel).

The LIAB data links annual information on establishments with information on all individuals employed at those establishments.⁵¹ Surveyed establishments (the ‘panel cases’) are followed between 2009-2016 and we observe individual-level information for *all* their employees. This

⁵¹Because we only observe data at the level of the establishment, we use ‘establishments’ and ‘firms’ interchangeably.

individual-level data, which includes workers’ gender, education, full-time employment status, gross daily wages and work district, is assembled from official social security records. For more information, see [Ruf et al. \(2021a\)](#) and [Ruf et al. \(2021b\)](#).

We augment these datasets with the Establishment History Panel 7518 (BHP), a 50% random sample of all German establishments with at least one employee subject to social security as of June 30 in any given year. In addition to standard information like total employment and average wages, it contains total inflows and outflows of workers at the establishment level.

Moreover, linking the BHP to the LIAB, we observe basic information for *any* employer in workers’ entire employment history between 1975 and 2018. While the original data is in a spell format, we transform it into a monthly panel.⁵²

We complement these main data sources with information on firm-level sales and costs of inputs from the Establishment Panel (EP). The EP is a nationally representative survey of about 15,000 firms that reports standard balance-sheet information on sales, inputs and employment, as well as information on a variety of survey questions on topics related to employment policy.

Sample Restrictions. Our baseline sample pools the years 2010-2017. We focus on full-time employees. We drop establishments with less than 5 employees and establishments whose mean real daily wage across the sample period is lower than 15 Euros, measured in 2015 euros (this wage restriction is based on [Card et al. \(2013\)](#)).

Defining Important Variables. *Monthly Real Wage.* To compute an individual’s *monthly real wage*, we multiply daily wages by 30, and deflate these nominal wages using the German CPI (Table 61111-0001 in the GENESIS database of the German Federal Statistical Office). The CPI base year is 2015. Data Source: Establishment History Panel (BHP).

Value Added per Full-Time Employee. We measure value added at the firm-level as the difference between sales and input costs as reported in the Establishment Panel, divided by the number of full-time employees. See also [Bruns \(2019\)](#). We deflate these variables using the same CPI as above. Data Source: Establishment Panel (EP).

Employment-to-Employment (EE) Transition. We say a worker made an EE move in month t in any of the following scenarios: (i) if they were employed at some establishment in month $t - 1$ and are employed at a different establishment in month t ; (ii) if they are employed at some establishment in month $t - 3$ (or $t - 2$), disappear from the sample during months $t - 2$ and $t - 1$ (or only $t - 1$) without claiming unemployment benefits, and are employed again at a different

⁵²If a new spell starts in the middle of a month, we assign the month to the longest spell within the month.

establishment in month t . In scenario (ii), we consider it likely that the new job was already lined up when the worker left the previous one. Data Source: Linked Employer-Employee Data (LIAB).

Unemployment-to-Employment (UE) Transition. A worker made a UE move in month t if they were unemployed—that is, collecting unemployment benefits—in month $t - 1$ and are employed at some establishment in t . Data Source: Linked Employer-Employee Data (LIAB).

Employment-to-Unemployment (EU) Transition. A worker made an EU move in month t if they were employed at some establishment in month $t - 1$ and are (officially) unemployed in t or permanently disappear from the sample (we exclude December 2017 from this count, since it is the last month in our panel). Data Source: Linked Employer-Employee Data (LIAB).

Labor Market Transition Rates. In our regression analysis, we construct measures of workers' monthly transition rates from the data: We proxy the contact rate of employed workers λ^E by the realized EE transition rate in the data. For the contact rate λ^U and the rate of job destruction δ —since in the model, unemployed workers accept all offers and separations to unemployment are exogenous—they are equal to the realized rates. Specifically, in each t :

$$\begin{aligned}\lambda_t^E &= \frac{\# \text{ Employed workers in } t - 1 \text{ working in another firm in } t}{\# \text{ Employed workers in } t - 1} \\ \lambda_t^U &= \frac{\# \text{ Unemployed workers in } t - 1 \text{ who are employed in } t}{\# \text{ Unemployed workers in } t - 1} \\ \delta_t &= \frac{\# \text{ Employed workers in } t - 1 \text{ who are unemployed in } t}{\# \text{ Employed workers in } t - 1}.\end{aligned}$$

We measure these flows at the monthly frequency in each local labor market and then take the average over years 2010-2017 to obtain one number per local labor market. Data Source: Linked Employer-Employee Data (LIAB).

Poaching Share. To measure job flows and poaching at the firm level, we follow Moscarini and Postel-Vinay (2018) and Bagger and Lentz (2019) and measure firms' *poaching shares*, which we define as the ratio of EE inflows relative to all inflows.⁵³ Given our focus on *local* labor markets, we also compute firms' share of EE inflows and UE inflows that are local, i.e., from within the same commuting zone. Data Source: Linked Employer-Employee Data (LIAB).

C.3 Variables from Other Data Sources

Residential Housing Prices. We use residential rental rates predicted for the centroids of postal codes (provided to us by Gabriel Ahlfeldt based on Ahlfeldt et al. (2022)) and aggregate them

⁵³In the terminology of Moscarini and Postel-Vinay (2018) and Bagger and Lentz (2019), this object refers to the poaching inflow share and the poaching index, respectively.

to the commuting-zone level. The model counterpart is $d(\ell)$ for each CZ ℓ .

Replacement Rate. We use the unemployment insurance net replacement rate. This variable is based on data from the Out-of-Work Benefits Dataset (OUTWB), provided as part of the Social Policy Indicator (SPIN) database (Nelson et al., 2020). Depending on household composition and earnings, replacement rates vary and we take 60% as a reference point.

Commercial Real Estate Prices. We use price data (EUR/ m^2) for commercial properties 2012/13 from the German Real Estate Association (*Deutscher Immobilienverband*). We aggregate prices from the city to the commuting-zone level. The model counterpart is $k(\ell)$.

Distance to Highway. Distance to highway is proxied by the area-weighted average car driving time to the next federal motorway junction in minutes. We obtain this variable from the German Federal Office for Building and Regional Planning. Data is available only for 2020.

C.4 Defining Locations

Local Labor Markets. We consider 257 commuting zones (*Arbeitsmarktregionen*)—our local labor markets. These are defined for the year 2017 by the Federal Office for Building and Regional Planning of Germany (*Bundesinstitut für Bau-, Stadt- und Raumforschung*—BBSR).

For the supplementary empirical analysis of firm sorting, Appendix SA.4.1, we consider a different, more aggregate definition of local labor markets (38 NUTS2 regions, defined by the European Union), because the number of observations per CZ is too small.

D Empirical Analysis

D.1 Spatial Wage Inequality: Real versus Nominal

In our main analysis, we use—consistent with our theory—nominal wages to measure inequality both within and across locations. In Table A.1 we report average wage premia after deflating wages with local CPIs. Doing so decreases spatial inequality by about one third.

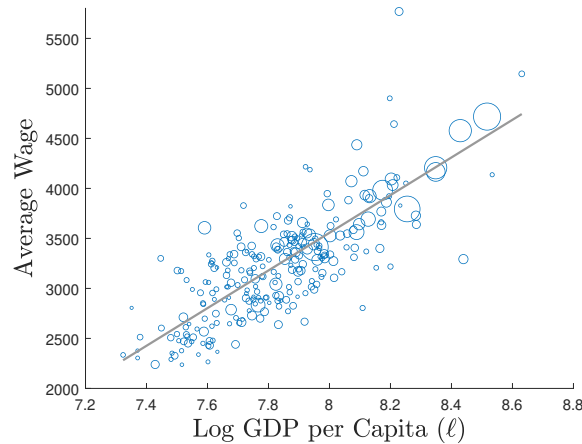
Table A.1: Spatial Inequality (Monthly, €): Real versus Nominal

	German CPI		Local CPI	
	Wage	Value Added	Wage	Value Added
<i>Rich-Poor Inequality</i>				
Rich	3784.6	5820.2	3899.6	5993.3
Poor	2755.2	4034.2	3135.2	4591.8
Rich/Poor	1.37	1.44	1.24	1.31
<i>West-East Inequality</i>				
West	3491.13	5237.02	3704.85	5552.25
East	2731.63	4045.24	3122.56	4624.49
West/East	1.28	1.30	1.19	1.20
<i>Urban-Rural Inequality</i>				
Urban	3510.01	5270.60	3701.94	5552.52
Rural	2984.37	4429.01	3372.72	5007.15
Urban/Rural	1.18	1.19	1.10	1.11

Notes: Data source: German Federal Statistical Office. With some abuse, we denote by ‘Nominal’ those variables that are deflated using the *Germany-wide* CPI in 2015; and by ‘Real’ we denote the variables that are deflated using the *Local* CPI, i.e., using commuting zone-level price deflators (computed from district-level price deflators from BBSR). “Rich-Poor Inequality” refers to the comparison of the bottom and top quartile of CZs when grouped according to their GDP per capita. We categorize CZs into Rural or Urban based on their districts, following the categorization provided by the BBSR for the year 2018. We categorize CZs into East or West Germany based on whether the districts they consist of belong to Eastern or Western states.

D.2 The Spatial Wage Premium

Figure A.1: Implication of Spatial Firm Sorting: Spatial Variation in Wages



Notes: Data source: BHP. The figure shows a scatter plot between local average firm wages and local log GDPpc. The size of the markers indicates the size of the CZ (i.e., the number of firms in each CZ).

D.3 Spatial Decomposition of Life-Time Earnings

In this exercise, we study the impact of heterogeneous wage ladders across space on spatial inequality in lifetime earnings. We compare lifetime earnings in two regions, ‘rich’ and ‘poor’ locations (where ‘rich’ and ‘poor’ locations again refer to the top and bottom 25% of commuting zones in terms of GDP per capita). We focus on a single cohort of workers in each region: They are 25-30 years old in 2002, and we follow them over 15 years, from 2002 to 2017. We restrict the sample to those workers who remain in the region where we first observe them.

Spatial Decomposition of Between-Location Wage Inequality. First, we measure average starting wages of workers in rich and poor locations, before they climb the wage ladder. Second, we compute average wages of workers in each region after 15 years. Third, we decompose the total average wage growth within regions into three parts: (i) the average wage growth of workers who never changed jobs nor experienced unemployment for more than four months (i.e., the ‘stayers’), (ii) the average wage growth of workers who changed jobs at least once and did not experience unemployment for more than four months (i.e., the ‘EE movers’), and (iii) the average wage growth of workers who have been unemployed at least once for more than four months (which we call the ‘unemployed’). Average wage growth of a region equals the weighted average of wage growth in these three categories, with the weights being equal to the number of workers in each category.

This decomposition of average wage growth allows us to assess the contribution of heterogeneous wage ladders across space to spatial wage inequality as follows. We compute wage growth in rich locations imposing the (counterfactual) wage growth of EE movers from poor locations, while keeping the number of EE movers fixed. This way, we get a measure of lifetime income inequality across space keeping wage ladders *the same* across regions.

The results are in Table A.2. If poor and rich regions had the same wage ladder, spatial inequality in lifetime income would be around 19% lower than under the heterogeneous wage ladders that we observe in the data, i.e., the rich region would be characterized by only 66.4% higher wages than the poor one, instead of the observed 81.7% ($\approx 5854.35/3221.44-1$).

Spatial Decomposition of Within-Location Wage Inequality. We further assess the impact of heterogeneous wage ladders on differences in within-location wage inequality. First, we compute the 90-10 log wage difference within each CZ and then take the average for CZs within rich and within poor locations, based on their wages after 15 years (i.e., in 2017). Second, to compute a measure of wage inequality within labor markets under the counterfactual of identical EE wage growth across space, we assume that the wage growth distribution of EE movers in rich locations follows that of poor locations, while keeping the number of EE movers and the wage growth of

Table A.2: Decomposition of Lifetime Earnings Inequality in Top versus Bottom 25% of Local Labor Markets

	Top 25%	Bottom 25%
Wage Growth (total)	0.595	0.362
Wage Growth of Stayers	0.540	0.364
Wage Growth of Unemployed	0.407	0.313
Wage Growth of EE Movers	0.712	0.381
Starting Wage (Monthly, €)	3670.44	2365.23
Wage after 15 Years (Monthly, €)	5854.35	3221.44
After-15-years Spatial Wage Inequality (data)		1.817
After-15-years Spatial Wage Inequality (counterfactual, same EE wage growth)		1.664
After-15-years Spatial Wage Inequality (counterfactual, same stayer wage growth)		1.720
Contribution of Job Ladder Differences to Spatial Wage Inequality		0.187
Contribution of Stayer Wage Growth Differences to Spatial Wage Inequality		0.119

Notes: Data source: LIAB. Top and bottom 25% of local labor markets (CZs) are categorized based on GDP per capita. The last two rows report the percentage differences between the actual inequality in lifetime income (row 7) and counterfactual inequality in lifetime income (row 8 and 9), that is, $19\% \sim (81.7 - 66.4)/81.7$, and $12\% \sim (81.7 - 72.0)/81.7$.

other workers unchanged in each location. We then compute the counterfactual wage distribution and the corresponding counterfactual 90-10 log wage difference in rich locations.

We find that if rich locations had the same wage ladder as poor locations, the spatial gap in within-location wage inequality between the two regions would be 22% smaller. In the actual data, the 90-10 log wage difference is 1.145 in rich locations and 0.869 in poor regions, resulting in a gap of 0.276 log points. In turn, in our counterfactual scenario, the 90-10 log wage difference in rich regions would decline from 1.145 to 1.083, reducing the spatial gap in within-location inequality to 0.214 log points, and thus by 22% (i.e., $(0.276-0.214)/0.276 \approx 0.22$).

E Identification

We prove identification of our model under the following assumption:

Assumption A1. *We assume the following functional forms and normalizations:*

1. *The labor market matching function is given by $M(\mathcal{V}(\ell), \mathcal{U}(\ell)) = \mathcal{A}\sqrt{\mathcal{V}(\ell)\mathcal{U}(\ell)}$.*
2. *Workers' flow utility function over housing and consumption is Cobb Douglas with share parameters ω and $1 - \omega$.*
3. *The ex post firm productivity distribution is given by $\Gamma(y \mid p) = 1 - y^{-\frac{1}{p}}$.*
4. *The production function is given by $z(y, A(\ell)) = yA(\ell)$.*
5. *Land distribution R and housing expenditure share ω are observed.*
6. *Normalize the value of unemployment such that $\rho V^U = 1$.*

Proof of Proposition 6. We need to identify the ranking of locations $[\underline{\ell}, \bar{\ell}]$; functions (Q, A, B) ; the tail parameters p of the ex post productivity distribution; the separation rate schedule δ ; the relative matching efficiency κ and the overall efficiency of the matching function \mathcal{A} ; as well as the parameters pertaining to the housing market (τ, h) .

First, we can assign $\ell \in [\underline{\ell}, \bar{\ell}]$ to each location, based on any observed statistic that—according to our model—is increasing in ℓ .

Second, $\mu(\ell)$ (and thus $p = \mu(\ell)$) can be obtained from a location's labor share, $LS(\ell) = 1 - \mu(\ell)$, see Corollary 1.

Third, we obtain κ as described in the text. In turn, equation (25), which allows us to back out the overall matching efficiency, is derived as follows. First, note that:

$$\begin{aligned}
\lambda^E(\ell) &= \frac{M(\mathcal{V}(\ell), \mathcal{U}(\ell))}{(u(\ell) + \kappa(1 - u(\ell)))L(\ell)} \kappa \\
&= M(\mathcal{V}(\ell), \mathcal{U}(\ell)) \frac{\kappa(1 - u(\ell))}{u(\ell) + \kappa(1 - u(\ell))} \frac{1}{(1 - u(\ell))L(\ell)} \\
&= \mathcal{A}\mathcal{V}(\ell)^{\frac{1}{2}} (u(\ell) + \kappa(1 - u(\ell))L(\ell))^{\frac{1}{2}} \frac{\lambda^E(\ell)}{\delta(\ell) + \lambda^E(\ell)} \frac{\delta(\ell) + \lambda^U(\ell)}{\lambda^U(\ell)} \frac{1}{L(\ell)} \\
\Rightarrow L(\ell) &= \mathcal{A}^2 \frac{\delta(\ell) + \lambda^U(\ell)}{\delta(\ell) + \kappa\lambda^U(\ell)} \left(\frac{1}{\lambda^U(\ell)} \right)^2. \tag{A.19}
\end{aligned}$$

Next, note that average firm size in location ℓ is given by $\bar{l}(\ell) = (1 - u(\ell))L(\ell)$, and thus,

$$L(\ell) = \left(1 + \frac{\delta(\ell)}{\lambda^U(\ell)} \right) \bar{l}(\ell). \tag{A.20}$$

Equalizing (A.19) and (A.20), and solving for \mathcal{A} gives (25) in the text, where we treat $(\lambda^U(\ell), u(\ell), \bar{l}(\ell))$ as observed for all ℓ .

Fourth, we obtain $\delta(\ell)$ in each ℓ from local unemployment and job-finding rates, see (24).

Fifth, given $\mu(\ell)$, we obtain the A -schedule from how average value added varies across space:

$$\mathbb{E}[z(y, A(\ell))|\ell] = A(\ell)\mathbb{E}[y|\ell] = A(\ell)\frac{1}{1-\mu(\ell)} \Rightarrow A(\ell) = (1 - \mu(\ell))\mathbb{E}[z(y, A(\ell))|\ell].$$

Sixth, regarding the housing market parameters, we treat $(u(\ell), L(\ell), d(\ell), \mathbb{E}[w(y, \ell)|\ell], \mathcal{T}, \omega)$ as observed for all ℓ (where \mathcal{T} is the economy-wide replacement rate of the unemployed) and obtain $(\tau, h(\cdot), w^U(\cdot))$ from a system of three equations:

We use government budget constraint,

$$\tau d(\ell)h(\ell) = w^U(\ell)u(\ell)L(\ell), \quad (\text{A.21})$$

housing market clearing,

$$h(\ell) = \omega \frac{w^U(\ell)}{d(\ell)} u(\ell)L(\ell) + \omega \frac{\mathbb{E}[w(y, \ell)|\ell]}{d(\ell)} (1 - u(\ell))L(\ell), \quad (\text{A.22})$$

and an equation obtained from the definition of replacement rate \mathcal{T}

$$\tau = \frac{1}{\omega \frac{\sum_{\hat{\ell}} (1 - u(\hat{\ell}))L(\hat{\ell})}{\sum_{\hat{\ell}} u(\hat{\ell})L(\hat{\ell})} + \mathcal{T}}. \quad (\text{A.23})$$

We computed (A.23) as follows. The replacement rate \mathcal{T} satisfies

$$\mathcal{T} \sum_{\ell} \mathbb{E}[w(y, \ell)|\ell] \frac{(1 - u(\ell))L(\ell)}{\sum_{\hat{\ell}} (1 - u(\hat{\ell}))L(\hat{\ell})} = \sum_{\ell} w^U(\ell) \frac{u(\ell)L(\ell)}{\sum_{\hat{\ell}} u(\hat{\ell})L(\hat{\ell})},$$

where the RHS is the aggregate unemployment benefit. Note that

$$\sum_{\ell} w^U(\ell) \frac{u(\ell)L(\ell)}{\sum_{\hat{\ell}} u(\hat{\ell})L(\hat{\ell})} = \tau \sum_{\ell} \frac{d(\ell)h(\ell)}{\sum_{\hat{\ell}} u(\hat{\ell})L(\hat{\ell})} = \frac{\omega\tau}{1 - \tau\omega} \sum_{\ell} \frac{\mathbb{E}[w(y, \ell)|\ell](1 - u(\ell))L(\ell)}{\sum_{\hat{\ell}} u(\hat{\ell})L(\hat{\ell})},$$

where the first equality uses government budget constraint (A.21) and the second one uses a combination of (A.21) and housing market clearing (A.22), which gives $\mathbb{E}[w(y, \ell)|\ell](1 - u(\ell))L(\ell) = (1 - \omega\tau)d(\ell)h(\ell)$. Equalizing the last two equations and solving for τ yields (A.23).

Equation (A.23) pins down τ given the observed housing expenditure share and replacement rate (ω, \mathcal{T}) . For each location ℓ and given τ , equations (A.22)–(A.21) are then a system of two equations and two unknowns $(w^U(\ell), h(\ell))$, which can be solved uniquely.

Last, to identify amenity schedule B , our starting point is the value of unemployment:

$$\rho V^U(\ell) = d(\ell)^{-\omega} B(\ell) w^U(\ell) + \tilde{b}(\ell) + d(\ell)^{-\omega} B(\ell) \lambda^U(\ell) \left[\int_{w^R(\ell)}^{\bar{w}} \frac{1 - F_{\ell}(t)}{\delta(\ell) + \lambda^E(\ell)(1 - F_{\ell}(t))} dt \right], \quad (\text{A.24})$$

which is the same as in the baseline model, except that unemployed workers receive unemployment benefit $w^U(\ell)$ and enjoy local amenity $B(\ell)$, but suffer from unemployment stigma, captured by $\tilde{b}(\ell)$. Next, as before, reservation wage w^R is implicitly defined by a condition that equalizes the value of holding a job with the value of unemployment:

$$d(\ell)^{-\omega} B(\ell) w^R(\ell) = d(\ell)^{-\omega} B(\ell) w^U(\ell) + \tilde{b}(\ell) \\ + d(\ell)^{-\omega} B(\ell) (\lambda^U(\ell) - \lambda^E(\ell)) \left[\int_{w^R(\ell)}^{\bar{w}} \frac{1 - F_\ell(t)}{\delta(\ell) + \lambda^E(\ell)(1 - F_\ell(t))} dt \right],$$

where, to satisfy Assumption 1, we now set $\tilde{b}(\ell)$ such that $w^R(\ell) = z(\underline{y}, A(\ell))$:

$$\tilde{b}(\ell) = d(\ell)^{-\omega} B(\ell) \left(z(\underline{y}, A(\ell)) - w^U(\ell) - (\lambda^U(\ell) - \lambda^E(\ell)) \left[\int_{w^R(\ell)}^{\bar{w}} \frac{1 - F_\ell(t)}{\delta(\ell) + \lambda^E(\ell)(1 - F_\ell(t))} dt \right] \right). \quad (\text{A.25})$$

Plug $\tilde{b}(\ell)$ back into V^U in (A.24), and use a change of variable (to re-express F_ℓ using Γ_ℓ , where $\Gamma_\ell(y) = \Gamma(y|\mu(\ell))$) and make use of the Pareto assumption on Γ to obtain

$$\rho V^U = d(\ell)^{-\omega} B(\ell) A(\ell) \left(1 + 2 (\lambda^E(\ell))^2 \int_1^\infty y^{-\frac{1}{\mu(\ell)}} \frac{1}{\mu(\ell)} y^{-\frac{1}{\mu(\ell)}-1} \int_1^y \frac{dt}{[\delta(\ell) + \lambda^E(\ell) t^{-\frac{1}{\mu(\ell)}}]^2} dy \right),$$

which allows us to back out $B(\ell)$ for each ℓ , given the normalization $\rho V^U = 1$ and given (A, μ) (obtained above) as well as observed rental rates d and transition rates (λ^E, δ) . \square

F Estimation Results

Table A.3: Calibrated Parameters

Parameter	Value	Calibration
κ	0.253	monthly UE and EE transition rate (LIAB)
\mathcal{A}	0.276	monthly UE transition rate (LIAB) and average firm size (GFSO)
ω	0.272	rent-to-income of main tenant households (GFSO)
τ	0.164	replacement rate of unemployed workers (SPIN)

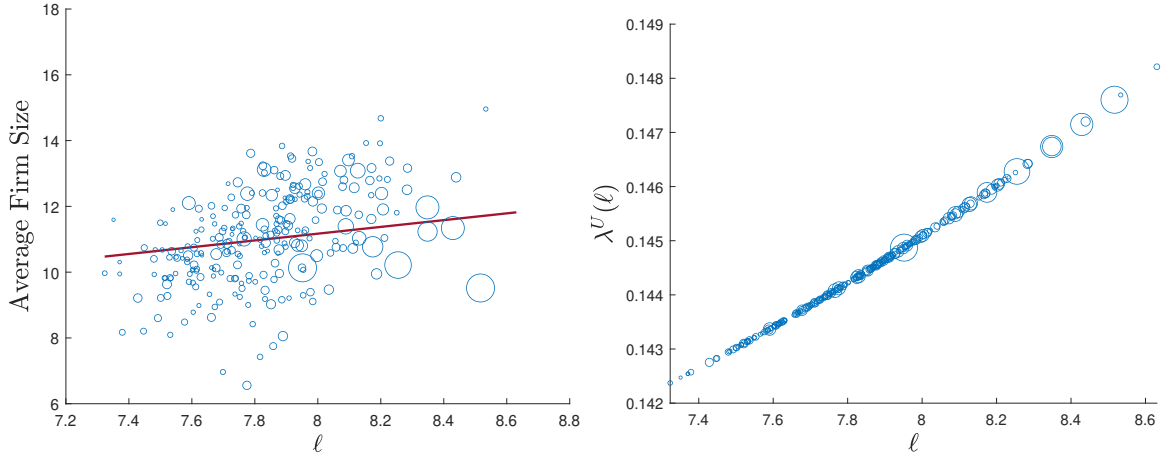
Notes: For details on data sources and variable definitions, see Appendix C. To obtain the relative matching efficiency of employed workers, κ , we use Germany-wide job-finding rates λ^E based on monthly EE transition rates, counting a transition only for those workers who move to a new firm with higher wage.

Table A.4: Determinants of Local TFP $A(\ell)$

	(1)	(2)	(3)	(4)	(5)
Trade Tax	0.1298* (0.0600)				-0.0854 (0.0458)
Distance to Highway		-0.0536*** (0.0123)			-0.0424*** (0.0093)
% of Employees with a College Degree			0.8765*** (0.1423)		0.6466*** (0.1275)
Net Business Registration Intensity				0.0511*** (0.0063)	0.0376*** (0.0064)
N	257	257	257	257	257
R ²	.0346	.153	.356	.274	.562

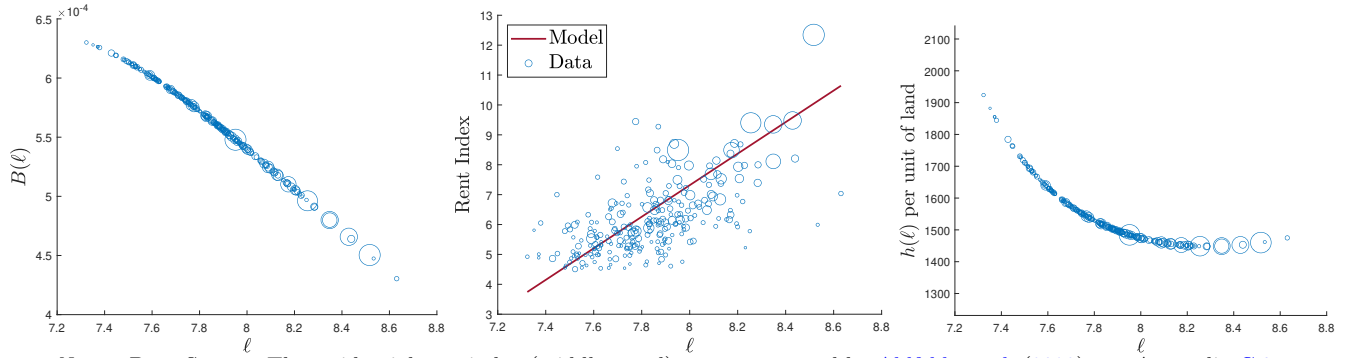
Notes: Data Sources: German Federal Statistical Office and Federal Office for Building and Regional Planning. All regressions are run at the CZ level and weighted by the number of establishments in each CZ. Data is averaged across years for the period 2010-2017. The dependent variable in columns (1)-(5) is ‘Log Local TFP, $\log A(\ell)$ ’ obtained from our estimation for each ℓ ; see Section 6.3. See Appendices C.1 and C.3 for the definition of the independent variables.

Figure A.2: Additional Parameter Estimates I: Local Worker Meeting Rate (right) Informed by Firm Size (left)



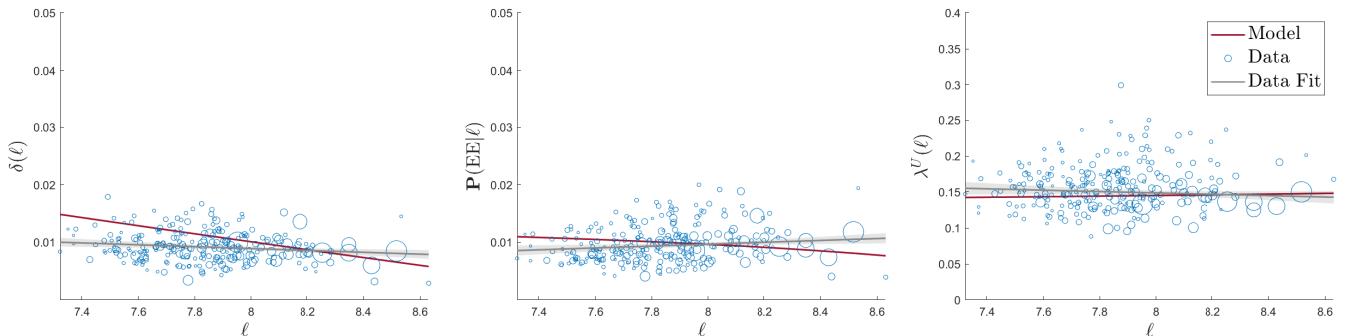
Notes: Data Source: Local firm size comes from the German Federal Statistical Office. Observations are weighted by the number of establishments in each CZ, indicated by the size of the markers.

Figure A.3: Additional Parameter Estimates II: Location Preference Schedule (left); Housing Supply (right) Informed by Residential Rents (middle)



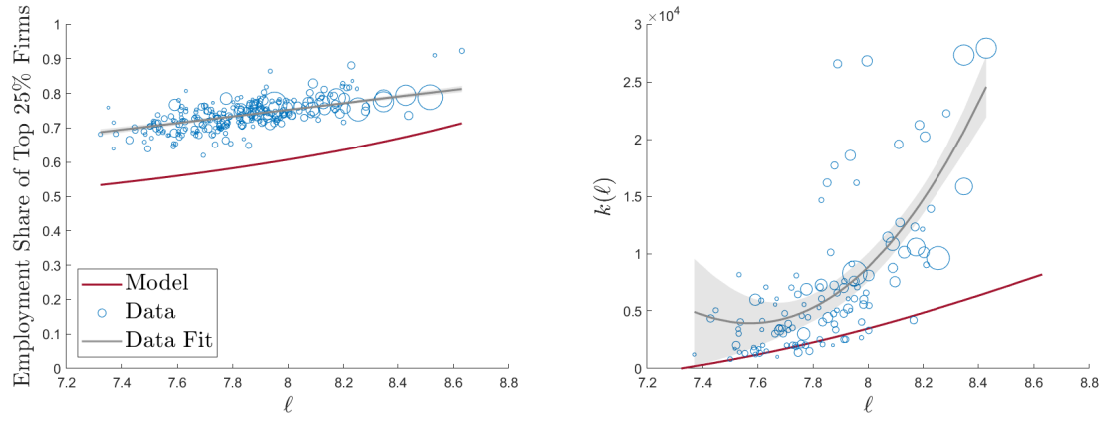
Notes: Data Source: The residential rent index (middle panel) was constructed by Ahlfeldt et al. (2022), see Appendix C.3, and the red line indicates the linear (model) fit as a function of ℓ . Observations are weighted by the number of establishments in each CZ, indicated by the size of the markers.

Figure A.4: Over-Identification: Transition Rates δ , $\mathbb{P}(EE|\ell)$ and λ^U in Data and Model



Notes: Data Source: LIAB. For details on how job destruction rate δ (left panel) and UE transition rate λ^U (right panel) are constructed, see Appendix C.2. In the middle panel, we display the probability of an EE transition, $\mathbb{P}(EE|\ell)$. In the model, workers change jobs if they receive an offer (which happens at rate $\lambda^E(\ell)$) and if its wage exceeds their current one. To compute the data counterpart, we thus multiply the observed EE probability by 0.597, which represents the fraction of EE moves associated with a wage gain. Red and gray lines indicate the linear model and data fit of each variable as a function of ℓ . Observations are weighted by the number of establishments in each CZ, indicated by the size of the markers. 95% confidence intervals are displayed in gray.

Figure A.5: Model Fit: Additional Non-Targeted Moments



Notes: Data Sources: Left panel is based on BHP; right panel is based on price data for commercial properties 2012/13 from the German Real Estate Association; see Appendices C.2 and C.3 for details. Red (gray) lines indicate the linear or quadratic model (data) fit of each variable as a function of ℓ . Observations are weighted by the number of establishments in each CZ, indicated by different sizes of the markers. 95% confidence intervals are displayed in gray.